5th EDITION

PHYSICS

for SCIENTISTS and ENGINEERS



DOUGLAS

GIANCOLI

5th EDITION

PHYSICS

for SCIENTISTS and ENGINEERS

DOUGLAS GIANCOLI

Pearson

Editorial Director: Jeanne Zalesky
Content Development: Margy Kuntz, Andrea Giancoli
Project Managers: Cynthia Rae Abbott, Elisa Mandelbaum, Francesca Monaco, Karen Misler, Rebecca Dunn
Production Vendor: CodeMantra
Interior Composition: Preparé Italia, Battipaglia (SA), Italy
Copyeditor: Joanna Dinesmore
Proofreaders: Andrea Giancoli, Carol Reitz, and Clare Romeo
Art House: Lachina Creative
Design Managers: Mark Ong, Derek Bacchus, Emily Friel, SPi Global
Rights & Permissions Manager: Ben Ferrini
SPi Global Photo Researcher: Eric Schrader and Mary Teresa Giancoli
Manufacturing Buyer: Stacey Wienberger

Copyright © 2020, 2008, 2000, 1989, 1984 by Douglas C. Giancoli. Published by Pearson Education, Inc. All Rights Reserved. Printed in the United States of America. This publication is protected by copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise. For information regarding permissions, request forms and the appropriate contacts within the Pearson Education Global Rights & Permissions department, please visit www.pearsoned.com/permissions/.

Photo credits appear on page A–77, which constitutes a continuation of this copyright page.

PEARSON, ALWAYS LEARNING and Mastering™ Physics are exclusive trademarks in the U.S. and/or other countries owned by Pearson Education, Inc. or its affiliates.

Unless otherwise indicated herein, any third-party trademarks that may appear in this work are the property of their respective owners and any references to third-party trademarks, logos or other trade dress are for demonstrative or descriptive purposes only. Such references are not intended to imply any sponsorship, endorsement, authorization, or promotion of Pearson's products by the owners of such marks, or any relationship between the owner and Pearson Education, Inc. or its affiliates, authors, licensees or distributors.

Library of Congress Cataloging-in-Publication Data

Giancoli, Douglas C., author.
Title: Physics for scientists & engineers with modern physics/Douglas C. Giancoli.
Other titles: Physics for scientists and engineers with modern physics
Description: Fifth edition. | Upper Saddle River, N.J. : Pearson Education, Inc., [2019] | Includes bibliographical references and index.
Contents: Introduction, measurement, estimating — Describing motion: kinematics in one dimension — Kinematics in two or three dimensions; vectors — Dynamics: Newton's laws of motion — Using Newton's laws: friction, circular motion, drag forces — Gravitation and Newton's synthesis — Work and energy — Conservation of energy — Linear momentum — Rotational motion — Angular momentum; general rotation — Static equilibrium; elasticity and fracture — Fluids — Oscillations — Wave motion — Sound — Temperature, thermal expansion, and the ideal gas law — Kinetic theory of gases — Heat and the first law of thermodynamics — Second law of thermodynamics — Electric charge and electric field — Gauss's law — Electric potential — Capacitance, dielectrics, electric energy storage — Electric currents and resistance — DC circuits — Magnetism — Sources of magnetic field — Electromagnetic induction and Faraday's law — Inductance, electromagnetic oscillations, and AC circuits — Maxwell's equations and electromagnetic waves — Light: reflection and refraction — Lenses and optical instruments — The wave nature of light: interference and polarization — Diffraction — The special theory of relativity — Early quantum theory and models of the atom — Quantum mechanics — Quantum mechanics of atoms — Molecules and solids — Nuclear physics and radioactivity — Nuclear energy; effects and uses of radiation — Elementary particles — Astrophysics and cosmology.
Identifiers: LCCN 2019015435 | ISBN 9780134378053 (v.1) | ISBN 0134378059 (v.1)
Subjects: LCSH: Physics--Textbooks.
Classification: LCC QC21.3 .G539 2019 | DDC 530--dc23 LC record available at https://lccn.loc.gov/2019015435

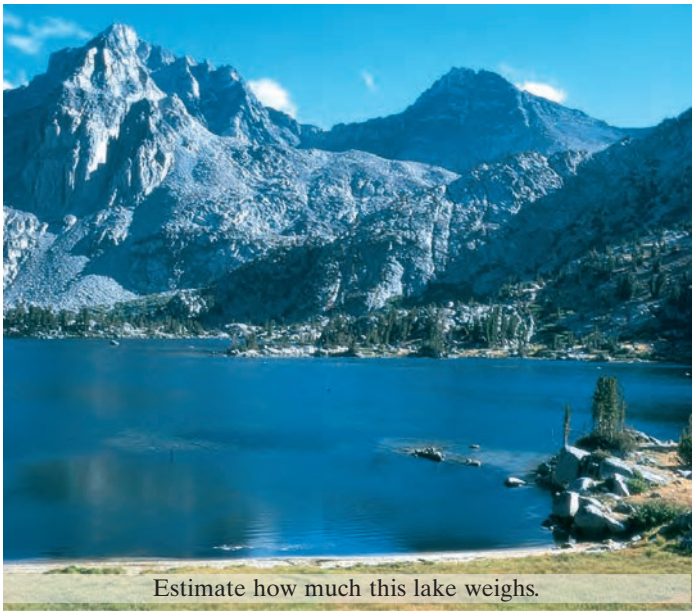
ISBN 10: 0-321-99227-X; ISBN 13: 978-0-32-199227-7 (Student Edition)
ISBN 10: 0-134-37806-7; ISBN 13: 978-13-437806-0 (Classic Student Edition)
ISBN 10: 0-134-37808-3; ISBN 13: 978-0-13-437808-4 (Looseleaf Edition)

Pearson

www.pearson.com

Contents

Applications List xii
Preface xvi
To Students xx
Use of Color xxi



Estimate how much this lake weighs.

1 Introduction, Measurement, Estimating 1

1–1 How Science Works 2
1–2 Models, Theories, and Laws 3
1–3 Measurement and Uncertainty; Significant Figures 3
1–4 Units, Standards, and the SI System 6
1–5 Converting Units 9
1–6 Order of Magnitude: Rapid Estimating 11
*1–7 Dimensions and Dimensional Analysis 14
Questions, MisConceptions, Problems 15–19

2 Describing Motion: Kinematics in One Dimension 20

2–1 Reference Frames and Displacement 21
2–2 Average Velocity 22
2–3 Instantaneous Velocity 24
2–4 Acceleration 27
2–5 Motion at Constant Acceleration 30
2–6 Solving Problems 33
2–7 Freely Falling Objects 37
*2–8 Variable Acceleration; Integral Calculus 43
Questions, MisConceptions, Problems 45–53

3 Kinematics in Two or Three Dimensions; Vectors 54

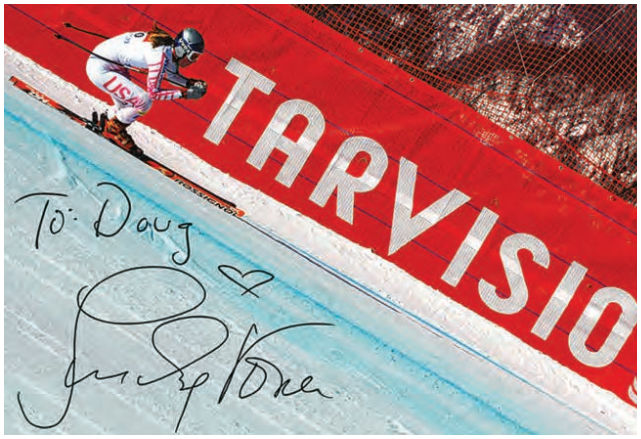
3–1 Vectors and Scalars 55
3–2 Addition of Vectors—Graphical Methods 55
3–3 Subtraction of Vectors, and Multiplication of a Vector by a Scalar 57
3–4 Adding Vectors by Components 58
3–5 Unit Vectors 62
3–6 Vector Kinematics 62
3–7 Projectile Motion 65
3–8 Solving Problems Involving Projectile Motion 67
3–9 Relative Velocity 73
Questions, MisConceptions, Problems 76–84

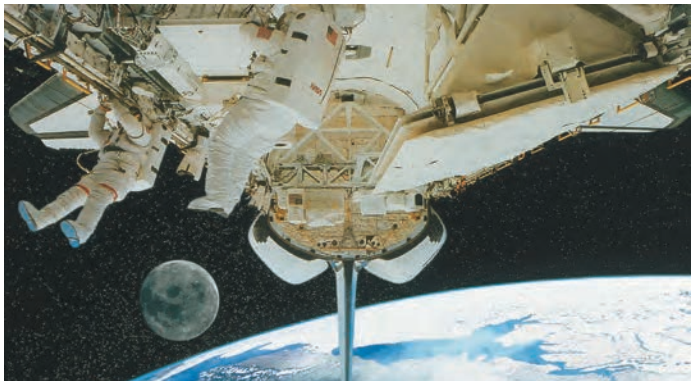
4 Dynamics: Newton's Laws of Motion 85

4–1 Force 86
4–2 Newton's First Law of Motion 86
4–3 Mass 88
4–4 Newton's Second Law of Motion 88
4–5 Newton's Third Law of Motion 91
4–6 Weight—the Force of Gravity; and the Normal Force 94
4–7 Solving Problems with Newton's Laws: Free-Body Diagrams 97
4–8 Problem Solving—A General Approach 104
Questions, MisConceptions, Problems 105–115

5 Using Newton's Laws: Friction, Circular Motion, Drag Forces 116

5–1 Using Newton's Laws with Friction 117
5–2 Uniform Circular Motion—Kinematics 123
5–3 Dynamics of Uniform Circular Motion 126
5–4 Highway Curves: Banked and Unbanked 130
5–5 Nonuniform Circular Motion 133
*5–6 Velocity-Dependent Forces: Drag and Terminal Velocity 134
Questions, MisConceptions, Problems 136–144





6 Gravitation and Newton's Synthesis 145

6–1 Newton's Law of Universal Gravitation 146
6–2 Vector Form of Newton's Law of Universal Gravitation 149
6–3 Gravity Near the Earth's Surface 149
6–4 Satellites and "Weightlessness" 152
6–5 Planets, Kepler's Laws, and Newton's Synthesis 155
6–6 Moon Rises an Hour Later Each Day 161
6–7 Types of Forces in Nature 161
*6–8 Gravitational Field 162
*6–9 Principle of Equivalence; Curvature of Space; Black Holes 163
Questions, MisConceptions, Problems 165–171

7 Work and Energy 172

7–1 Work Done by a Constant Force 173
7–2 Scalar Product of Two Vectors 176
7–3 Work Done by a Varying Force 177
7–4 Kinetic Energy and the Work-Energy Principle 181
Questions, MisConceptions, Problems 186–193

8 Conservation of Energy 194

8–1 Conservative and Nonconservative Forces 195
8–2 Potential Energy 197
8–3 Mechanical Energy and Its Conservation 200
8–4 Problem Solving Using Conservation of Mechanical Energy 201
8–5 The Law of Conservation of Energy 207
8–6 Energy Conservation with Dissipative Forces: Solving Problems 208
8–7 Gravitational Potential Energy and Escape Velocity 210
8–8 Power 213
8–9 Potential Energy Diagrams; Stable and Unstable Equilibrium 215
*8–10 Gravitational Assist (Slingshot) 216
Questions, MisConceptions, Problems 218–226

9 Linear Momentum 227

9–1 Momentum and Its Relation to Force 228
9–2 Conservation of Momentum 230
9–3 Collisions and Impulse 234
9–4 Conservation of Energy and Momentum in Collisions 235
9–5 Elastic Collisions in One Dimension 236
9–6 Inelastic Collisions 239
9–7 Collisions in 2 or 3 Dimensions 241
9–8 Center of Mass (CM) 244
9–9 Center of Mass and Translational Motion 248
*9–10 Systems of Variable Mass; Rocket Propulsion 251
Questions, MisConceptions, Problems 254–263

10 Rotational Motion 264

10–1 Angular Quantities 265
10–2 Vector Nature of Angular Quantities 270
10–3 Constant Angular Acceleration 270
10–4 Torque 271
10–5 Rotational Dynamics; Torque and Rotational Inertia 274
10–6 Solving Problems in Rotational Dynamics 276
10–7 Determining Moments of Inertia 279
10–8 Rotational Kinetic Energy 281
10–9 Rotation plus Translational Motion; Rolling 283
*10–10 Why Does a Rolling Sphere Slow Down? 289
Questions, MisConceptions, Problems 291–301

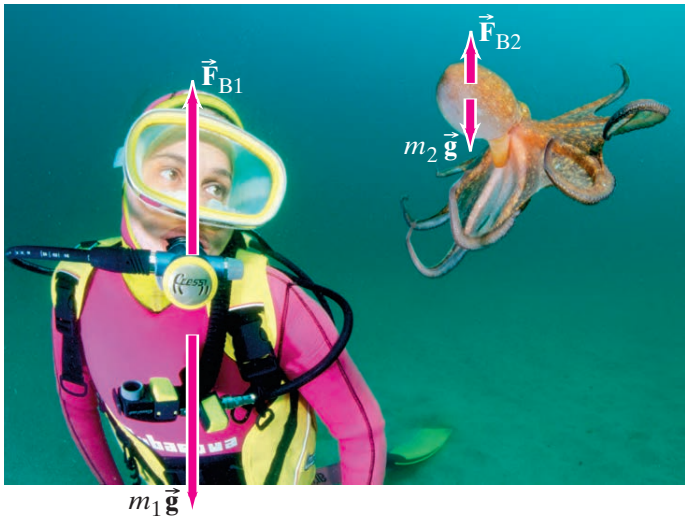
11 Angular Momentum; General Rotation 302

11–1 Angular Momentum—Objects Rotating About a Fixed Axis 303
11–2 Vector Cross Product; Torque as a Vector 307
11–3 Angular Momentum of a Particle 309
11–4 Angular Momentum and Torque for a System of Particles; General Motion 310
11–5 Angular Momentum and Torque for a Rigid Object 312
11–6 Conservation of Angular Momentum 315
*11–7 The Spinning Top and Gyroscope 317
11–8 Rotating Frames of Reference; Inertial Forces 318
*11–9 The Coriolis Effect 319
Questions, MisConceptions, Problems 322–330



12 Static Equilibrium; Elasticity and Fracture 331

12–1 The Conditions for Equilibrium 332
12–2 Solving Statics Problems 334
*12–3 Applications to Muscles and Joints 339
12–4 Stability and Balance 341
12–5 Elasticity; Stress and Strain 342
12–6 Fracture 345
*12–7 Trusses and Bridges 347
*12–8 Arches and Domes 350
Questions, MisConceptions, Problems 353–364



13 Fluids 365

13–1 Phases of Matter 366
13–2 Density and Specific Gravity 366
13–3 Pressure in Fluids 367
13–4 Atmospheric Pressure and Gauge Pressure 371
13–5 Pascal's Principle 371
13–6 Measurement of Pressure; Gauges and the Barometer 372
13–7 Buoyancy and Archimedes' Principle 374
13–8 Fluids in Motion; Flow Rate and the Equation of Continuity 378
13–9 Bernoulli's Equation 380
13–10 Applications of Bernoulli's Principle: Torricelli, Airplanes, Baseballs, Blood Flow 382
13–11 Viscosity 385
*13–12 Flow in Tubes: Poiseuille's Equation, Blood Flow 385
*13–13 Surface Tension and Capillarity 386
*13–14 Pumps, and the Heart 388
Questions, MisConceptions, Problems 390–398

14 Oscillations 399

14–1 Oscillations of a Spring 400
14–2 Simple Harmonic Motion 402
14–3 Energy in the Simple Harmonic Oscillator 408
14–4 Simple Harmonic Motion Related to Uniform Circular Motion 410
14–5 The Simple Pendulum 411
*14–6 The Physical Pendulum and the Torsion Pendulum 412
14–7 Damped Harmonic Motion 414
14–8 Forced Oscillations; Resonance 417
Questions, MisConceptions, Problems 420–427

15 Wave Motion 428

15–1 Characteristics of Wave Motion 429
15–2 Types of Waves: Transverse and Longitudinal 431
15–3 Energy Transported by Waves 435
15–4 Mathematical Representation of a Traveling Wave 437
*15–5 The Wave Equation 440
15–6 The Principle of Superposition 441
15–7 Reflection and Transmission 443
15–8 Interference 444
15–9 Standing Waves; Resonance 446
15–10 Refraction 449
15–11 Diffraction 450
Questions, MisConceptions, Problems 452–459

16 Sound 460

16–1 Characteristics of Sound 461
16–2 Mathematical Representation of Longitudinal Waves 462
16–3 Intensity of Sound: Decibels 464
16–4 Sources of Sound: Vibrating Strings and Air Columns 467
*16–5 Quality of Sound, and Noise; Superposition 472
16–6 Interference of Sound Waves; Beats 473
16–7 Doppler Effect 476
*16–8 Shock Waves and the Sonic Boom 480
*16–9 Applications: Sonar, Ultrasound, and Medical Imaging 481
Questions, MisConceptions, Problems 484–491

17 Temperature, Thermal Expansion, and the Ideal Gas Law 492

17–1 Atomic Theory of Matter 493
17–2 Temperature and Thermometers 495
17–3 Thermal Equilibrium and the Zeroth Law of Thermodynamics 497
17–4 Thermal Expansion 497
*17–5 Thermal Stresses 501
17–6 The Gas Laws and Absolute Temperature 502
17–7 The Ideal Gas Law 503
17–8 Problem Solving with the Ideal Gas Law 504
17–9 Ideal Gas Law in Terms of Molecules: Avogadro's Number 506
*17–10 Ideal Gas Temperature Scale—a Standard 507
Questions, MisConceptions, Problems 509–515

18 Kinetic Theory of Gases 516

18–1 The Ideal Gas Law and the Molecular Interpretation of Temperature 516
18–2 Distribution of Molecular Speeds 520
18–3 Real Gases and Changes of Phase 522
18–4 Vapor Pressure and Humidity 524
18–5 Temperature Decrease of Boiling Water with Altitude 526
18–6 Van der Waals Equation of State 527
18–7 Mean Free Path 528
18–8 Diffusion 530
Questions, MisConceptions, Problems 532–537

19 Heat and the First Law of Thermodynamics 538

19–1 Heat as Energy Transfer 539
19–2 Internal Energy 540
19–3 Specific Heat 541
19–4 Calorimetry—Solving Problems 542
19–5 Latent Heat 545
19–6 The First Law of Thermodynamics 549
19–7 Thermodynamic Processes and the First Law 551
19–8 Molar Specific Heats for Gases, and the Equipartition of Energy 556
19–9 Adiabatic Expansion of a Gas 559
19–10 Heat Transfer: Conduction, Convection, Radiation 560
Questions, MisConceptions, Problems 568–575



20 Second Law of Thermodynamics 576

20–1 The Second Law of Thermodynamics—Introduction 577
20–2 Heat Engines 578
20–3 The Carnot Engine; Reversible and Irreversible Processes 580
20–4 Refrigerators, Air Conditioners, and Heat Pumps 584
20–5 Entropy 587
20–6 Entropy and the Second Law of Thermodynamics 590
20–7 Order to Disorder 593
20–8 Unavailability of Energy; Heat Death 594
20–9 Statistical Interpretation of Entropy and the Second Law 595
*20–10 Thermodynamic Temperature; Third Law of Thermodynamics 597
20–11 Thermal Pollution, Global Warming, and Energy Resources 598
Questions, MisConceptions, Problems 601–608





21 Electric Charge and Electric Field 609

21–1 Static Electricity; Electric Charge and Its Conservation 610
21–2 Electric Charge in the Atom 611
21–3 Insulators and Conductors 611
21–4 Induced Charge; the Electroscope 612
21–5 Coulomb's Law 613
21–6 The Electric Field 618
21–7 Electric Field Calculations for Continuous Charge Distributions 622
21–8 Field Lines 626
21–9 Electric Fields and Conductors 627
21–10 Motion of a Charged Particle in an Electric Field 628
21–11 Electric Dipoles 629
*21–12 Electric Forces in Molecular Biology: DNA Structure and Replication 631
Questions, MisConceptions, Problems 634–642

22 Gauss's Law 643

22–1 Electric Flux 644
22–2 Gauss's Law 645
22–3 Applications of Gauss's Law 647
*22–4 Experimental Basis of Gauss's and Coulomb's Laws 652
Questions, MisConceptions, Problems 653–659

23 Electric Potential 660

23–1 Electric Potential Energy and Potential Difference 661
23–2 Relation between Electric Potential and Electric Field 664
23–3 Electric Potential Due to Point Charges 666
23–4 Potential Due to Any Charge Distribution 669
23–5 Equipotential Lines and Surfaces 670
23–6 Potential Due to Electric Dipole; Dipole Moment 671
23–7 $\vec{\mathbf{E}}$ Determined from V 672
23–8 Electrostatic Potential Energy; the Electron Volt 674
23–9 Digital; Binary Numbers; Signal Voltage 676
*23–10 TV and Computer Monitors 679
*23–11 Electrocardiogram (ECG or EKG) 682
Questions, MisConceptions, Problems 684–691

24 Capacitance, Dielectrics, Electric Energy Storage 692

24–1 Capacitors 692
24–2 Determination of Capacitance 694
24–3 Capacitors in Series and Parallel 698
24–4 Storage of Electric Energy 700
24–5 Dielectrics 703
*24–6 Molecular Description of Dielectrics 706
Questions, MisConceptions, Problems 708–716

25 Electric Current and Resistance 717

25–1 The Electric Battery 718
25–2 Electric Current 720
25–3 Ohm's Law: Resistance and Resistors 722
25–4 Resistivity 724
25–5 Electric Power 726
25–6 Power in Household Circuits 729
25–7 Alternating Current 730
25–8 Microscopic View of Electric Current 732
*25–9 Superconductivity 735
*25–10 Electrical Conduction in the Human Nervous System 736
Questions, MisConceptions, Problems 739–746



26 DC Circuits 747

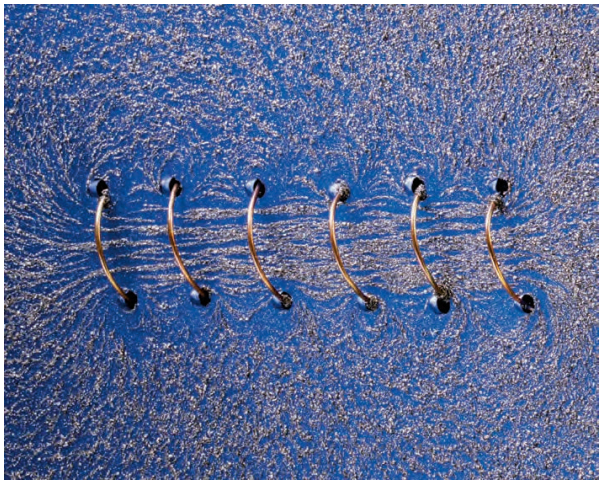
26–1 EMF and Terminal Voltage 748
26–2 Resistors in Series and in Parallel 749
26–3 Kirchhoff's Rules 754
26–4 EMFs in Series and in Parallel; Charging a Battery 757
26–5 *RC* Circuits—Resistor and Capacitor in Series 759
26–6 Electric Hazards and Safety 764
26–7 Ammeters and Voltmeters—Measurement Affects Quantity Measured 767
Questions, MisConceptions, Problems 771–781

27 Magnetism 782

27–1 Magnets and Magnetic Fields 782
27–2 Electric Currents Produce Magnetic Fields 785
27–3 Force on an Electric Current in a Magnetic Field; Definition of $\vec{\mathbf{B}}$ 786
27–4 Force on an Electric Charge Moving in a Magnetic Field 788
27–5 Torque on a Current Loop; Magnetic Dipole Moment 793
27–6 Applications: Motors, Loudspeakers, Galvanometers 795
27–7 Discovery and Properties of the Electron 797
27–8 The Hall Effect 799
27–9 Mass Spectrometer 800
Questions, MisConceptions, Problems 802–810

28 Sources of Magnetic Field 811

28–1 Magnetic Field Due to a Straight Wire 812
28–2 Force between Two Parallel Wires 813
28–3 Definitions of the Ampere and the Coulomb 814
28–4 Ampère's Law 815
28–5 Magnetic Field of a Solenoid and a Toroid 819
28–6 Biot-Savart Law 821
28–7 Magnetic Field Due to a Single Moving Charge 824
28–8 Magnetic Materials—Ferromagnetism 824
28–9 Electromagnets and Solenoids—Applications 826
28–10 Magnetic Fields in Magnetic Materials; Hysteresis 827
*28–11 Paramagnetism and Diamagnetism 828
Questions, MisConceptions, Problems 830–837



29 Electromagnetic Induction and Faraday's Law 838

29–1 Induced EMF 839
29–2 Faraday's Law of Induction; Lenz's Law 840
29–3 EMF Induced in a Moving Conductor 845
29–4 Electric Generators 846
29–5 Back EMF and Counter Torque; Eddy Currents 848
29–6 Transformers and Transmission of Power 851
29–7 A Changing Magnetic Flux Produces an Electric Field 854
*29–8 Information Storage: Magnetic and Semiconductor 856
*29–9 Applications of Induction: Microphone, Seismograph, GFCI 858
Questions, MisConceptions, Problems 860–868

30 Inductance, Electromagnetic Oscillations, and AC Circuits 869

30–1 Mutual Inductance 870
30–2 Self-Inductance; Inductors 872
30–3 Energy Stored in a Magnetic Field 874
30–4 *LR* Circuits 875
30–5 *LC* Circuits and Electromagnetic Oscillations 877
30–6 *LC* Oscillations with Resistance (*LRC* Circuit) 880
30–7 AC Circuits and Reactance 881
30–8 *LRC* Series AC Circuit; Phasor Diagrams 885
30–9 Resonance in AC Circuits 887
30–10 Impedance Matching 888
*30–11 Three-Phase AC 889
Questions, MisConceptions, Problems 890–897

31 Maxwell's Equations and Electromagnetic Waves 898

31–1 Changing Electric Fields Produce Magnetic Fields; Displacement Current 899
31–2 Gauss's Law for Magnetism 902
31–3 Maxwell's Equations 903
31–4 Production of Electromagnetic Waves 903
31–5 Electromagnetic Waves, and Their Speed, Derived from Maxwell's Equations 905
31–6 Light as an Electromagnetic Wave and the Electromagnetic Spectrum 909
31–7 Measuring the Speed of Light 912
31–8 Energy in EM Waves; the Poynting Vector 913
31–9 Radiation Pressure 915
31–10 Radio and Television; Wireless Communication 917
Questions, MisConceptions, Problems 921–925



32 Light: Reflection and Refraction 926

32–1 The Ray Model of Light 927
32–2 Reflection; Image Formation by a Plane Mirror 927
32–3 Formation of Images by Spherical Mirrors 931
32–4 Seeing Yourself in a Magnifying Mirror (Concave) 936
32–5 Convex (Rearview) Mirrors 938
32–6 Index of Refraction 939
32–7 Refraction: Snell's Law 939
32–8 The Visible Spectrum and Dispersion 941
32–9 Total Internal Reflection; Fiber Optics 943
*32–10 Refraction at a Spherical Surface 946
Questions, MisConceptions, Problems 949–957

33 Lenses and Optical Instruments 958

33–1 Thin Lenses; Ray Tracing and Focal Length 959
33–2 The Thin Lens Equation 962
33–3 Combinations of Lenses 966
33–4 Lensmaker's Equation 968
33–5 Cameras: Film and Digital 970
33–6 The Human Eye; Corrective Lenses 975
33–7 Magnifying Glass 979
33–8 Telescopes 980
33–9 Compound Microscope 983
33–10 Aberrations of Lenses and Mirrors 984
Questions, MisConceptions, Problems 986–994

34 The Wave Nature of Light: Interference and Polarization 995

34–1 Waves vs. Particles; Huygens' Principle and Diffraction 996
34–2 Huygens' Principle and the Law of Refraction; Mirages 997
34–3 Interference–Young's Double-Slit Experiment 998
*34–4 Intensity in the Double-Slit Interference Pattern 1002
34–5 Interference in Thin Films 1004
34–6 Michelson Interferometer 1010
34–7 Polarization 1010
*34–8 Liquid Crystal Displays (LCD) 1014
*34–9 Scattering of Light by the Atmosphere 1015
34–10 Brightness: Lumens and Luminous Intensity 1016
*34–11 Efficiency of Lightbulbs 1016
Questions, MisConceptions, Problems 1018–1024

35 Diffraction 1025

35–1 Diffraction by a Single Slit or Disk 1026
*35–2 Intensity in Single-Slit Diffraction Pattern 1028
*35–3 Diffraction in the Double-Slit Experiment 1031
35–4 Interference vs. Diffraction 1033
35–5 Limits of Resolution; Circular Apertures 1033
35–6 Resolution of Telescopes and Microscopes; the λ Limit 1035
35–7 Resolution of the Human Eye and Useful Magnification 1037
35–8 Diffraction Grating 1037
35–9 The Spectrometer and Spectroscopy 1040
*35–10 Peak Widths and Resolving Power for a Diffraction Grating 1041
35–11 X-Rays and X-Ray Diffraction 1043
*35–12 X-Ray Imaging and Computed Tomography (CT Scan) 1045
*35–13 Specialty Microscopes and Contrast 1048
Questions, MisConceptions, Problems 1049–1054



36 The Special Theory of Relativity 1055

36–1 Galilean–Newtonian Relativity 1056
36–2 The Michelson–Morley Experiment 1058
36–3 Postulates of the Special Theory of Relativity 1061
36–4 Simultaneity 1062
36–5 Time Dilation and the Twin Paradox 1064
36–6 Length Contraction 1070
36–7 Four-Dimensional Space–Time 1072
36–8 Galilean and Lorentz Transformations 1072
36–9 Relativistic Momentum 1077
36–10 The Ultimate Speed 1079
36–11 $E = mc^2$; Mass and Energy 1080
36–12 Doppler Shift for Light 1085
36–13 The Impact of Special Relativity 1086
Questions, MisConceptions, Problems 1088–1094

37 Early Quantum Theory and Models of the Atom 1095

37–1 Blackbody Radiation; Planck's Quantum Hypothesis 1096
37–2 Photon Theory of Light and the Photoelectric Effect 1098
37–3 Energy, Mass, and Momentum of a Photon 1101
37–4 Compton Effect 1102
37–5 Photon Interactions; Pair Production 1104
37–6 Wave–Particle Duality; the Principle of Complementarity 1105
37–7 Wave Nature of Matter 1106
37–8 Electron Microscopes 1108
37–9 Early Models of the Atom 1110
37–10 Atomic Spectra: Key to the Structure of the Atom 1111
37–11 The Bohr Model 1113
37–12 de Broglie's Hypothesis Applied to Atoms 1120
Questions, MisConceptions, Problems 1121–1127

Appendices

A Mathematical Formulas A–1
B Derivatives and Integrals A–6
C Numerical Integration A–8
D More on Dimensional Analysis A–12
E Gravitational Force Due to a Spherical Mass Distribution A–13
F Differential Form of Maxwell's Equations A–16
G Selected Isotopes A–18

Answers to Odd-Numbered Problems A–23
Index A–47
Photo Credits A–77

38 Quantum Mechanics 1128

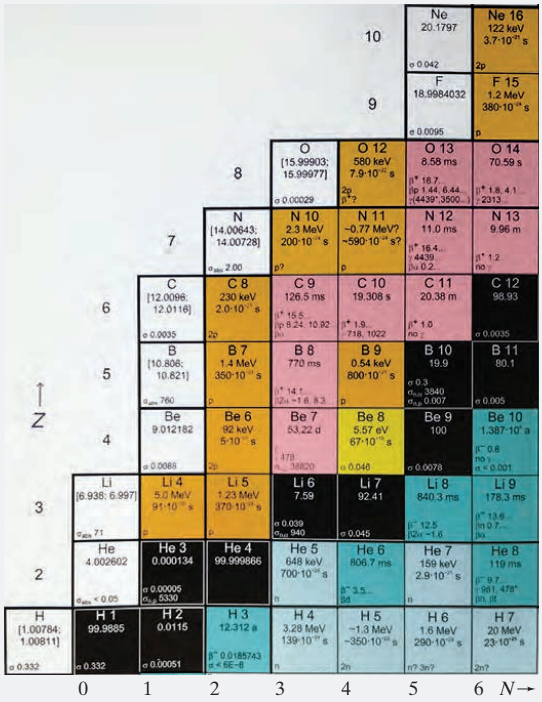
38–1 Quantum Mechanics—A New Theory 1129
38–2 The Wave Function and Its Interpretation; the Double-Slit Experiment 1129
38–3 The Heisenberg Uncertainty Principle 1131
38–4 Philosophic Implications; Probability Versus Determinism 1135
38–5 The Schrödinger Equation in One Dimension—Time-Independent Form 1136
*38–6 Time-Dependent Schrödinger Equation 1138
38–7 Free Particles; Plane Waves and Wave Packets 1140
38–8 Particle in an Infinitely Deep Square Well Potential (a Rigid Box) 1142
38–9 Finite Potential Well 1147
38–10 Tunneling through a Barrier 1149
Questions, MisConceptions, Problems 1152–1157

39 Quantum Mechanics of Atoms 1158

39–1 Quantum-Mechanical View of Atoms 1159
39–2 Hydrogen Atom: Schrödinger Equation and Quantum Numbers 1159
39–3 Hydrogen Atom Wave Functions 1163
39–4 Multielectron Atoms; the Exclusion Principle 1166
39–5 Periodic Table of Elements 1167
39–6 X-Ray Spectra and Atomic Number 1169
*39–7 Magnetic Dipole Moment; Total Angular Momentum 1171
39–8 Fluorescence and Phosphorescence 1174
39–9 Lasers 1175
*39–10 Holography 1178
Questions, MisConceptions, Problems 1180–1185

40 Molecules and Solids 1186

40–1 Bonding in Molecules 1187
40–2 Potential-Energy Diagrams for Molecules 1189
40–3 Weak (van der Waals) Bonds 1192
40–4 Molecular Spectra 1196
40–5 Bonding in Solids 1202
40–6 Free-Electron Theory of Metals; Fermi Energy 1203
40–7 Band Theory of Solids 1208
40–8 Semiconductors and Doping 1210
40–9 Semiconductor Diodes, LEDs, OLEDs 1212
40–10 Transistors: Bipolar and MOSFETs 1218
40–11 Integrated Circuits, 10-nm Technology 1219
Questions, MisConceptions, Problems 1220–1225



41 Nuclear Physics and Radioactivity 1226

41–1 Structure and Properties of the Nucleus 1227
41–2 Binding Energy and Nuclear Forces 1230
41–3 Radioactivity 1233
41–4 Alpha Decay 1234
41–5 Beta Decay 1237
41–6 Gamma Decay 1239
41–7 Conservation of Nucleon Number and Other Conservation Laws 1240
41–8 Half-Life and Rate of Decay 1240
41–9 Decay Series 1245
41–10 Radioactive Dating 1246
41–11 Detection of Particles 1248
Questions, MisConceptions, Problems 1250–1255

42 Nuclear Energy; Effects and Uses of Radiation 1256

42–1 Nuclear Reactions and the Transmutation of Elements 1257
42–2 Cross Section 1260
42–3 Nuclear Fission; Nuclear Reactors 1261
42–4 Nuclear Fusion 1266
42–5 Passage of Radiation Through Matter; Biological Damage 1271
42–6 Measurement of Radiation—Dosimetry 1272
*42–7 Radiation Therapy 1276
*42–8 Tracers in Research and Medicine 1277
*42–9 Emission Tomography: PET and SPECT 1278
*42–10 Nuclear Magnetic Resonance (NMR); Magnetic Resonance Imaging (MRI) 1279
Questions, MisConceptions, Problems 1283–1288

43 Elementary Particles 1289

43–1 High-Energy Particles and Accelerators 1290
43–2 Beginnings of Elementary Particle Physics—Particle Exchange 1296
43–3 Particles and Antiparticles 1299
43–4 Particle Interactions and Conservation Laws 1300
43–5 Neutrinos 1302
43–6 Particle Classification 1304
43–7 Particle Stability and Resonances 1306
43–8 Strangeness? Charm? Towards a New Model 1307
43–9 Quarks 1308
43–10 The Standard Model: QCD and Electroweak Theory 1311
43–11 Grand Unified Theories 1314
43–12 Strings and Supersymmetry 1317
Questions, MisConceptions, Problems 1318–1231

44 Astrophysics and Cosmology 1322

44–1 Stars and Galaxies 1323
44–2 Stellar Evolution: Birth and Death of Stars, Nucleosynthesis 1326
44–3 Distance Measurements 1332
44–4 General Relativity: Gravity and the Curvature of Space 1334
44–5 The Expanding Universe: Redshift and Hubble's Law 1338
44–6 The Big Bang and the Cosmic Microwave Background 1342
44–7 The Standard Cosmological Model: Early History of the Universe 1345
44–8 Inflation: Explaining Flatness, Uniformity, and Structure 1348
44–9 Dark Matter and Dark Energy 1350
44–10 Large-Scale Structure of the Universe 1353
44–11 Gravitational Waves—LIGO 1354
44–12 Finally . . . 1354
Questions, MisConceptions, Problems 1356–1360

Appendices

A Mathematical Formulas A–1
B Derivatives and Integrals A–6
C Numerical Integration A–8
D More on Dimensional Analysis A–12
E Gravitational Force Due to a Spherical Mass Distribution A–13
F Differential Form of Maxwell's Equations A–16
G Selected Isotopes A–18

Answers to Odd-Numbered Problems A–23
Index
Photo Credits

Applications (Selected) to Medicine and Biology and to Engineering, Environment, Everyday Life, Etc.

(Entries with a star * include material new to this edition)

Chapter 1
Viruses attack cell 7
Heartbeats in a lifetime 12
Number of nucleons in human body 17
Lung capacity 19
Building collapse 2, 332, 346–7
The 8000-m peaks 9
Making estimates: volume of a lake 11
Page thickness 12
Building height by triangulation 12
Earth radius estimate 13, 18
Fermi estimates 13
Particulate pollution 18
Global positioning satellite 18
Computer chips 18

Chapter 2
Airport runway design 32
*Car air bag inflation time 32, 254
Car braking distance 35, 183
CD bit size, bit rate, playing time 48, 53
*Baseball 49, 82, 83, 84, 172
Basketball 50, 83, 109
Golf putt, uphill or down 52
Rapid transit system 53

Chapter 3
Helicopter supply drop 54, 72, 83
*Sports 54, 65, 69, 71, 76, 77, 79, 80, 81, 82, 83, 84
Kicked football 69, 71
Truck escape lane 79, 110
Golf on the Moon 82
Extreme sports 83

Chapter 4
How we can walk 92
Whiplash 106
*Force heart exerts 107
Rocket 85, 92, 108, 233, 252, 395
Skater pushoff 91
What force accelerates a car 92
You weigh less in a falling elevator 96
Hockey 98
Elevator, discomfort 101, 108
Mechanical advantage, pulley 102, 188
Accelerometer 102
*Sports 106, 107, 108, 109, 110, 112
*Bear sling 106, 355
Tug of war 106, 255
Car accident "*g*'s" 107
Optical tweezers 108, 916, 923
*Tightrope walker 109
Basketball shot 109
Mountain climbers 110, 114, 115, 193, 370
City planning, cars on hill 112
Bicycling 112, 114
Supermarket ramp design 113
Doomsday asteroid 114, 262
*Car stuck in mud 115

Chapter 5
Centrifugation 126
Skiing 116, 121, 136
Push or pull a sled? 120
Skier speed in air vs. on snow 121
*Simulating gravity 126, 136, 141, 168, 171
*Uranium enrichment, reactor, bomb 126
Ferris wheel 129
Avoid skidding on a curve 130–2
Banked highway curves 132
Cross-country skiing friction 136
Rotating space station 136, 141, 168
*Rotor ride 137, 143
Airplane bank/turn 137, 144
Roller coaster upside down 141
Car flying up off road 141
Rock climbing friction 143

Chapter 6
Weightlessness 154–5
*Astronauts in orbit 145, 155, 165
Gravity on tall peaks 150
Oil and mineral exploration 150, 165, 167
Satellites, spacecraft 145, 152–5, 168, 169
Geostationary satellites 153
Free fall, for athletes 155
Planets 155–8, 167
Determining the Sun's mass 158
Planets around other stars 158, 250, 262
*Ocean tides 159, 165, 170
Lagrange point 160
*Moon's orbit, periods, phases, diagram 161, 169
*Eclipses 161
Curved space 163–4
Black holes 164, 167
White dwarfs 167
Comets, asteroids, moons 168, 169, 171
GPS 169
Milky Way Galaxy 171

Chapter 7
Baseball pitch 172
Car stopping distance $\propto v^2$ 183
Lever 187, 334
*Pulley 188
Jet catapults 189
Bicycle, sprockets (teeth) 192, 299
Climbing rope stretch 193

Chapter 8
Stair-climbing power 213
*ATP 216
Hike *over* logs 218
Pole vault 194, 203–4, (189)
Downhill ski runs 194
Roller coaster 198, 202, 209
Escape velocity from Earth or Moon 212
Power needs of car 214
Efficiency of engine 215
*Gravitational assist 216–7, 224, 263
High jump 220
Bungee jump 221
Lunar module landing 222
Escape velocity from solar system 223
Ski jump 225
Long jump 225

Chapter 9
Impulse in fall: break a leg? 257
Billiard balls 227, 230, 237, 242
Tennis serve 229, 234
Rocket propulsion 233, 252, 395
Rifle recoil 233
Karate blow 235
Nuclear reactors 238
Nuclear collisions 238, 239, 241, 243
Ballistic pendulum, speed measured 240
Distant planet discovery 250, 262
Conveyor belt 253
Car crashworthiness 261
Asteroid danger 262
Force wind exerts 263
Bowling 263

Chapter 10
Acuity of bird's eye 266
Centrifuge 271
*Biceps, triceps, torque 273, 295, 339
Situps 291
Fast mammal 291
Rotating carnival rides 264, 267, 268
Tire iron extension 272
Flywheel, energy 282, 301
Yo-yo 287
Braking forces on a car 288–9
Bicycle odometer 291
Tightrope walking 291
*Total solar eclipse 293
Wrench torque 294
Hammer throw 296
CD rotation frequency 298
Bicycle gears 299
Cue stick, ball roll 300
*Bicycle turn angle 301

Chapter 11
Rotating skaters/divers 302, 304, 330
Neutron star collapse 305, 330
Strange spinning bike wheel 307, 314

Automobile wheel balancing 314–5
Precessing top 317–8
Gyroscope 318
Hurricanes, cyclones, typhoons 321, 394
*Anticyclonic weather 321
Precession of the equinoxes 327
SUV rollover 328
Baseball bat sweet spot 330

Chapter 12
*Forces in muscles & joints 339, 358, (273)
*What can make an athlete 339
*Forces on the spine and back pain 340
Human balance with loads 342
*Bone fracture 346, 359, 364
Buildings, statics 331–352
Lever, mechanical advantage 334
Balancing a seesaw 335
Cantilever 336
Fracture 345–7
Tragic collapse 346–7, (332)
Trusses and bridges 347–9, 363
Architecture: arches and domes 350–2
*Forces in a dome 352

Chapter 13
Pressure in cells 371
Blood flow 380, 384, 386
Human circulatory system 380
Blood loss to brain, TIA 384
*Air flow in animal burrow 384
Heart disease, artery clogging 386
Walking on water, insect 387
Heart as a pump 388–9
*Blood pressure measurement 389
Blood transfusion 395, 396
Water supply pressure 369
Atmospheric pressure decrease with elevation 370
Altitude where air pressure is half 370
Finger holds water in straw 371
Hydraulic lift 372
Hydraulic brakes 372
Pressure gauges 372–3
Barometer 373
Suction 374
Hydrometer 377
Continental drift, plate tectonics 378
*Lake level change, rock thrown overboard 378, 390
Helium balloon lift 378
Heating duct 380
Hot-water heating system flow 382
Perfume atomizer 383
Airplane wing lift 383
Sailing upwind 383
Baseball curve 384
Why smoke goes up a chimney 384
Soaps and detergents 387
Pumps 388–9, (374)
Siphon 390
Hydraulic press 393
Rocket thrust 395
Reynolds number 395
Barrel broken by thin liquid column 397

Chapter 14
Spider web oscillations 405
Human leg as pendulum 424
Shock absorbers 399, 415
Unwanted floor vibrations 406
Loudspeaker 406–7
*Pendulum clock 412, 421, 424
Geology 412, 415
Measure *g* with pendulum 412
Earthquake dampers 415
Child on a swing, resonance 417–8
Resonance damage 418
Q-value 419, 425, 896
Bungee jumper 422
*Metronome 424
Natural stride 424
Tall building sway 426

Chapter 15
Echolocation by bats, dolphins, whales 434
Water waves 428, 435
Sound wave 431, 460 *ff*
Geology 435, 452, 457
Earthquake waves 435, 437, 450, 453
Square wave 442
*Cell phone signal 451
AM and FM radio wave bending 452
Fish and fisher: internal reflection 456
Seismic reflection: oil prospecting 457
Coffee spill 457
Tsunami 459

Chapter 16
Wide range of human hearing 464
Sensitivity of the ear 467, (466)
Bats use Doppler 479
Doppler blood-flow meter 479, 491
Ultrasound medical imaging 482–3
*Doppler ultrasound imaging 483
Stringed instruments 460, 468–9
Wind instruments 460, 469–72
Piano strings 460, 468, 469
Distance from lightning, seconds 461
Autofocusing camera 462
Loudspeaker output 465
Musical scale 468
Guitar, violin 468, 469, 484, 487
Organ pipes 471–2
Tuning with beats 475–6
Doppler in weather forecasting 480
*Radar speed gun 480
Galaxy redshift 480
Sonic boom; sound barrier 481, 489
Sonar: depth in sea, Earth "soundings" 481–2, 489
Signal-to-noise ratio 486, 490, 679
Quartz oscillator clock 487
Motion sensor 489
Audio gain 490

Chapter 17
Life under ice 500–1
Molecules in one breath 507, 514
Snorkels are short 515
*Hot air balloon 492, 515
Expansion joints 495, 498, 501
Do holes expand? 499
Opening a tight lid 499
Gas tank overflow 500
Highway buckling 501
Closed jars in fires 503
Mass (weight) of air in a room 505
Cold and hot tire pressure 506
Thermostat 509
Pyrex glass 509
*Tape measure inaccuracy 510, 513
Scuba 512, 513, 514, 515
Potato chip bag puff up 513

Chapter 18
KE of molecules in cells 519
Humidity, and comfort 525–6
Chromatography 531
Diffusion in living organisms 531–2, 536
Temperature effect on chemical reactions 521
Evaporation cools 524, 548
Humidity, weather 526
*Temperature decrease of boiling water with altitude 526–7
Pressure cooker 535

Chapter 19
Working off Calories 540
Measuring Calorie content 545, 570
Evaporation and body temperature 548–9, (524)
Body heat: convection by blood 563, 574
Body's radiative heat loss 564
Room comfort: cool air, warm walls 565
Medical thermography 567
Avoid plants freezing 568
Eating snow makes you colder 571
Heat conduction to skin, blood capillaries 573
Leaf's energy absorption 575
Metabolizing fat 575
Cold tile, warm rugs 561
Heat loss through windows 562
Thermal windows (two panes) 562
How clothing insulates 562
R-values of thermal insulation 562
Ocean currents and wind 563
Convective home heating 563
Dark vs. light clothing 564
Radiation from the sun, seasons 566
Astronomy—size of a star 566
Goose down loft 568
Thermos bottle 568
Emergency blanket 568
Air parcels, weather, adiabatic lapse rate 573

Chapter 20
Biological development, evolution 594
*Trees offsetting CO_2 buildup 608
Steam engine 576, 578, 582, 606
Internal combustion engines 578–80, 583–4
Engine efficiency 582–3
Refrigerators, air conditioners 584–6, 603
Heat pump 586–7, 603
*SEER rating 587
Thermal pollution, climate 598–600
*Carbon footprint 598
Energy resources 599, 605–6
Solar, thermal, wind energy 599, 605
Diesel engine 607, (575)
Stirling cycle 607
Jet engine, Brayton cycle 607
Dehumidifier 608, (537)

Chapter 21
Inside a cell: kinetic theory plus electrostatic force 631
DNA structure, replication 631–3, 640
Static electricity 609, 610, 635, 640
Photocopiers and printers 619
Electrical shielding, safety 628

Chapter 23
Electrocardiogram (ECG) 660, 682–3, 779
Dipoles in molecular biology 672
Heart beat, depolarization process 682–3
Common voltages 10^{-4} V to 10^{8} V 663
Breakdown voltage 666
Lightning rods 666
*Supply voltage, signal voltage 676
*Digital, bits, bytes, binary numbers 676
*Analog-to-digital converter (ADC) 676
*Morse code 676
*Bit-rate, TV transmission 676, 678–9, 682
*Data compression, jpeg 677–8
*Quantization error 677
*Sampling rate, bit depth 677
*Digital-to-analog converter (DAC) 677, 780
*Bandwidth 678
*Noise, bit flips 678–9
*Digital error correction, parity bit 678
*Bit error rate 679
*Signal-to-noise ratio (S/N) 679
*TV and computer monitors 679–82
*Digital TV, pixels, subpixels 680
*Flat screens, HD 680–1
*Addressing pixels 680–1
*Data stream 681
*Active matrix, TFT, data lines 681–2
*TV refresh rate 682
Oscilloscope 682
*ASCII code 688
Photocell 689

Chapter 24
Capacitor shocks, burns 703
Heart defibrillator 703, 712, 764
Capacitor use as power backup, surge protector, memory 692, 695
Condenser microphone 695
Computer key 695
Camera flash energy 701
Electrostatic air cleaner 710
Tiny distance measurement 710
Coaxial cable 714, 818, 874, 911
*Dynamic random access memory (DRAM) 716, 857

Chapter 25
Electrical conduction in human nervous system, neurons 736–8
Action potential 737
Battery construction, terminals 718–9
*Electric cars 720, 744
Battery connections 721, 724
Loudspeaker wire thickness 725
Heating element 726–8
Resistance thermometer 726
Lightning bolt 728, (690, 716)
Household circuits, shorts 729–30
Fuses, circuit breakers 729, 766
Safety—wires getting hot 729, 764–6
Extension cord danger 730
Hair dryer 732
Strain gauge 746

Chapter 26
*Blood sugar phone app 747
Heart pacemaker 764
Electricity dangers to humans 764–6
Ventricular fibrillation 764
Two-speed fan 752–3
Car battery charging 757
*Jump-starting a car, safely 758–9
RC: sawtooth, flashers, wipers 763, 780
Hazards, electric safety 764–6
Proper grounding, plugs 765–6
Leakage current 766
Dangerous downed power line 766
Ammeters, voltmeters, ohmmeters 767–9
Meter connection, corrections 768–9, 781
*Measurement affects quantity measured 769
Voltage divider 774
Solar panel 778
Potentiometer and bridge circuits 778–9
Car battery corrosion 780
Digital-to-analog converter (DAC) 780, (677)

Chapter 27
Electromagnetic blood pump 802
Blood flow rate, Hall effect 807
Use of a compass 784
Magnetic declination 784
Maps and true north 784
Aurora borealis 792
Electric motors, DC and AC 795–6
Loudspeakers and headsets 796

Chapter 28
Coaxial cable 818, 874, 911
Solenoid switches: doorbell, car starter 826
Magnetic circuit breakers 826
Relay (magnetic) 830

Chapter 29
EM blood-flow measurement 845
Induction stove 842
Generators, power plants 846–7
Alternators, in cars 848
Motor overload 849
Eddy-current damping 850, 861
Airport metal detector 850
Transformers, power transmission 851–3
Cell phone charger 852
Car ignition system 852
*Wireless electric power transmission 854
Inductive charger 854
Magnetic information storage 856
*Semiconductor memory 857–8
*RAM, DRAM 857
*Bit-line & word-line 857
*Writing and reading memory 857
*Volatile and non volatile memory 858
*Flash memory, MOSFET, MRAM 858
Microphone 858
Card reader, magnetic strip 858
Seismograph 859
Ground fault circuit interrupter (GFCI) 859
Shielded cable 861
Recycling solid waste 861

Chapter 30
*Electric car inductive charging 869
Surge protection 877
Capacitors as filters 884, 896, 897
Loudspeaker cross-over 884
Impedance matching 888
3-phase AC 889
Q-value 896, (419, 425)
Filter circuit 896

Chapter 31
Optical tweezers 916, 923
*TV from the Moon 898, 920, 924
Wireless devices, transmission 898, 917–20
Antennas 911, 919
Phone call time lag 912
*Solar sail 916, 925
Radio and TV 917–9
AM and FM 918
Cell phones, remotes, cable TV, satellite TV 920
*GPS 924
Solar power use 924

Chapter 32
Medical endoscope, bronchoscope, colonoscope 945
How tall a mirror do you need 930
Seeing yourself in a magnifying mirror (concave) 936–7
Convex (rearview mirrors) 938
Optical illusions 939, 998
Apparent water depth 939–40, 941
Rainbows 942, 957
Colors underwater 943
Diamonds sparkle 944
Prism binoculars 944
Fiber optic cables 945, 954, 956
*High-frequency trading, interception 945
Solar cooker 951
Washing machine water level detector 956
Road reflectors 957

Chapter 33
Human eye 975–8
Fovea, denser in cones 976, 1037
Near- and far-sighted 976–8
Corrective lenses 976–7, 987
Contact lenses 978
Seeing underwater 978
Light microscopes 983–4, 1048
Where your eye can see a lens image 961
Cameras, film and digital 970–5
*CCD, CMOS sensors, potential well 970–1
*Bayer pixels, Fovean 971
Digital artifacts 971
Camera adjustments, *f*-stop 971–3
Depth of field 973
*Resolution, compression, JPEG, raw 973–4
Telephoto, wide angle 975
Optical vs. digital zoom 975
Magnifying glass 979–80
Telescopes 980–2
*Microscopes 983–4, 1048
Lens aberrations 984–5
Film projector 989
Pinhole camera 990

Chapter 34
Soap bubbles, oil films, colors 995, 1004–8
Highway mirages 998
Lens coatings 1008–9
Polarizing sunglasses 1012–13
Liquid crystal displays, TV and computer screens 1014–5
Sky color 1015–6
*Lightbulb efficiency, LED 1016–7
Stealth aircraft coating 1022
CD bits, pits & lands 1024

Chapter 35
Resolution of eye 1035, 1037
Useful magnification 1037
Spectroscopy in biology 1041
X-ray diffraction in biology 1044
Medical imaging: X-rays, CT 1045–7
*Interference microscope 1048
*Phase-contrast microscope 1048
Hubble space telescope 1034–5
Telescope and microscope resolution 1035–7
X-rays 1043–7
Tomography 1045–7

Chapter 36
Space travel 1067–8
Global position system (GPS) 1068–9
Fantasy supertrain 1071
Radar speed gun 1092

Chapter 37
Electron microscope image: blood vessel, clot, retina, viruses 1095, 1109, (7)
Photosynthesis 1102
Measuring bone density 1103
Electron microscopes (EM), TEM, SEM 1109, 1151, (1095)
Photocells 1098
Photodiodes, soundtracks 1101

Chapter 38
Scanning tunneling electron microscope 1151
Atomic force microscope 1151

Chapter 39
Fluorescence analysis 1174–5
Medical uses of lasers, surgery 1178, 1183
Neon lights 1158
Fluorescent lightbulbs 1175
Lasers 1175–9, 1216
Bar code readers 1177
DVD, CD, Blu-ray 1177–8
Holography 1178–9

Chapter 40
Cell energy—ATP 1192
Weak bonds, DNA 1192–4
Protein synthesis 1194–6
*Pulse oximeter 1216
Computer processor chips 1186
Transparent objects 1210
Zener diode voltage regulator 1213–4, 1225
Rectifiers 1214
*Photovoltaic cells 1214–5
*LED displays, bulbs 1215–6
TV remote 1215, 1225
*Solid-state lighting 1215–6
**pn* diode laser 1216
*OLED, AMOLED displays 1216–7
Amplifiers 1218
*MOSFET switch 1218–9
*Technology generation 1219

Chapter 41
Earliest life 1248
Radiation film badges 1249, 1274
Smoke detector 1237
Radioactive activity and safety 1243–4
Carbon-14 dating 1246–7
Archeological & geological dating 1246–8
Oldest Earth rocks 1248
Geiger counter 1248
Rubidium-strontium dating 1253
Tritium dating 1254
*Mass excess, mass defect 1254

Chapter 42
Biological radiation damage 1271–6
Radiation dosimetry, RBE 1272–6
Radon exposure 1274, 1276
Natural radioactive background 1274
Radiation exposure, film badge 1274
Radiation sickness 1274
Whole-body dose 1275
Radiation therapy 1276–7
Proton therapy 1277
Radioactive tracers 1277–8
Gamma camera 1278
Medical imaging, PET, SPECT, MRI 1278–82
*Brain PET scan using cell phone 1279
Imaging resolutions compared 1282
Radiation and thyroid 1286
Nuclear reactors, power plants 1256, 1263–5, 1269–71
Breeder reactors 1265
Manhattan project 1266
Nuclear fusion 1266–71
Why stars shine 1267–9
Thermonuclear devices 1269
Fusion energy reactors 1269–71

Chapter 43
Linacs and tumor irradiation 1294

Chapter 44
Stars and galaxies 1323–32
Black holes 1331, 1337–8
Big Bang storyline 1345–8

Preface

New Stuff!

1. **MisConceptual Questions**, 10 or 15 at the end of each chapter. The multiple-choice answers include common misconceptions as well as correct responses. Pedagogically, asking students to think, to consider the options, is more effective than just telling them what is valid and what is wrong. (These are in addition to the one at the start of each chapter.
2. **Digital** is all around us. Yet that word is not always used carefully. In this new edition we have 20 new pages describing the basics from the ground up. **Binary** numbers, *bits* and *bytes*, are introduced in Chapter 23 along with analog-to-digital conversion (ADC), and vice versa, including *digital audio* and how video screens work. Also information **compression**, *sampling rate*, *bit depth*, *pixel addressing*, *digital transmission* and, in later chapters, information **storage** (RAM, DRAM, flash), *digital cameras* and their *sensors* (CCD, CMOS).
3. **Gravitational Assist** (Slingshot) to accelerate spacecraft (Chapter 8).
4. **Magnetic field** of a **single moving charge**, rarely treated (and if it is, maybe not well), and it shows the need for relativity theory.
5. Seeing **yourself** in a **magnifying mirror** (concave), angular magnification and blurriness with a paradox. Also **convex** (rearview) **mirrors** (Chapter 32).
6. Pedagogical clarification on defining **potential energy**, and energy itself (Chapter 8), and on hundreds of other topics.
7. The **Moon** rises an hour later each day (Chapter 6), its *phases*, *periods*, and diagram.
8. Efficiency of **lightbulbs** (Chapter 34).
9. **Idealization** vs. reality emphasized—such as PV diagrams (Chapter 19) as an idealized approximation.
10. Many new Problems ($\sim$ 500) plus new Questions as well as the 500 or so MisConceptual Questions (point 1 above).
11. Many new worked-out Examples.
12. More **math** steps included in derivations and Examples.
13. **State** of a system and *state variables* clarified (Chapter 17).
14. Contemporary physics: Gravitational waves, LIGO and Virgo, Higgs, WIMPS, OLEDS and other semiconductor physics, nuclear fusion updates, neutrino-less double beta decay.
15. New SI units (Chapter 1, Chapter 21, Tables).
16. *Boiling* temperature of water vs. *elevation* (Chapter 18).
17. Modern physics in earlier classical Chapters (sometimes in Problems): Light-years, observable universe (Chapter 1); optical tweezers (Chapter 4); uranium enrichment (Chapter 5); black holes and curved space, white dwarfs (Chapter 6); crystal structure (Chapter 7); Yukawa potential, Lennard-Jones potential (Chapter 8); neutrons, nuclear reactors, moderator, nuclear collisions, radioactive decay, neutron star collapse (Chapter 9); galaxy redshift (Chapter 16); gas diffusion of uranium (Chapter 18); quarks (Chapter 21); liquid-drop model of nucleus, Geiger counter, Van de Graaff (Chapter 23); transistors (Chapters 23, 29); isotopes, cyclotron (Chapter 27); MOSFET (Chapter 29); semiconductor (camera sensor), photon (Chapter 33); line spectra, X-ray crystallography (Chapter 35).
18. Second law of thermodynamics and heat energy reorganized (Chapter 20).
19. **Symmetry** emphasized throughout.
20. *Uranium enrichment*, % needed in reactors, bombs (Chapters 5, 42).
21. Mass excess, mass defect (Chapter 41).
22. The *mole*, more careful definition (Chapter 17).
23. Liquid-gas ambiguity above critical temperature (Chapter 18).
24. Measurement affects quantity measured, new emphasis.

25. New **Applications**:

- Ocean Tides (Chapter 6)
- Anticyclonic weather (Chapter 11)
- Jump starting a car safely (Chapter 26)
- Light bulb efficiency (Chapter 34)
- Specialty microscopes and contrast (Chapter 35)
- Forces on Muscles and Joints (Chapter 12)
- Doppler ultrasound imaging (Chapter 16)
- Lake level change when rock thrown from boat (Chapter 13)
- Skier speed on snow vs. flying through the air (Chapter 5)
- Inductive charging (Chapter 29)
- Human body internal heat transfer is convection (blood) (Chapter 19)
- Blood pressure measurement (Chapter 13)
- Sports (lots)
- Voltage divider (Chapter 26, Problems)
- Flat screen TV (Chapters 23, 34, 40)
- Carbon footprint and climate (Chapter 20)
- Electrocardiogram (Chapter 23)
- Wireless from the Moon unimaginable (Chapter 31)
- Why snorkels are short (Chapter 17 Problem)
- Electric cars (Chapter 25)
- Digital (Chapters 23, 29, 33, 40) includes (in addition to details in point 2 above) quantization error, digital error correction, noise, bit error rate, digital TV data stream, refresh rate, active matrix, thin film transistors, digital memory, bit-line, reading and writing of memory cells (MOSFET), floating gate, volatile and nonvolatile memory, Bayer, JPEG, ASCII code, and more.

Seeing the World through Eyes that Know Physics

I was motivated to write a textbook different from others which typically present physics as a sequence of facts, like a catalog. Instead of beginning formally and dogmatically, I aim to begin each topic with everyday observations and experiences the students can relate to: start with specifics, the real world, and then go to the great generalizations and the more formal aspects of the physics, showing why we believe what we believe. This approach reflects how science is actually practiced.

The aim is to give students a thorough understanding of the basic concepts of physics in all its aspects, from mechanics to modern physics. Also important is to show students how useful physics is in their own everyday lives and in their future professions by means of interesting applications to biology, medicine, engineering, architecture, and more.

Much effort has gone into approaches for the practical techniques of solving problems: worked-out Examples, Problem Solving sections, and Problem Solving Strategies.

Chapter 1 is *not* a throwaway. It is fundamental to physics to realize that every measurement has an *uncertainty*, and how significant figures are used. Being able to make rapid *estimates* is a powerful tool useful for every student, and used throughout the book starting in Chapter 1 (you can estimate the Earth's radius!).

Mathematics can be an obstacle to students. I have aimed at including all steps in a derivation. Important mathematical tools, such as addition of vectors and vector product, are incorporated in the text where first needed, so they come with a context rather than in a forbidding introductory Chapter. Appendices contain a basic math review, derivatives and integrals, plus some more advanced topics including numerical integration, gravitational field of spherical mass distribution, Maxwell's equations in differential form, and a Table of selected nuclear isotopes (carefully updated, as are the Periodic Table and the Fundamental Constants found inside the back and front covers).

Some instructors may find this book contains more material than can be covered in their courses. The text offers great flexibility. Sections marked with a star * may be considered optional. These contain slightly more advanced

Versions of this Book

Complete version: 44 Chapters including 9 Chapters of modern physics.

Classic version: 37 Chapters, 35 on classical physics, plus one each on relativity and quantum theory.

3 Volume version: Available separately or packaged together

Volume 1: Chapters 1–20 on mechanics, including fluids, oscillations, waves, plus heat and thermodynamics.

Volume 2: Chapters 21–35 on electricity and magnetism, plus light and optics.

Volume 3: Chapters 36–44 on modern physics: relativity, quantum theory, atomic physics, condensed matter, nuclear physics, elementary particles, cosmology and astrophysics.

physics material, or material not usually covered in typical courses, or interesting applications; they contain no material needed in later Chapters (except perhaps in later optional Sections). For a brief course, all optional material could be dropped as well as significant parts of Chapters 13, 16, 26, 30, and 35, and selected parts of Chapters 9, 12, 19, 20, 33. Topics not covered in class can be a valuable resource for outside study by students. Indeed, this text can serve as a useful reference for years because of its wide range of coverage.

Thanks

Many physics professors provided input or direct feedback on every aspect of this textbook. They are listed below, and I owe each a debt of gratitude.

Edward Adelson, The Ohio State University
Lorraine Allen, United States Coast Guard Academy
Zaven Altounian, McGill University
Leon Amstutz, Taylor University
Kim Arvidsson, Schreiner University
Philip S. Baringer, Kansas University
Bruce Barnett, Johns Hopkins University
Michael Barnett, Lawrence Berkeley Lab
Anand Batra, Howard University
David Branning, Trinity College
Bruce Bunker, University of Notre Dame
Wayne Carr, Stevens Institute of Technology
Charles Chiu, University of Texas Austin
Roger N. Clark, U. S. Geological Survey
Russell Clark, University of Pittsburgh
Robert Coakley, University of Southern Maine
David Curott, University of North Alabama
Biman Das, SUNY Potsdam
Bob Davis, Taylor University
Kaushik De, University of Texas Arlington
Michael Dennin, University of California Irvine
Kathryn Dimiduk, Cornell University
John DiNardo, Drexel University
Scott Dudley, United States Air Force Academy
John Essick, Reed College
Cassandra Fesen, Dartmouth College
Leonard Finegold, Drexel University
Alex Filippenko, University of California Berkeley
Richard Firestone, Lawrence Berkeley Lab
Tom Furtak, Colorado School of Mines
Gill Gabelmann, Washburn University
Gabriel Orebi Gann, University of California Berkeley
Edward Gibson, California State University Sacramento
John Hamilton, University of Hawai'i – Hilo
John Hardy, Texas A&M
J. Erik Hendrickson, University of Wisconsin-Eau Claire
Charles Hibbard, Lowell High School
Dr. Laurent Hodges, Iowa State University
David Hogg, New York University
Mark Hollabaugh, Normandale Community College
Russell Holmes, University of Minnesota Twin Cities
William Holzapfel, University of California Berkeley
Bob Jacobsen, University of California Berkeley
Arthur W. John, Northeastern University
David Jones, Florida International University
Andrew N. Jordan, University of Rochester
Teruki Kamon, Texas A&M
Thomas Hemmick, State University of New York Stonybrook
Daryao Khatri, University of the District of Columbia
Woo-Joong Kim, Seattle University
John Kinard, Greenwood High School
Jay Kunze, Idaho State University
Jim LaBelle, Dartmouth College
Andrei Linde, Stanford University
M.A.K. Lodhi, Texas Tech
Lisa Madewell, University of Wisconsin
Ponn Maheswaranathan, Winthrop University
Bruce Mason, University of Oklahoma
Mark Mattson, James Madison University
Linda McDonald, North Park College
Raj Mohanty, Boston University
Giuseppe Molesini, Isituto Nazionale di ottica Florence
Lisa K. Morris, Washington State University
Richard Muller, University of California Berkeley
Blaine Norum, University of Virginia
Lauren Movatne, Reedley College
Alexandria Oakes, Eastern Michigan University
Ralph Oberly, Marshall University
Michael Ottinger, San Juan College
Lyman Page, Princeton
Laurence Palmer, University of Maryland
Bruce Partridge, Haverford College
R. Daryl Pedigo, University of Washington
Robert Pelcovitz, Brown University
Saul Perlmutter, University of California Berkeley
Vahe Peroomian, UCLA
Harvey Picker, Trinity College
Amy Pope, Clemson University
James Rabchuk, Western Illinois University
Michele Rallis, Ohio State University
Andrew Resnick, Cleveland State University
Paul Richards, University of California Berkeley
Peter Riley, University of Texas Austin
Dennis Rioux, University of Wisconsin Oshkosh
John Rollino, Rutgers University
Larry Rowan, University of North Carolina Chapel Hill
Arthur Schmidt, Northwestern University
Cindy Schwarz, Vassar College
Peter Sheldon, Randolph-Macon Woman's College
James Siegrist, University of California Berkeley
Christopher Sirola, University of Southern Mississippi
Earl Skelton, Georgetown University
George Smoot, University of California Berkeley
Stanley Sobolewski, Indiana University of Pennsylvania
Mark Sprague, East Carolina University
Michael Strauss, University of Oklahoma
Leo Takahashi, Pennsylvania State University
Richard Taylor, University of Oregon
Oswald Tekyi-Mensah, Alabama State University
Ray Turner, Clemson University
Som Tyagi, Drexel University
David Vakil, El Camino College
Robert Webb, Texas A&M
Robert Weidman, Michigan Technological University
Edward A. Whittaker, Stevens Institute of Technology
Lisa M. Will, San Diego City College
Suzanne Willis, Northern Illinois University
Michael Winokur, University of Wisconsin-Madison
Stanley George Wojcicki, Stanford University
Mark Worthy, Mississippi State University
Edward Wright, UCLA
Todd Young, Wayne State College

I owe special thanks to Prof. Bob Davis for much valuable input, and especially for working out all the Problems and producing the Solutions Manual for all Problems, as well as for providing the answers to odd-numbered Problems at the back of the book. Many thanks also to J. Erik Hendrickson who collaborated with Bob Davis on the solutions, and to the team they managed (Michael Ottinger, John Kinard, David Jones, Kristi Hatch, Lisa Will).

I am especially grateful to Profs. Lorraine Allen, Kathryn Dimiduk, Michael Strauss, Cindy Schwarz, Robert Coakley, Robert Pelcovitz, Mark Hollabaugh, Charles Hibbard, and Michael Winokur, who helped root out errors and offered significant improvements and clarifications.

For Chapters 43 and 44 on Particle Physics and Cosmology and Astrophysics, I was fortunate to receive generous input from some of the top experts in the field, to whom I owe a debt of gratitude: Saul Perlmutter, George Smoot, Richard Muller, Alex Filippenko, Paul Richards, Gabriel Orebi Gann, James Siegrist, and William Holzapfel (UC Berkeley), Andreí Linde (Stanford U.), Lyman Page (Princeton), Edward Wright (UCLA), Michael Strauss (University of Oklahoma), and Bob Jacobsen (UC Berkeley).

I also wish to thank many others at the University of California, Berkeley, Physics Department for helpful discussions, and for hospitality. Thanks also to Prof. Tito Arecchi at the Istituto Nazionale di Ottica, Florence, Italy.

Finally, I am grateful to the many people at Pearson Education with whom I worked on this project, especially Jeanne Zalesky and Paul Corey, and the perspicacious editors Margy Kuntz and Andrea Giancoli.

The final responsibility for all errors lies with me. I welcome comments, corrections, and suggestions as soon as possible to benefit students for the next reprint.

D.G.

email: jeanne.zalesky@pearson.com
paper mail: Jeanne Zalesky
Pearson Education
501 Boylston Street
Boston, MA 020116

About the Author

Doug Giancoli obtained his BA in physics (summa cum laude) from UC Berkeley, his MS in physics at MIT, and his PhD in elementary particle physics back at UC Berkeley. He spent 2 years as a post-doctoral fellow at UC Berkeley's Virus Lab developing skills in molecular biology and biophysics.

His mentors include Nobel winners Emilio Segrè, Barry Barish, and Donald Glaser.

He has taught a wide range of undergraduate courses, traditional as well as innovative ones, and works to improve his textbooks meticulously, seeking ways to provide a better understanding of physics for students.

Doug loves the outdoors, especially climbing peaks. He says climbing peaks is like learning physics: it takes effort and the rewards are great.



Students Advice

HOW TO STUDY

1. Read the Chapter. Learn new vocabulary and notation. Respond to questions and exercises as they occur. Follow carefully the steps of worked-out Examples and derivations. Avoid time looking at a screen. Paper is better than pixels when it comes to learning and thinking.
2. Attend all class meetings. Listen. Take notes. Ask questions (everyone wants to, but maybe you will have the courage). You will get more out of class if you read the Chapter first.
3. Read the Chapter again, paying attention to details. Follow derivations and worked-out Examples. Absorb their logic. Answer Exercises and as many of the end-of-Chapter Questions as you can, and all MisConceptual Questions.
4. Solve at least 10 to 20 end-of-Chapter Problems, especially those assigned. In doing Problems you may find out what you learned and what you didn't. Discuss them with other students. Problem solving is one of the great learning tools. Don' t just look for a formula—it might be the wrong one.

NOTES ON THE FORMAT AND PROBLEM SOLVING

1. Sections marked with a star (*) may be considered optional or advanced. They can be omitted without interrupting the main flow of topics. No later material depends on them except possibly later starred Sections. They may be fun to read, though.
2. The customary **conventions** are used: symbols for quantities (such as m for mass) are italicized, whereas units (such as m for meter) are not italicized. Symbols for vectors are shown in boldface with a small arrow above: $\vec{\mathbf{F}}$.
3. Few equations are valid in all situations. Where practical, the **range of validity** of important equations are stated in square brackets next to the equation. The equations that represent the great laws of physics are displayed with a tan background, as are a few other indispensable equations.
4. At the end of each Chapter is a set of **Questions** you should try to answer. Attempt all the multiple-choice **MisConceptual Questions**, which are intendend to get common misconceptions "out on the table" by including them as responses (temptations) along with correct answers. Most important are **Problems** which are ranked as Level I, II, or III, according to estimated difficulty. Level I Problems are easiest, Level II are standard Problems, and Level III are "challenge problems." These ranked Problems are arranged by Section, but Problems for a given Section may depend on earlier material too. There follows a group of **General Problems**, not arranged by Section or ranked. Problems that relate to optional Sections are starred (*). Answers to odd-numbered Problems are given at the end of the book.
5. Being able to solve **Problems** is a crucial part of learning physics, and provides a powerful means for understanding the concepts and principles. This book contains many aids to problem solving: (a) worked-out **Examples**, including an Approach and a Solution, which should be studied as an integral part of the text; (b) some of the worked-out Examples are **Estimation Examples**,which show how rough or approximate results can be obtained even if the given data are sparse (see Section 1-6); (c) **Problem Solving Strategies** placed throughout the text to suggest a step-by-step approach to problem solving for a particular topic—but the basics remain the same; most of these "Strategies" are followed by an Example that is solved by explicitly following the suggested steps; (d) special problem-solving Sections; (e) "Problem Solving" marginal notes which refer to hints within the text for solving Problems; (f) **Exercises** within the text that you should work out immediately, and then check your response against the answer given at the bottom of the last page of that Chapter; (g) the Problems themselves at the end of each Chapter.
6. **Conceptual Examples** pose a question which hopefully starts you to think about a response. Give yourself a little time to come up with your own response before reading the Response given.
7. Math review, plus additional topics, are found in **Appendices**. **Useful data**, **conversion factors**, and math **formulas** are found inside the front and back covers.

USE OF COLOR

Vectors

A general vector

resultant vector (sum) is slightly thicker

components of any vector are dashed

Displacement ($\vec{\mathbf{D}}$, $\vec{\mathbf{r}}$)

Velocity ($\vec{\mathbf{v}}$)

Acceleration ($\vec{\mathbf{a}}$)

Force ($\vec{\mathbf{F}}$)

Force on second object

or third object in same figure

Momentum ($\vec{\mathbf{p}}$ or $m\vec{\mathbf{v}}$)

Angular momentum ($\vec{\mathbf{L}}$)

Angular velocity ($\vec{\boldsymbol{\omega}}$)

Torque ($\vec{\boldsymbol{\tau}}$)

Electric field ($\vec{\mathbf{E}}$)

Magnetic field ($\vec{\mathbf{B}}$)

Electricity and magnetism

Electric field lines

Equipotential lines

Magnetic field lines

Electric charge (+) or +

Electric charge (–) or –

Electric circuit symbols

Wire, with switch S

Resistor

Capacitor

Inductor

Battery

Ground

Optics

Light rays

Object

Real image (dashed)

Virtual image (dashed and paler)

Other

Energy level (atom, etc.)

Measurement lines 1.0 m

Path of a moving object

Direction of motion or current

This page intentionally left blank

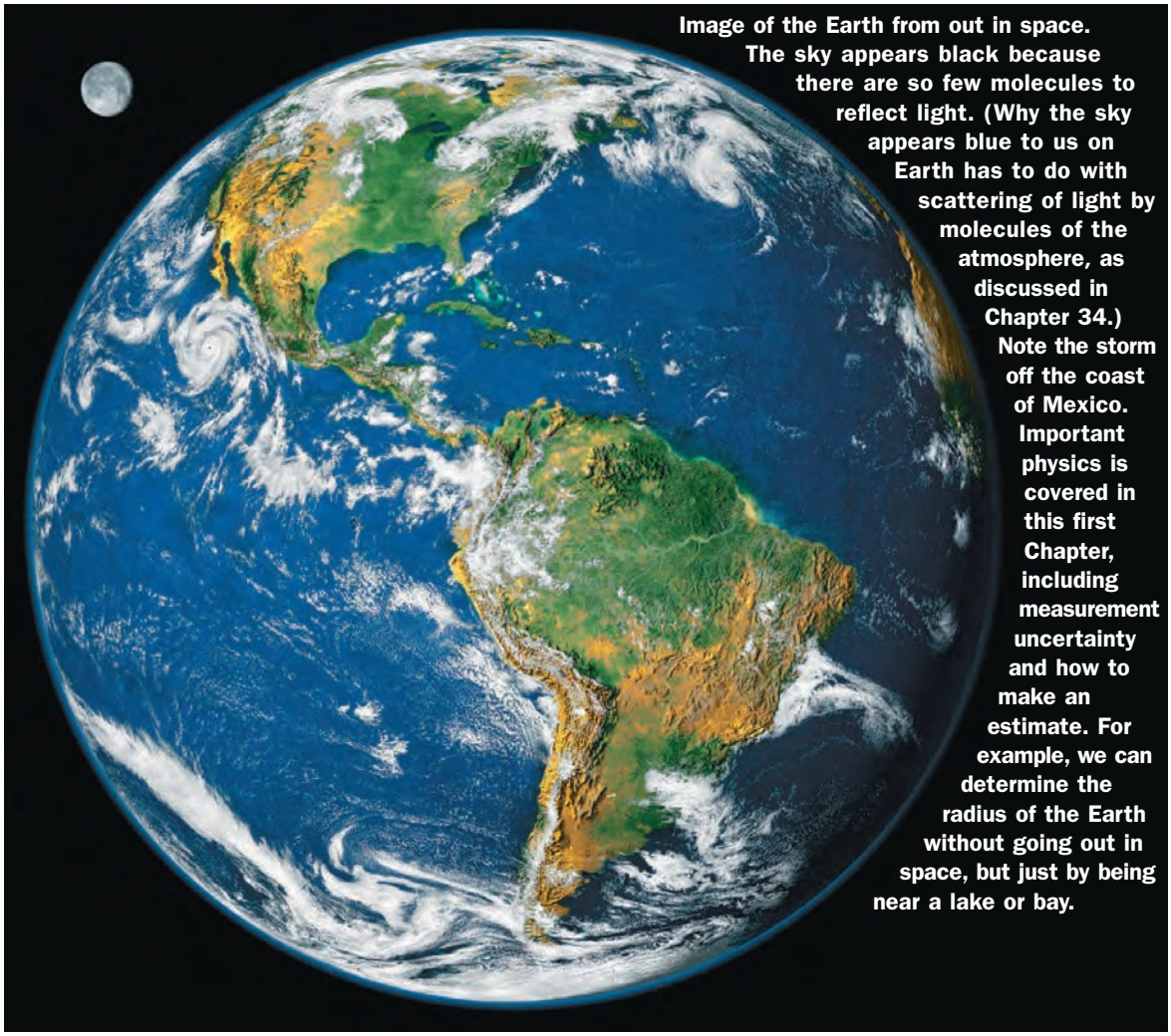


Image of the Earth from out in space. The sky appears black because there are so few molecules to reflect light. (Why the sky appears blue to us on Earth has to do with scattering of light by molecules of the atmosphere, as discussed in Chapter 34.) Note the storm off the coast of Mexico. Important physics is covered in this first Chapter, including measurement uncertainty and how to make an estimate. For example, we can determine the radius of the Earth without going out in space, but just by being near a lake or bay.

CHAPTER 1

Introduction, Measurement, Estimating

CHAPTER-OPENING QUESTIONS—Guess now!

1. How many cm^3 are in $1.0\,m^3$?

(a) 10. **(b)** 100. **(c)** 1000. **(d)** 10,000. **(e)** 100,000. **(f)** 1,000,000.

2. Suppose you wanted to actually measure the radius of the Earth, at least roughly, rather than taking other people's word for what it is. Which response below describes the best approach?

(a) Use an extremely long measuring tape.
(b) It is only possible by flying high enough to see the actual curvature of the Earth.
(c) Use a standard measuring tape, a stepladder, and a large smooth lake.
(d) Use a laser and a mirror on the Moon or on a satellite.
(e) Give up; it is impossible using ordinary means.

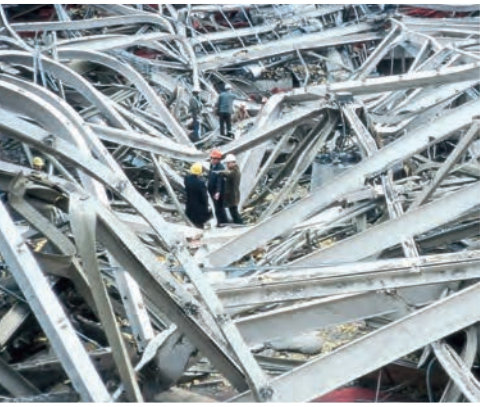
[*We start each Chapter with a Question—sometimes two. Try to answer right away. Don't worry about getting the right answer now—the idea is to get your preconceived notions out on the table. If they are misconceptions, we expect them to be cleared up as you read the Chapter. You will get another chance at the Question later in the Chapter when the appropriate material has been covered. These Chapter-Opening Questions will also help you see the power and usefulness of physics.*]

CONTENTS

1–1 How Science Works
1–2 Models, Theories, and Laws
1–3 Measurement and Uncertainty; Significant Figures
1–4 Units, Standards, and the SI System
1–5 Converting Units
1–6 Order of Magnitude: Rapid Estimating
*1–7 Dimensions and Dimensional Analysis



(a)



(b)

FIGURE 1–1 (a) This bridge over the River Tiber in Rome was built 2000 years ago and still stands. (b) The Hartford Civic Center collapsed in 1978, just two years after it was built.

⚠ CAUTION

Science is not static. It changes and develops

Physics is the most basic of the sciences. It deals with the behavior and structure of matter. The field of physics is usually divided into *classical physics* which includes motion, fluids, heat, sound, light, electricity and magnetism; and *modern physics* which includes the topics of relativity, atomic structure, condensed matter, nuclear physics, elementary particles, and cosmology and astrophysics. We will cover all these topics in this book, beginning with motion (or mechanics, as it is often called) and ending with the most recent results in our study of the cosmos.

An understanding of physics is wonderfully useful for anyone making a career in science or technology. Engineers, for example, must know how to calculate the forces within a structure to design it so that it remains standing (Fig. 1–1a). Indeed, in Chapter 12 we will see a worked-out Example of how a simple physics calculation—or even intuition based on understanding the physics of forces—would have saved hundreds of lives (Fig. 1–1b). We will see many examples in this book of how physics is useful in many fields, and in everyday life.

1–1 How Science Works

There is a real physical world out there. We could just walk through it, not thinking much about it. Or, we can instead examine it carefully. That is what scientists do. The aim of science is the search for order in our observations of the physical world so as to provide a deeper picture or description of this world around us. Sometimes we just want to understand how things work.

Some people seem to think that science is a mechanical process of collecting facts and devising theories. But it is not so simple. Science is a creative activity, and in many ways resembles other creative activities of the human mind.

One important aspect of science is **observation** of events (which great writers and artists also do), and includes the design and carrying out of experiments. But observation and experiment require imagination, because scientists can never include everything in a description of what they observe. In other words, scientists must make judgments about what is relevant in their observations and experiments.

Consider, for example, how two great minds, Aristotle (384–322 B.C.) and Galileo (1564–1642), interpreted motion along a horizontal surface. Aristotle noted that objects given an initial push along the ground (or on a level tabletop) always slow down and stop. Consequently, Aristotle argued, the natural state of an object is to be at rest. Galileo, in his reexamination of horizontal motion in the 1600s, had the idea that friction is a kind of force like a push or a pull; and he imagined that if friction could be eliminated, an object given an initial push along a horizontal surface would continue to move indefinitely without stopping. He concluded that for an object to be in motion was *just as natural* as for it to be at rest. By inventing a new approach, Galileo founded our modern view of motion (Chapters 2, 3, and 4), and he did so with a leap of the imagination. Galileo made this leap conceptually, without actually eliminating friction.

Observation, with careful experimentation and measurement, is one side of the scientific process. The other side is the invention or creation of **theories** to explain and order the observations. Theories are never derived directly from observations. Observations may help inspire a theory, and theories are accepted or rejected based on the results of observation and experiment.

Theories are inspirations that come from the minds of humans. For example, the idea that matter is made up of atoms (the atomic theory) was not arrived at by direct observation of atoms. Rather, the idea sprang from creative minds. The theory of relativity, the electromagnetic theory of light, and Newton's law of universal gravitation were likewise the result of human imagination.

The great theories of science may be compared, as creative achievements, with great works of art or literature. But how does science differ from these other creative activities? One important difference is that science requires **testing** of its ideas or theories to see if their predictions are borne out by experiment.

But theories are not "proved" by testing. First of all, no measuring instrument is perfect, so exact confirmation is not possible. Furthermore, it is not possible to test a theory in every single possible circumstance. Hence a theory cannot be absolutely verified.

Indeed, the history of science tells us that long-held theories can often be replaced by new ones.

1–2 Models, Theories, and Laws

When scientists are trying to understand a particular aspect of the physical world, they often make use of a **model**. A model, in the scientist's sense, is a kind of analogy or mental image of the phenomena in terms of something we are familiar with. One example is the wave model of light. We cannot see waves of light as we can water waves. But it is valuable to think of light as made up of waves because experiments indicate that light behaves in many respects as water waves do.

The purpose of a model is to give us an approximate mental or visual picture—something to hold on to—when we cannot see what actually is happening in the real world. Models often give us a deeper understanding: the analogy to a known system (for instance, water waves in the above example) can suggest new experiments to perform and can provide ideas about what other related phenomena might occur.

You may wonder what the difference is between a theory and a model. Usually a model is relatively simple and provides a structural similarity to the phenomena being studied. A **theory** is broader, more detailed, and can give quantitatively testable predictions, often with great precision.

It is important not to confuse a model or a theory with the real world and the phenomena themselves. Theories are descriptions of the physical world, and they are made up by us. Theories are *invented*—usually by very smart people.

CAUTION
Theories and laws are NOT discovered. They are invented

Scientists give the title **law** to certain concise but general statements about how nature behaves (that energy is conserved, for example). Sometimes the statement takes the form of a relationship or equation between quantities (such as Newton's second law, $F = ma$).

To be called a law, a statement must be found experimentally valid over a wide range of observed phenomena. For less general statements, the term **principle** is often used (such as Archimedes' principle). We use "theory" to describe a more general picture of a large group of phenomena.

Scientific laws are different from political laws, which are *prescriptive*: they tell us how we ought to behave. Scientific laws are *descriptive*: they do not say how nature *should* behave, but rather are meant to describe how nature *does* behave. As with theories, laws cannot be tested in the infinite variety of cases possible. So we cannot be sure that any law is absolutely true. We use the term "law" when its validity has been tested over a wide range of situations, and when any limitations and the range of validity are clearly understood.

Scientists normally do their research as if the accepted laws and theories were true. But they are obliged to keep an open mind in case new information should alter the validity of any given law or theory. In other words, laws of physics, or the "laws of nature", represent our descriptions of reality and are not inalterable facts that last forever. Laws are not lying there in nature, waiting to be discovered. We humans, the brightest humans, invent the laws using observations and intuition as a basis. And we hope our laws provide a good description of nature, and at a minimum give us a reliable approximation of how nature really behaves.

1–3 Measurement and Uncertainty; Significant Figures

In the quest to understand the world around us, scientists seek to find relationships among physical quantities that can be measured.

Uncertainty

Reliable measurements are an important part of physics. But no measurement is absolutely precise. There is an uncertainty associated with every measurement. Among the most important sources of uncertainty, other than blunders, are the limited accuracy of every measuring instrument and the inability to read



FIGURE 1–2 Measuring the width of a board with a centimeter ruler. The uncertainty is about ± 1 mm.

an instrument (such as a ruler) beyond some fraction of the smallest division shown. For example, if you were to use a centimeter ruler to measure the width of a board (Fig. 1–2), the result could be claimed to be precise to about 0.1 cm (1 mm), the smallest division on the ruler, although half of this value might be a valid claim as well. The reason is that it is difficult for the observer to estimate (or *interpolate*) between the smallest divisions. Furthermore, the ruler itself may not have been manufactured to an accuracy very much better than this.

When giving the result of a measurement, it is important to state the **estimated uncertainty** in the measurement. For example, the width of a board might be written as 8.8 ± 0.1 cm. The ± 0.1 cm ("plus or minus 0.1 cm") represents the estimated uncertainty in the measurement, so that the actual width most likely lies between 8.7 and 8.9 cm. The **percent uncertainty** is the ratio of the uncertainty to the measured value, multiplied by 100. For example, if the measurement is 8.8 and the uncertainty about 0.1 cm, the percent uncertainty is

$$\frac{0.1}{8.8} \times 100\% \approx 1\%,$$

where $\approx$ means "is approximately equal to."

Often the uncertainty in a measured value is not specified explicitly. In such cases, scientists follow a general rule that

> **uncertainty in a numerical value is assumed to be *one or a few units* in the last digit specified.**

For example, if a length is given as 5.6 cm, the uncertainty is assumed to be about 0.1 cm or 0.2 cm, or possibly 0.3 cm. It is important in this case that you do not write 5.60 cm, for this implies an uncertainty on the order of 0.01 or 0.02 cm; it assumes that the length is probably between about 5.58 cm and 5.62 cm, when actually you believe it is between about 5.4 and 5.8 cm.

Significant Figures

The number of reliably known digits in a number is called the number of **significant figures**. Thus there are four significant figures in the number 23.21 cm and two in the number 0.062 cm (the zeros in the latter are merely place holders that show where the decimal point goes). The number of significant figures may not always be clear. Take, for example, the number 80. Are there one or two significant figures? We need words here: If we say it is *roughly* 80 km between two cities, there is only one significant figure (the 8) since the zero is merely a place holder. If there is no suggestion that the 80 is a rough approximation, then we can often assume (as we will in this book) that it has two significant figures: so it is 80 km within an accuracy of about 1 or 2 km. If it is precisely 80 km, to within ± 0.1 or ± 0.2 km, then we need to write 80.0 km (three significant figures).

When specifying numerical results, you should avoid the temptation to keep more digits in the final answer than is justified: see boldface statement above. For example, to calculate the area of a rectangle 11.3 cm by 6.8 cm, the result of multiplication would be 76.84 cm^2. But this answer can not be accurate to the implied 0.01 cm^2 uncertainty. Why? Because (using the outer limits of the assumed uncertainty for each measurement) the result could be between $11.2 \text{ cm} \times 6.7 \text{ cm} = 75.04 \text{ cm}^2$ and $11.4 \text{ cm} \times 6.9 \text{ cm} = 78.66 \text{ cm}^2$. At best, we can quote the answer as 77 cm^2, which implies an uncertainty of about 1 or 2 cm^2. The other two digits (in the number 76.84 cm^2) must be dropped (rounded off) because they are not significant. As a rough general **significant figures rule**,

> *the final result of a multiplication or division should have no more digits than the numerical value with the fewest significant figures.*

In our example, 6.8 cm has the least number of significant figures, namely two. Thus the result 76.84 cm^2 needs to be rounded off to 77 cm^2.

PROBLEM SOLVING

Significant figures rule: Number of significant figures in final result should be same as the least significant input value

EXERCISE A The area of a rectangle 4.5 cm by 3.25 cm is correctly given by (*a*) 14.625 cm^2; (*b*) 14.63 cm^2; (*c*) 14.6 cm^2; (*d*) 15 cm^2.

When *adding* or *subtracting* numbers, the final result should contain no more decimal places than the number with the fewest decimal places. For example, the result of subtracting 0.57 from 3.6 is 3.0 (not 3.03). Similarly $36 + 8.2 = 44$, not 44.2.

Be careful not to confuse significant figures with the number of decimal places. Significant figures are related to the expected uncertainty in any measured quantity.

PROBLEM SOLVING
Significant figures when adding and subtracting

EXERCISE B For each of the following numbers, state the number of significant figures and the number of decimal places: (*a*) 1.23; (*b*) 0.123; (*c*) 0.0123.

Keep in mind when you use a calculator that all the digits it produces may not be significant. When you divide 2.0 by 3.0, the proper answer is 0.67, and not 0.666666666 as calculators give (Fig. 1–3a). Digits should not be quoted in a result unless they are truly significant figures. However, to obtain the most accurate result, you should normally *keep one or more extra significant figures throughout a calculation, and round off only in the final result*. (With a calculator, you can keep all its digits in intermediate results.) Calculators can also give too few significant figures. For example, when you multiply 2.5×3.2, a calculator may give the answer as simply 8. See Fig. 1–3b. But the answer is accurate to two significant figures, so the proper answer is 8.0.[†]

CAUTION
Calculators err with significant figures

PROBLEM SOLVING
Report only the proper number of significant figures in the final result. But keep extra digits during the calculation



(a) (b)

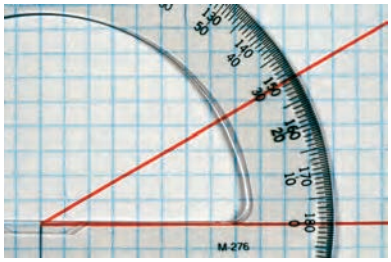
FIGURE 1–3 These two calculators show the wrong number of significant figures. In (a), 2.0 was divided by 3.0. The correct final result should be stated as 0.67. In (b), 2.5 was multiplied by 3.2. The correct result is 8.0.

CONCEPTUAL EXAMPLE 1–1 **Significant figures.** Using a protractor (Fig. 1–4), you measure an angle to be 30°. (*a*) How many significant figures should you quote in this measurement? (*b*) Use a calculator to find the cosine of the angle you measured.

RESPONSE (*a*) If you look at a protractor, you will see that the precision with which you can measure an angle is about one degree (certainly not 0.1°). So you can quote two significant figures, namely 30° (not 30.0°). (*b*) If you enter cos 30° in your calculator, you will get a number like 0.866025403. But the angle you entered is known only to two significant figures, so its cosine is correctly given by 0.87; you must round your answer to two significant figures.

NOTE Trigonometric functions, like cosine, are reviewed in Appendix A.

FIGURE 1–4 Example 1–1. A protractor used to measure an angle.



EXERCISE C Do 0.00324 and 0.00056 have the same number of significant figures?

Scientific Notation

We commonly write numbers in "powers of ten," or "scientific" notation—for instance 36,900 as 3.69×10^4, or 0.0021 as 2.1×10^{-3}. One advantage of scientific notation is that it allows the number of significant figures to be clearly expressed. For example, it is not clear whether 36,900 has three, four, or five significant figures. With powers of ten notation the ambiguity can be avoided: if the number is known to three significant figures, we write 3.69×10^4, but if it is known to four, we write 3.690×10^4.

[†]Be careful also about other digital read-outs. If a digital bathroom scale shows 85.6, do not assume the uncertainty is ± 0.1 or ± 0.2; the scale was likely manufactured with an accuracy of perhaps only 1% or so: that is, ± 1 or ± 2. For digital scientific instruments, also be careful: the instruction manual should state the accuracy.

EXERCISE D Write each of the following in scientific notation and state the number of significant figures for each: (*a*) 0.0258, (*b*) 42,300, (*c*) 344.50.

Percent Uncertainty versus Significant Figures

The significant figures rule is only approximate, and in some cases may underestimate the accuracy (or uncertainty) of the answer. Suppose for example we divide 97 by 92:

$$\frac{97}{92} = 1.05 \approx 1.1.$$

Both 97 and 92 have two significant figures, so the rule says to give the answer as 1.1. Yet the numbers 97 and 92 both imply an uncertainty of ± 1 if no other uncertainty is stated. Both 92 ± 1 and 97 ± 1 imply an uncertainty of about 1% ($1/92 \approx 0.01 = 1\%$). But the final result to two significant figures is 1.1, with an implied uncertainty of ± 0.1, which is an uncertainty of $0.1/1.1 \approx 0.1 \approx 10\%$. In this case it is better to give the answer as 1.05 (which is three significant figures). Why? Because 1.05 implies an uncertainty of ± 0.01 which is $0.01/1.05 \approx 0.01 \approx 1\%$, just like the uncertainty in the original numbers 92 and 97.

SUGGESTION: Use the significant figures rule, but consider the % uncertainty too, and add an extra digit if it gives a more realistic estimate of uncertainty.

Approximations

Much of physics involves approximations, often because we do not have the means to solve a problem precisely. For example, we may choose to ignore air resistance or friction in doing a Problem even though they are present in the real world, and then our calculation is only an estimate or approximation. In doing Problems, we should be aware of what approximations we are making, and be aware that the precision of our answer may not be nearly as good as the number of significant figures given in the result.

Accuracy versus Precision

There is a technical difference between "precision" and "accuracy." **Precision** in a strict sense refers to the repeatability of the measurement using a given instrument. For example, if you measure the width of a board many times, getting results like 8.81 cm, 8.85 cm, 8.78 cm, 8.82 cm (interpolating between the 0.1 cm marks as best as possible each time), you could say the measurements give a *precision* a bit better than 0.1 cm. **Accuracy** refers to how close a measurement is to the true value. For example, if the ruler shown in Fig. 1–2 was manufactured with a 2% error, the accuracy of its measurement of the board's width (about 8.8 cm) would be about 2% of 8.8 cm or about ± 0.2 cm. Estimated uncertainty is meant to take both accuracy and precision into account.

1–4 Units, Standards, and the SI System

The measurement of any quantity is made relative to a particular standard or **unit**, and this unit must be specified along with the numerical value of the quantity. For example, we can measure length in British units such as inches, feet, or miles, or in the metric system in centimeters, meters, or kilometers. To specify that the length of a particular object is 18.6 is insufficient. The unit *must* be given, because 18.6 meters is very different from 18.6 inches or 18.6 millimeters.

For any unit we use, such as the meter for distance or the second for time, we need to define a **standard** which defines exactly how long one meter or one second is. It is important that standards be chosen that are readily reproducible so that anyone needing to make a very accurate measurement can refer to the standard in the laboratory and communicate results with other scientists.

Length

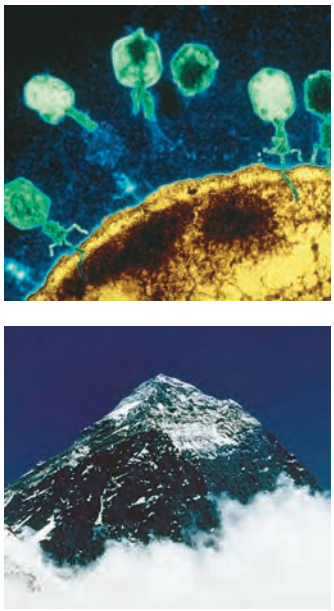
The first truly international standard was the **meter** (abbreviated m) established as the standard of **length** by the French Academy of Sciences in the 1790s. The standard meter was originally chosen to be one ten-millionth of the distance from the Earth's equator to either pole,[†] and a platinum rod to represent this length was made. (One meter is, very roughly, the distance from the tip of your nose to the tip of your finger, with arm and hand stretched out horizontally.) In 1889, the meter was defined more precisely as the distance between two finely engraved marks on a particular bar of platinum–iridium alloy. In 1960, to provide greater precision and reproducibility, the meter was redefined as 1,650,763.73 wavelengths of a particular orange light emitted by the gas krypton-86.

In 1983 the meter was again redefined, this time in terms of the speed of light (whose best measured value in terms of the older definition of the meter was 299,792,458 m/s, with an uncertainty of 1 m/s). The new definition reads: "The meter is the length of path traveled by light in vacuum during a time interval of 1/299,792,458 of a second." The new definition of the meter has the effect of giving the speed of light the exact value of 299,792,458 m/s. [The newer definitions provided greater precision than the 2 marks on the old platinum bar.]

New definition of the meter

British units of length (inch, foot, mile) are now defined in terms of the meter. The **inch** (in.) is defined as exactly 2.54 centimeters (cm; 1 cm = 0.01 m). One **foot** is exactly 12 in., and 1 **mile** is 5280 ft. Other conversion factors are given in the Table on the inside of the front cover of this book. Table 1–1 below presents some typical lengths, from very small to very large, rounded off to the nearest power of 10. (We call this rounded off value the **order of magnitude**.) See also Fig. 1–5. (Note that the abbreviation for inches (in.) is the only one with a period, to distinguish it from the word "in".) [The **nautical mile** = 6076 ft = 1852 km is used by ships on the open sea and was originally defined as 1/60 of a degree latitude on Earth's surface. A speed of 1 **knot** is 1 nautical mile per hour.]

FIGURE 1–5 Some lengths: (a) viruses (about 10^{-7} m long) attacking a cell; (b) Mt. Everest's height is on the order of 10^4 m (8850 m, to be precise).



(a)

(b)

Time

The standard unit of **time** is the **second** (s). For many years, the second was defined as 1/86,400 of a mean solar day (24 h/day × 60 min/h × 60 s/min = 86,400 s/day). The standard second can be defined more precisely in terms of the frequency of radiation emitted by cesium atoms when they pass between two particular states. [Specifically, one second is the time required for 9,192,631,770 periods of this radiation. This number was chosen to keep "one second" the same as in the old definition.] There are, by definition, 60 s in one minute (min) and 60 minutes in one hour (h). Table 1–2 presents a range of time intervals, rounded off to the nearest power of 10.

[†]Modern measurements of the Earth's circumference reveal that the intended length is off by about one-fiftieth of 1%. Not bad!

TABLE 1–1 Some Typical Lengths or Distances (order of magnitude)

Length (or Distance)	Meters (approximate)
Neutron or proton (diameter)	10^{-15} m
Atom (diameter)	10^{-10} m
Virus [see Fig. 1–5a]	10^{-7} m
Sheet of paper (thickness)	10^{-4} m
Finger width	10^{-2} m
Football field length	10^{2} m
Height of Mt. Everest [see Fig. 1–5b]	10^{4} m
Earth diameter	10^{7} m
Earth to Sun	10^{11} m
Earth to nearest star	10^{16} m
Earth to nearest galaxy	10^{22} m
Earth to farthest galaxy visible	10^{26} m

TABLE 1–2 Some Typical Time Intervals (order of magnitude)

Time Interval	Seconds (approximate)
Lifetime of very unstable subatomic particle	10^{-23} s
Lifetime of radioactive elements	10^{-22} s to 10^{28} s
Lifetime of muon	10^{-6} s
Time between human heartbeats	10^{0} s (= 1 s)
One day	10^{5} s
One year	3×10^{7} s
Human life span	2×10^{9} s
Length of recorded history	10^{11} s
Humans on Earth	10^{14} s
Life on Earth	10^{17} s
Age of Universe	4×10^{17} s

TABLE 1–3 Some Masses

Object	Kilograms (approximate)
Electron	10^{-30} kg
Proton, neutron	10^{-27} kg
DNA molecule	10^{-17} kg
Bacterium	10^{-15} kg
Mosquito	10^{-5} kg
Plum	10^{-1} kg
Human	10^{2} kg
Ship	10^{8} kg
Earth	6×10^{24} kg
Sun	2×10^{30} kg
Galaxy	10^{41} kg

Mass

The standard unit of **mass** is the **kilogram** (kg). The standard mass has been, since 1889, a particular platinum–iridium cylinder, kept at the International Bureau of Weights and Measures near Paris, France, whose mass is defined as exactly 1 kg. A range of masses is presented in Table 1–3. [For practical purposes, a 1 kg mass weighs about 2.2 pounds on Earth.]

1 metric **ton** is 1000 kg. In the British system of units, 1 ton is 2000 pounds.

When dealing with atoms and molecules, we usually use the **unified atomic mass unit** (u or amu). In terms of the kilogram,

$$1\,\text{u} = 1.6605 \times 10^{-27}\,\text{kg}.$$

(Precise values of this and other numbers are given inside the front cover.) The **density** of a uniform object is its mass divided by its volume, commonly expressed in kg/m^3.

Unit Prefixes

In the metric system, the larger and smaller units are defined in multiples of 10 from the standard unit, and this makes calculation particularly easy. Thus 1 kilometer (km) is 1000 m, 1 centimeter is $\frac{1}{100}$ m, 1 millimeter (mm) is $\frac{1}{1000}$ m or $\frac{1}{10}$ cm, and so on. The *prefixes* "centi-," "kilo-," and others are listed in Table 1–4 and can be applied not only to units of length but to units of volume, mass, or any other unit. For example, a centiliter (cL) is $\frac{1}{100}$ liter (L), and a kilogram (kg) is 1000 grams (g). An 8.2-megapixel camera has a detector with 8,200,000 pixels (individual "picture elements").

In common usage, $1\,\mu\text{m}$ $(= 10^{-6}\,\text{m})$ is called 1 **micron**.

PROBLEM SOLVING

Always use a consistent set of units

Systems of Units

When dealing with the laws and equations of physics it is very important to use a consistent set of units. Several systems of units have been in use over the years. Today the most important is the **Système International** (French for International System), which is abbreviated SI. In SI units, the standard of length is the meter, the standard for time is the second, and the standard for mass is the kilogram. This system used to be called the MKS (meter-kilogram-second) system.

A second metric system is the **cgs system**, in which the centimeter, gram, and second are the standard units of length, mass, and time, as abbreviated in the title. The **British engineering system** (although more used in the U.S. than Britain) has as its standards the foot for length, the pound for force, and the second for time.

We use SI units almost exclusively in this book, although we often define the cgs and British units when a new quantity is introduced. In the SI, there have traditionally been seven *base* quantities, each defined in terms of a standard; seven is the smallest number of base quantities consistent with a full description of the physical world. See Table 1–5. All other quantities[†] can be defined in terms of seven base quantities; see the Table inside the front cover which lists many quantities and their units in terms of base units.

TABLE 1–4 Metric (SI) Prefixes

Prefix	Abbreviation	Value
yotta	Y	10^{24}
zetta	Z	10^{21}
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deka	da	10^{1}
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro[†]	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}
zepto	z	10^{-21}
yocto	y	10^{-24}

[†]μ is the Greek letter "mu."

*A New SI

As always in science, new ideas and approaches can produce better precision and closer correspondence with the real world. Even for units and standards.

International organizations on units have proposed further changes that should make standards more readily available and reproducible. To cite one example, the standard kilogram (see above) has been found to have changed slightly in mass (contamination is one cause).

The new redefinition of SI standards follows the method already used for the meter as being related to the defined value of the speed of light, as we mentioned on page 7 under "Length." For example, the charge on the electron, e, instead of being a measured value, becomes *defined* as a certain value (its current value), and the unit of electric charge (the coulomb) follows from that. All units then become based on

[†]Some exceptions are for angle (radians—see Chapter 10), solid angle (steradian), and sound level (bel or decibel, Chapter 16).

*Some Sections of this book, such as this subsection, may be considered *optional* at the discretion of the instructor and they are marked with an asterisk (*). See the Preface for more details.

defined fundamental constants like e and the speed of light. Seven is still the number of basic standards. The new definitions maintain the values of the traditional definitions: the "new" meter is the same length as the "old" meter. The new definitions do not change our understanding of what length, time, or mass means.

For us, using this book, the difference between the new SI and the traditional SI is highly technical and does not affect the physics we study. We include the traditional SI because there is some good physics in explaining it. [The Table of Fundamental Constants inside the front cover would look slightly different using the new SI. The value of the charge e on the electron, for example, is *defined*, and so would have no uncertainty attached to it; instead, our Table inside the front cover includes the traditional SI measured uncertainty (updated) of $\pm 98 \times 10^{-29}$ C.]

TABLE 1–5
Traditional SI Base Quantities

Quantity	Unit	Unit Abbreviation
Length	meter	m
Time	second	s
Mass	kilogram	kg
Electric current	ampere	A
Temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd

1–5 Converting Units

Any quantity we measure, such as a length, a speed, or an electric current, consists of a number *and* a unit. Often we are given a quantity in one set of units, but we want it expressed in another set of units. For example, suppose we measure that a shelf is 21.5 inches wide, and we want to express this in centimeters. We must use a **conversion factor**, which in this case is, *by definition*, exactly

$$1 \text{ in.} = 2.54 \text{ cm}$$

or, written another way,

$$1 = 2.54 \text{ cm/in.}$$

Since multiplying by the number one does not change anything, the width of our shelf, in cm, is

$$21.5 \text{ inches} = (21.5 \cancel{\text{in.}}) \times \left(2.54 \frac{\text{cm}}{\cancel{\text{in.}}}\right) = 54.6 \text{ cm}.$$

Note how the units (inches in this case) cancelled out (thin red lines). A Table containing many unit conversions is found inside the front cover of this book. Let's consider some Examples.

EXAMPLE 1–2 **The 8000-m peaks.** There are only 14 peaks whose summits are over 8000 m above sea level. They are the highest peaks in the world (Fig. 1–6 and Table 1–6) and are referred to as "eight-thousanders." What is the elevation, in feet, of an elevation of 8000 m?

APPROACH We need to convert meters to feet, and we can start with the conversion factor 1 in. = 2.54 cm, which is exact. That is, 1 in. = 2.5400 cm to any number of significant figures, because it is *defined* to be.

SOLUTION One foot is defined to be 12 in., so we can write

$$1 \text{ ft} = (12 \cancel{\text{in.}})\left(2.54 \frac{\text{cm}}{\cancel{\text{in.}}}\right) = 30.48 \text{ cm} = 0.3048 \text{ m},$$

which is exact. Note how the units cancel (colored slashes). We can rewrite this equation to find the number of feet in 1 meter:

$$1 \text{ m} = \frac{1 \text{ ft}}{0.3048} = 3.28084 \text{ ft}.$$

(We could carry the result to 6 significant figures because 0.3048 is exact, 0.304800 · · · .) We multiply this equation by 8000.0 (to have five significant figures):

$$8000.0 \text{ m} = (8000.0 \cancel{\text{m}})\left(3.28084 \frac{\text{ft}}{\cancel{\text{m}}}\right) = 26{,}247 \text{ ft}.$$

An elevation of 8000 m is 26,247 ft above sea level.

NOTE We could have done the unit conversions all in one line:

$$8000.0 \text{ m} = (8000.0 \cancel{\text{m}})\left(\frac{100 \cancel{\text{cm}}}{1 \cancel{\text{m}}}\right)\left(\frac{1 \cancel{\text{in.}}}{2.54 \cancel{\text{cm}}}\right)\left(\frac{1 \text{ ft}}{12 \cancel{\text{in.}}}\right) = 26{,}247 \text{ ft}.$$

The key is to multiply conversion factors, each equal to one (= 1.0000), and to make sure which units cancel.



FIGURE 1–6 The world's second highest peak, K2, whose summit is considered the most difficult of the "8000-ers." Example 1–2.

PHYSICS APPLIED
The world's tallest peaks

TABLE 1–6 The 8000-m Peaks

Peak	Height (m)
Mt. Everest	8850
K2	8611
Kangchenjunga	8586
Lhotse	8516
Makalu	8462
Cho Oyu	8201
Dhaulagiri	8167
Manaslu	8156
Nanga Parbat	8125
Annapurna	8091
Gasherbrum I	8068
Broad Peak	8047
Gasherbrum II	8035
Shisha Pangma	8013

The first two equations in Example 1–2 on the previous page show how to change from feet to meters, or meters to feet. For practical purposes

$$1\,\text{m} \;=\; 3.28\,\text{ft} \;\approx\; 3.3\,\text{ft}$$

which means that we can change any distance or height in meters to feet by multiplying by 3 and adding 10% (0.1). For example, a 3000-m-high peak in feet is 9000 ft + 900 ft ≈ 10,000 ft.

TABLE 1–6 The 8000-m Peaks

Peak	Height (m)
Mt. Everest	8850
K2	8611
Kangchenjunga	8586
Lhotse	8516
Makalu	8462
Cho Oyu	8201
Dhaulagiri	8167
Manaslu	8156
Nanga Parbat	8125
Annapurna	8091
Gasherbrum I	8068
Broad Peak	8047
Gasherbrum II	8035
Shisha Pangma	8013

EXERCISE E The names and elevations of the 14 eight-thousand-meter peaks in the world (see Example 1–2) are given in Table 1–6, repeated here. They are all in the Himalaya mountain range in India, Pakistan, Tibet, and China. Determine the elevation of the world's three highest peaks in feet.

EXAMPLE 1–3 Apartment area. You have seen a nice apartment whose floor area is 880 square feet (ft^2). What is its area in square meters?

APPROACH We use the same conversion factor, 1 in. = 2.54 cm, but this time we have to use it twice.

SOLUTION Because 1 in. = 2.54 cm = 0.0254 m, then

$$1\,\text{ft}^2 \;=\; (12\,\text{in.})^2(0.0254\,\text{m/in.})^2 \;=\; 0.0929\,\text{m}^2.$$

So

$$880\,\text{ft}^2 \;=\; (880\,\text{ft}^2)(0.0929\,\text{m}^2/\text{ft}^2) \;\approx\; 82\,\text{m}^2.$$

NOTE As a rule of thumb, an area given in ft^2 is roughly 10 times the number of square meters (more precisely, about 10.8×).

Rule of thumb: Floor area in ft^2 is about 10× area in m^2: $100\,\text{m}^2 \approx 1000\,\text{ft}^2$

EXERCISE F One **hectare** is defined as $1.000 \times 10^4\,\text{m}^2$. There are 640 **acres** in a square mile. Both units are used for land area. (*a*) How many acres are in one hectare? (*b*) What would be an easy everyday rule-of-thumb conversion factor?

EXAMPLE 1–4 Speeds. Where the posted speed limit is 55 miles per hour (mi/h or mph), what is this speed (*a*) in meters per second (m/s) and (*b*) in kilometers per hour (km/h)?

APPROACH We again use the conversion factor 1 in. = 2.54 cm, and we recall that there are 5280 feet in a mile and 12 inches in a foot; also, one hour contains (60 min/h) × (60 s/min) = 3600 s/h.

SOLUTION (*a*) We can write 1 mile as

$$1\,\text{mi} \;=\; (5280\,\cancel{\text{ft}})\left(12\,\frac{\cancel{\text{in.}}}{\cancel{\text{ft}}}\right)\left(2.54\,\frac{\cancel{\text{cm}}}{\cancel{\text{in.}}}\right)\left(\frac{1\,\text{m}}{100\,\cancel{\text{cm}}}\right)$$
$$= \;1609\,\text{m}.$$

We also know that 1 hour contains 3600 s, so

$$55\,\frac{\text{mi}}{\text{h}} \;=\; \left(55\,\frac{\cancel{\text{mi}}}{\cancel{\text{h}}}\right)\left(1609\,\frac{\text{m}}{\cancel{\text{mi}}}\right)\left(\frac{1\,\cancel{\text{h}}}{3600\,\text{s}}\right) \;=\; 25\,\frac{\text{m}}{\text{s}},$$

where we rounded off to two significant figures.

(*b*) Now we use 1 mi = 1609 m = 1.609 km; then

$$55\,\frac{\text{mi}}{\text{h}} \;=\; \left(55\,\frac{\cancel{\text{mi}}}{\text{h}}\right)\left(1.609\,\frac{\text{km}}{\cancel{\text{mi}}}\right) \;=\; 88\,\frac{\text{km}}{\text{h}}.$$

NOTE Each conversion factor is equal to one. You can look up most conversion factors in the Table inside the front cover.

PROBLEM SOLVING
Conversion factors = 1

EXERCISE G Return to the first Chapter-Opening Question, page 1, and answer it again now. Try to explain why you may have answered differently the first time.

PROBLEM SOLVING
Unit conversion is wrong if units do not cancel

When changing units, you can avoid making an error in the use of conversion factors by checking that units cancel out properly. For example, in our conversion of 1 mi to 1609 m in Example 1–4(*a*), if we had incorrectly used the factor $\left(\frac{100\,\text{cm}}{1\,\text{m}}\right)$ instead of $\left(\frac{1\,\text{m}}{100\,\text{cm}}\right)$, the centimeter units would not have cancelled out; we would not have ended up with meters.

1–6 Order of Magnitude: Rapid Estimating

This is an exciting and powerful Section that will be useful throughout this book, and in real life. We will see how to make approximate calculations of quantities you may never have dreamed you could do.

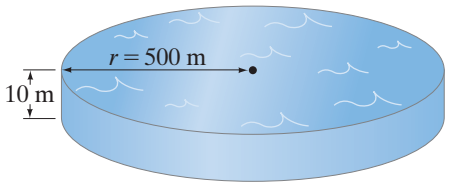
Also, we are sometimes interested only in an approximate value for a quantity, maybe because an accurate calculation would take more time than it is worth or requires data that are not available. In other cases, we may want to make a rough estimate in order to check a calculation made on a calculator, to make sure that no blunders were made when the numbers were entered.

PROBLEM SOLVING
How to make a rough estimate

A rough estimate can be made by rounding off all numbers to one significant figure and its power of 10, and after the calculation is made, again keeping only one significant figure. Such an estimate is called an **order-of-magnitude estimate** and can be accurate within a factor of 10, and often better. In fact, the phrase "order of magnitude" is sometimes used to refer simply to the power of 10.



(a)



(b)

FIGURE 1–7 Example 1–5. (a) How much water is in this lake? (Photo is one of the Rae Lakes in the Sierra Nevada of California.) (b) Model of the lake as a cylinder. [We could go one step further and estimate the mass or weight of this lake. We will see later that water has a density of 1000 kg/m^3, so this lake has a mass of about $(10^3\,\text{kg/m}^3)(10^7\,\text{m}^3) \approx 10^{10}\,\text{kg}$, which is about 10 billion kg or 10 million metric tons. (A metric ton is 1000 kg, about 2200 lb, slightly larger than a British ton, 2000 lb.)]

EXAMPLE 1–5 **ESTIMATE** **Volume of a lake.** Estimate how much water there is in a particular lake, Fig. 1–7a, which is roughly circular, about 1 km across, and you guess it has an average depth of about 10 m.

APPROACH No lake is a perfect circle, nor can lakes be expected to have a perfectly flat bottom. We are only estimating here. To estimate the volume, we can use a simple model of the lake as a cylinder: we multiply the average depth of the lake times its roughly circular surface area, as if the lake were a cylinder (Fig. 1–7b).

PHYSICS APPLIED
Estimating the volume (or mass) of a lake; see also Fig. 1–7

SOLUTION The volume V of a cylinder is the product of its height h times the area of its base: $V = h\pi r^2$, where r is the radius of the circular base.[†] The radius r is $\frac{1}{2}\,\text{km} = 500\,\text{m}$, so the volume is approximately

$$V = h\pi r^2 \approx (10\,\text{m}) \times (3) \times (5 \times 10^2\,\text{m})^2 \approx 8 \times 10^6\,\text{m}^3 \approx 10^7\,\text{m}^3,$$

where π was rounded off to 3. So the volume is on the order of $10^7\,\text{m}^3$, ten million cubic meters. Because of all the estimates that went into this calculation, the order-of-magnitude estimate $(10^7\,\text{m}^3)$ is probably better to quote than the $8 \times 10^6\,\text{m}^3$ figure.

NOTE To express our result in U.S. gallons, we see in the Table on the inside front cover that $1\,\text{liter} = 10^{-3}\,\text{m}^3 \approx \frac{1}{4}\,\text{gallon}$. Hence, the lake contains about $(8 \times 10^6\,\text{m}^3)(1\,\text{gallon}/4 \times 10^{-3}\,\text{m}^3) \approx 2 \times 10^9$ gallons of water.

[†]Formulas like this for volume, area, etc., are found inside the back cover of this book.

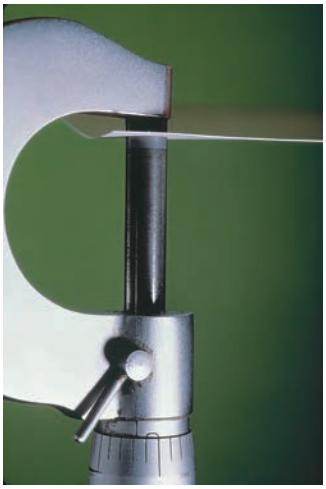


FIGURE 1–8 Example 1–6. Micrometer used for measuring small thicknesses.

EXAMPLE 1–6 ESTIMATE **Thickness of a sheet of paper.** Estimate the thickness of a page of this book.

APPROACH At first you might think that a special measuring device, a micrometer (Fig. 1–8), is needed to measure the thickness of one page since an ordinary ruler can not be read so finely. But we can use a trick or, to put it in physics terms, make use of a *symmetry*: we can make the reasonable assumption that all the pages of this book are equal in thickness.

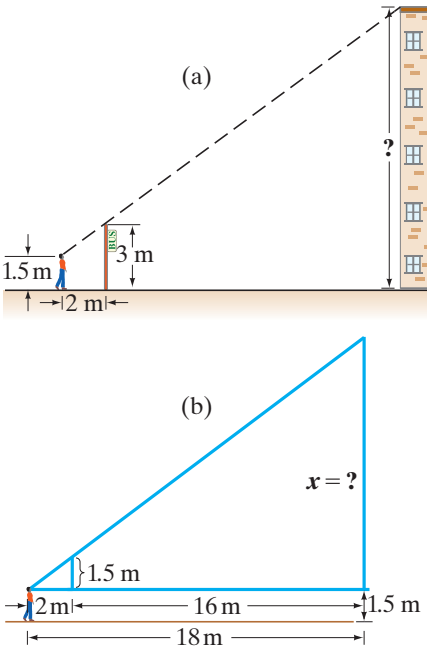
SOLUTION We can use a ruler to measure hundreds of pages at once. If you measure the thickness of the first 500 pages of this book (page 1 to page 500), you might get something like 1.5 cm. Note that 500 numbered pages, counted front and back, is 250 separate pieces of paper. So one sheet must have a thickness of about

$$\frac{1.5\text{ cm}}{250\text{ sheets}} \approx 6 \times 10^{-3}\text{ cm} = 6 \times 10^{-2}\text{ mm},$$

or less than a tenth of a millimeter (0.1 mm).

It cannot be emphasized enough how important it is to draw a diagram when solving a physics Problem, as the next Example shows.

FIGURE 1–9 Example 1–7. Diagrams are really useful!



EXAMPLE 1–7 ESTIMATE **Height by triangulation.** Estimate the height of the building shown in Fig. 1–9, by "triangulation," with the help of a bus-stop pole and a friend.

APPROACH By standing your friend next to the pole, you estimate the height of the pole to be 3 m. You next step away from the pole until the top of the pole is in line with the top of the building, Fig. 1–9a. You are 5 ft 6 in. tall, so your eyes are about 1.5 m above the ground. Your friend is taller, and when she stretches out her arms, one hand touches you and the other touches the pole, so you estimate that distance as 2 m (Fig. 1–9a). You then pace off the distance from the pole to the base of the building with big, 1-m-long steps, and you get a total of 16 steps or 16 m.

SOLUTION Now you draw, to scale, the diagram shown in Fig. 1–9b using these measurements. You can measure, right on the diagram, the last side of the triangle to be about $x \approx 13$ or 14 m. Alternatively, you can use similar triangles to obtain the height x:

$$\frac{1.5\text{ m}}{2\text{ m}} = \frac{x}{18\text{ m}},$$

so

$$x \approx 13\tfrac{1}{2}\text{ m}.$$

Finally you add in your eye height of 1.5 m above the ground to get your final result: the building is about 15 m tall.

EXAMPLE 1–8 ESTIMATE **Total number of heartbeats.** Estimate the total number of beats a typical human heart makes in a lifetime.

APPROACH A typical resting heart rate is 70 beats/min. But during exercise it can be a lot higher. A reasonable average might be 80 beats/min.

SOLUTION One year, in seconds, is $(24\text{ h/d})(3600\text{ s/h})(365\text{ d}) \approx 3 \times 10^7\text{ s}$. If an average person lives 70 years $= (70\text{ yr})(3 \times 10^7\text{ s/yr}) \approx 2 \times 10^9\text{ s}$, then the total number of heartbeats would be about

$$\left(80\frac{\text{beats}}{\text{min}}\right)\left(\frac{1\text{ min}}{60\text{ s}}\right)(2 \times 10^9\text{ s}) \approx 3 \times 10^9,$$

or 3 billion.

EXAMPLE 1–9 ESTIMATE **Estimating the radius of Earth.** Believe it or not, you can estimate the radius of the Earth without having to go into space (see the photograph on page 1). If you have ever been on the shore of a large lake, you may have noticed that you cannot see the beaches, piers, or rocks at water level across the lake on the opposite shore. The lake seems to bulge out between you and the opposite shore—a good clue that the Earth is round. Suppose you climb a stepladder and discover that when your eyes are 10 ft (3.0 m) above the water, you can just see the rocks at water level on the opposite shore. From a map, you estimate the distance to the opposite shore as $d \approx 6.1$ km. Use Fig. 1–10 with $h = 3.0$ m to estimate the radius R of the Earth.

APPROACH We use simple geometry, including the theorem of Pythagoras, $c^2 = a^2 + b^2$, where c is the length of the hypotenuse of any right triangle, and a and b are the lengths of the other two sides.

SOLUTION For the right triangle of Fig. 1–10, the two sides are the radius of the Earth R and the distance $d = 6.1$ km $= 6100$ m. The hypotenuse is approximately the length $R + h$, where $h = 3.0$ m. By the Pythagorean theorem,

$$\begin{aligned} R^2 + d^2 &\approx (R + h)^2 \\ &\approx R^2 + 2hR + h^2. \end{aligned}$$

We solve algebraically for R, after cancelled R^2 on both sides:

$$\begin{aligned} R \approx \frac{d^2 - h^2}{2h} &= \frac{(6100\,\text{m})^2 - (3.0\,\text{m})^2}{6.0\,\text{m}} \\ &= 6.2 \times 10^6\,\text{m} \\ &= 6200\,\text{km}. \end{aligned}$$

NOTE Precise measurements give 6380 km. But look at your achievement! With a few simple rough measurements and simple geometry, you made a good estimate of the Earth's radius. You did not need to go out in space, nor did you need a very long measuring tape.[†]

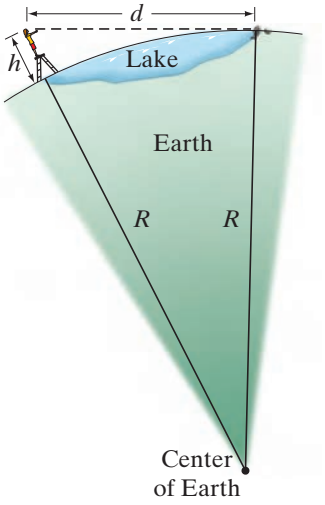


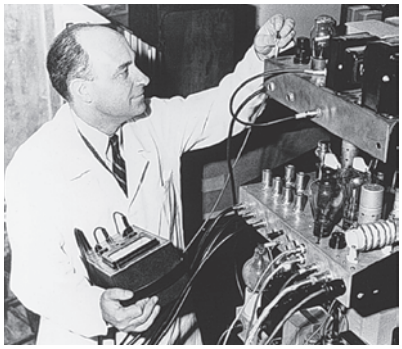
FIGURE 1–10 Example 1–9, but not to scale. You can just barely see rocks at water level on the opposite shore of a lake 6.1 km wide if you stand on a stepladder.

EXERCISE H Return to the second Chapter-Opening Question, page 1, and answer it again now. Try to explain why you may have answered differently the first time.

Another type of estimate, this one made famous by Enrico Fermi (1901–1954, Fig. 1–11), was to show his students how to estimate the number of piano tuners in a city, such as Chicago or San Francisco. To get a rough order-of-magnitude estimate of the number of piano tuners today in San Francisco, a city of about 800,000 inhabitants, we can proceed by estimating the number of functioning pianos, how often each piano is tuned, and how many pianos each tuner can tune. To estimate the number of pianos in San Francisco, we note that certainly not everyone has a piano. A guess of 1 family in 3 having a piano would correspond to 1 piano per 12 persons, assuming an average family of 4 persons.

As an order of magnitude, let's say 1 piano per 10 people. This is certainly more reasonable than 1 per 100 people, or 1 per every person, so let's proceed with the estimate that 1 person in 10 has a piano, or about 80,000 pianos in San Francisco. Now a piano tuner needs an hour or two to tune a piano. So let's estimate that a tuner can tune 4 or 5 pianos a day. A piano ought to be tuned every 6 months or a year—let's say once each year. A piano tuner tuning 4 pianos a day, 5 days a week, 50 weeks a year can tune about 1000 pianos a year. So San Francisco, with its (very) roughly 80,000 pianos, needs about 80 piano tuners. This is, of course, only a rough estimate.[‡] It tells us that there must be many more than 10 piano tuners, and surely not as many as 1000.

FIGURE 1–11 Enrico Fermi. Fermi contributed significantly to both theoretical and experimental physics, a feat almost unique in modern times.



PROBLEM SOLVING
Estimating how many piano tuners there are in a city

[†]As a teenager I had a summer job washing dishes at a camp located 350 m above famous Lake Tahoe in California. Starting the drive down to Lake Tahoe, the beaches across the lake were visible. But approaching the level of Lake Tahoe, the beaches across the lake were no longer visible! I realized that Lake Tahoe was bulging up in the middle, blocking the view. ("The Earth is round.")

[‡]A search on the internet (done after this calculation) reveals over 50 listings. Each of these listings may employ more than one tuner, but on the other hand, each may also do repairs as well as tuning. In any case, our estimate is reasonable.

*1–7 Dimensions and Dimensional Analysis

When we speak of the **dimensions** of a quantity, we are referring to the type of base units that make it up. The dimensions of area, for example, are always length squared, abbreviated $[L^2]$ using square brackets; the units can be square meters, square feet, cm^2, and so on. Velocity, on the other hand, can be measured in units of km/h, m/s, or mi/h, but the dimensions are always a length $[L]$ divided by a time $[T]$: that is, $[L/T]$.

The formula for a quantity may be different in different cases, but the dimensions remain the same. For example, the area of a triangle of base b and height h is $A = \frac{1}{2}bh$, whereas the area of a circle of radius r is $A = \pi r^2$. The formulas are different in the two cases, but the dimensions of area are always $[L^2]$.

Dimensions can be used as a help in working out relationships, a procedure referred to as **dimensional analysis**. One useful technique is the use of dimensions to check if a relationship is *incorrect*. Note that we add or subtract quantities only if they have the same dimensions (we don't add centimeters and hours); and the quantities on each side of an equals sign must have the same dimensions. (In numerical calculations, the units must also be the same on both sides of an equation.)

For example, suppose you derived the equation $v = v_0 + \frac{1}{2}at^2$, where v is the velocity of an object after a time t, v_0 is the object's initial velocity, and the object undergoes an acceleration a. Let's do a dimensional check to see if this equation could be correct or is surely incorrect. Note that numerical factors, like the $\frac{1}{2}$ here, do not affect dimensional checks. We write a dimensional equation as follows, remembering that the dimensions of velocity are $[L/T]$ and (as we shall see in Chapter 2) the dimensions of acceleration are $[L/T^2]$:

$$\left[\frac{L}{T}\right] \stackrel{?}{=} \left[\frac{L}{T}\right] + \left[\frac{L}{T^2}\right][T^2]$$
$$\stackrel{?}{=} \left[\frac{L}{T}\right] + [L].$$

The dimensions are incorrect: on the right side, we have the sum of quantities whose dimensions are not the same. Thus we conclude that an error was made in the derivation of the original equation.

A dimensional check can only tell you when a relationship is wrong. It can not tell you if it is completely right. For example, a dimensionless numerical factor (such as $\frac{1}{2}$ or 2π) could be missing.

Dimensional analysis can also be used as a quick check on an equation you are not sure about. For example, consider a simple pendulum of length ℓ. Suppose that you can't remember whether the equation for the period T (the time to make one back-and-forth swing) is $T = 2\pi\sqrt{\ell/g}$ or $T = 2\pi\sqrt{g/\ell}$, where g is the acceleration due to gravity and, like all accelerations, has dimensions $[L/T^2]$. (Do not worry about these formulas—the correct one will be derived in Chapter 11; what we are concerned about here is a person's recalling whether it contains ℓ/g or g/ℓ.) A dimensional check shows that the former (ℓ/g) is correct:

$$[T] = \sqrt{\frac{[L]}{[L/T^2]}} = \sqrt{[T^2]} = [T],$$

whereas the latter (g/ℓ) is not:

$$[T] \neq \sqrt{\frac{[L/T^2]}{[L]}} = \sqrt{\frac{1}{[T^2]}} = \frac{1}{[T]}.$$

The constant 2π has no dimensions and so can't be checked using dimensions.

Further uses of dimensional analysis are found in Appendix D.

*Some Sections of this book, such as this one, may be considered *optional* at the discretion of the instructor, and they are marked with an asterisk (*). See the Preface for more details.

EXAMPLE 1–10 Planck length. The smallest meaningful measure of length is called the "Planck length," and is defined in terms of three fundamental constants in nature: the speed of light $c = 3.00 \times 10^8\,\text{m/s}$, the gravitational constant $G = 6.67 \times 10^{-11}\,\text{m}^3/\text{kg}\cdot\text{s}^2$, and Planck's constant $h = 6.63 \times 10^{-34}\,\text{kg}\cdot\text{m}^2/\text{s}$. The Planck length λ_P (λ is the Greek letter "lambda") is given by the following combination of these three constants:

$$\lambda_P = \sqrt{\frac{Gh}{c^3}}.$$

Show that the dimensions of λ_P are length $[L]$, and find the order of magnitude of λ_P.

APPROACH We rewrite the above equation in terms of dimensions. The dimensions of c are $[L/T]$, of G are $[L^3/MT^2]$, and of h are $[ML^2/T]$.

SOLUTION The dimensions of λ_P are

$$\sqrt{\frac{[L^3/MT^2][ML^2/T]}{[L^3/T^3]}} = \sqrt{[L^2]} = [L]$$

which is a length. Good. The value of the Planck length is

$$\lambda_P = \sqrt{\frac{Gh}{c^3}} = \sqrt{\frac{(6.67 \times 10^{-11}\,\text{m}^3/\text{kg}\cdot\text{s}^2)(6.63 \times 10^{-34}\,\text{kg}\cdot\text{m}^2/\text{s})}{(3.00 \times 10^8\,\text{m/s})^3}} \approx 4 \times 10^{-35}\,\text{m},$$

which is on the order of 10^{-34} or $10^{-35}\,\text{m}$.

NOTE Some recent theories (Chapters 43 and 44) suggest that the smallest particles (quarks, leptons) have sizes on the order of the Planck length, $10^{-35}\,\text{m}$. These theories also suggest that the "Big Bang," with which the Universe is believed to have begun, started from an initial size on the order of the Planck length.

Summary

[*The Summary that appears at the end of each Chapter in this book gives a brief overview of the main ideas of the Chapter. The Summary* cannot *serve to give an understanding of the material, which can be accomplished only by a detailed reading of the Chapter.*]

Physics, like other sciences, is a creative endeavor. It is not simply a collection of facts. Important **theories** are created with the idea of explaining **observations**. To be accepted, theories are **tested** by comparing their predictions with the results of actual experiments. Note that, in general, a theory cannot be "proved" in an absolute sense.

Scientists often devise models of physical phenomena. A **model** is a kind of picture or analogy that helps to describe the phenomena in terms of something we already know about. A **theory**, often developed from a model, is usually deeper and more complex than a simple model.

A scientific **law** is a concise statement, often expressed in the form of an equation, which quantitatively describes a wide range of phenomena.

Measurements play a crucial role in physics, but can never be perfectly precise. It is important to specify the **uncertainty** of a measurement either by stating it directly using the $\pm$ notation, and/or by keeping only the correct number of **significant figures**.

Physical quantities are always specified relative to a particular standard or **unit**, and the unit used should always be stated. The commonly accepted set of units today is the **Système International** (SI), in which the standard units of length, mass, and time are the **meter**, **kilogram**, and **second**.

When converting units, check all **conversion factors** for correct cancellation of units.

Making rough, **order-of-magnitude estimates** is a very useful technique in science as well as in everyday life.

[*The **dimensions** of a quantity refer to the combination of base quantities that comprise it. Velocity, for example, has dimensions of [length/time] or $[L/T]$. Working with only the dimensions of the various quantities in a given relationship—this technique is called **dimensional analysis**—makes it possible to check a relationship for correct form.]

Questions

1. What are the merits and drawbacks of using a person's foot as a standard? Consider both (*a*) a particular person's foot, and (*b*) any person's foot. Keep in mind that it is advantageous that fundamental standards be accessible (easy to compare to), invariable (do not change), indestructible, and reproducible.
2. What is wrong with this road sign:

   Memphis 7 mi (11.263 km)?
3. Why is it incorrect to think that the more digits you include in your answer, the more accurate it is?
4. For an answer to be complete, units need to be specified. Why?
5. You measure the radius of a wheel to be 4.16 cm. If you multiply by 2 to get the diameter, should you write the result as 8 cm or as 8.32 cm? Justify your answer.
6. Express the sine of 30.0° with the correct number of significant figures.
7. List assumptions useful to estimate the number of car mechanics in (*a*) San Francisco, (*b*) your hometown, and then make the estimates.

MisConceptual Questions

[List all answers that are valid.]

1. The laws of physics
(*a*) are permanent and unalterable.
(*b*) are part of nature and are waiting to be discovered.
(*c*) can change, but only because of evidence that convinces the community of physicists.
(*d*) apply to physics but not necessarily to chemistry or other fields.
(*e*) were basically complete by 1900, and have undergone only minor revisions since.
(*f*) are accepted by all major world countries, and cannot be changed without international treaties.

2. How should we write the result of the following calculation, being careful about significant figures?

$$(3.84\text{ s})(37\text{ m/s}) + (5.3\text{ s})(14.1\text{ m/s}) =$$

(*a*) 200 m.
(*b*) 210 m.
(*c*) 216.81 m.
(*d*) 217 m.
(*e*) 220 m.

3. Four students use different instruments to measure the length of the same pen. Which measurement implies the greatest precision?
(*a*) 160.0 mm.
(*b*) 16.0 cm.
(*c*) 0.160 m.
(*d*) 0.00016 km.
(*e*) Need more information.

4. The number 0.0078 has how many significant figures?
(*a*) 1.
(*b*) 2.
(*c*) 3.
(*d*) 4.

5. How many significant figures does $1.362 + 25.2$ have?
(*a*) 2.
(*b*) 3.
(*c*) 4.
(*d*) 5.

6. Accuracy represents
(*a*) repeatability of a measurement, using a given instrument.
(*b*) how close a measurement is to the true value.
(*c*) an ideal number of measurements to make.
(*d*) how poorly an instrument is operating.

7. Precision represents
(*a*) repeatability of a measurement, using a given instrument.
(*b*) how close a measurement is to the true value.
(*c*) an ideal number of measurements to make.
(*d*) how poorly an instrument is operating.

8. To convert from ft^2 to yd^2, you should
(*a*) multiply by 3.
(*b*) multiply by 1/3.
(*c*) multiply by 9.
(*d*) multiply by 1/9.
(*e*) multiply by 6.
(*f*) multiply by 1/6.

9. Which is *not* true about an order-of-magnitude estimation?
(*a*) It gives you a rough idea of the answer.
(*b*) It can be done by keeping only one significant figure.
(*c*) It can be used to check if an exact calculation is reasonable.
(*d*) It may require making some reasonable assumptions in order to calculate the answer.
(*e*) It will always be accurate to at least two significant figures.

***10.** $[L^2]$ represents the dimensions for which of the following?
(*a*) cm^2.
(*b*) square feet.
(*c*) m^2.
(*d*) All of the above.

Problems

[*The Problems at the end of each Chapter are ranked* I, II, *or* III *according to estimated difficulty, with* (I) *Problems being easiest. Level* III *are meant as challenges for the best students. The Problems are arranged by Section, meaning that the reader should have read up to and including that Section, but not only that Section—Problems often depend on earlier material. Next is a set of "General Problems" not arranged by Section and not ranked.*]

1–3 Measurement, Uncertainty, Significant Figures

(Note: *In Problems, assume a number like* 6.4 *is accurate to* ± 0.1; *and* 950 *is accurate to* 2 *significant figures* (± 10) *unless* 950 *is said to be "precisely" or "very nearly"* 950, *in which case assume* 950 ± 1.)

1. (I) How many significant figures do each of the following numbers have: (*a*) 777, (*b*) 81.60, (*c*) 7.03, (*d*) 0.03, (*e*) 0.0086, (*f*) 6465, and (*g*) 8700?

2. (I) Write the following numbers in powers of 10 notation: (*a*) 5.859, (*b*) 21.8, (*c*) 0.0068, (*d*) 328.65, (*e*) 0.219, (*f*) 444.

3. (I) Write out the following numbers in full with the correct number of zeros: (*a*) 8.69×10^5, (*b*) 9.1×10^3, (*c*) 2.5×10^{-1}, (*d*) 4.76×10^2, and (*e*) 3.62×10^{-5}.

4. (II) What is the percent uncertainty in the measurement 3.25 ± 0.35 m?

5. (II) Time intervals measured with a physical stopwatch typically have an uncertainty of about 0.2 s, due to human reaction time at the start and stop moments. What is the percent uncertainty of a hand-timed measurement of (*a*) 4.5 s, (*b*) 45 s, (*c*) 4.5 min?

6. (II) Add $(9.2 \times 10^3\text{ s}) + (6.3 \times 10^4\text{ s}) + (0.008 \times 10^6\text{ s})$.

7. (II) Multiply 4.079×10^2 m by 0.057×10^{-1} m, taking into account significant figures.

8. (II) What, approximately, is the percent uncertainty for a measurement given as 1.27 m^2?

9. (II) For small angles θ, the numerical value of $\sin\theta$ is approximately the same as the numerical value of $\tan\theta$. Find the largest angle for which sine and tangent agree to within two significant figures.

10. (II) A report stated that "a survey of 215 students found that 37.2% had bought a sugar-rich soft drink the day before." (*a*) How many students bought a soft drink? (*b*) What is wrong with the original statement?

11. (II) A watch manufacturer claims that its watches gain or lose no more than 9 seconds in a year. How accurate are these watches, expressed as a percentage?

12. (III) What is the area, and its approximate uncertainty, of a circle of radius 5.1×10^4 cm?

13. (III) What, roughly, is the percent uncertainty in the volume of a spherical beach ball of radius $r = 0.64 \pm 0.04$ m?

1–4 and 1–5 Units, Standards, SI, Converting Units

14. (I) Write the following as full (decimal) numbers without prefixes on the units: (*a*) 286.6 mm, (*b*) 74 μV, (*c*) 430 mg, (*d*) 47.2 ps, (*e*) 22.5 nm, (*f*) 2.50 gigavolts.

15. (I) Express the following using the prefixes of Table 1–4: (*a*) 3×10^6 volts, (*b*) 2×10^{-6} meters, (*c*) 5×10^3 days, (*d*) 18×10^2 bucks, and (*e*) 9×10^{-7} seconds.

16. (I) Determine your own height in meters, and your mass in kg.

17. (II) To the correct number of significant figures, use the information inside the front cover of this book to determine the ratio of (*a*) the surface area of Earth compared to the surface area of the Moon, (*b*) the volume of Earth compared to the volume of the Moon.

18. (II) Would a driver traveling at 15 m/s in a 35 mi/h zone be exceeding the speed limit? Why or why not?

19. (II) The age of the universe is thought to be about 14 billion years. Assuming two significant figures, write this in powers of 10 in (*a*) years, (*b*) seconds.

20. (II) The Sun, on average, is 93 million miles from Earth. How many meters is this? Express (*a*) using powers of 10, and (*b*) using a metric prefix (km).

21. (II) Express the following sum with the correct number of significant figures: $1.90\,\text{m} + 142.5\,\text{cm} + 6.27 \times 10^5\,\mu\text{m}$.

22. (II) A typical atom has a diameter of about 1.0×10^{-10} m. (*a*) What is this in inches? (*b*) Approximately how many atoms are along a 1.0-cm line, assuming they just touch?

23. (II) Determine the conversion factor between (*a*) km/h and mi/h, (*b*) m/s and ft/s, and (*c*) km/h and m/s.

24. (II) What is the conversion factor between (*a*) ft^2 and yd^2, (*b*) m^2 and ft^2?

25. (II) A **light-year** is the distance light travels in one year $\left(\text{at speed} = 2.998 \times 10^8\,\text{m/s}\right)$. (*a*) How many meters are there in 1.00 light-year? (*b*) An **astronomical unit** (AU) is the average distance from the Sun to Earth, 1.50×10^8 km. How many AU are there in 1.00 light-year?

26. (II) How much longer (percentage) is a one-mile race than a 1500-m race ("the metric mile")?

27. (II) How many wavelengths of orange krypton-86 light (Section 1–4) would fit into the thickness of one page of this book? See Example 1–6.

28. (II) Using the French Academy of Sciences' original definition of the meter, calculate Earth's circumference and radius in *those* meters. Give % error relative to today's accepted values (inside front cover).

29. (II) A passenger jet uses about 12 liters of fuel per km of flight. What is that value expressed as miles per gallon?

30. (II) American football uses a *field* that is 100.0 yd long, whereas a *soccer field* is 100.0 m long. Which field is longer, and by how much (give yards, meters, and percent)?

31. (II) (*a*) How many seconds are there in 1.00 year? (*b*) How many nanoseconds are there in 1.00 year? (*c*) How many years are there in 1.00 second?

32. (II) Use Table 1–3 to estimate the total number of protons or neutrons in (*a*) a bacterium, (*b*) a DNA molecule, (*c*) the human body, (*d*) our Galaxy.

33. (II) The diameter of the planet Mercury is 4879 km. (*a*) What is the surface area of Mercury? (*b*) How many times larger is the surface area of the Earth?

34. (III) A standard baseball has a circumference of approximately 23 cm. If a baseball had the same mass per unit volume (see Tables in Section 1–4) as a neutron or a proton, about what would its mass be?

1–6 Order-of-Magnitude Estimating

(Note: *Remember that for rough estimates, only round numbers are needed both as input to calculations and as final results.*)

35. (I) Estimate the order of magnitude (power of 10) of: (*a*) 3200, (*b*) 86.30×10^3, (*c*) 0.076, and (*d*) 15.0×10^8.

36. (II) Estimate how many books can be shelved in a college library with 6500 m^2 of floor space. Assume 8 shelves high, having books on both sides, with corridors 1.5 m wide. Assume books are about the size of this one, on average.

37. (II) Estimate how many hours it would take to run (at 10 km/h) across the U.S. from New York to California.

38. (II) Estimate the number of liters of water a human drinks in a lifetime.

39. (II) Estimate the number of *cells* in an adult human body, given that a typical cell has a diameter of about 10 μm, and the human body has a density of about 1000 kg/m^3.

40. (II) Estimate how long it would take one person to mow a football field using an ordinary home lawn mower (Fig. 1–12). (State your assumptions, such as the mower moves with a 1-km/h speed, and has a 0.5-m width.)



FIGURE 1–12
Problem 40.

41. (II) Estimate the number of gallons of gasoline consumed by the total of all automobile drivers in the U.S., per year.

42. (II) Estimate the number of dentists (*a*) in San Francisco and (*b*) in your town or city.

43. (II) Estimate how many kilograms of laundry soap are used in the U.S. in one year (and therefore pumped out of washing machines with the dirty water). Assume each load of laundry takes 0.1 kg of soap.

44. (II) How big is a *ton* (1000 kg)? That is, what is the volume of something that weighs a ton? To be specific, estimate the diameter of a 1-ton rock, but first make a wild guess: will it be 1 ft across, 3 ft, or the size of a car? [*Hint*: Rock has mass per volume about 3 times that of water, which is 1 kg per liter $\left(10^3\,\text{cm}^3\right)$ or 62 lb per cubic foot.]

45. (II) A hiking trail is 270 km long through varying terrain. A group of hikers cover the first 49 km in two and a half days. Estimate how much time they should allow for the rest of the trip.

46. (II) Estimate how many days it would take to walk around the circumference of the Earth, assuming 12 h walking per day at 4 km/h.

47. (II) Estimate the number of jelly beans in the jar of Fig. 1–13.



FIGURE 1–13
Problem 47. Estimate the number of jelly beans in the jar.

48. (II) Estimate the number of bus drivers (*a*) in Washington, D.C., and (*b*) in your town.

49. (III) You are in a hot air balloon, 300 m above the flat Texas plains. You look out toward the horizon. How far out can you see—that is, how far is your horizon? The Earth's radius is about 6400 km.

50. (III) I agree to hire you for 30 days. You can decide between two methods of payment: either (1) $1000 a day, or (2) one penny on the first day, two pennies on the second day and continue to double your daily pay each day up through day 30. Use quick estimation to make your decision, and justify it.

51. (III) The rubber worn from tires mostly enters the atmosphere as *particulate pollution*. Estimate how much rubber (in kg) is put into the air in the United States every year. To get started, a good estimate for a tire tread's depth is 1 cm when new, and rubber has a mass of about 1200 kg per m^3 of volume.

52. (III) Many sailboats are docked at a marina 4.4 km away on the opposite side of a lake. You stare at one of the sailboats because, when you are lying flat at the water's edge, you can just see its deck but none of the side of the sailboat. You then go to that sailboat on the other side of the lake and measure that the deck is 1.5 m above the level of the water. Using Fig. 1–14, where $h = 1.5$ m, estimate the radius R of the Earth.

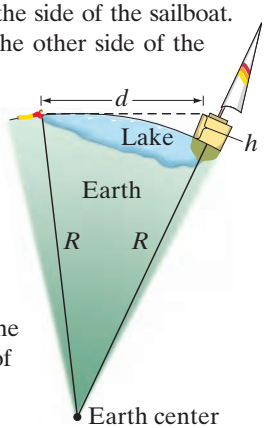


FIGURE 1–14 Problem 52. You see a sailboat across a lake (not to scale). R is the radius of the Earth. Because of the curvature of the Earth, the water "bulges out" between you and the boat.

53. (III) You are lying on a beach, your eyes 20 cm above the sand. Just as the Sun sets, fully disappearing over the horizon, you immediately jump up, your eyes now 150 cm above the sand, and you can again just see the top of the Sun. If you count the number of seconds $(= t)$ until the Sun fully disappears again, you can estimate the *Earth's radius*. But for this Problem, use the known radius of the Earth to calculate the time t.

*1–7 Dimensions

***54.** (I) What are the dimensions of density, which is mass per volume?

***55.** (II) The speed v of an object is given by the equation $v = At^3 - Bt$, where t refers to time. (*a*) What are the dimensions of A and B? (*b*) What are the SI units for the constants A and B?

***56.** (II) Three students derive the following equations in which x refers to distance traveled, v the speed, a the acceleration (m/s^2), t the time, and the subscript zero $(_0)$ means a quantity at time $t = 0$. Here are their equations: (*a*) $x = vt^2 + 2at$, (*b*) $x = v_0 t + \frac{1}{2}at^2$, and (*c*) $x = v_0 t + 2at^2$. Which of these could possibly be correct according to a dimensional check, and why?

***57.** (II) (*a*) Show that the following combination of the three fundamental constants of nature that we used in Example 1–10 (that is G, c, and h) forms a quantity with the dimensions of time:

$$t_P = \sqrt{\frac{Gh}{c^5}}.$$

This quantity, t_P, is called the **Planck time** and is thought to be the earliest time, after the creation of the Universe, at which the currently known laws of physics can be applied. (*b*) Estimate the order of magnitude of t_P using values given inside the front cover (or Example 1–10).

General Problems

58. **Global positioning satellites (GPS)** can be used to determine your position with great accuracy. If one of the satellites is 20,000 km from you, and you want to know your position to ± 2 m, what percent uncertainty in the distance is required? How many significant figures are needed in the distance?

59. One mole of atoms consists of 6.02×10^{23} individual atoms. If a mole of atoms were spread uniformly over the Earth's surface, how many atoms would there be per square meter?

60. **Computer chips** (Fig. 1–15) can be etched on circular silicon wafers of thickness 0.300 mm that are sliced from a solid cylindrical silicon crystal of length 25 cm. If each wafer can hold 750 chips, what is the maximum number of chips that can be produced from one entire cylinder?

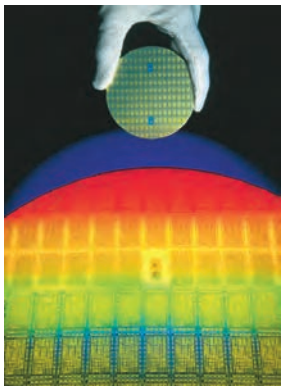


FIGURE 1–15 Problem 60. The wafer held by the hand is shown below, enlarged and illuminated by colored light. Visible are rows of integrated circuits (chips).

61. If you used only a keyboard to enter data, how many years would it take to fill up a *hard drive* in a computer that can store 1.0 terabytes $(1.0 \times 10^{12}$ bytes) of data? Assume 40-hour work weeks, and that you can type 150 characters per minute, and that one byte is one keyboard character.

62. An average family of four uses roughly 1200 L (about 300 gallons) of water per day $(1\text{ L} = 1000\text{ cm}^3)$. How much depth would a lake lose per year if it covered an area of 60 km^2 with uniform depth and supplied a local town with a population of 40,000 people? Consider only population uses, and neglect evaporation, rain, creeks and rivers.

63. A certain compact disc (CD) contains 783.216 megabytes of digital information. Each byte consists of exactly 8 bits. When played, a CD player reads the CD's information at a constant rate of 1.4 megabits per second. How many minutes does it take the player to read the entire CD?

64. An *angstrom* (symbol Å) is a unit of length, defined as 10^{-10} m, which is on the order of the diameter of an atom. (*a*) How many nanometers are in 1.0 angstrom? (*b*) How many femtometers or fermis (the common unit of length in nuclear physics) are in 1.0 angstrom? (*c*) How many angstroms are in 1.0 m? (*d*) How many angstroms are in 1.0 light-year (see Problem 25)?

65. A typical adult human lung contains about 300 million tiny cavities called alveoli. Estimate the average diameter of a single alveolus.

66. Hold a pencil in front of your eye at a position where its blunt end just blocks out the Moon (Fig. 1–16). Make appropriate measurements to estimate the diameter of the Moon, given that the Earth–Moon distance is 3.8×10^5 km.

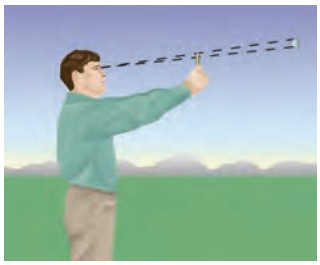


FIGURE 1–16
Problem 66. How big is the Moon?

67. A storm dumps 1.0 cm of rain on a city 5 km wide and 7 km long in a 2-h period. How many metric tons (1 metric ton $= 10^3$ kg) of water fell on the city? ($1\ \text{cm}^3$ of water has a mass of $1\ \text{g} = 10^{-3}$ kg.) How many gallons of water was this?

68. Greenland's ice sheet covers over $1.7 \times 10^6\ \text{km}^2$ and is approximately 2.5 km thick. If it were to melt completely then by how much would you expect the ocean to rise? Assume $\frac{2}{3}$ of Earth's surface is ocean. See Tables inside front and back covers.

69. Noah's ark was ordered to be 300 cubits long, 50 cubits wide, and 30 cubits high. The cubit was a unit of measure equal to the length of a human forearm, elbow to the tip of the longest finger. Express the dimensions of Noah's ark in meters, and estimate its volume (m^3).

70. One liter ($1000\ \text{cm}^3$) of oil is spilled onto a smooth lake. If the oil spreads out uniformly until it makes an oil slick just one molecule thick, with adjacent molecules just touching, estimate the diameter of the oil slick. Assume the oil molecules have a diameter of 2×10^{-10} m.

71. If you walked north along one of Earth's lines of longitude until you had changed latitude by 1 minute of arc (there are 60 minutes per degree), how far would you have walked (in miles)? This distance is a *nautical mile* (page 7).

72. Determine the percent uncertainty in θ, and in $\sin\theta$, when (*a*) $\theta = 15.0° \pm 0.5°$, (*b*) $\theta = 75.0° \pm 0.5°$.

73. Jim stands beside a wide river and wonders how wide it is. He spots a large rock on the bank directly across from him. He then walks upstream 85 strides and judges that the angle between him and the rock, which he can still see, is now at an angle of 30° downstream (Fig. 1–17). Jim measures his stride to be about 0.8 m long. Estimate the width of the river.

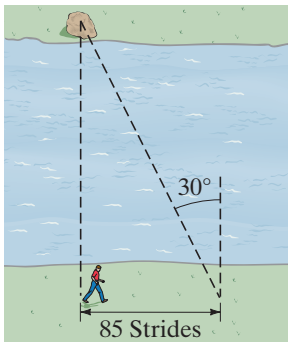


FIGURE 1–17
Problem 73.

74. Make a rough estimate of the volume of your body (in m^3).

75. Estimate the number of plumbers in San Francisco.

76. Estimate the ratio (order of magnitude) of the mass of a human to the mass of a DNA molecule. [*Hint*: Check the Tables in this Chapter.]

77. The following formula estimates an average person's lung capacity V (in liters, where $1\ \text{L} = 10^3\ \text{cm}^3$):

$$V = 4.1H - 0.018A - 2.7,$$

where H and A are the person's height (in meters) and age (in years), respectively. In this formula, what are the units of the numbers 4.1, 0.018, and 2.7?

78. The density of an object is defined as its mass divided by its volume. Suppose a rock's mass and volume are measured to be 6 g and $2.8325\ \text{cm}^3$. To the correct number of significant figures, determine the rock's density (mass/volume).

79. Recent findings in astrophysics suggest that the observable universe can be modeled as a sphere of radius $R = 13.7 \times 10^9$ light-years $= 13.0 \times 10^{25}$ m with an average total mass density of about $1 \times 10^{-26}\ \text{kg/m}^3$. Only about 4% of total mass is due to "ordinary" matter (such as protons, neutrons, and electrons). Estimate how much ordinary matter (in kg) there is in the observable universe. (For the light-year, see Problem 25.)

ANSWERS TO EXERCISES

A: (*d*).
B: All three have three significant figures; the number of decimal places is (*a*) 2, (*b*) 3, (*c*) 4.
C: No: they have three and two, respectively.
D: (*a*) 2.58×10^{-2}, 3; (*b*) 4.23×10^4, 3 (probably); (*c*) 3.4450×10^2, 5.
E: Mt. Everest, 29,035 ft; K2, 28,251 ft; Kangchenjunga, 28,169 ft.
F: (*a*) 2.47 acres in 1 hectare; (*b*) $2\frac{1}{2}$ or even just 2 acres in 1 hectare.
G: (*f*) 1,000,000; that is, one million.
H: (*c*).

A space shuttle has released a parachute to reduce its speed quickly. The directions of the shuttle's velocity and acceleration are shown by the green ($\vec{\mathbf{v}}$) and gold ($\vec{\mathbf{a}}$) arrows.

Motion is described using the concepts of velocity and acceleration. In the case shown here, the velocity $\vec{\mathbf{v}}$ is to the right, in the direction of motion. The acceleration $\vec{\mathbf{a}}$ is in the opposite direction from the velocity $\vec{\mathbf{v}}$, which means the object is slowing down.

We examine in detail motion with constant acceleration, including the vertical motion of objects falling under gravity.



CHAPTER 2

Describing Motion: Kinematics in One Dimension

CONTENTS

2–1 Reference Frames and Displacement
2–2 Average Velocity
2–3 Instantaneous Velocity
2–4 Acceleration
2–5 Motion at Constant Acceleration
2–6 Solving Problems
2–7 Freely Falling Objects
*2–8 Variable Acceleration; Integral Calculus

CHAPTER-OPENING QUESTION—Guess now!

[*Don't worry about getting the right answer now—you will get another chance later in the Chapter. See also page 1 of Chapter 1 for more explanation.*]

Two small heavy balls have the same diameter but one weighs twice as much as the other. The balls are dropped from a second-story balcony at the exact same time. The time to reach the ground below will be:

(a) twice as long for the lighter ball as for the heavier one.
(b) longer for the lighter ball, but not twice as long.
(c) twice as long for the heavier ball as for the lighter one.
(d) longer for the heavier ball, but not twice as long.
(e) nearly the same for both balls.

The motion of objects—baseballs, automobiles, joggers, and even the Sun and Moon—is an obvious part of everyday life. It was not until the sixteenth and seventeenth centuries that our modern understanding of motion was established. Many individuals contributed to this understanding, particularly Galileo Galilei (1564–1642) and Isaac Newton (1642–1727).

The study of the motion of objects, and the related concepts of force and energy, form the field called **mechanics**. Mechanics is customarily divided into two parts: **kinematics**, which is the description of how objects move, and **dynamics**, which deals with force and why objects move as they do. This Chapter and the next deal with kinematics.

For now we only discuss objects that move without rotating (Fig. 2–1a). Such motion is called **translational motion**. In this Chapter we will be concerned with describing an object that moves along a straight-line path, which is one-dimensional translational motion. In Chapter 3 we will describe translational motion in two (or three) dimensions along paths that are not straight. (Rotation, shown in Fig. 2–1b, is discussed in Chapters 10 and 11.)

We will often use the concept, or *model*, of an idealized **particle** which is considered to be a mathematical **point** with no spatial extent (no size). A point particle can undergo only translational motion. The particle model is useful in many real situations where we are interested only in translational motion and the object's size is not significant. For example, we might consider a billiard ball, or even a spacecraft traveling toward the Moon, as a particle for many purposes.

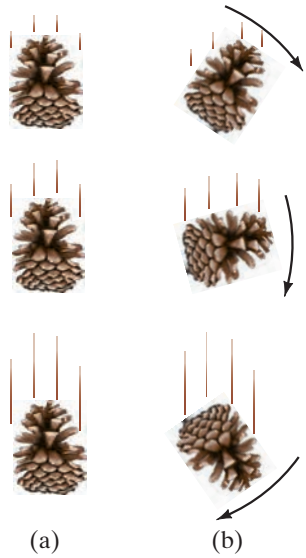


FIGURE 2–1 A falling pinecone undergoes (a) pure translation; (b) it is rotating as well as translating.

2–1 Reference Frames and Displacement

Any measurement of position, distance, or speed must be made with respect to a **reference frame**, or **frame of reference**. For example, while you are on a train traveling at 80 km/h, suppose a person walks past you toward the front of the train at a speed of, say, 5 km/h (Fig. 2–2). This 5 km/h is the person's speed with respect to the train as frame of reference. With respect to the ground, that person is moving at a speed of 80 km/h + 5 km/h = 85 km/h. It is always important to specify the frame of reference when stating a speed. In everyday life, we usually mean "with respect to the Earth" without even thinking about it, but the reference frame must be specified whenever there might be confusion.

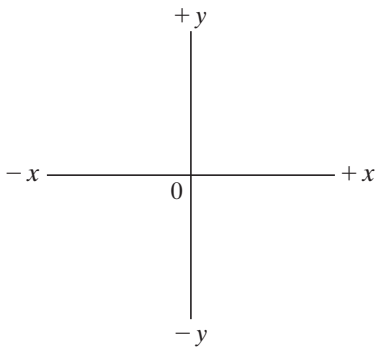
FIGURE 2–2 A person walks toward the front of a train at 5 km/h. The train is moving at 80 km/h with respect to the ground, so the walking person's speed, relative to the ground, is 85 km/h.



When specifying the motion of an object, it is important to specify not only the speed but also the direction of motion. Often we can specify a direction by using north, east, south, and west, and by "up" and "down." In physics, we often draw a set of **coordinate axes**, as shown in Fig. 2–3, to represent a frame of reference. We can always place the origin 0, and the directions of the x and y axes, as we like for convenience. The x and y axes are always perpendicular to each other. The **origin** is where $x = 0$, $y = 0$. Objects positioned to the right of the origin of coordinates (0) on the x axis have an x coordinate which we almost always choose to be positive; objects at points to the left of 0 have a negative x coordinate. The position along the y axis is usually considered positive when above 0, and negative when below 0, although the reverse convention can be used if convenient. Any point on the xy plane can be specified by giving its x and y coordinates. In three dimensions, a z axis perpendicular to the x and y axes is added.

For one-dimensional motion, we often choose the x axis as the line along which the motion takes place. Then the **position** of an object at any moment is given by its x coordinate. If the motion is vertical, as for a dropped object, we usually use the y axis.

FIGURE 2–3 Standard set of xy coordinate axes, sometimes called "rectangular coordinates." [Also called *Cartesian coordinates*, after René Descartes (1596–1650), who invented them.]



⚠ CAUTION

The displacement may not equal the total distance traveled

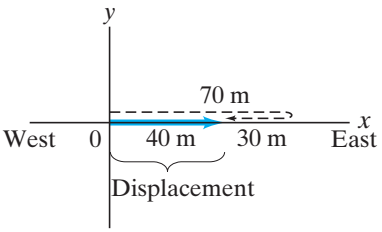


FIGURE 2–4 A person walks 70 m east, then 30 m west. The total distance traveled is 100 m (path is shown dashed in black); but the displacement, shown as a solid blue arrow, is 40 m to the east.

FIGURE 2–5 The arrow represents the displacement $x_2 - x_1$. Distances are in meters.

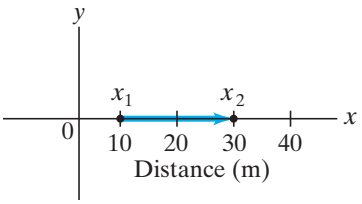
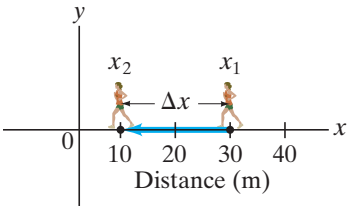


FIGURE 2–6 For the displacement $\Delta x = x_2 - x_1 = 10.0\,\text{m} - 30.0\,\text{m}$, the displacement vector points left.



We need to make a distinction between the **distance** an object has traveled and its **displacement**, which is defined as the *change in position* of the object.

That is, *displacement is how far the object is from its starting point*. To see the distinction between total distance and displacement, imagine a person walking 70 m to the east and then turning around and walking back (west) a distance of 30 m (see Fig. 2–4). The total distance traveled is 70 m + 30 m = 100 m, but the displacement is only 40 m since the person is now only 40 m from the starting point.

Displacement is a quantity that has both *magnitude* and *direction*. Such quantities are called **vectors**, and are represented by arrows in diagrams. For example, in Fig. 2–4, the blue arrow represents the displacement whose magnitude is 40 m and whose direction is to the right (east).

We will deal with vectors more fully in Chapter 3. For now, we deal only with motion in one dimension, along a line. In this case, vectors which point in one direction will be positive (usually to the right along the x axis). Vectors that point in the opposite direction will have a negative sign in front of their magnitude.

Consider the motion of an object over a particular time interval. Suppose that at some initial time, call it t_1, the object is on the x axis at the position x_1 in the coordinate system shown in Fig. 2–5. At some later time, t_2, suppose the object has moved to position x_2. The displacement of our object is $x_2 - x_1$, and is represented by the arrow pointing to the right in Fig. 2–5. It is convenient to write

$$\Delta x = x_2 - x_1,$$

where the symbol Δ (Greek letter delta) means "change in." Then Δx means "the change in x," or "change in position," which is in fact the displacement. The **change in** any quantity means the *final* value of that quantity, minus the *initial* value. Suppose $x_1 = 10.0\,\text{m}$ and $x_2 = 30.0\,\text{m}$, as in Fig. 2–5. Then

$$\Delta x = x_2 - x_1 = 30.0\,\text{m} - 10.0\,\text{m} = 20.0\,\text{m},$$

so the displacement is 20.0 m in the positive direction, Fig. 2–5.

Now consider an object moving to the left as shown in Fig. 2–6. Here the object, a person, starts at $x_1 = 30.0\,\text{m}$ and walks to the left to the point $x_2 = 10.0\,\text{m}$. In this case her displacement is

$$\Delta x = x_2 - x_1 = 10.0\,\text{m} - 30.0\,\text{m} = -20.0\,\text{m},$$

and the blue arrow representing the vector displacement points to the left. For one-dimensional motion along the x axis, a vector pointing to the right is positive, whereas a vector pointing to the left has a negative sign.

EXERCISE A An ant starts at $x = 20\,\text{cm}$ on a piece of graph paper and walks along the x axis to $x = -20\,\text{cm}$. It then turns around and walks back to $x = -10\,\text{cm}$. Determine (a) the ant's displacement and (b) the total distance traveled.

2–2 Average Velocity

An important aspect of the motion of a moving object is how *fast* it is moving—its speed or velocity.

The term "speed" refers to how far an object travels in a given time interval, regardless of direction. If a car travels 240 kilometers (km) in 3 hours (h), we say its average speed was 80 km/h. In general, the **average speed** of an object is defined as *the total distance traveled along its path divided by the time it takes to travel this distance*:

$$\text{average speed} = \frac{\text{distance traveled}}{\text{time elapsed}}. \qquad \textbf{(2–1)}$$

The terms "velocity" and "speed" are often used interchangeably in ordinary language. But in physics we make a distinction between the two. Speed is simply a positive number, with units. **Velocity**, on the other hand, is used to signify both the *magnitude* (numerical value) of how fast an object is moving and also the *direction* in which it is moving. Velocity is therefore a *vector*.

There is a second difference between speed and velocity: namely, the *average velocity* is defined in terms of *displacement*, rather than total distance traveled:

$$\text{average velocity} = \frac{\text{displacement}}{\text{time elapsed}} = \frac{\text{final position} - \text{initial position}}{\text{time elapsed}}.$$

CAUTION
Average speed is not necessarily equal to the magnitude of the average velocity

Average speed and average velocity have the same magnitude when the motion is all in one direction. In other cases, they may differ: recall the walk we described earlier, in Fig. 2–4, where a person walked 70 m east and then 30 m west. The total distance traveled was 70 m + 30 m = 100 m, but the displacement was 40 m. Suppose this walk took 70 s to complete. Then the average speed was:

$$\frac{\text{distance}}{\text{time elapsed}} = \frac{100\text{ m}}{70\text{ s}} = 1.4\text{ m/s}.$$

The magnitude of the average velocity, on the other hand, was:

$$\frac{\text{displacement}}{\text{time elapsed}} = \frac{40\text{ m}}{70\text{ s}} = 0.57\text{ m/s}.$$

In everyday life, we are usually interested in average speed. If this second equation on average velocity seems strange, we will see its usefulness in the next Section.

To discuss one-dimensional motion of an object in general, suppose that at some moment in time, call it t_1, the object is on the x axis at position x_1 in a coordinate system, and at some later time, t_2, suppose it is at position x_2. The **elapsed time** (= change in time) is $\Delta t = t_2 - t_1$. During this time interval the displacement of our object is $\Delta x = x_2 - x_1$. Then the **average velocity**, defined as *the displacement divided by the elapsed time*, can be written

$$\bar{v} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{\Delta x}{\Delta t}, \qquad \text{[average velocity]} \quad \textbf{(2–2)}$$

where v stands for velocity and the bar $(^{-})$ over the v is a standard symbol meaning "average."

CAUTION
Time interval = elapsed time

It is always important to choose (and state) the *elapsed time*, or **time interval**, $t_2 - t_1$, the time that passes during our chosen period of observation.

EXAMPLE 2–1 **Runner's average velocity.** The position of a runner is plotted as moving along the x axis of a coordinate system. During a 3.00-s time interval, the runner's position changes from $x_1 = 50.0\text{ m}$ to $x_2 = 30.5\text{ m}$, as shown in Fig. 2–7. What is the runner's average velocity?

APPROACH We want to find the average velocity, which is the displacement divided by the elapsed time.

SOLUTION The displacement is

$$\Delta x = x_2 - x_1 = 30.5\text{ m} - 50.0\text{ m} = -19.5\text{ m}.$$

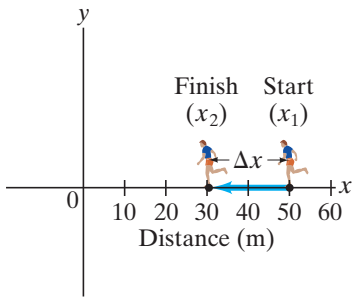
In this case the displacement is negative.

The elapsed time, or time interval, is given as $\Delta t = 3.00\text{ s}$. The average velocity (Eq. 2–2) is

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{-19.5\text{ m}}{3.00\text{ s}} = -6.50\text{ m/s}.$$

The displacement and average velocity are negative: that is, the runner is moving to the left along the x axis, as indicated by the arrow in Fig. 2–7. The runner's average velocity is 6.50 m/s to the left.

FIGURE 2–7 Example 2–1. A person runs from $x_1 = 50.0\text{ m}$ to $x_2 = 30.5\text{ m}$. The displacement is -19.5 m.



For one-dimensional motion in the usual case of the $+x$ axis to the right, if x_2 is less than x_1, then the object is moving to the left, and $\Delta x = x_2 - x_1$ is less than zero. The sign of the displacement, and thus of the average velocity, indicates the direction: the average velocity is positive for an object moving to the right along the x axis and negative when the object moves to the left. The direction of the average velocity is always the same as the direction of the displacement.

PROBLEM SOLVING
+ or − sign can signify the direction for linear motion

EXAMPLE 2–2 **Distance a cyclist travels.** How far can a cyclist travel in 2.5 h along a straight road if her average velocity is 18 km/h?

APPROACH We want to find the distance traveled, which in this case equals the displacement Δx, so we solve Eq. 2–2 for Δx.

SOLUTION In Eq. 2–2, $\bar{v} = \Delta x/\Delta t$, we multiply both sides by Δt and obtain

$$\Delta x = \bar{v}\,\Delta t = (18\,\text{km/h})(2.5\,\text{h}) = 45\,\text{km}.$$

EXAMPLE 2–3 **Car changes speed.** A car travels at a constant 50 km/h for 100 km. It then speeds up to 100 km/h and is driven another 100 km. What is the car's average speed for the 200-km trip?

APPROACH At 50 km/h, the car takes 2.0 h to travel 100 km. At 100 km/h, it takes only 1.0 h to travel 100 km. We use the definition of average velocity, Eq. 2–2.

SOLUTION Average velocity (Eq. 2–2) is

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{100\,\text{km} + 100\,\text{km}}{2.0\,\text{h} + 1.0\,\text{h}} = 67\,\text{km/h}.$$

NOTE Averaging the two speeds, $(50\,\text{km/h} + 100\,\text{km/h})/2 = 75\,\text{km/h}$, gives a wrong answer. Can you see why? You must use the definition of $\bar{v}$, Eq. 2–2.

2–3 Instantaneous Velocity



FIGURE 2–8 Car speedometer showing mi/h in white, and km/h in orange.

If you drive a car along a straight road for 150 km in 2.0 h, the magnitude of your average velocity is 75 km/h. It is unlikely, though, that you were moving at precisely 75 km/h at every instant. To describe this situation we need the concept of *instantaneous velocity*, which is the velocity at any instant of time. (Its magnitude is the number, with units, indicated by a speedometer, Fig. 2–8.) More precisely, the **instantaneous velocity** at any moment is defined as *the average velocity over an infinitesimally short time interval*. That is, Eq. 2–2 is to be evaluated in the limit of Δt becoming extremely small, approaching zero. We can write the definition of instantaneous velocity, v, for one-dimensional motion as

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}. \qquad \text{[instantaneous velocity]} \quad \textbf{(2–3)}$$

The notation $\lim_{\Delta t \to 0}$ means the ratio $\Delta x/\Delta t$ is to be evaluated in the limit of Δt approaching zero. But we do not simply set $\Delta t = 0$ in this definition, for then Δx would also be zero, and we would not be able to evaluate it. Rather, we consider the *ratio* $\Delta x/\Delta t$, as a whole. As we let Δt approach zero, Δx approaches zero as well. But the ratio $\Delta x/\Delta t$ approaches some definite value, which is the instantaneous velocity at a given instant.

In Eq. 2–3, the limit as $\Delta t \to 0$ is written in calculus notation as dx/dt and is called the **derivative** of x with respect to t:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}. \qquad \textbf{(2–4)}$$

This equation is the definition of instantaneous velocity for one-dimensional motion.

For instantaneous velocity we use the symbol v, whereas for average velocity we use $\bar{v}$, with a bar above. In the rest of this book, when we use the term "velocity" it will refer to instantaneous velocity. When we want to speak of the average velocity, we will make this clear by including the word "average."

Note that the *instantaneous speed* always equals the magnitude of the instantaneous velocity. Why? Because as the time interval becomes infinitesimally small ($\Delta t \to 0$), an object has no time to change speed or direction, and so the distance traveled and the magnitude of the displacement have to be the same.

If an object moves at a uniform (that is, constant) velocity during a particular time interval, then its instantaneous velocity at any instant is the same as its average velocity (see Fig. 2–9a). But in many situations this is not the case. For example, a car may start from rest, speed up to 50 km/h, remain at that velocity for a time, then slow down to 20 km/h in a traffic jam, and finally stop at its destination after traveling a total of 15 km in 30 min. This trip is plotted on the graph of Fig. 2–9b. Also shown on the graph is the average velocity (dashed line), which is $\bar{v} = \Delta x/\Delta t = 15\,\text{km}/0.50\,\text{h} = 30\,\text{km/h}$.

To better understand instantaneous velocity, let us consider a graph of the position versus time (x vs. t) of a particle moving along the x axis, as shown in Fig. 2–10. (Note that this is different from showing the "path" of a particle moving in two dimensions on an x vs. y plot.) The particle is at position x_1 at time t_1, and at position x_2 at time t_2. P_1 and P_2 represent these two points on the graph. A straight line drawn from point $P_1\,(x_1, t_1)$ to point $P_2\,(x_2, t_2)$ forms the hypotenuse of a right triangle whose sides are Δx and Δt. The ratio $\Delta x/\Delta t$ is the **slope** of the straight line P_1P_2. But $\Delta x/\Delta t$ is also the average velocity of the particle during the time interval $\Delta t = t_2 - t_1$. Therefore, we conclude that the average velocity of a particle during any time interval $\Delta t = t_2 - t_1$ is equal to the slope of the straight line (or *chord*) connecting the two points (x_1, t_1) and (x_2, t_2) on an x vs. t graph.

Consider now a time t_i, intermediate between t_1 and t_2, at which time the particle is at x_i (Fig. 2–11). The slope of the straight line P_1P_i is less than the slope of P_1P_2 in this case. Thus the average velocity during the time interval $t_i - t_1$ is less than during the time interval $t_2 - t_1$.

Now let us imagine that we take the point P_i in Fig. 2–11 to be closer and closer to point P_1. That is, we let the interval $t_i - t_1$, which we now call Δt, become smaller and smaller. The slope of the line connecting the two points becomes closer and closer to the slope of a line tangent to the curve at point P_1. The average velocity (equal to the slope of the chord) thus approaches the slope of the tangent at point P_1. The definition of the instantaneous velocity (Eq. 2–3) is the limiting value of the average velocity as Δt approaches zero. Thus the *instantaneous velocity equals the slope of the tangent to the x vs. t curve* at that point (which we can simply call "the slope of the curve" at that point).

Because the velocity at any instant equals the slope of the tangent to the x vs. t graph at that instant, we can obtain the velocity at any instant from such a graph. For example, in Fig. 2–12 (which shows the same curve as in Figs. 2–10 and 2–11), the slope continually increases as our object moves from x_1 to x_2, so the velocity is increasing. For times after t_2, however, the slope begins to decrease and in fact reaches zero (so $v = 0$) where x has its maximum value, at point P_3 in Fig. 2–12. Beyond this point, the slope is negative, as for point P_4. The velocity is therefore negative, which makes sense since x is now decreasing—the particle is moving to the left on a standard xy plot, toward decreasing values of x.

If an object moves with constant velocity over a particular time interval, its instantaneous velocity is equal to its average velocity. The graph of x vs. t in this case will be a straight line whose slope equals the velocity. The curve of Fig. 2–10 has no straight sections, so there are no time intervals when the velocity is constant.

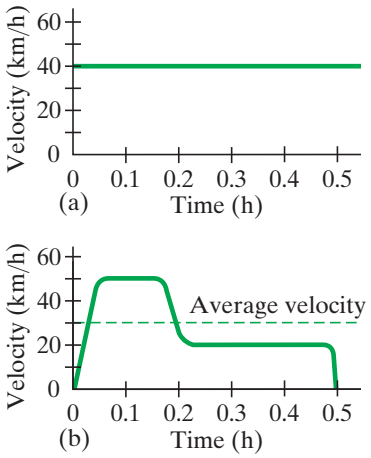


FIGURE 2–9 Velocity of a car as a function of time: (a) at constant velocity; (b) with velocity varying in time.

FIGURE 2–10 Graph of a particle's position x vs. time t. The slope of the straight line P_1P_2 represents the average velocity of the particle during the time interval $\Delta t = t_2 - t_1$.

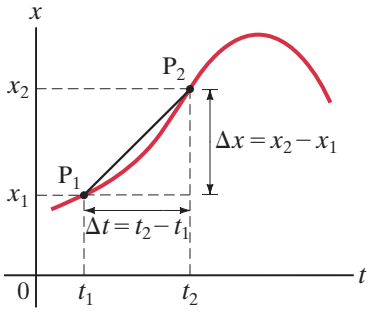


FIGURE 2–11 Same position vs. time curve as in Fig. 2–10, but including an intermediate time t_i. Note that the average velocity over the time interval $t_i - t_1$ (which is the slope of P_1P_i) is less than the average velocity over the time interval $t_2 - t_1$. The slope of the thin line tangent to the curve at point P_1 equals the instantaneous velocity at time t_1.

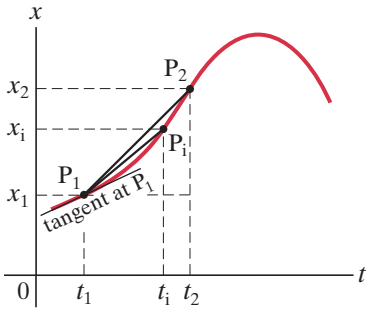
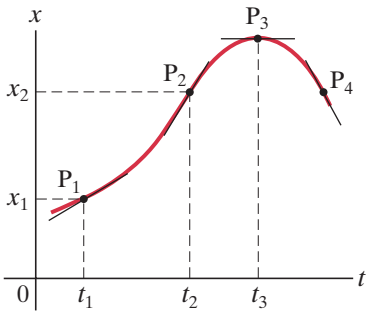


FIGURE 2–12 Same x vs. t curve as in Figs. 2–10 and 2–11, but here showing the slope at four different points: At P_3, the slope is zero, so $v = 0$. At P_4 the slope is negative, so $v < 0$.



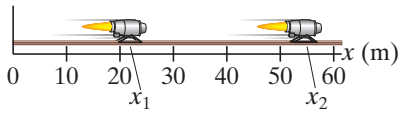
EXERCISE B What is your speed at the instant you turn around to move in the opposite direction? (*a*) Depends on how quickly you turn around; (*b*) always zero; (*c*) always negative; (*d*) none of the above.

The derivatives of various functions are studied in calculus courses, and you can find a review in this book in Appendix B. The derivatives of polynomial functions (which we use a lot) are:

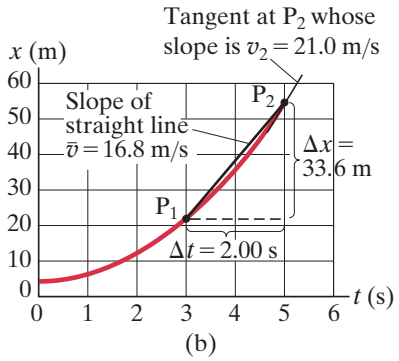
$$\frac{d}{dt}\left(Ct^n\right) = nCt^{n-1} \quad \text{and} \quad \frac{dC}{dt} = 0,$$

where C is any constant.

FIGURE 2–13 Example 2–4.
(a) Engine traveling on a straight track.
(b) Graph of x vs. t: $x = At^2 + B$.



(a)



(b)

EXAMPLE 2–4 **Given x as a function of t.** A jet engine moves along an experimental track (which we call the x axis) as shown in Fig. 2–13a. We will treat the engine as if it were a particle. Its position as a function of time is given by the equation $x = At^2 + B$, where $A = 2.10\,\text{m/s}^2$ and $B = 2.80\,\text{m}$, and this equation is plotted in Fig. 2–13b. (*a*) Determine the displacement of the engine during the time interval from $t_1 = 3.00\,\text{s}$ to $t_2 = 5.00\,\text{s}$. (*b*) Determine the average velocity during this time interval. (*c*) Determine the magnitude of the instantaneous velocity at $t = 5.00\,\text{s}$.

APPROACH (*a*) We substitute values for t_1 and t_2 in the given equation for x to obtain x_1 and x_2. (*b*) The average velocity can be found from Eq. 2–2. (*c*) To find the instantaneous velocity, we take the derivative of the given x equation with respect to t using the formulas given above.

SOLUTION (*a*) At $t_1 = 3.00\,\text{s}$, the position (point P_1 in Fig. 2–13b) is

$$x_1 = At_1^2 + B = (2.10\,\text{m/s}^2)(3.00\,\text{s})^2 + 2.80\,\text{m} = 21.7\,\text{m}.$$

At $t_2 = 5.00\,\text{s}$, the position (P_2 in Fig. 2–13b) is

$$x_2 = (2.10\,\text{m/s}^2)(5.00\,\text{s})^2 + 2.80\,\text{m} = 55.3\,\text{m}.$$

The displacement is thus

$$x_2 - x_1 = 55.3\,\text{m} - 21.7\,\text{m} = 33.6\,\text{m}.$$

(*b*) The magnitude of the average velocity can then be calculated as

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{33.6\,\text{m}}{2.00\,\text{s}} = 16.8\,\text{m/s}.$$

This equals the slope of the straight line joining points P_1 and P_2 shown in Fig. 2–13b.

(*c*) The instantaneous velocity at $t = t_2 = 5.00\,\text{s}$ equals the slope of the tangent to the curve at point P_2 shown in Fig. 2–13b. We could measure this slope off the graph to obtain v_2. But we can calculate v more precisely for any time t, using the given formula

$$x = At^2 + B,$$

which is the engine's position x as a function of time t. We take the derivative of x with respect to time (see formulas at top of this page):

$$v = \frac{dx}{dt} = \frac{d}{dt}\left(At^2 + B\right) = 2At.$$

We are given $A = 2.10\,\text{m/s}^2$, so for $t = t_2 = 5.00\,\text{s}$,

$$v_2 = 2At = 2(2.10\,\text{m/s}^2)(5.00\,\text{s}) = 21.0\,\text{m/s}.$$

2–4 Acceleration

An object whose velocity is changing is said to be accelerating. For instance, a car whose velocity increases in magnitude from zero to 80 km/h is accelerating. Acceleration specifies how rapidly the velocity of an object is changing.

Average Acceleration

Average acceleration is defined as the change in velocity divided by the time taken to make this change:

$$\text{average acceleration} = \frac{\text{change of velocity}}{\text{time elapsed}}.$$

In symbols, the average acceleration over a time interval $\Delta t = t_2 - t_1$ during which the velocity changes by $\Delta v = v_2 - v_1$ is defined as

$$\bar{a} = \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t}. \qquad \textbf{(2–5)}$$

Because velocity is a vector, acceleration is a vector too. But for one-dimensional motion, we need only use a plus or minus sign to indicate acceleration direction relative to a chosen coordinate axis.

EXAMPLE 2–5 **Average acceleration.** A car accelerates along a straight road from rest to 90 km/h in 5.0 s, Fig. 2–14. What is the magnitude of its average acceleration?

APPROACH Average acceleration is the change in velocity divided by the elapsed time, 5.0 s. The car starts from rest, so $v_1 = 0$. The final velocity is $v_2 = 90\,\text{km/h} = 90 \times 10^3\,\text{m}/3600\,\text{s} = 25\,\text{m/s}$.

SOLUTION From Eq. 2–5, the average acceleration is

$$\bar{a} = \frac{v_2 - v_1}{t_2 - t_1} = \frac{25\,\text{m/s} - 0\,\text{m/s}}{5.0\,\text{s}} = 5.0\,\frac{\text{m/s}}{\text{s}}.$$

This is read as "five meters per second per second" and means that, on average, the velocity changed by 5.0 m/s during each second. That is, assuming the acceleration was constant, during the first second the car's velocity increased from zero to 5.0 m/s. During the next second its velocity increased by another 5.0 m/s, reaching a velocity of 10.0 m/s at $t = 2.0$ s, and so on. See Fig. 2–14.

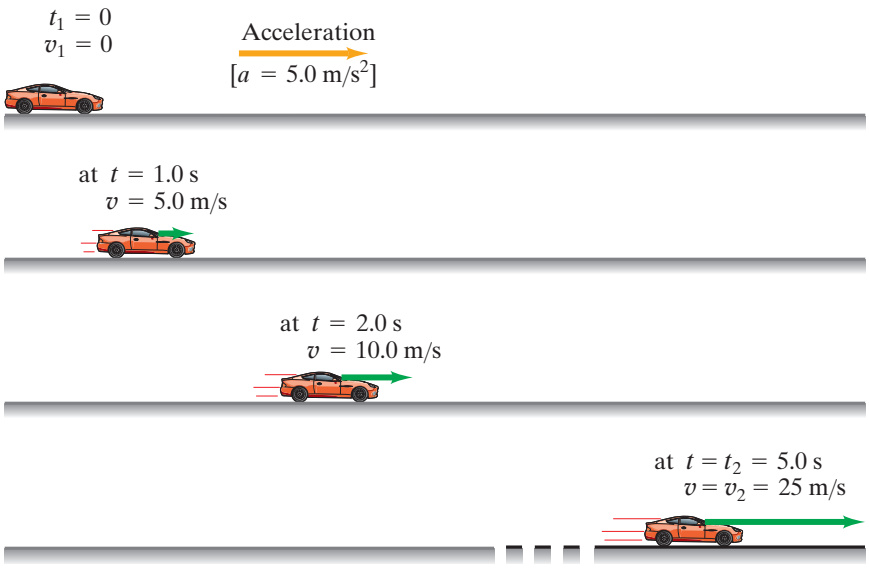


FIGURE 2–14 Example 2–5. The car is shown at the start with $v_1 = 0$ at $t_1 = 0$. The car is shown three more times, at $t = 1.0$ s, $t = 2.0$ s, and at the end of our time interval, $t_2 = 5.0$ s. The green arrows represent the velocity vectors, whose length represents the magnitude of the velocity at that moment and get longer with time. The acceleration vector is the orange arrow, whose magnitude is constant and is found to equal 5.0 m/s^2. Distances are not to scale.

We almost always write the units for acceleration as m/s^2 (meters per second squared) instead of m/s/s. This is possible because:

$$\frac{m/s}{s} = \frac{m}{s \cdot s} = \frac{m}{s^2}.$$

According to the calculation in Example 2–5, the velocity changed on average by 5.0 m/s during each second, for a total change of 25 m/s over the 5.0 s; the average acceleration was $5.0\ m/s^2$.

Note that *acceleration* tells us how quickly the *velocity* changes, whereas *velocity* tells us how quickly the *position* changes.

⚠ CAUTION
Distinguish velocity from acceleration

⚠ CAUTION
If v or a is zero, is the other zero too?

CONCEPTUAL EXAMPLE 2–6 **Velocity and acceleration.** (*a*) If the velocity of an object is zero, does it mean that the acceleration is zero? (*b*) If the acceleration is zero, does it mean that the velocity is zero? Think of some examples.

RESPONSE A zero velocity does not necessarily mean that the acceleration is zero, nor does a zero acceleration mean that the velocity is zero. (*a*) For example, when you put your foot on the gas pedal of your car which is at rest, the velocity starts from zero but the acceleration is not zero since the velocity of the car changes. (How else could your car start forward if its velocity weren't changing—that is, accelerating?) (*b*) As you cruise along a straight highway at a constant velocity of 100 km/h, your acceleration is zero: $a = 0$, $v \neq 0$.

EXERCISE C A powerful car is advertised to go from zero to 60 mi/h in 5.4 s. What does this say about the car: (*a*) it is fast (high speed); or (*b*) it accelerates well?

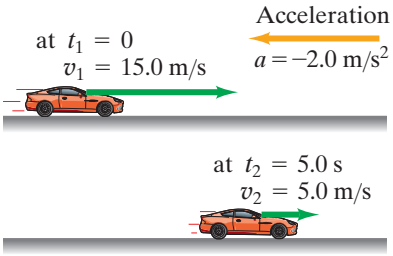


FIGURE 2–15 Example 2–7, showing the position of the car at times t_1 and t_2, as well as the car's velocity represented by the green arrows. We calculate that the acceleration vector (orange) points to the left as the car slows down while moving to the right.

EXAMPLE 2–7 **Car slowing down.** An automobile is moving to the right along a straight highway, which we choose to be the positive *x* axis (Fig. 2–15). Then the driver puts on the brakes. If the initial velocity (when the driver hits the brakes) is $v_1 = 15.0$ m/s, and it takes 5.0 s to slow down to $v_2 = 5.0$ m/s, what was the car's average acceleration?

APPROACH We put the given initial and final velocities, and the elapsed time, into Eq. 2–5 for $\bar{a}$.

SOLUTION In Eq. 2–5, we call the initial time $t_1 = 0$, and set $t_2 = 5.0$ s. (Note that our choice of $t_1 = 0$ doesn't affect the calculation of $\bar{a}$ because only $\Delta t = t_2 - t_1$ appears in Eq. 2–5.) Then

$$\bar{a} = \frac{5.0\ m/s - 15.0\ m/s}{5.0\ s} = -2.0\ m/s^2.$$

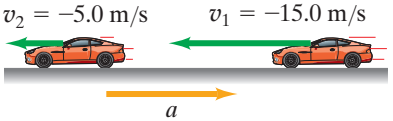
The negative sign appears because the final velocity is less than the initial velocity. In this case the direction of the acceleration is to the left (in the negative *x* direction)—even though the velocity is always pointing to the right. We say that the acceleration is $2.0\ m/s^2$ to the left, and it is shown in Fig. 2–15 as an orange arrow.

FIGURE 2–16 The car of Example 2–7, now moving to the *left* and decelerating. The acceleration is

$$a = \frac{v_2 - v_1}{\Delta t}$$

$$a = \frac{(-5.0\ m/s) - (-15.0\ m/s)}{5.0\ s}$$

$$= \frac{-5.0\ m/s + 15.0\ m/s}{5.0\ s} = +2.0\ m/s^2.$$



"Deceleration"

When an object is slowing down, we sometimes say it is **decelerating**. In physics, the concept of acceleration is all we need: it can be + or −. But if the word "deceleration" is used, be careful: deceleration does *not* mean that the acceleration is necessarily negative, as in Example 2–7. The velocity of an object moving to the right along the positive *x* axis is positive; if the object is slowing down (as in Fig. 2–15), the acceleration *is* negative. But the same car moving to the left (decreasing *x*), and slowing down, has positive acceleration that points to the right, as shown in Fig. 2–16. We have a deceleration whenever the magnitude of the velocity is decreasing; thus the *velocity and acceleration point in opposite directions* when there is deceleration.

EXERCISE D A car moves along the x axis. What is the sign of the car's acceleration if it is moving in the positive x direction with (a) increasing speed or (b) decreasing speed? What is the sign of the acceleration if the car moves in the negative x direction with (c) increasing speed or (d) decreasing speed?

Instantaneous Acceleration

The **instantaneous acceleration**, a, is defined as the *limiting value of the average acceleration as we let* Δt *approach zero*:

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt}. \qquad \textbf{(2–6)}$$

This limit, dv/dt, is the derivative of v with respect to t. We will use the term "acceleration" to refer to the instantaneous value. If we want to discuss the average acceleration, we will always include the word "average."

If we draw a graph of the velocity, v, vs. time, t, as shown in Fig. 2–17, then the *average* acceleration over a time interval $\Delta t = t_2 - t_1$ is represented by the slope of the straight line connecting the two points P_1 and P_2 in Fig. 2–17. [Compare this to the position vs. time graph of Fig. 2–10 for which the slope of the straight line represents the average velocity.] The *instantaneous* acceleration at any time, say t_1, is the slope of the tangent to the v vs. t curve at time t_1, which is also shown in Fig. 2–17. In Fig. 2–17, as we go from time t_1 to time t_2 the velocity continually increases, but the acceleration (the rate at which the velocity changes) is decreasing since the slope of the curve is decreasing.

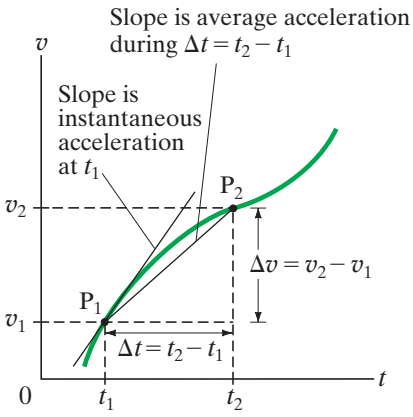


FIGURE 2–17 A graph of velocity v vs. time t. The average acceleration over a time interval $\Delta t = t_2 - t_1$ is the slope of the straight line P_1P_2: $\bar{a} = \Delta v/\Delta t$. The instantaneous acceleration at time t_1 is the slope of the v vs. t curve at that instant.

EXAMPLE 2–8 **Acceleration given $x(t)$.** A particle is moving in a straight line so that its position is given by the relation $x = (2.10\,\text{m/s}^2)t^2 + (2.80\,\text{m})$, as in Example 2–4. Calculate (a) its average acceleration during the time interval from $t_1 = 3.00\,\text{s}$ to $t_2 = 5.00\,\text{s}$, and (b) its instantaneous acceleration as a function of time.

APPROACH To determine acceleration, we first must find the velocity at t_1 and t_2 by differentiating x: $v = dx/dt$. Then we use Eq. 2–5 to find the average acceleration, and Eq. 2–6 to find the instantaneous acceleration.

SOLUTION (a) The velocity at any time t is

$$v = \frac{dx}{dt} = \frac{d}{dt}\left[(2.10\,\text{m/s}^2)t^2 + 2.80\,\text{m}\right] = (4.20\,\text{m/s}^2)t,$$

as we already saw in Example 2–4*c*. Therefore, at time $t_1 = 3.00\,\text{s}$, $v_1 = (4.20\,\text{m/s}^2)(3.00\,\text{s}) = 12.6\,\text{m/s}$ and at $t_2 = 5.00\,\text{s}$, $v_2 = 21.0\,\text{m/s}$. Therefore,

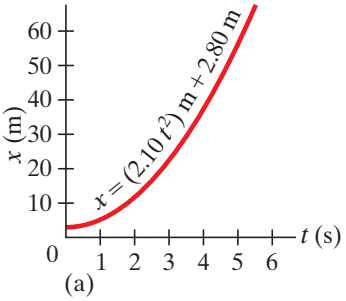
$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{21.0\,\text{m/s} - 12.6\,\text{m/s}}{5.00\,\text{s} - 3.00\,\text{s}} = 4.20\,\text{m/s}^2.$$

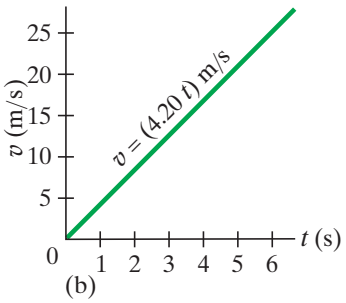
(b) With $v = (4.20\,\text{m/s}^2)t$, the instantaneous acceleration at any time is

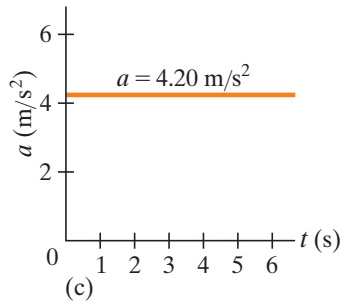
$$a = \frac{dv}{dt} = \frac{d}{dt}\left[(4.20\,\text{m/s}^2)t\right] = 4.20\,\text{m/s}^2.$$

The acceleration in this case is constant; it does not depend on time. Figure 2–18 shows graphs of (a) x vs. t (the same as Fig. 2–13b), (b) v vs. t, which is linearly increasing as calculated above, and (c) a vs. t, which is a horizontal straight line because a = constant.

FIGURE 2–18 Example 2–8. Graphs of (a) x vs. t, (b) v vs. t, and (c) a vs. t for the motion $x = At^2 + B$. Note that v increases linearly with t and that the acceleration a is constant. Also, v is the slope of the x vs. t curve, whereas a is the slope of the v vs. t curve.







Like velocity, acceleration is a rate. The velocity of an object is the rate at which its displacement changes with time; its acceleration, on the other hand, is the rate at which its velocity changes with time. In a sense, acceleration is a "rate of a rate."

This can be expressed in equation form: since $a = dv/dt$ and $v = dx/dt$, then

$$a = \frac{dv}{dt} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}.$$

Here d^2x/dt^2 is the *second derivative* of x with respect to time: we first take the derivative of x with respect to time (dx/dt), and then we again take the derivative with respect to time, $(d/dt)(dx/dt)$, to get the acceleration.

EXERCISE E The position of a particle is given by the following equation:

$$x = (2.00\,\text{m/s}^3)t^3 + (2.50\,\text{m/s})t.$$

What is the acceleration of the particle at $t = 2.00$ s? Choose one: (*a*) 13.0 m/s^2; (*b*) 22.5 m/s^2; (*c*) 24.0 m/s^2; (*d*) 2.00 m/s^2; (*e*) 21.0 m/s^2.

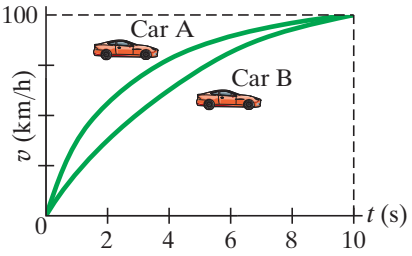


FIGURE 2–19 Example 2–9.

CONCEPTUAL EXAMPLE 2–9 **Analyzing with graphs.** Figure 2–19 shows the velocity as a function of time for two cars accelerating from 0 to 100 km/h in a time of 10.0 s. Compare for the two cars: (*a*) the average acceleration; (*b*) the instantaneous acceleration; and (*c*) the total distance traveled.

RESPONSE (*a*) Average acceleration is $\Delta v/\Delta t$. Both cars have the same Δv (100 km/h) and the same Δt (10.0 s), so the average acceleration is the same for both cars. (*b*) Instantaneous acceleration is the slope of the tangent to the v vs. t curve. For about the first 4 s, the top curve is steeper than the bottom curve, so car A has a greater instantaneous acceleration during this interval. The bottom curve is steeper during the last 6 s, so car B has the larger acceleration during this period. (*c*) Except at $t = 0$ and $t = 10.0$ s, car A is always going faster than car B. Since it is going faster, it will go farther in the same time. Notice what marvelous information we can get from a graph.

2–5 Motion at Constant Acceleration

We now examine motion in a straight line when the magnitude of the acceleration is constant. In this case, the instantaneous and average acceleration are equal. We use the definitions of average velocity and acceleration to derive a set of valuable equations that relate x, v, a, and t when a is constant, allowing us to determine any one of these variables if we know the others.

Notation in physics varies from book to book; and different instructors use different notation. We are now going to change our notation, to simplify it a bit for our discussion here of motion at **constant acceleration**. First we choose the initial time in any discussion to be zero, and we call it t_0. That is, $t_1 = t_0 = 0$. (This is effectively starting a stopwatch at t_0.) We can then let $t_2 = t$ be the elapsed time. The initial position (x_1) and the initial velocity (v_1) of an object will now be represented by x_0 and v_0, since they represent x and v at $t = 0$. At time t the position and velocity will be called x and v (rather than x_2 and v_2). The average velocity during the time interval $t - t_0$ will be (Eq. 2–2)

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{x - x_0}{t - t_0} = \frac{x - x_0}{t}$$

since we chose $t_0 = 0$. The acceleration, assumed constant in time, is $a = \Delta v/\Delta t$ (Eq. 2–5), so

$$a = \frac{v - v_0}{t}.$$

A common problem is to determine the velocity of an object after any elapsed time t, when we are given the object's constant acceleration. We can solve such problems by solving for v in the last equation: first we multiply both sides by t, which gives $at = v - v_0$, and then

$$v = v_0 + at. \qquad \text{[constant acceleration]} \quad \textbf{(2–7)}$$

If an object, such as a motorcycle, starts from rest $(v_0 = 0)$ and accelerates

at $4.0\,\text{m/s}^2$, then after an elapsed time $t = 6.0\,\text{s}$ its velocity will be $v = 0 + at = (4.0\,\text{m/s}^2)(6.0\,\text{s}) = 24\,\text{m/s}$.

Next, let us see how to calculate the position x of an object after a time t when it undergoes constant acceleration. The definition of average velocity (Eq. 2–2) is $\bar{v} = (x - x_0)/t$, which we can rewrite by multiplying both sides by t:

$$x = x_0 + \bar{v}t. \qquad \textbf{(2–8)}$$

Because the velocity increases at a uniform rate, the average velocity, $\bar{v}$, will be midway between the initial and final velocities:

$$\bar{v} = \frac{v_0 + v}{2}. \qquad \text{[constant acceleration]} \quad \textbf{(2–9)}$$

CAUTION
Average velocity, but only if a = constant

(Careful: Equation 2–9 is not necessarily valid if the acceleration is not constant.) We combine the last two Equations with Eq. 2–7 and find, starting with Eq. 2–8,

$$\begin{aligned} x &= x_0 + \bar{v}t \\ &= x_0 + \left(\frac{v_0 + v}{2}\right)t \\ &= x_0 + \left(\frac{v_0 + v_0 + at}{2}\right)t \end{aligned}$$

or

$$x = x_0 + v_0 t + \tfrac{1}{2}at^2. \qquad \text{[constant acceleration]} \quad \textbf{(2–10)}$$

Equations 2–7, 2–9, and 2–10 are three of the four most useful equations for motion at constant acceleration. We now derive the fourth equation, which is useful in situations where the time t is not known. We substitute Eq. 2–9 into Eq. 2–8:

$$x = x_0 + \bar{v}t = x_0 + \left(\frac{v + v_0}{2}\right)t.$$

Next we solve Eq. 2–7 for t, obtaining

$$t = \frac{v - v_0}{a},$$

and substituting this into the previous equation we have

$$x = x_0 + \left(\frac{v + v_0}{2}\right)\left(\frac{v - v_0}{a}\right) = x_0 + \frac{v^2 - v_0^2}{2a}.$$

We solve this for v^2 and obtain

$$v^2 = v_0^2 + 2a(x - x_0), \qquad \text{[constant acceleration]} \quad \textbf{(2–11)}$$

which is the other useful equation we sought.

We now have four equations relating position, velocity, acceleration, and time, when the acceleration a is constant. We collect these *kinematic equations for constant acceleration* here in one place for further reference (the tan background is used to emphasize their importance):

$$v = v_0 + at \qquad [a = \text{constant}] \quad \textbf{(2–12a)}$$

$$x = x_0 + v_0 t + \tfrac{1}{2}at^2 \qquad [a = \text{constant}] \quad \textbf{(2–12b)}$$

$$v^2 = v_0^2 + 2a(x - x_0) \qquad [a = \text{constant}] \quad \textbf{(2–12c)}$$

$$\bar{v} = \frac{v + v_0}{2}. \qquad [a = \text{constant}] \quad \textbf{(2–12d)}$$

Kinematic equations for constant acceleration (we'll use them a lot)

These useful equations are not valid unless a is a constant. In many cases we can set $x_0 = 0$, and this simplifies the above equations a bit. Note that x represents position (not distance), and that $x - x_0$ is the displacement, whereas t is the elapsed time.

Equations 2–12 are useful also when a is approximately constant, in order to obtain reasonable estimates.

PHYSICS APPLIED
Airport design

EXAMPLE 2–10 **Runway design.** You are designing an airport for small planes. One kind of airplane that might use this airfield must reach a speed before takeoff of at least 27.8 m/s (100 km/h), and can accelerate at 2.00 m/s^2. (*a*) If the runway is 150 m long, can this airplane reach the required speed for takeoff? (*b*) If not, what minimum length must the runway have?

APPROACH Assuming the plane's acceleration is constant, we use the kinematic equations for constant acceleration. In (*a*), we want to find v, and we are given:

Known	Wanted
$x_0 = 0$	v
$v_0 = 0$	
$x = 150\text{ m}$	
$a = 2.00\text{ m/s}^2$	

SOLUTION (*a*) Of the four kinematic equations on page 31, Eq. 2–12c will give us v when we know v_0, a, x, and x_0:

$$\begin{aligned} v^2 &= v_0^2 + 2a(x - x_0) \\ &= 0 + 2(2.00\text{ m/s}^2)(150\text{ m}) = 600\text{ m}^2/\text{s}^2 \\ v &= \sqrt{600\text{ m}^2/\text{s}^2} = 24.5\text{ m/s}. \end{aligned}$$

PROBLEM SOLVING
Equations 2–12 are valid only when the acceleration is constant, which we assume in this Example

This runway length is *not* sufficient, because the minimum speed is not reached. (*b*) Now we want to find the minimum runway length, $x - x_0$, for a plane to reach $v = 27.8\text{ m/s}$, given $a = 2.00\text{ m/s}^2$. We again use Eq. 2–12c, but rewritten as

$$(x - x_0) = \frac{v^2 - v_0^2}{2a} = \frac{(27.8\text{ m/s})^2 - 0}{2(2.00\text{ m/s}^2)} = 193\text{ m}.$$

A 200-m runway is more appropriate for this plane.

NOTE We did this Example as if the plane were a particle, so we round off our answer to 200 m.

PHYSICS APPLIED
Car safety—air bags

EXAMPLE 2–11 **ESTIMATE** **Air bags.** Suppose you want to design an air bag system that can protect the driver at a speed of 100 km/h (60 mph) if the car hits a brick wall. Estimate how fast the air bag must inflate (Fig. 2–20) to effectively protect the driver. How does the use of a seat belt help the driver?

FIGURE 2–20 Example 2–11. An air bag deploying on impact.



APPROACH We assume the acceleration is roughly constant, so we can use Eqs. 2–12. Both Eqs. 2–12a and 2–12b contain t, our desired unknown. They both contain a, so we must first find a, which we can do using Eq. 2–12c if we know the distance x over which the car crumples. A rough estimate might be about 1 meter. We choose the time interval to start at the instant of impact with the car moving at $v_0 = 100\text{ km/h}$, and to end when the car comes to rest ($v = 0$) after traveling 1 m.

SOLUTION We convert the given initial speed to SI units: $100\text{ km/h} = 100 \times 10^3\text{ m}/3600\text{ s} = 28\text{ m/s}$. We then find the acceleration from Eq. 2–12c:

$$a = -\frac{v_0^2}{2x} = -\frac{(28\text{ m/s})^2}{2.0\text{ m}} = -390\text{ m/s}^2.$$

This enormous acceleration takes place in a time given by (Eq. 2–12a):

$$t = \frac{v - v_0}{a} = \frac{0 - 28\text{ m/s}}{-390\text{ m/s}^2} = 0.07\text{ s}.$$

To be effective, the air bag would need to inflate faster than this.

What does the air bag do? It spreads the force over a large area of the chest (to avoid puncture of the chest by the steering wheel). The seat belt keeps the person in a stable position directly in front of the expanding air bag.

EXERCISE F A car starts from rest and accelerates at a constant $10\,\text{m/s}^2$ during a $\frac{1}{4}$-mile (402 m) race. How fast is the car going at the finish line? (*a*) 8040 m/s; (*b*) 90 m/s; (*c*) 81 m/s; (*d*) 804 m/s.

2–6 Solving Problems

Before doing more worked-out Examples, let us look at how to approach problem solving. First, it is important to note that physics is *not* a collection of equations to be memorized. Simply searching for an equation that might work can lead you to a wrong result and will not help you understand physics (Fig. 2–21). A better approach is to use the following (rough) procedure, which we present as a special "Problem Solving Strategy." (Other such Problem Solving Strategies will be found throughout the book.)



FIGURE 2–21 Read each Chapter of this book, study it by reading it again carefully, and work the Problems using your reasoning abilities.

PROBLEM SOLVING

1. Read and **reread** the whole problem carefully before trying to solve it.

2. Decide what **object** (or objects) you are going to study, and for what **time interval**. You can often choose the initial time to be $t = 0$.

3. **Draw** a **diagram** or picture of the situation, with coordinate axes wherever applicable. [You can place the origin of coordinates and the axes wherever you like to make your calculations easier. You also choose which direction is positive and which is negative. Usually we choose the x axis to the right as positive.]

4. Write down what quantities are "**known**" or "given," and then what you *want* to know ("unknowns"). Consider quantities both at the beginning and at the end of the chosen time interval. You may need to "translate" language into physical terms, such as "starts from rest" means $v_0 = 0$.

5. Think about which **principles of physics** apply in this problem. Use common sense and your own experiences. Then plan an approach.

6. Consider which **equations** (and/or definitions) relate the quantities involved. Before using them, be sure their **range of validity** includes your problem (for example, Eqs. 2–12 are valid only when the acceleration is constant). If you find an applicable equation that involves only known quantities and one desired unknown, **solve** the equation algebraically for the unknown. Sometimes several sequential calculations, or a combination of equations, may be needed. It is often preferable to solve algebraically for the desired unknown before putting in numerical values.

7. Carry out the **calculation** if it is a numerical problem. Keep one or two extra digits during the calculations, but round off the final answer(s) to the correct number of significant figures (Section 1–3).

8. Think carefully about the result you obtain: Is it **reasonable**? Does it make sense according to your own intuition and experience? A good check is to do a rough **estimate** using only powers of 10, as discussed in Section 1–6. Often it is preferable to do a rough estimate at the *start* of a numerical problem because it can help you focus your attention on finding a path toward a solution.

9. A very important aspect of doing problems is keeping track of **units**. An equals sign implies the units on each side must be the same, just as the numbers must. If the units do not balance, a mistake has been made. This can serve as a **check** on your solution (but it only tells you if you're wrong, not if you're right). Always use a consistent set of units.

PROBLEM SOLVING
"Starting from rest" means $v = 0$ at $t = 0$ [i.e., $v_0 = 0$]

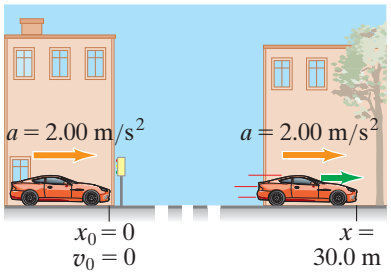


FIGURE 2–22 Example 2–12. (Not to scale.)

Known	Wanted
$x_0 = 0$	t
$x = 30.0\ \text{m}$	
$a = 2.00\ \text{m/s}^2$	
$v_0 = 0$	

EXAMPLE 2–12 **Acceleration of a car.** How long does it take a 4.0-m-long car to cross a 26.0-m-wide intersection after the light turns green, if the car accelerates from rest at a constant $2.00\ \text{m/s}^2$? The car has to travel $26.0\ \text{m} + 4.0\ \text{m} = 30.0\ \text{m}$ to clear the intersection.

APPROACH We follow the Problem Solving Strategy on the previous page.

SOLUTION

1. **Reread** the problem. Be sure you understand what it asks for (here, a time interval: "how long does it take").
2. The **object** under study is the car. We need to choose the **time interval** during which we look at the car's motion: we choose $t = 0$, the initial time, to be the moment the car starts to accelerate from rest $(v_0 = 0)$; the time t is the instant the car has traveled the full 30.0 m of the intersection.
3. **Draw a diagram**: the situation is shown in Fig. 2–22 where the car is shown moving along the positive x axis. We choose $x_0 = 0$ at the front bumper of the car before it starts to move.
4. The "**knowns**" and the "wanted" information are shown in the Table in the margin. Note that "starting from rest" means $v = 0$ at $t = 0$; that is, $v_0 = 0$. The wanted time t is how long it takes the car to travel 30.0 m.
5. The **physics**: the car, starting from rest $(\text{at } t_0 = 0)$, increases in speed as it covers more distance. The acceleration is constant, so we can use the kinematic equations, Eqs. 2–12.
6. **Equations**: we want to find the time, given the distance and acceleration; Eq. 2–12b $\left(x = x_0 + v_0 t + \frac{1}{2}at^2\right)$ is perfect since the only unknown quantity is t. Setting $v_0 = 0$ and $x_0 = 0$ in Eq. 2–12b, we have
$$x = \tfrac{1}{2}at^2.$$
We solve for t
$$t = \sqrt{\frac{2x}{a}}.$$
7. The **calculation**:
$$t = \sqrt{\frac{2x}{a}} = \sqrt{\frac{2(30.0\ \text{m})}{2.00\ \text{m/s}^2}} = 5.48\ \text{s}.$$
This is our answer. Note that the units come out correctly.

PROBLEM SOLVING
Check your answer

8. We can check the **reasonableness** of the answer by doing an alternate calculation: we use our result in step 7 and check if the distance traveled turns out to be 30.0 m. First we find the final velocity (Eq. 2–12a),
$$v = at = (2.00\ \text{m/s}^2)(5.48\ \text{s}) = 10.96\ \text{m/s},$$
and then find the distance traveled using the definition of average velocity (see Eq. 2–8).
$$x = x_0 + \bar{v}t = 0 + \tfrac{1}{2}(10.96\ \text{m/s} + 0)(5.48\ \text{s}) = 30.0\ \text{m},$$
which checks with our given distance.
9. We checked the **units** in step 7, and they came out correctly (seconds).

NOTE In steps 6 and 7, when we took the square root, we should have written $t = \pm\sqrt{2x/a} = \pm 5.48\ \text{s}$. Mathematically there are two solutions. But the second solution, $t = -5.48\ \text{s}$, is a time *before* our chosen time interval and makes no sense physically. We say it is "unphysical" and ignore it.

PROBLEM SOLVING
"Unphysical" solutions

We explicitly followed the steps of the Problem Solving Strategy in Example 2–12. In upcoming Examples, we will use our usual "Approach" and "Solution" to avoid being wordy.

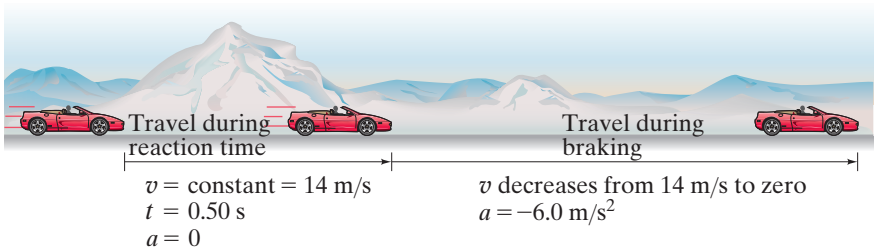


FIGURE 2–23 Example 2–13: stopping distance for a braking car.

EXAMPLE 2–13 ESTIMATE **Braking distances.** Estimate the minimum stopping distance for a car, which is important for traffic safety and traffic design. The problem is best dealt with in two parts, two separate time intervals. (1) The first time interval begins when the driver decides to hit the brakes, and ends when the foot touches the brake pedal. This is the "reaction time" during which the speed is constant, so $a = 0$. (2) The second time interval is the actual braking period when the vehicle slows down ($a \neq 0$) and comes to a stop. The stopping distance depends on the reaction time of the driver, the initial speed of the car (the final speed is zero), and the deceleration of the car. For a dry road and good tires, good brakes can decelerate a car at a rate of about 5 m/s² to 8 m/s². Calculate the total stopping distance for an initial velocity of 50 km/h ($= 14$ m/s ≈ 31 mi/h) and assume the acceleration of the car is -6.0 m/s² (the minus sign appears because the velocity is taken to be in the positive x direction and its magnitude is decreasing). Reaction time for typical drivers varies from perhaps 0.3 s to about 1.0 s; take it to be 0.50 s.

PHYSICS APPLIED
Car stopping distances

APPROACH During the "reaction time," part (1), the car moves at constant speed of 14 m/s, so $a = 0$. Once the brakes are applied, part (2), the acceleration is $a = -6.0$ m/s² and is constant over this time interval. For both parts a is constant, so we can use Eqs. 2–12.

SOLUTION Part (1). We take $x_0 = 0$ for the first time interval, when the driver is reacting (0.50 s): the car travels at a constant speed of 14 m/s so $a = 0$. See Fig. 2–23 and the Table in the margin. To find x, the position of the car at $t = 0.50$ s (when the brakes are first applied), we cannot use Eq. 2–12c because x is multiplied by a, which is zero. But Eq. 2–12b works:

$$x = v_0 t + 0 = (14\text{ m/s})(0.50\text{ s}) = 7.0\text{ m}.$$

Thus the car travels 7.0 m during the driver's reaction time, until the instant the brakes are applied. We will use this result as input to part (2).

Part (2). During the second time interval, the brakes are applied and the car is brought to rest. The initial position is $x_0 = 7.0$ m (result of part (1)), and other variables are shown in the second Table in the margin. Equation 2–12a doesn't contain x; Eq. 2–12b contains x but also the unknown t. Equation 2–12c, $v^2 - v_0^2 = 2a(x - x_0)$, is what we want; after setting $x_0 = 7.0$ m, we solve for x, the final position of the car (where it stops):

$$x = x_0 + \frac{v^2 - v_0^2}{2a}$$

$$= 7.0\text{ m} + \frac{0 - (14\text{ m/s})^2}{2(-6.0\text{ m/s}^2)} = 7.0\text{ m} + \frac{-196\text{ m}^2/\text{s}^2}{-12\text{ m/s}^2}$$

$$= 7.0\text{ m} + 16\text{ m} = 23\text{ m}.$$

The car traveled 7.0 m while the driver was reacting and another 16 m during the braking period before coming to a stop, for a total distance traveled of 23 m. Figure 2–24 shows graphs of (a) v vs. t (we were given that v is constant from $t = 0$ until $t = 0.50$ s, and after $t = 0.50$ s it decreases linearly to zero), and (b) x vs. t.

NOTE From the equation above for x, we see that the stopping distance after the driver hit the brakes $(= x - x_0)$ increases with the *square* of the initial speed, not just linearly with speed. If you are traveling twice as fast, it takes *four* times the distance to stop (when braking at the same rate, a).

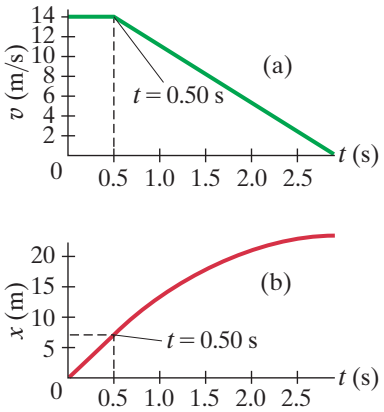
Part 1: Reaction time

Known	Wanted
$t = 0.50$ s	x
$v_0 = 14$ m/s	
$v = 14$ m/s	
$a = 0$	
$x_0 = 0$	

Part 2: Braking

Known	Wanted
$x_0 = 7.0$ m	x
$v_0 = 14$ m/s	
$v = 0$	
$a = -6.0$ m/s²	

FIGURE 2–24 Example 2–13. Graphs of (*a*) v vs. t and (*b*) x vs. t.



EXAMPLE 2–14 **ESTIMATE** **Two Moving Objects: Police and Speeder.** A car speeding at 150 km/h (over 90 mph) passes a still police car which immediately takes off in hot pursuit. Using simple assumptions, such as that the speeder continues at constant speed, estimate how long it takes the police car to overtake the speeder. Then estimate the police car's speed at that moment and decide if the assumptions were reasonable.

PROBLEM SOLVING
Guess the acceleration

APPROACH When the police car takes off, it accelerates, and the simplest assumption is that its acceleration is constant. This may not be reasonable, but let's see what happens. We can estimate the acceleration if we have noticed automobile ads, which claim cars can accelerate from rest to 100 km/h in 5.0 s. So the average acceleration of the police car could be approximately

$$a_P = \frac{100\,\text{km/h}}{5.0\,\text{s}} = 20\,\frac{\text{km/h}}{\text{s}}\left(\frac{1000\,\text{m}}{1\,\text{km}}\right)\left(\frac{1\,\text{h}}{3600\,\text{s}}\right) = 5.6\,\text{m/s}^2.$$

SOLUTION We need to set up the kinematic equations to determine the unknown quantities, and since there are two moving objects, we need two separate sets of equations. We denote the speeding car's position by x_S and the police car's position by x_P. Because we are interested in solving for the time when the two vehicles arrive at the same position on the road, we use Eq. 2–12b for each car ($x_0 = 0$ for both cars):

$$x_S = v_{0S}t + \tfrac{1}{2}a_S t^2 = (150\,\text{km/h})t + 0 = (42\,\text{m/s})t$$
$$x_P = v_{0P}t + \tfrac{1}{2}a_P t^2 = 0 + \tfrac{1}{2}(5.6\,\text{m/s}^2)t^2,$$

where we have set $v_{0P} = 0$ and $a_S = 0$ (speeder assumed to move at constant speed). We want the time when the cars meet, so we set $x_S = x_P$ and solve for t:

$$(42\,\text{m/s})t = (2.8\,\text{m/s}^2)t^2.$$

The solutions are

$$t = 0 \quad \text{and} \quad t = \frac{42\,\text{m/s}}{2.8\,\text{m/s}^2} = 15\,\text{s}.$$

The first solution corresponds to the instant the speeder passed the police car. The second solution tells us when the police car catches up to the speeder, 15 s later. This is our answer, but is it reasonable? The police car's speed at $t = 15$ s is

$$v_P = v_{0P} + a_P t = 0 + (5.6\,\text{m/s}^2)(15\,\text{s}) = 84\,\text{m/s}$$

or 300 km/h ($\approx$ 190 mi/h). Not reasonable, and highly dangerous.

CAUTION
Initial assumptions need to be checked out for reasonableness

NOTE More reasonable is to give up the assumption of constant acceleration. The police car surely cannot maintain constant acceleration at high speed. Also, the speeder, if a reasonable person, would slow down upon hearing the police siren. Figure 2–25 shows (a) x vs. t and (b) v vs. t graphs, based on the original assumption of a_P = constant, whereas (c) shows v vs. t for more reasonable assumptions.

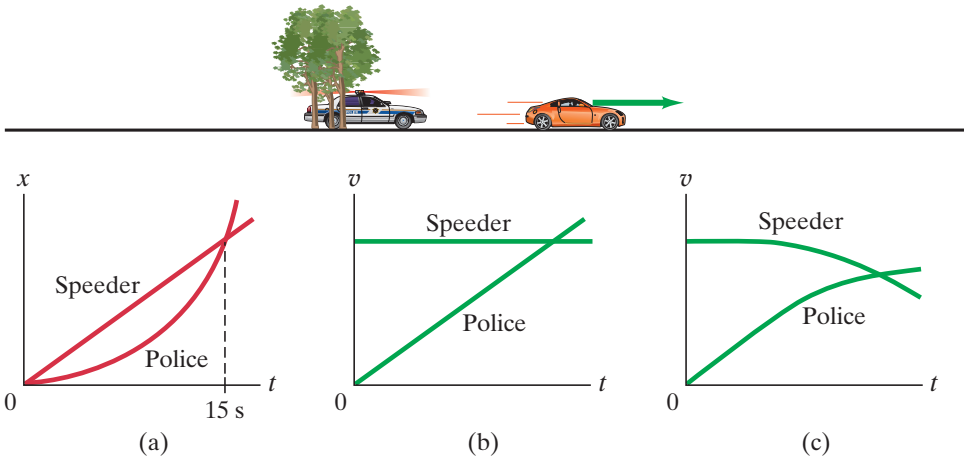


FIGURE 2–25 Example 2–14.



FIGURE 2–26 Painting of Galileo demonstrating to the Grand Duke of Tuscany his argument for the action of gravity being uniform acceleration. He used a wooden inclined plane (center) to slow down the action. A ball rolling down the plane still accelerates. Tiny bells placed at equal distances along the inclined plane would ring at shorter time intervals as the ball "fell," indicating that the speed was increasing.



FIGURE 2–27 Multiflash photograph of a falling apple, at equal time intervals. The apple falls farther during each successive interval, which means it is accelerating.

2–7 Freely Falling Objects

One of the most common examples of uniformly accelerated motion is that of an object allowed to fall freely near the Earth's surface. That a falling object is accelerating may not be obvious at first. And beware of thinking, as was widely believed before the time of Galileo (Fig. 2–26), that heavier objects fall faster than lighter objects and that the speed of fall is proportional to how heavy the object is. Wrong. *The speed of a falling object is* not *proportional to its mass*.

Galileo made use of his new technique of imagining what would happen in idealized (simplified) cases. For free fall, he postulated that *all objects would fall with the same constant acceleration in the absence of air or other resistance*. He showed that this postulate predicts that for an object falling from rest, the distance traveled will be proportional to the square of the time (Fig. 2–27); that is, $d \propto t^2$. We can see this from Eq. 2–12b for constant acceleration; but Galileo was the first to deduce this mathematical relation.

To support his claim that falling objects increase in speed as they fall, Galileo made use of a clever argument: a heavy stone dropped from a height of 2 m will drive a stake into the ground much further than if the same stone is dropped from a height of only 0.2 m. Clearly, the stone must be moving faster in the former case.

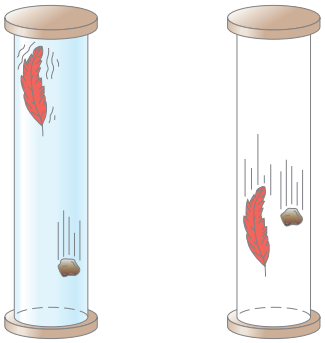
Galileo claimed that *all* objects, light or heavy, fall with the *same* acceleration, at least in the absence of air. If you hold a piece of paper flat and horizontal in one hand, and a heavier object like a baseball in the other, and release them at the same time as in Fig. 2–28a, the heavier object will reach the ground first. But if you repeat the experiment, this time crumpling the paper into a small wad, you will find (see Fig. 2–28b) that the two objects reach the floor at nearly the same time.

Galileo was sure that air acts as a resistance to very light objects that have a large surface area. But in many ordinary circumstances this air resistance is negligible. In a chamber from which the air has been removed, even light objects like a feather or a horizontally held piece of paper will fall with the same acceleration as any other object (see Fig. 2–29). Such a demonstration in vacuum was not possible in Galileo's time, which makes Galileo's achievement all the greater. Galileo is often called the "father of modern science," not only for the *content* of his science: astronomical discoveries, inertia, free fall; but also for his new methods of *doing* science: idealization and simplification, mathematization of theory, theories that have testable consequences, experiments to test theoretical predictions.

FIGURE 2–28 (a) A ball and a light piece of paper are dropped at the same time. (b) Repeated, with the paper wadded up.



FIGURE 2–29 A rock and a feather are dropped simultaneously (a) in air, (b) in a vacuum.



Galileo's specific contribution to our understanding of the motion of falling objects can be summarized as follows:

> **at a given location on the Earth and in the absence of air resistance, all objects fall with the same constant acceleration.**

We call this acceleration the **acceleration due to gravity** at the surface of the Earth, and we give it the symbol g. Its magnitude is approximately

$$g = 9.80\ \text{m/s}^2. \qquad \left[\begin{array}{l}\text{acceleration due to gravity}\\ \text{at surface of Earth}\end{array}\right]$$

In British units g is about $32\ \text{ft/s}^2$. Actually, g varies slightly according to latitude and elevation above sea level on the Earth's surface, but these variations are so small that we will ignore them for most purposes. (Acceleration of gravity out in space beyond the Earth's surface is treated in Chapter 6.) Air resistance acts to reduce the speed of a falling object, but this effect is often small, and we will neglect it for the most part. However, air resistance will be noticeable even on a reasonably heavy object if the velocity becomes large.[†]

Acceleration due to gravity is a vector, as is any acceleration, and its direction is downward toward the center of the Earth.

When dealing with freely falling objects we can make use of Eqs. 2–12, where for a we use the value of g given above. Also, since the motion is vertical we will substitute y in place of x, and y_0 in place of x_0. We take $y_0 = 0$ unless otherwise specified. *It is arbitrary whether we choose y to be positive in the upward direction or in the downward direction; but we must be consistent about it throughout a problem's solution.*

PROBLEM SOLVING
You can choose y to be positive either up or down

EXERCISE G Return to the Chapter-Opening Question, page 20, and answer it again now, assuming minimal air resistance. Try to explain why you may have answered differently the first time.

EXAMPLE 2–15 **Falling from a tower.** Suppose that a ball is dropped ($v_0 = 0$) from a tower. How far will it have fallen after a time $t_1 = 1.00\,\text{s}$, $t_2 = 2.00\,\text{s}$, and $t_3 = 3.00\,\text{s}$? Ignore air resistance.

APPROACH Let us take y as positive downward, so the acceleration is $a = g = +9.80\ \text{m/s}^2$. We set $v_0 = 0$ and $y_0 = 0$. We want to find the position y of the ball after three different time intervals. Equation 2–12b, with x replaced by y, relates the given quantities (t, a, and v_0) to the unknown y.

SOLUTION We set $t = t_1 = 1.00\,\text{s}$ in Eq. 2–12b:

$$y_1 = v_0 t_1 + \tfrac{1}{2}at_1^2$$
$$= 0 + \tfrac{1}{2}at_1^2 = \tfrac{1}{2}(9.80\,\text{m/s}^2)(1.00\,\text{s})^2 = 4.90\,\text{m}.$$

The ball has fallen a distance of 4.90 m during the time interval $t = 0$ to $t_1 = 1.00\,\text{s}$. Similarly, after $2.00\,\text{s}$ $(= t_2)$, the ball's position is

$$y_2 = \tfrac{1}{2}at_2^2 = \tfrac{1}{2}(9.80\,\text{m/s}^2)(2.00\,\text{s})^2 = 19.6\,\text{m}.$$

Finally, after $3.00\,\text{s}$ $(= t_3)$, the ball's position is (see Fig. 2–30a)

$$y_3 = \tfrac{1}{2}at_3^2 = \tfrac{1}{2}(9.80\,\text{m/s}^2)(3.00\,\text{s})^2 = 44.1\,\text{m}.$$

NOTE Whenever we say "dropped," it means $v_0 = 0$. Note also the graph of y vs. t (Fig. 2–30b): the curve is not straight but bends upward because y is proportional to t^2.

NOTE Because of air resistance, all of the distances in Fig. 2–30, y_1, y_2, y_3, would be smaller than shown (and as just calculated); but the difference will be small for a reasonably heavy (but small) object.

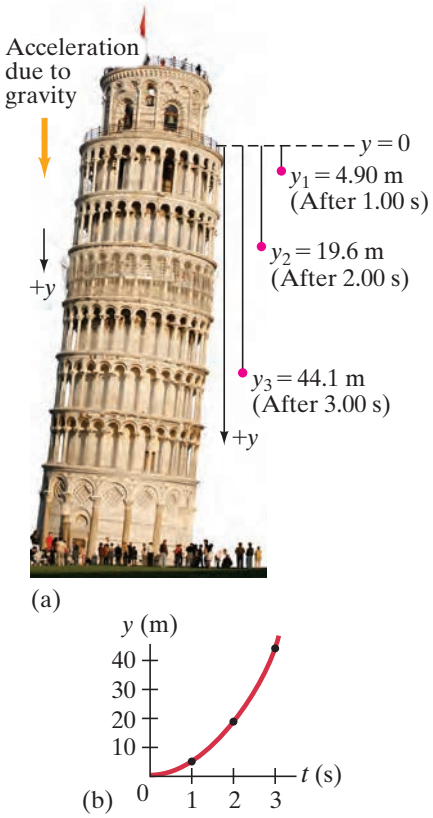


FIGURE 2–30 Example 2–15. (a) An object dropped from a tower falls with progressively greater speed and covers greater distance with each successive second. (See also Fig. 2–27.) (b) Graph of y vs. t (y is + downward).

[†]The speed of an object falling in air (or other fluid) does not increase indefinitely. If the object falls far enough, it will reach a maximum velocity called the **terminal velocity** due to air resistance. (Section 5–6 deals directly with air resistance.)

EXAMPLE 2–16 **Thrown down from a tower.** Suppose the ball in Example 2–15 is *thrown* downward with an initial velocity of 3.00 m/s, instead of being dropped. (*a*) What then would be its position after 1.00 s and 2.00 s? (*b*) What would its speed be after 1.00 s and 2.00 s? Compare with the speeds of a dropped ball.

APPROACH Again we use Eq. 2–12b, but now v_0 is not zero, it is $v_0 = 3.00\,\text{m/s}$.

SOLUTION (*a*) At $t_1 = 1.00\,\text{s}$, the position of the ball as given by Eq. 2–12b is

$$y = v_0 t + \tfrac{1}{2}at^2 = (3.00\,\text{m/s})(1.00\,\text{s}) + \tfrac{1}{2}(9.80\,\text{m/s}^2)(1.00\,\text{s})^2 = 7.90\,\text{m}.$$

At $t_2 = 2.00\,\text{s}$ (time interval $t = 0$ to $t = 2.00\,\text{s}$), the position is

$$y = v_0 t + \tfrac{1}{2}at^2 = (3.00\,\text{m/s})(2.00\,\text{s}) + \tfrac{1}{2}(9.80\,\text{m/s}^2)(2.00\,\text{s})^2 = 25.6\,\text{m}.$$

As expected, the ball falls farther each second than when it is dropped with $v_0 = 0$, Example 2–15.

(*b*) The velocity is obtained from Eq. 2–12a:

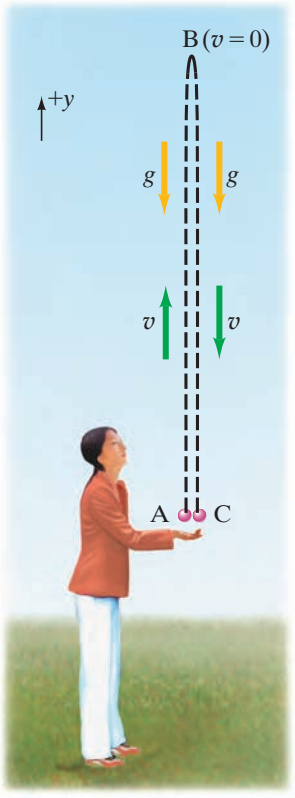
$$\begin{aligned} v &= v_0 + at \\ &= 3.00\,\text{m/s} + (9.80\,\text{m/s}^2)(1.00\,\text{s}) = 12.8\,\text{m/s} \quad [\text{at } t_1 = 1.00\,\text{s}] \\ &= 3.00\,\text{m/s} + (9.80\,\text{m/s}^2)(2.00\,\text{s}) = 22.6\,\text{m/s}. \quad [\text{at } t_2 = 2.00\,\text{s}] \end{aligned}$$

In Example 2–15, when the ball was dropped $(v_0 = 0)$, the first term (v_0) in these equations was zero, so

$$\begin{aligned} v &= 0 + at \\ &= (9.80\,\text{m/s}^2)(1.00\,\text{s}) = 9.80\,\text{m/s} \quad [\text{at } t_1 = 1.00\,\text{s}] \\ &= (9.80\,\text{m/s}^2)(2.00\,\text{s}) = 19.6\,\text{m/s}. \quad [\text{at } t_2 = 2.00\,\text{s}] \end{aligned}$$

NOTE For both Examples 2–15 and 2–16 the speed increases linearly in time by 9.80 m/s during each second. But the speed of the downwardly thrown ball at any instant is always 3.00 m/s (its initial speed) greater than that of a dropped ball.

FIGURE 2–31 An object thrown into the air leaves the thrower's hand at A, reaches its maximum height at B, and returns to the original position at C. Examples 2–17, 2–18, 2–19, 2–20, and 2–21.



EXAMPLE 2–17 **Ball thrown upward.** A person throws a ball *upward* into the air with an initial velocity of 15.0 m/s. Calculate how high it goes. Ignore air resistance.

APPROACH We are not concerned here with the throwing action, but only with the motion of the ball *after* it leaves the thrower's hand (Fig. 2–31) at point A. Let us choose y to be positive in the upward direction and negative in the downward direction. (This is a different convention from that used in Examples 2–15 and 2–16, and so illustrates our options.) The acceleration due to gravity is downward and so will have a negative sign, $a = -g = -9.80\,\text{m/s}^2$. As the ball rises, its speed decreases until it reaches the highest point (B in Fig. 2–31), where its speed is zero for an instant; then it descends, with increasing speed.

SOLUTION We consider the time interval from when the ball leaves the thrower's hand until the ball reaches the highest point. To determine the maximum height, we calculate the position of the ball when its velocity equals zero ($v = 0$ at the highest point). At $t = 0$ (point A in Fig. 2–31) we set $y_0 = 0$, $v_0 = 15.0\,\text{m/s}$, and $a = -9.80\,\text{m/s}^2$. At time t (maximum height), $v = 0$, $a = -9.80\,\text{m/s}^2$, and we wish to find y. We use Eq. 2–12c, replacing x with y: $v^2 = v_0^2 + 2ay$. We solve this equation for y:

$$y = \frac{v^2 - v_0^2}{2a} = \frac{0 - (15.0\,\text{m/s})^2}{2(-9.80\,\text{m/s}^2)} = 11.5\,\text{m}.$$

The ball reaches a height of 11.5 m above the hand.

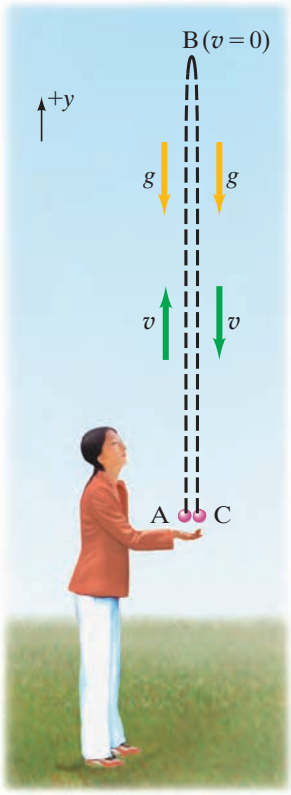


FIGURE 2–31 (Repeated.) An object thrown into the air leaves the thrower's hand at A, reaches its maximum height at B, and returns to the original position at C. Examples 2–17, 2–18, 2–19, 2–20, and 2–21.

EXAMPLE 2–18 **Ball thrown upward, II.** In Example 2–17, Fig. 2–31 (shown here again), how long is the ball in the air before it comes back to the hand?

APPROACH We need to choose a time interval to calculate how long the ball is in the air before it returns to the hand. We could do this calculation in two parts by first determining the time required for the ball to reach its highest point, and then determining the time it takes to fall back down. However, it is simpler to consider the time interval for the entire motion from A to B to C (Fig. 2–31) in one step and use Eq. 2–12b. We can do this because y is position (or displacement from the origin), and not the total distance traveled. Thus, at both points A and C we have $y = 0$.

SOLUTION We use Eq. 2–12b with $a = -9.80\,\text{m/s}^2$ and find

$$y = y_0 + v_0 t + \tfrac{1}{2}at^2$$
$$0 = 0 + (15.0\,\text{m/s})t + \tfrac{1}{2}(-9.80\,\text{m/s}^2)t^2.$$

This equation can be factored (we factor out one t):

$$(15.0\,\text{m/s} - 4.90\,\text{m/s}^2\, t)\, t = 0.$$

There are two solutions:

$$t = 0 \quad \text{and} \quad t = \frac{15.0\,\text{m/s}}{4.90\,\text{m/s}^2} = 3.06\,\text{s}.$$

The first solution ($t = 0$) corresponds to the initial point (A) in Fig. 2–31, when the ball was first thrown from $y_0 = 0$. The second solution, $t = 3.06\,\text{s}$, corresponds to point C, when the ball has returned to $y = 0$. Thus the ball is in the air 3.06 s.

NOTE We have ignored air resistance in these last two Examples, which could be significant, so our result is only an approximation to a real, practical situation.

We did not consider the throwing action in these Examples. Why? Because during the throw, the thrower's hand is touching the ball and accelerating the ball at a rate unknown to us—the acceleration is *not* g. We consider only the time when the ball is in the air and the acceleration is equal to g.

Every quadratic equation (where the variable is squared) mathematically produces two solutions. In physics, sometimes only one solution corresponds to the real situation, as we saw in Example 2–12, in which case we ignore the "unphysical" solution. But here in Example 2–18, both solutions to our equation in t^2 are physically meaningful: $t = 0$ and $t = 3.06\,\text{s}$.

CAUTION

Quadratic equations have two solutions. Sometimes only one corresponds to reality, sometimes both

CONCEPTUAL EXAMPLE 2–19 **Two possible misconceptions.** Give examples to show the error in these two common misconceptions: (1) that acceleration and velocity are always in the same direction, and (2) that an object thrown upward has zero acceleration at the highest point (B in Fig. 2–31).

RESPONSE Both are wrong. (1) Velocity and acceleration are *not* necessarily in the same direction. When the ball in Fig. 2–31 is moving upward, its velocity is positive (upward), whereas the acceleration is negative (downward). (2) At the highest point (B in Fig. 2–31), the ball has zero velocity for an instant. Is the acceleration also zero at this point? No. The velocity near the top of the arc points upward, then becomes zero for an instant (zero time) at the highest point, and then points downward. Gravity does not stop acting, so $a = -g = -9.80\,\text{m/s}^2$ even there. Thinking that $a = 0$ at point B would lead to the conclusion that upon reaching point B, the ball would stay there: if the acceleration (= rate of change of velocity) were zero, the velocity would stay zero at the highest point, and the ball would stay up there without falling. Remember: the acceleration of gravity always points down toward the center of the Earth, even when the object is moving up.

CAUTION

(1) Velocity and acceleration are not always in the same direction; the acceleration (of gravity) always points down
(2) $a \neq 0$ even at the highest point of a trajectory

EXAMPLE 2–20 **Ball thrown upward, III.** Let us consider again the ball thrown upward of Examples 2–17 and 2–18, and make more calculations. Calculate (*a*) how much time it takes for the ball to reach the maximum height (point B in Fig. 2–31), and (*b*) the velocity of the ball when it returns to the thrower's hand (point C).

APPROACH Again we assume the acceleration is constant, so we can use Eqs. 2–12. We have the maximum height of 11.5 m and initial speed of 15.0 m/s from Example 2–17. Again we take y as positive upward.

SOLUTION (*a*) We consider the time interval between the throw $(t_0 = 0, v_0 = 15.0\,\text{m/s})$ and the top of the path ($y = +11.5\,\text{m}$, $v = 0$), and we want to find t. The acceleration is constant at $a = -g = -9.80\,\text{m/s}^2$. Both Eqs. 2–12a and 2–12b contain the time t with other quantities known. Let us use Eq. 2–12a with $a = -9.80\,\text{m/s}^2$, $v_0 = 15.0\,\text{m/s}$, and $v = 0$:

$$v = v_0 + at;$$

setting $v = 0$ gives $0 = v_0 + at$, which we rearrange to solve for t: $at = -v_0$ or

$$t = -\frac{v_0}{a} = -\frac{15.0\,\text{m/s}}{-9.80\,\text{m/s}^2} = 1.53\,\text{s}.$$

This is just half the time it takes the ball to go up and fall back to its original position [3.06 s, calculated in Example 2–18]. Thus it takes the *same time* to reach the maximum height as to fall back to the starting point. We might have guessed this from the *symmetry* of the motion. But be careful. When air resistance cannot be neglected, the symmetry is no longer perfect.

(*b*) To find the ball's velocity when it returns to the hand (point C), we consider the time interval from the throw $(t_0 = 0, v_0 = 15.0\,\text{m/s})$ until the ball's return to the hand, which occurs at $t = 3.06\,\text{s}$ (as calculated in Example 2–18). We use Eq. 2–12a again to find v when $t = 3.06\,\text{s}$:

$$\begin{aligned} v &= v_0 + at \\ &= 15.0\,\text{m/s} - (9.80\,\text{m/s}^2)(3.06\,\text{s}) = -15.0\,\text{m/s}. \end{aligned}$$

NOTE The ball has the same speed (magnitude of velocity) when it returns to the starting point as it did initially, but in the opposite direction (this is the meaning of the negative sign). And, as we saw in part (*a*), the time is the same up as down. Thus the motion is *symmetrical* about the maximum height.

PROBLEM SOLVING
Acceleration in g's

The acceleration of objects such as rockets and fast airplanes is often given as a multiple of $g = 9.80\,\text{m/s}^2$. For example, an airplane pulling out of a dive (see Fig. 2–32) and undergoing 3.00 *g*'s would have an acceleration of $(3.00)(9.80\,\text{m/s}^2) = 29.4\,\text{m/s}^2$.



FIGURE 2–32 Several airplanes, in formation, are just coming out of a downward dive.

EXAMPLE 2–21 **Ball thrown upward, IV; the quadratic formula.** For the ball in Example 2–20, calculate at what time t the ball passes a point 8.00 m above the person's hand. (See Fig. 2–31, repeated below.)

APPROACH We choose the time interval from the throw $(t_0 = 0,\ v_0 = 15.0\ \text{m/s})$ until the time t (to be determined) when the ball is at position $y = 8.00\ \text{m}$, using Eq. 2–12b.

SOLUTION We want to find t, given $y = 8.00\ \text{m}$, $y_0 = 0$, $v_0 = 15.0\ \text{m/s}$, and $a = -9.80\ \text{m/s}^2$. We use Eq. 2–12b:

$$\begin{aligned} y &= y_0 + v_0 t + \tfrac{1}{2}at^2 \\ 8.00\ \text{m} &= 0 + (15.0\ \text{m/s})t + \tfrac{1}{2}(-9.80\ \text{m/s}^2)t^2. \end{aligned}$$

To solve any quadratic equation of the form $at^2 + bt + c = 0$, where a, b, and c are constants (a is not acceleration here), we use the **quadratic formula**:

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

We rewrite our y equation just above in standard form, $at^2 + bt + c = 0$:

$$(4.90\ \text{m/s}^2)t^2 - (15.0\ \text{m/s})t + (8.00\ \text{m}) = 0.$$

So the coefficient a is $4.90\ \text{m/s}^2$, b is $-15.0\ \text{m/s}$, and c is 8.00 m. Putting these into the quadratic formula, we obtain

$$t = \frac{15.0\ \text{m/s} \pm \sqrt{(-15.0\ \text{m/s})^2 - 4(4.90\ \text{m/s}^2)(8.00\ \text{m})}}{2(4.90\ \text{m/s}^2)},$$

which gives us $t = 0.69\ \text{s}$ and $t = 2.37\ \text{s}$. Are both solutions valid? Yes, because the ball passes $y = 8.00\ \text{m}$ when it goes up $(t = 0.69\ \text{s})$ and again when it comes down $(t = 2.37\ \text{s})$.

NOTE Figure 2–33 shows graphs of (a) y vs. t and (b) v vs. t for the ball thrown upward in Fig. 2–31, incorporating the results of Examples 2–17, 2–18, 2–20, and 2–21.

PROBLEM SOLVING
Quadratic formula is a very useful tool

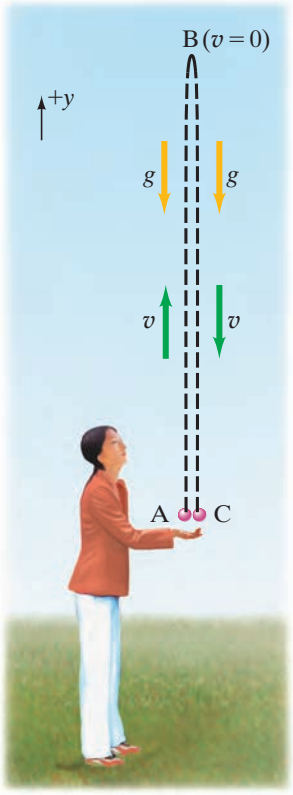
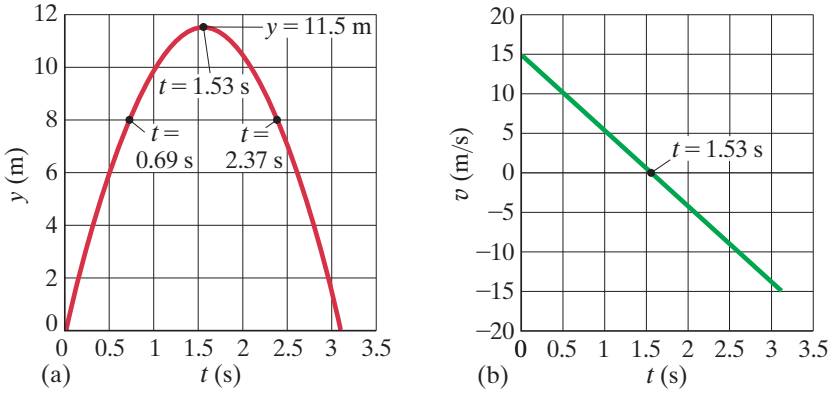


FIGURE 2–31 (Repeated.) An object thrown into the air leaves the thrower's hand at A, reaches its maximum height at B, and returns to the original position at C. Examples 2–17, 2–18, 2–19, 2–20, and 2–21.

FIGURE 2–33 Graphs of (a) y vs. t, (b) v vs. t, for a ball thrown upward, Examples 2–17, 2–18, 2–20, and 2–21.



EXAMPLE 2–22 **Ball thrown upward at edge of cliff.** Suppose that the person of Examples 2–17, 2–18, 2–20, and 2–21 is standing on the edge of a cliff, so that the ball can fall to the base of the cliff 50.0 m below, as shown in Fig. 2–34. (a) How long does it take the ball to reach the base of the cliff? (b) What is the total distance traveled by the ball? Ignore air resistance (likely to be significant, so our result is an approximation).

APPROACH We again use Eq. 2–12b, but this time we set $y = -50.0\ \text{m}$, the bottom of the cliff, which is 50.0 m below the initial position $(y_0 = 0)$.

SOLUTION (*a*) We use Eq. 2–12b with $a = -9.80\,\text{m/s}^2$, $v_0 = 15.0\,\text{m/s}$, $y_0 = 0$, and $y = -50.0\,\text{m}$:

$$\begin{aligned} y &= y_0 + v_0 t + \tfrac{1}{2}at^2 \\ -50.0\,\text{m} &= 0 + (15.0\,\text{m/s})t - \tfrac{1}{2}(9.80\,\text{m/s}^2)t^2. \end{aligned}$$

Rewriting in the standard form we have

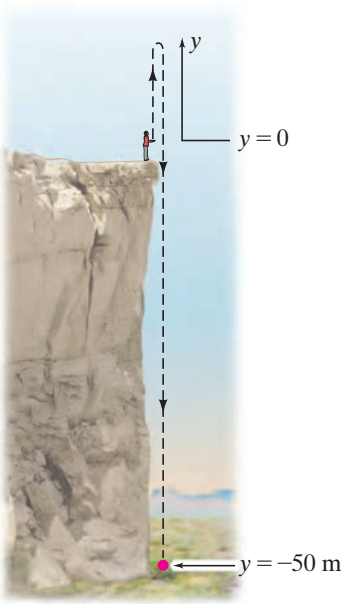
$$(4.90\,\text{m/s}^2)t^2 - (15.0\,\text{m/s})t - (50.0\,\text{m}) = 0.$$

Using the quadratic formula, we find as solutions $t = 5.07\,\text{s}$ and $t = -2.01\,\text{s}$. The first solution, $t = 5.07\,\text{s}$, is the answer we are seeking: the time it takes the ball to rise to its highest point and then fall to the base of the cliff. To rise and fall back to the top of the cliff took 3.06 s (Example 2–18); so it took an additional 2.01 s to fall to the base. But what is the meaning of the other solution, $t = -2.01\,\text{s}$? This is a time before the throw, when our calculation begins, so it isn't relevant here. It is outside our chosen time interval, and so is an *unphysical* solution just as in Example 2–12.

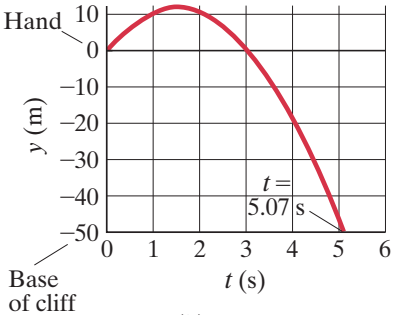
(*b*) For the total distance traveled, Example 2–17 told us that the ball moves up 11.5 m, falls 11.5 m back down to the top of the cliff; it then falls down another 50.0 m to the base of the cliff, for a total distance traveled of 2 (11.5 m) + 50.0 m = 73.0 m. Note that the *displacement*, however, was −50.0 m. Figure 2–34b shows the y vs. t graph for this situation.

NOTE The solution $t = -2.01\,\text{s}$ in part (*a*) could be meaningful in a different physical situation. Suppose that a person standing on top of a 50.0-m-high cliff sees a rock pass by her at $t = 0$ moving upward at 15.0 m/s; at what time did the rock leave the base of the cliff, and when did it arrive back at the base of the cliff? The equations will be precisely the same as for our original Example, and the answers $t = -2.01\,\text{s}$ and $t = 5.07\,\text{s}$ will be the correct answers. Note that we cannot put *all* the information for a problem into the mathematics, so we have to use common sense in interpreting results.

EXERCISE H Two balls are thrown from a cliff. One is thrown directly up, the other directly down, each with the same initial speed, and both hit the ground below the cliff. Which ball hits the ground at the greater speed: (*a*) the ball thrown upward, (*b*) the ball thrown downward, or (*c*) both the same? Ignore air resistance.



(a)



(b)

FIGURE 2–34 Example 2–22. (*a*) A person stands on the edge of a cliff. A ball is thrown upward, and then falls back down past the thrower to the base of the cliff, 50.0 m below. (*b*) The y vs. t graph.

*2–8 Variable Acceleration; Integral Calculus

In this brief optional Section we use integral calculus to derive the kinematic equations for constant acceleration, Eqs. 2–12a and b. We also show how calculus can be used when the acceleration is not constant. If you have not yet studied simple integration in your calculus course, you may want to postpone reading this Section until you have. We discuss integration in more detail in Section 7–3, where we begin to use it in the physics.

First we derive Eq. 2–12a, assuming as we did in Section 2–5 that an object has velocity v_0 at $t = 0$ and a constant acceleration a. We start with the definition of instantaneous acceleration, $a = dv/dt$, which we rewrite as

$$dv = a\,dt.$$

We take the definite integral of both sides of this equation, using the same notation we did in Section 2–5, from $v = v_0$ at $t_0 = 0$ to some velocity v at time t:

$$\int_{v=v_0}^{v} dv = \int_{t_0=0}^{t} a\,dt$$

which gives, since a = constant,

$$v - v_0 = at.$$

This is Eq. 2–12a, $v = v_0 + at$.

Next we derive Eq. 2–12b starting with the definition of instantaneous velocity, Eq. 2–4, $v = dx/dt$. We rewrite this as

$$dx = v\,dt$$

or

$$dx = (v_0 + at)dt$$

where we substituted in Eq. 2–12a.

Now we integrate from $x = x_0$ at $t_0 = 0$ to an arbitrary position x at time t:

$$\int_{x=x_0}^{x} dx = \int_{t_0=0}^{t} (v_0 + at)dt$$

$$x - x_0 = \int_{t_0=0}^{t} v_0\,dt + \int_{t_0=0}^{t} at\,dt$$

$$x - x_0 = v_0 t + \tfrac{1}{2}at^2$$

since v_0 and a are constants. This result is just Eq. 2–12b, $x = x_0 + v_0 t + \frac{1}{2}at^2$.

Finally let us use calculus to find velocity and displacement, given an acceleration that is not constant but varies in time.

EXAMPLE 2–23 **Integrating a time-varying acceleration.** An experimental vehicle starts from rest $(v_0 = 0)$ at $t = 0$ and accelerates for a few seconds at a rate given by $a = (7.00\,\text{m/s}^3)t$. What is (*a*) its velocity and (*b*) its displacement 2.00 s later?

APPROACH We cannot use Eqs. 2–12 because a is not constant. We integrate the acceleration $a = dv/dt$ over time to find v as a function of time; and then integrate $v = dx/dt$ to get the displacement.

SOLUTION From the definition of acceleration, $a = dv/dt$, we have

$$dv = a\,dt.$$

We take the integral of both sides from $v = 0$ at $t = 0$ to velocity v at an arbitrary time t:

$$\int_0^v dv = \int_0^t a\,dt$$

$$v = \int_0^t (7.00\,\text{m/s}^3)t\,dt$$

$$= (7.00\,\text{m/s}^3)\left(\frac{t^2}{2}\right)\Bigg|_0^t = (7.00\,\text{m/s}^3)\left(\frac{t^2}{2} - 0\right) = (3.50\,\text{m/s}^3)t^2.$$

At $t = 2.00$ s, $v = (3.50\,\text{m/s}^3)(2.00\,\text{s})^2 = 14.0$ m/s.

(*b*) To get the displacement, we assume $x_0 = 0$ and start with $v = dx/dt$ which we rewrite as $dx = v\,dt$. Then we integrate from $x = 0$ at $t = 0$ to position x at time t:

$$\int_0^x dx = \int_0^t v\,dt$$

$$x = \int_0^{2.00\,\text{s}} (3.50\,\text{m/s}^3)t^2\,dt = (3.50\,\text{m/s}^3)\,\frac{t^3}{3}\Bigg|_0^{2.00\,\text{s}} = 9.33\,\text{m}.$$

In sum, at $t = 2.00$ s, $v = 14.0$ m/s and $x = 9.33$ m.

Some problems in kinematics can be solved using Numerical Integration, which is discussed in Appendix C.

Summary

[*The Summary that appears at the end of each Chapter in this book gives a brief overview of the main ideas of the Chapter. The Summary* cannot *serve to give an understanding of the material, which can be accomplished only by a detailed reading of the Chapter.*]

Kinematics deals with the description of how objects move. The description of the motion of any object must always be given relative to some particular **reference frame**.

The **displacement** of an object is the change in position of the object.

Average speed is the distance traveled divided by the **elapsed time** or **time interval**, Δt, the time period over which we choose to make our observations. An object's **average velocity** over a particular time interval Δt is its displacement Δx during that time interval, divided by Δt:

$$\bar{v} = \frac{\Delta x}{\Delta t}. \qquad \textbf{(2–2)}$$

The **instantaneous velocity**, whose magnitude is the same as the *instantaneous speed*, is defined as the average velocity taken over an infinitesimally short time interval $(\Delta t \to 0)$:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}, \qquad \textbf{(2–4)}$$

where dx/dt is the derivative of x with respect to t.

On a graph of position vs. time, the **slope** is equal to the instantaneous velocity.

Acceleration is the change of velocity per unit time. An object's **average acceleration** over a time interval Δt is

$$\bar{a} = \frac{\Delta v}{\Delta t}, \qquad \textbf{(2–5)}$$

where Δv is the change of velocity during the time interval Δt.

Instantaneous acceleration is the average acceleration taken over an infinitesimally short time interval:

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt}. \qquad \textbf{(2–6)}$$

On a graph of velocity vs. time, the slope is equal to the instantaneous acceleration.

If an object has position x_0 and velocity v_0 at time $t = 0$ and moves in a straight line with **constant acceleration**, the velocity v and position x at a later time t are related to the acceleration a, the initial position x_0, and the initial velocity v_0 by Eqs. 2–12:

$$\begin{aligned} v &= v_0 + at, \\ x &= x_0 + v_0 t + \tfrac{1}{2}at^2, \\ v^2 &= v_0^2 + 2a(x - x_0), \\ \bar{v} &= \frac{v + v_0}{2}. \end{aligned} \qquad \textbf{(2–12)}$$

Objects that move vertically near the surface of the Earth, either falling or having been projected vertically up or down, move with the constant downward **acceleration due to gravity**, whose magnitude is $g = 9.80\ \text{m/s}^2$ if air resistance can be ignored.

[*Integral calculus can be used to derive the kinematic Equations 2–12 and to solve Problems involving varying acceleration. Numerical integration is also a useful tool.]

Questions

1. Does a car speedometer measure speed, velocity, or both? Explain.
2. Can an object have a varying speed if its velocity is constant? Can it have varying velocity if its speed is constant? If yes, give examples in each case.
3. When an object moves with constant velocity, does its average velocity during any time interval differ from its instantaneous velocity at any instant? Explain.
4. If one object has a greater speed than a second object, does the first necessarily have a greater acceleration? Explain, using examples.
5. Compare the acceleration of a motorcycle that accelerates from 80 km/h to 90 km/h with the acceleration of a bicycle that accelerates from rest to 10 km/h in the same time.
6. Can an object have a northward velocity and a southward acceleration? Explain.
7. Can the velocity of an object be negative when its acceleration is positive? What about vice versa? If yes, give examples.
8. Give an example where both the velocity and acceleration are negative.
9. Two cars emerge side by side from a tunnel. Car A is traveling with a speed of 60 km/h and has an acceleration of 40 km/h/min. Car B has a speed of 40 km/h and has an acceleration of 60 km/h/min. Which car is passing the other as they come out of the tunnel? Explain your reasoning.
10. Can an object be increasing in speed as its acceleration decreases? If so, give an example. If not, explain.
11. A baseball player hits a ball straight up into the air. It leaves the bat with a speed of 120 km/h. Ignoring air resistance, how fast would the ball be traveling when the catcher catches it (at the same height it left the bat)? Explain.
12. As a freely falling object speeds up, what is happening to its acceleration—does it increase, decrease, or stay the same? (*a*) Ignore air resistance. (*b*) Consider air resistance.
13. You travel from point A to point B in a car moving at a constant speed of 70 km/h. Then you travel the same distance from point B to another point C, moving at a constant speed of 90 km/h. Is your average speed 80 km/h for the entire trip from A to C? Explain why or why not.
14. Can an object have zero velocity and nonzero acceleration at the same time? Give examples.
15. Can an object have zero acceleration and nonzero velocity at the same time? Give examples.
16. Which of these motions is *not* at constant acceleration: a rock falling from a cliff, an elevator moving from the second floor to the fifth floor making stops along the way, a dish resting on a table? Explain your answers.
17. Discuss two conditions given in Section 2–7 for being able to use a constant acceleration of magnitude $g = 9.8\ \text{m/s}^2$. Give an example in which one of these conditions would not be met and would not even be a reasonable approximation of motion. [*Hint*: Carefully read Section 2–7, especially page 38.]

18. Describe in words the motion plotted in Fig. 2–35 in terms of velocity, acceleration, etc. [*Hint*: First try to duplicate the motion plotted by walking or moving your hand.]

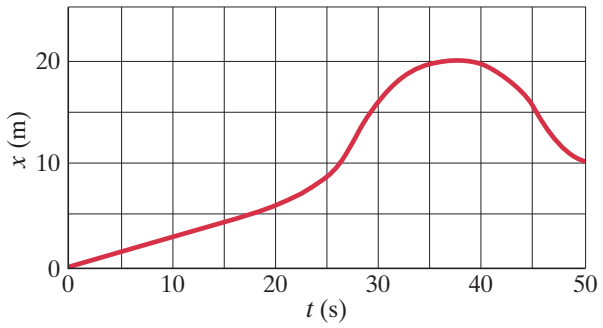


FIGURE 2–35 Question 18.

19. Describe in words the motion of the object graphed in Fig. 2–36.

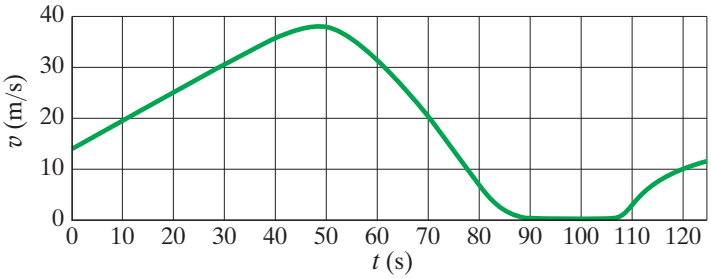


FIGURE 2–36 Question 19.

MisConceptual Questions

[List all answers that are valid, and ignore air resistance.]

1. In which of the following cases does a car have a negative velocity and a positive acceleration? A car that is traveling in the
(*a*) $-x$ direction at a constant 20 m/s.
(*b*) $-x$ direction increasing in speed.
(*c*) $+x$ direction increasing in speed.
(*d*) $-x$ direction decreasing in speed.
(*e*) $+x$ direction decreasing in speed.

2. At time $t = 0$ an object is traveling to the right along the $+x$ axis at a speed of 10.0 m/s with constant acceleration of $-2.0\ \text{m/s}^2$. Which statement is true?
(*a*) The object will slow down, eventually coming to a complete stop.
(*b*) The object cannot have a negative acceleration and be moving to the right.
(*c*) The object will continue to move to the right, slowing down but never coming to a complete stop.
(*d*) The object will slow down, momentarily stopping, then pick up speed moving to the left.

3. You drive 4 km at 30 km/h and then another 4 km at 50 km/h. What is your average speed for the whole 8-km trip?
(*a*) More than 40 km/h. (*c*) Less than 40 km/h.
(*b*) Equal to 40 km/h. (*d*) Not enough information.

4. Two cars start from rest and travel a distance d with constant acceleration. The acceleration of car B is four times that of car A. After each has traveled distance d,
(*a*) car B is moving 16 times as fast as car A.
(*b*) car B is moving 8 times as fast as car A.
(*c*) car B is moving 4 times as fast as car A.
(*d*) car B is moving 2 times as fast as car A.

5. A rock is thrown straight up rising to a maximum height before falling back down.
(*a*) The acceleration is constant for the entire trip.
(*b*) The velocity is constant for the entire trip.
(*c*) The magnitudes of both the velocity and acceleration decrease on the way up and increase on the way down.
(*d*) The acceleration is constant for the entire trip except at the top where it is 0.

6. A ball is dropped from the top of a tall building. At the same instant, a second ball is thrown upward from ground level. When the two balls pass one another, one on the way up, the other on the way down, compare the magnitudes of their acceleration:
(*a*) The acceleration of the dropped ball is greater.
(*b*) The acceleration of the ball thrown upward is greater.
(*c*) The acceleration of both balls is the same.
(*d*) The acceleration changes during the motion, so you cannot predict the exact value when the two balls pass each other.
(*e*) The accelerations are in opposite directions.

7. You drop a rock off a bridge. When the rock has fallen 4 m, you drop a second rock. As the two rocks continue to fall, what happens to their velocities?
(*a*) Both increase at the same rate.
(*b*) The velocity of the first rock increases faster than the velocity of the second.
(*c*) The velocity of the second rock increases faster than the velocity of the first.
(*d*) Both velocities stay constant.

8. Two objects are dropped from a bridge, an interval of 1.0 s apart. During the time that both objects continue to fall, their separation
(*a*) decreases at first, but then stays constant.
(*b*) increases at first, but then stays constant.
(*c*) decreases.
(*d*) stays constant.
(*e*) increases.

9. A ball is thrown downward at a speed of 20 m/s. Choosing the $+y$ axis pointing up and neglecting air resistance, which equation(s) would correctly describe the motions? The acceleration due to gravity is $g = 9.8\ \text{m/s}^2$ downward.
(*a*) $v = (20\ \text{m/s}) - gt$.
(*b*) $y = y_0 + (-20\ \text{m/s})t - (1/2)gt^2$.
(*c*) $v^2 = (20\ \text{m/s})^2 - 2g(y - y_0)$.
(*d*) $(20\ \text{m/s}) = (v + v_0)/2$.
(*e*) All of the above.

10. A car travels along the x axis with increasing speed. We are not sure if it is moving to the left or to the right. Which of the graphs in Fig. 2–37 could possibly represent the motion of the car?

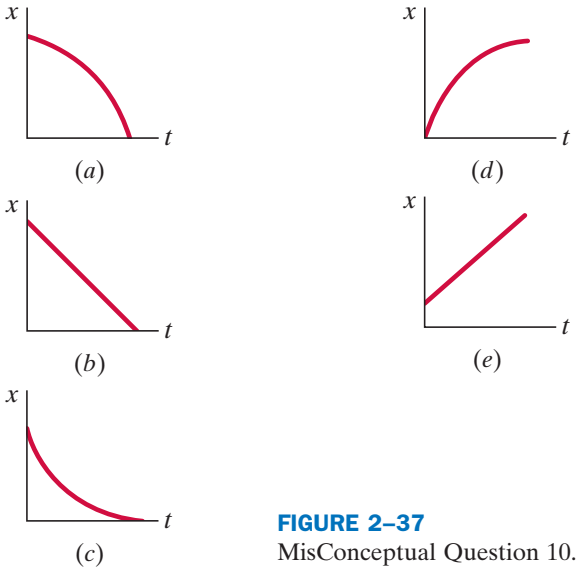


FIGURE 2–37
MisConceptual Question 10.

11. Two objects start at the same place at the same time and move along the same straight line (the x axis). Figure 2–38 shows the position x as a function of time t for each object. At point A, what must be true about the motion of the objects? (More than one statement may be correct.)
(*a*) Both have the same instantaneous speed.
(*b*) Both have the same instantaneous velocity.
(*c*) Both are at the same position.
(*d*) Both have traveled the same total distance.
(*e*) With respect to the start, both have the same average velocity.

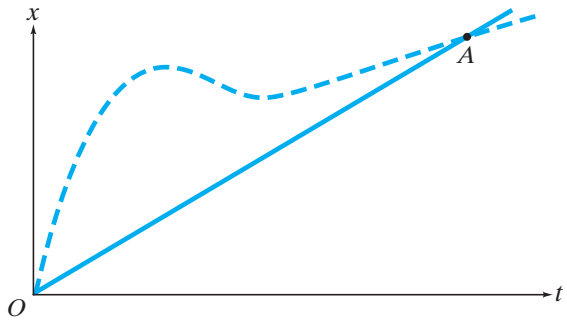


FIGURE 2–38 MisConceptual Question 11.

Problems

[*The Problems at the end of each Chapter are ranked* I, II, *or* III *according to estimated difficulty, with level* I *Problems being easiest. Level* III *are meant as challenges for the best students. The Problems are arranged by Section, meaning that the reader should have read up to and including that Section, but not only that Section—Problems often depend on earlier material. Next is a set of "General Problems" not arranged by Section and not ranked.*]

(Note: *In Problems, assume a number like* 6.4 *is accurate to* ±0.1; *and* 950 *is* ±10 *unless* 950 *is said to be "precisely" or "very nearly"* 950, *in which case assume* 950 ± 1. *See Section 1–3.*)

2–1 to 2–3 Speed and Velocity

1. (I) If you are driving 85 km/h along a straight road and you look to the side for 2.0 s, how far do you travel during this inattentive period?

2. (I) What must your car's average speed be in order to travel 235 km in 2.85 h?

3. (I) A particle at $t_1 = -2.0$ s is at $x_1 = 5.2$ cm and at $t_2 = 3.4$ s is at $x_2 = 8.5$ cm. What is its average velocity over this time interval? Can you calculate its average speed from these data? Why or why not?

4. (II) According to a rule-of-thumb, each five seconds between a lightning flash and the following thunder gives the distance to the flash in miles. (*a*) Assuming that the light from the flash arrives in essentially no time at all, estimate the speed of sound in m/s from this rule. (*b*) What would be the rule for kilometers?

5. (II) You are driving home from school steadily at 95 km/h for 210 km. It then begins to rain and you slow to 65 km/h. You arrive home after driving 4.5 h. (*a*) How far is your hometown from school? (*b*) What was your average speed?

6. (II) A horse trots away from its trainer in a straight line, moving 38 m away in 7.4 s. It then turns abruptly and gallops halfway back in 1.8 s. Calculate (*a*) its average speed and (*b*) its average velocity for the entire trip, using "away from the trainer" as the positive direction.

7. (II) A person jogs eight complete laps around a 400-m track in a total time of 14.5 min. Calculate (*a*) the average speed and (*b*) the average velocity, in m/s.

8. (II) Every year the Earth travels about 10^9 km as it orbits the Sun. What is Earth's average speed in km/h?

9. (II) A car traveling 95 km/h is 310 m behind a truck traveling 75 km/h. How long will it take the car to reach the truck?

10. (II) Calculate (*a*) the average speed and (*b*) average velocity of a round trip: the outgoing 280 km is covered at 95 km/h, followed by a 1.0-h lunch break, and the return 280 km is covered at 55 km/h.

11. (II) Two locomotives approach each other on parallel tracks. Each has a speed of 155 km/h with respect to the ground. If they are initially 9.5 km apart, how long will it be before they reach each other? (See Fig. 2–39.)

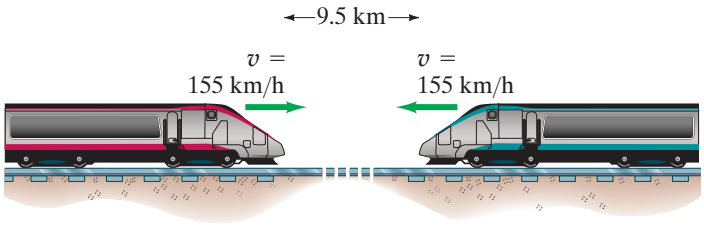


FIGURE 2–39 Problem 11.

12. (II) Digital bits on a 12.0-cm diameter audio CD are encoded along an outward spiraling path that starts at radius $R_1 = 2.5\,\text{cm}$ and finishes at radius $R_2 = 5.8\,\text{cm}$. The distance between the centers of neighboring spiral-windings is $1.6\,\mu\text{m}$ $(= 1.6 \times 10^{-6}\,\text{m})$. (*a*) Determine the total length of the spiraling path. [*Hint*: Imagine "unwinding" the spiral into a straight path of width $1.6\,\mu\text{m}$, and note that the original spiral and the straight path both occupy the same area.] (*b*) The CD player adjusts the rotation of the CD so that the player's readout laser reads along the spiral path at a constant rate of about 1.2 m/s. Estimate the maximum playing time of such a CD.

13. (II) The position of a small object is given by $x = 27 + 10t - 2t^3$, where t is in seconds and x in meters. (*a*) Plot x as a function of t from $t = 0$ to $t = 3.0\,\text{s}$. (*b*) Find the average velocity of the object between 0 and 3.0 s. (*c*) At what time between 0 and 3.0 s is the instantaneous velocity zero?

14. (II) An airplane travels 1900 km at a speed of 720 km/h, and then encounters a tailwind that boosts its speed to 990 km/h for the next 2700 km. What was the total time for the trip? What was the average speed of the plane for this trip? [*Hint*: Does Eq. 2–12d apply?]

15. (II) Two ships need to arrive at a site in the middle of the ocean at the same time. They start out at the same time from positions equally distant from the arrival site. They travel at different velocities but both go in a straight line. The first ship travels at an average velocity of 20 km/h for the first 600 km, 40 km/h for the next 800 km, and 20 km/h for the final 600 km. The second ship can only sail at constant velocity. What is the magnitude of that velocity?

16. (II) The position of an object along a straight tunnel as a function of time is plotted in Fig. 2–40. What is its instantaneous velocity (*a*) at $t = 10.0\,\text{s}$ and (*b*) at $t = 30.0\,\text{s}$? What is its average velocity (*c*) between $t = 0$ and $t = 5.0\,\text{s}$, (*d*) between $t = 25.0\,\text{s}$ and $t = 30.0\,\text{s}$, and (*e*) between $t = 40.0\,\text{s}$ and $t = 50.0\,\text{s}$?

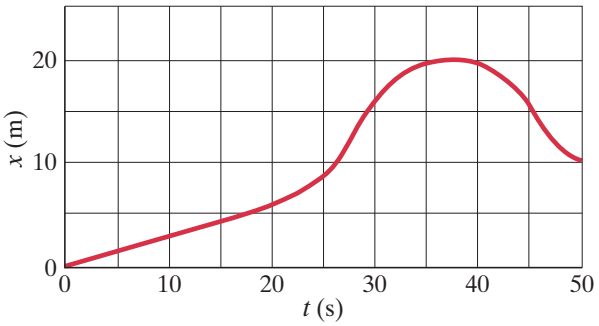


FIGURE 2–40 Problems 16, 17, and 18.

17. (II) In Fig. 2–40, (*a*) during what time intervals, if any, is the velocity constant? (*b*) At what time is the velocity greatest? (*c*) At what time, if any, is the velocity zero? (*d*) Does the object move in one direction or in both directions during the time shown?

18. (III) Sketch the v vs. t graph for the object whose displacement as a function of time is given by Fig. 2–40.

19. (III) A bowling ball traveling with constant speed hits the pins at the end of a bowling lane 16.5 m long. The bowler hears the sound of the ball hitting the pins 2.75 s after the ball is released from his hands. What is the speed of the ball, assuming the speed of sound is 340 m/s?

20. (III) An automobile traveling 95 km/h overtakes a 1.50-km-long train traveling in the same direction on a track parallel to the road. If the train's speed is 75 km/h, how long does it take the car to pass it, and how far will the car have traveled in this time? See Fig. 2–41. What are the results if the car and train are traveling in opposite directions?

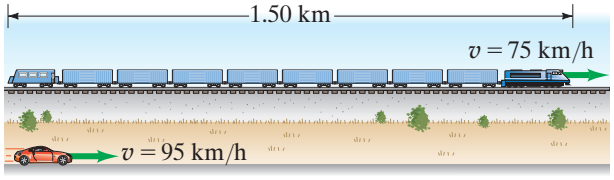


FIGURE 2–41 Problem 20.

2–4 Acceleration

21. (I) A sprinter accelerates from rest to 9.00 m/s in 1.48 s. What is her acceleration in (*a*) m/s^2; (*b*) km/h^2?

22. (I) A bicyclist in the Tour de France crests a mountain pass as he moves at 15 km/h. At the bottom, 4.0 km farther, his speed is 65 km/h. Estimate his average acceleration (in m/s^2) while riding down the mountain.

23. (II) A sports car moving at constant velocity travels 120 m in 5.0 s. If it then brakes and comes to a stop in 3.7 s, what is the magnitude of its acceleration (assumed constant) in m/s^2, and in g's $(g = 9.80\,\text{m/s}^2)$?

24. (II) At highway speeds, a particular automobile is capable of an acceleration of about $1.8\,\text{m/s}^2$. At this rate, how long does it take to accelerate from 65 km/h to 120 km/h?

25. (II) A car moving in a straight line starts at $x = 0$ when $t_0 = 0$. It passes the point $x = 25.0\,\text{m}$ with a speed of 11.0 m/s at $t = 3.00\,\text{s}$. It passes the point $x = 385\,\text{m}$ with a speed of 45.0 m/s at $t = 20.0\,\text{s}$. Find (*a*) the average velocity, and (*b*) the average acceleration, between $t = 3.00\,\text{s}$ and $t = 20.0\,\text{s}$.

26. (II) A particle moves along the x axis. Its position as a function of time is given by $x = 4.8t + 7.3t^2$, where t is in seconds and x is in meters. What is the acceleration as a function of time?

27. (II) The position of an object is given by $x = At + Bt^2$, where x is in meters and t is in seconds. (*a*) What are the units of A and B? (*b*) What is the acceleration as a function of time? (*c*) What are the velocity and acceleration at $t = 6.0\,\text{s}$? (*d*) What is the velocity as a function of time if $x = At + Bt^{-3}$?

28. (II) The position of a race car, which starts from rest at $t = 0$ and moves in a straight line, is given as a function of time in the following Table. Estimate (*a*) its velocity and (*b*) its acceleration as a function of time. Display each in a Table and on a graph.

t(s)	0	0.25	0.50	0.75	1.00	1.50	2.00	2.50
x(m)	0	0.11	0.46	1.06	1.94	4.62	8.55	13.79

t(s)	3.00	3.50	4.00	4.50	5.00	5.50	6.00
x(m)	20.36	28.31	37.65	48.37	60.30	73.26	87.16

29. (II) A car traveling 25.0 m/s passes a second car which is at rest. When the cars are right next to each other, the first car slows down at a constant rate of $2.0\,\text{m/s}^2$ and the second car starts to accelerate at the same constant rate. When will the two cars be next to each other again?

30. (II) Figure 2–42 shows the velocity of a train as a function of time. (*a*) At what time was its velocity greatest? (*b*) During what periods, if any, was the velocity constant? (*c*) During what periods, if any, was the acceleration constant? (*d*) When was the magnitude of the acceleration greatest?

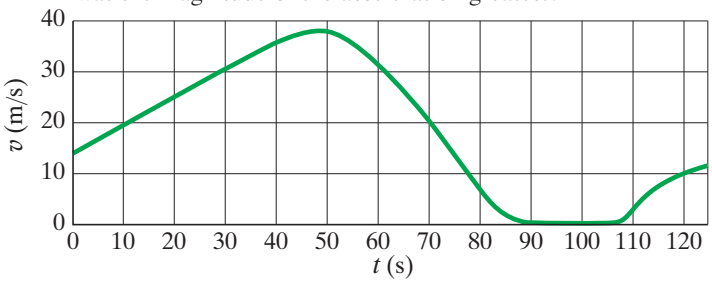


FIGURE 2–42 Problem 30.

31. (II) A sports car accelerates approximately as shown in the velocity–time graph of Fig. 2–43. (The short flat spots in the curve represent manual shifting of the gears.) Estimate the car's average acceleration in (*a*) second gear and (*b*) fourth gear.

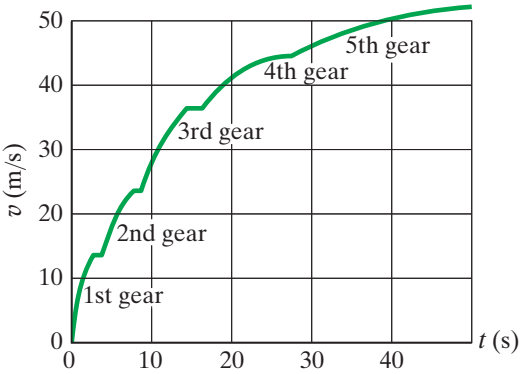


FIGURE 2–43 Problem 31. The velocity of a car as a function of time, starting from a dead stop. The flat spots in the curve represent gear shifts.

32. (III) A fugitive tries to hop on a freight train traveling at a constant speed of 5.0 m/s. Just as an empty box car passes him, the fugitive starts from rest and accelerates at $a = 1.4\ \text{m/s}^2$ to his maximum speed of 6.0 m/s, which he then maintains. (*a*) How long does it take him to catch up to the empty box car? (*b*) What is the distance traveled to reach the box car?

2–5 to 2–6 Motion at Constant Acceleration

33. (I) A car slows down from 26 m/s to rest in a distance of 88 m. What was its acceleration, assumed constant?

34. (I) A car slows down from 28 m/s to rest in 6.3 s. What was its (constant) acceleration?

35. (I) A car accelerates from 13 m/s to 22 m/s in 6.5 s. What was its acceleration? How far did it travel in this time? Assume constant acceleration.

36. (II) A world-class sprinter can reach a top speed (of about 11.5 m/s) in the first 18.0 m of a race. What is the average acceleration of this sprinter and how long does it take her to reach that speed?

37. (II) A car slows down uniformly from a speed of 28.0 m/s to rest in 8.60 s. How far did it travel in that time?

38. (II) In coming to a stop, an old truck leaves skid marks 45 m long on the highway. Assuming a deceleration of $6.00\ \text{m/s}^2$, estimate the speed of the truck just before braking.

39. (II) A baseball pitcher throws a baseball with a speed of 43 m/s. Estimate the average acceleration of the ball during the throwing motion. In throwing the baseball, the pitcher accelerates it through a displacement of about 3.5 m, from behind the body to the point where it is released (Fig. 2–44).

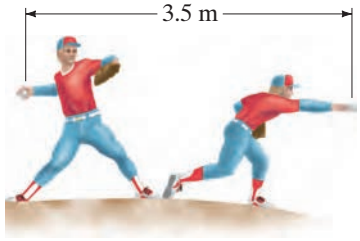


FIGURE 2–44 Problem 39.

40. (II) A car traveling at 95 km/h strikes a tree. The front end of the car compresses and the driver comes to rest after traveling 0.80 m. What was the magnitude of the average acceleration of the driver during the collision? Express the answer in terms of "*g*'s," where $1.00\,g = 9.80\ \text{m/s}^2$.

41. (II) A car traveling 85 km/h slows down at a constant $0.50\ \text{m/s}^2$ just by "letting up on the gas." Calculate (*a*) the distance the car coasts before it stops, (*b*) the time it takes to stop, and (*c*) the distance it travels during the first and fifth seconds.

42. (II) Determine the stopping distances for an automobile going a constant initial speed of 95 km/h in the $+x$ direction, and human reaction time of 0.40 s: (*a*) for an acceleration $a = -2.5\ \text{m/s}^2$; (*b*) for $a = -5.5\ \text{m/s}^2$.

43. (II) A driver is traveling 18.0 m/s when she sees a red light ahead. Her car is capable of decelerating at a rate of $3.65\ \text{m/s}^2$. If it takes her 0.380 s to get the brakes on and she is 24.0 m from the intersection when she sees the light, will she be able to stop in time? How far from the beginning of the intersection will she be, and in which direction?

44. (II) Show that $\bar{v} = (v + v_0)/2$ (see Eq. 2–12d) is not valid when the acceleration $a = A + Bt$, where A and B are non-zero constants.

45. (II) An 85-m-long train begins uniform acceleration from rest. The front of the train has a speed of 18 m/s when it passes a railway worker who is standing 180 m from where the front of the train started. What will be the speed of the last car as it passes the worker? (See Fig. 2–45.)

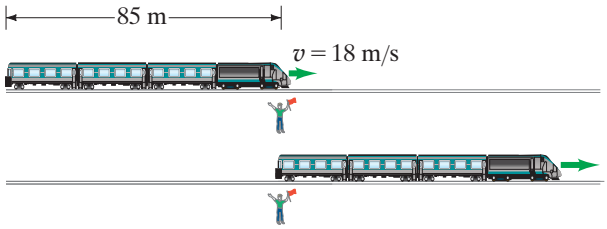


FIGURE 2–45 Problem 45.

46. (II) A space vehicle accelerates uniformly from 85 m/s at $t = 0$ to 162 m/s at $t = 10.0$ s. How far did it move between $t = 2.0$ s and $t = 7.0$ s?

47. (II) A runner hopes to complete the 10,000-m run in less than 30.0 min. After running at constant speed for exactly 27.0 min, there are still 1200 m to go. The runner must then accelerate at $0.20\ \text{m/s}^2$ for how many seconds in order to achieve the desired time?

48. (III) Mary and Sally are in a foot race (Fig. 2–46). When Mary is 22 m from the finish line, she has a speed of 4.0 m/s and is 5.0 m behind Sally, who has a speed of 5.0 m/s. Sally thinks she has an easy win and so, during the remaining portion of the race, slows down at a constant rate of 0.40 m/s^2 to the finish line. What constant acceleration does Mary now need during the remaining portion of the race, if she wishes to cross the finish line side-by-side with Sally?

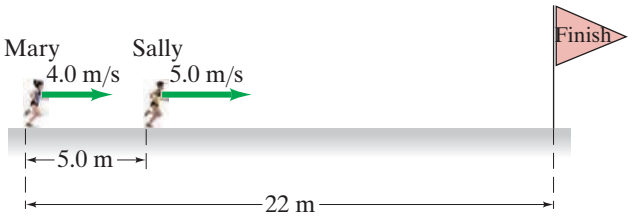


FIGURE 2–46 Problem 48.

49. (III) An unmarked police car traveling a constant 95 km/h is passed by a speeder traveling 135 km/h. Precisely 1.00 s after the speeder passes, the police officer steps on the accelerator; if the police car's acceleration is 2.60 m/s^2, how much time passes before the police car overtakes the speeder (assumed moving at constant speed)?

50. (III) Assume in Problem 49 that the speeder's speed is not known. If the police car accelerates uniformly at 2.60 m/s^2 and overtakes the speeder after accelerating for 7.00 s, what was the speeder's speed?

51. (III) A runner completes a 400-meter race in 55.0 s. The 55.0 seconds is made up of a 0.15 s reaction time from the starting sound until the runner starts moving followed by 30.0 m of constant acceleration and then 370 m at constant speed. What are the values of the acceleration and the constant speed?

2–7 Freely Falling Objects (neglect air resistance)

52. (I) A stone is dropped from the top of a cliff. It is seen to hit the ground below after 3.25 s. How high is the cliff?

53. (I) Estimate (*a*) how long it took King Kong to fall straight down from the top of the Empire State Building (380 m high), and (*b*) his velocity just before "landing."

54. (I) If a car rolls gently $(v_0 = 0)$ off a vertical cliff, how long does it take it to reach 55 km/h?

55. (II) A ball player catches a ball 2.6 s after throwing it vertically upward. With what speed did he throw it, and what height did it reach?

56. (II) A baseball is hit almost straight up into the air with a speed of 22 m/s. Estimate (*a*) how high it goes, and (*b*) how long it is in the air. (*c*) What factors make this an estimate?

57. (II) A kangaroo jumps straight up to a vertical height of 1.45 m. How long was it in the air before returning to Earth?

58. (II) The best *rebounders* in *basketball* have a vertical leap (that is, the vertical movement of a fixed point on their body) of about 120 cm. (*a*) What is their initial "launch" speed off the ground? (*b*) How long are they in the air?

59. (II) A stone is thrown vertically upward with a speed of 18.0 m/s. (*a*) How fast is it moving when it is at a height of 11.0 m? (*b*) How much time is required to reach this height? (*c*) Why are there two answers to (*b*)?

60. (II) For an object falling freely from rest, show that the distance traveled *during* each successive second increases in the ratio of successive odd integers (1, 3, 5, etc.). (This was first shown by Galileo.) See Figs. 2–27 and 2–30.

61. (II) If there were no air resistance, how long would it take a free-falling skydiver to fall from a plane at 3800 m to an altitude of 450 m, where she will open her parachute? What would her speed be at 450 m? (In reality, air resistance will restrict her speed to perhaps 150 km/h.)

62. (II) Pelicans tuck their wings and free-fall straight down when diving for fish. Suppose a pelican starts its dive from a height of 16.0 m and cannot change its path once committed. If it takes a fish 0.20 s to perform evasive action, at what minimum height must it spot the pelican to escape? Assume the fish is at the surface of the water.

63. (II) A stone is thrown vertically upward with a speed of 15.5 m/s from the edge of a cliff 75.0 m high (Fig. 2–47). (*a*) How much later does it reach the bottom of the cliff? (*b*) What is its speed just before hitting? (*c*) What total distance did it travel?

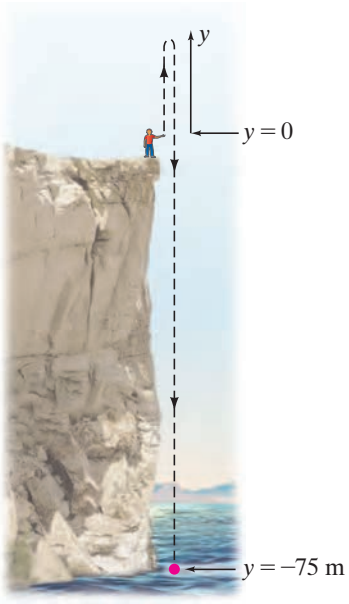


FIGURE 2–47 Problem 63.

64. (II) A rocket rises vertically, from rest, with an acceleration of 3.2 m/s^2 until it runs out of fuel at an altitude of 725 m. After this point, its acceleration is that of gravity, downward. (*a*) What is the velocity of the rocket when it runs out of fuel? (*b*) How long does it take to reach this point? (*c*) What maximum altitude does the rocket reach? (*d*) How much time (total) does it take to reach maximum altitude? (*e*) With what velocity does it strike the Earth? (*f*) How long (total) is it in the air?

65. (II) Suppose you adjust your garden hose nozzle for a fast stream of water. You point the nozzle vertically upward at a height of 1.8 m above the ground (Fig. 2–48). When you quickly turn off the nozzle, you hear the water striking the ground next to you for another 2.5 s. What is the water speed as it leaves the nozzle?



FIGURE 2–48 Problem 65.

66. (II) A helicopter is ascending vertically with a constant speed of 6.40 m/s. At a height of 105 m above the Earth, a package is dropped from the helicopter. How much time does it take for the package to reach the ground? [*Hint*: What is v_0 for the package?]

67. (II) Roger sees water balloons fall past his window. He notices that each balloon strikes the sidewalk 0.83 s after passing his window, 15 m above the sidewalk. (*a*) How fast are the balloons traveling when they pass Roger's window? (*b*) Assuming the balloons are being released from rest, from what height are they being released?

68. (II) A baseball is seen to pass upward by a window with a vertical speed of 13 m/s. If the ball was thrown by a person 18 m below on the street, (*a*) what was its initial speed, (*b*) what altitude does it reach, (*c*) when was it thrown, and (*d*) when does it reach the street again?

69. (III) A falling stone takes 0.28 s to travel past a window 2.2 m tall (Fig. 2–49). From what height above the top of the window did the stone fall?

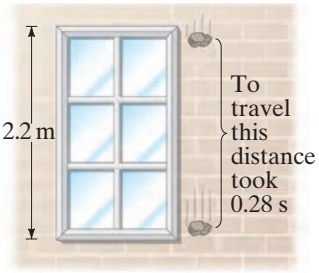


FIGURE 2–49
Problem 69.

70. (III) A toy rocket moving vertically upward passes by a 2.0-m-high window whose base is 8.0 m above the ground. The rocket takes 0.15 s to travel the 2.0 m height of the window. What was the launch speed of the rocket, and how high will it go? Assume the propellant is burned very quickly at blastoff.

71. (III) A ball is dropped from the top of a 55.0-m-high cliff. At the same time, a carefully aimed stone is thrown straight up from the bottom of the cliff with a speed of 24.0 m/s. The stone and ball collide part way up. How far above the base of the cliff does this happen?

*2–8 Variable Acceleration; Calculus

*72. (II) Given $v(t) = 25 + 18t$, where v is in m/s and t is in s, use calculus to determine the total displacement from $t_1 = 1.3$ s to $t_2 = 3.6$ s.

*73. (III) The acceleration of a particle is given by $a = A\sqrt{t}$ where $A = 3.0\,\text{m/s}^{5/2}$. At $t = 0$, $v = 7.5$ m/s and $x = 0$. (*a*) What is the velocity as a function of time? (*b*) What is the displacement as a function of time? (*c*) What are the acceleration, velocity, and displacement at $t = 5.0$ s?

*74. (III) Air resistance acting on a falling body can be taken into account by the approximate relation for the acceleration:

$$a = \frac{dv}{dt} = g - kv,$$

where k is a constant. (*a*) Derive a formula for the velocity of the body as a function of time assuming it starts from rest ($v = 0$ at $t = 0$). [*Hint*: Change variables by setting $u = g - kv$.] (*b*) Determine an expression for the terminal velocity, which is the maximum value the velocity reaches.

General Problems

75. The acceleration due to gravity on the Moon is about one-sixth what it is on Earth. If an object is thrown vertically upward on the Moon, how many times higher will it go than it would on Earth, assuming the same initial velocity?

76. A person jumps out a fourth-story window 18.0 m above a firefighter's safety net. The survivor stretches the net 1.0 m before coming to rest, Fig. 2–50. (*a*) What was the average deceleration experienced by the survivor when she was slowed to rest by the net? (*b*) What would you do to make it "safer" (that is, to generate a smaller deceleration): would you stiffen or loosen the net? Explain.

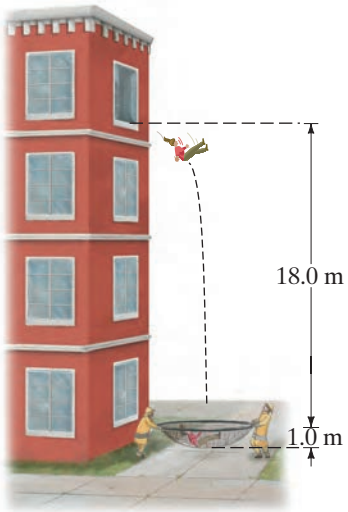


FIGURE 2–50
Problem 76.

77. A person who is properly restrained by an over-the-shoulder seat belt has a good chance of surviving a car collision if the deceleration does not exceed 30 "*g*'s" $(1.00\,g = 9.80\,\text{m/s}^2)$. Assuming uniform deceleration at 30 *g*'s, calculate the distance over which the front end of the car must be designed to collapse if a crash brings the car to rest from 95 km/h.

78. The position of a ball rolling in a straight line is given by $x = 2.0 - 3.6t + 1.7t^2$, where x is in meters and t in seconds. (*a*) What do the numbers 2.0, 3.6, and 1.7 refer to? (*b*) What are the units of each of these numbers? (*c*) Determine the position of the ball at $t = 1.0$ s, 2.0 s, and 3.0 s. (*d*) What is the average velocity over the interval $t = 1.0$ s to $t = 3.0$ s?

79. In a lecture demonstration, a 3.0-m-long vertical string with ten bolts tied to it at equal intervals is dropped from the ceiling of the lecture hall. The string falls on a tin plate, and the class hears the clink of each bolt as it hits the plate. (*a*) The sounds will not occur at equal time intervals. Why? (*b*) Will the time between clinks increase or decrease as the string falls? (*c*) How could the bolts be tied so that the clinks occur at equal intervals? (Assume the string is vertical with the bottom bolt touching the tin plate when the string is released.)

80. Two students are asked to find the height of a particular building using a barometer. Instead of using the barometer as an altitude-measuring device, they take it to the roof of the building and drop it off, timing its fall. One student reports a fall time of 2.0 s, and the other, 2.3 s. What % difference does the 0.3 s make for the estimates of the building's height?

81. Consider the street pattern shown in Fig. 2–51. Each intersection has a traffic signal, and the speed limit is 40 km/h. Suppose you are driving from the west at the speed limit. When you are 10.0 m from the first intersection, all the lights turn green. The lights are green for 13.0 s each. (*a*) Calculate the time needed to reach the third stoplight. Can you make it through all three lights without stopping? (*b*) Another car was stopped at the first light when all the lights turned green. It can accelerate at the rate of 2.00 m/s^2 to the speed limit. Can the second car make it through all three lights without stopping? By how many seconds would it make it, or not make it?

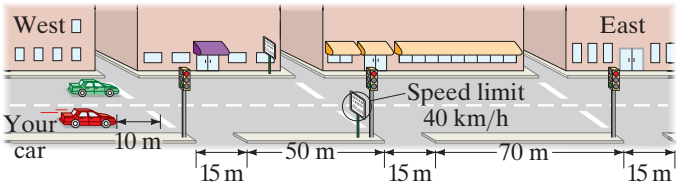


FIGURE 2–51 Problem 81.

82. Suppose a car manufacturer tested its cars for front-end collisions by hauling them up on a crane and dropping them from a certain height. (*a*) Show that the speed just before a car hits the ground, after falling from rest a vertical distance H, is given by $\sqrt{2gH}$. What height corresponds to a collision at (*b*) 35 km/h? (*c*) 95 km/h?

83. A stone is dropped from the roof of a high building. A second stone is dropped 1.50 s later. How far apart are the stones when the second one has reached a speed of 12.0 m/s?

84. A person jumps off a diving board 4.0 m above the water's surface into a deep pool. The person's downward motion stops 1.8 m below the surface of the water. Estimate the average deceleration of the person while under the water.

85. A police car at rest is passed by a speeder traveling at a constant 140 km/h. The police officer takes off in hot pursuit and catches up to the speeder in 850 m, maintaining a constant acceleration. (*a*) Qualitatively plot the position vs. time graph for both cars from the police car's start to the catch-up point. Calculate (*b*) how long it took the police officer to overtake the speeder, (*c*) the required police car acceleration, and (*d*) the speed of the police car at the overtaking point. (*e*) This last result is unrealistic—so which assumptions do we have to reconsider?

86. Agent Bond is standing on a bridge, 15 m above the road below, and his pursuers are getting too close for comfort. He spots a flatbed truck approaching at 25 m/s, which he measures by knowing that the telephone poles the truck is passing are 25 m apart in this region. The roof of the truck is 3.5 m above the road, and Bond quickly calculates how many poles away the truck should be when he drops down from the bridge onto the truck, making his getaway. How many poles is it?

87. Two children are playing on two trampolines. The first child bounces up one-and-a-half times higher than the second child. The initial speed upwards of the second child is 4.0 m/s. (*a*) Find the maximum height the second child reaches. (*b*) What is the initial speed of the first child? (*c*) How long was the first child in the air?

88. In putting, the force with which a golfer strikes a ball is planned so that the ball will stop within some small distance of the cup, say 1.0 m long or short, in case the putt is missed. Accomplishing this from an uphill lie (that is, putting the ball downhill, see Fig. 2–52) is more difficult than from a downhill lie. To see why, assume that on a particular "green" the ball decelerates constantly at 1.8 m/s^2 going downhill, and constantly at 2.6 m/s^2 going uphill. Suppose we have an uphill lie 7.0 m from the cup. Calculate the allowable range of initial velocities we may impart to the ball so that it stops in the range 1.0 m short to 1.0 m long of the cup. Do the same for a downhill lie 7.0 m from the cup. What in your results suggests that the downhill putt is more difficult?

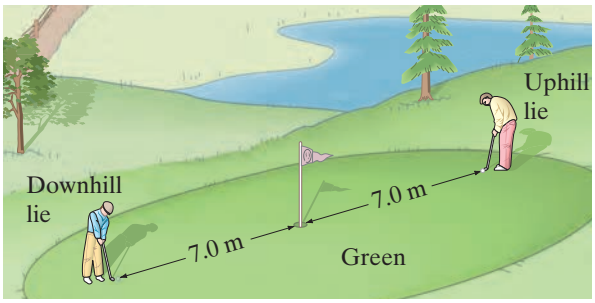


FIGURE 2–52 Problem 88.

89. A person driving her car at 35 km/h approaches an intersection just as the traffic light turns yellow. She knows that the yellow light lasts only 2.0 s before turning to red, and she is 28 m away from the near side of the intersection (Fig. 2–53). Should she try to stop, or should she speed up to cross the intersection before the light turns red? The intersection is 15 m wide. Her car's maximum deceleration is −5.8 m/s^2, whereas it can accelerate from 45 km/h to 65 km/h in 6.0 s. Ignore the length of her car and her reaction time.

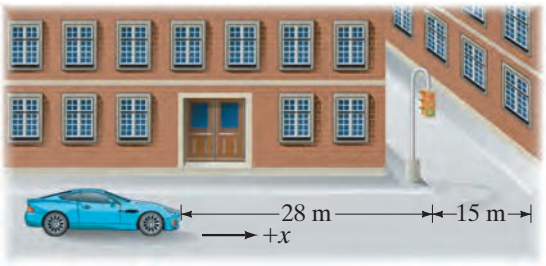


FIGURE 2–53 Problem 89.

90. A car is behind a truck going 18 m/s on the highway. The car's driver looks for an opportunity to pass, guessing that his car can accelerate at 0.60 m/s^2 and that he has to cover the 20-m length of the truck, plus 10 m of extra space at the rear of the truck and 10 m more at the front of it. In the oncoming lane, he sees a car approaching, probably at the speed limit, 25 m/s (55 mph). He estimates that the car is about 500 m away. Should he attempt the pass? Give details.

91. A rock is dropped from a sea cliff and the sound of it striking the ocean is heard 4.1 s later. If the speed of sound is 340 m/s, how high is the cliff?

92. A conveyor belt is used to send burgers through a grilling machine. If the grilling machine is 1.2 m long and the burgers require 2.8 min to cook, how fast must the conveyor belt travel? If the burgers are spaced 25 cm apart, what is the rate of burger production (in burgers/min)?

93. A rock is thrown vertically upward with a speed of 15.0 m/s. Exactly 1.00 s later, a ball is thrown up vertically along the same path with a speed of 22.0 m/s. (*a*) At what time will they strike each other? (*b*) At what height will the collision occur? (*c*) Answer (*a*) and (*b*) assuming that the order is reversed: the ball is thrown 1.00 s before the rock.

94. Figure 2–54 is a position versus time graph for the motion of an object along the x axis. Consider the time interval from A to B. (*a*) Is the object moving in the positive or negative x direction? (*b*) Is the object speeding up or slowing down? (*c*) Is the acceleration of the object positive or negative? Now consider the time interval from D to E. (*d*) Is the object moving in the positive or negative x direction? (*e*) Is the object speeding up or slowing down? (*f*) Is the acceleration of the object positive or negative? (*g*) Finally, answer these same three questions for the time interval from C to D.

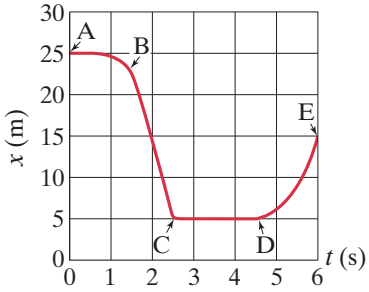


FIGURE 2–54
Problem 94.

95. In the design of a **rapid transit system**, it is necessary to balance the average speed of a train against the distance between station stops. The more stops there are, the slower the train's average speed. To get an idea of this problem, calculate the time it takes a train to make a 15.0-km trip in two situations: (*a*) the stations at which the trains must stop are 3.0 km apart (a total of 6 stations, including those at the ends); and (*b*) the stations are 5.0 km apart (4 stations total). Assume that at each station the train accelerates at a rate of 1.1 m/s^2 until it reaches 95 km/h, then stays at this speed until its brakes are applied for arrival at the next station, at which time it decelerates at -2.0 m/s^2. Assume it stops at each intermediate station for 22 s.

96. A race car driver must average 200.0 km/h over the course of a time trial lasting ten laps. If the first nine laps were done at an average speed of 196.0 km/h, what average speed must be maintained for the last lap?

97. A parachutist bails out of an airplane, and freely falls 75 m (ignore air friction). Then the parachute opens, and her acceleration is -1.5 m/s^2 (up). The parachutist reaches the ground with a speed of 1.5 m/s. (*a*) From how high did she bail out of the plane? (*b*) How much time did her fall take?

98. You stand at the top of a cliff while your friend stands on a beach below you. You drop a ball from rest and see that she catches it 1.4 s later. Your friend then throws the ball up to you, and it comes to rest just as it reaches your hand. What is the speed with which your friend threw the ball?

99. A robot used in a pharmacy picks up a medicine bottle at $t = 0$. It accelerates at 0.20 m/s^2 for 4.5 s, then travels without acceleration for 68 s and finally decelerates at -0.40 m/s^2 for 2.5 s to reach the counter where the pharmacist will take the medicine from the robot. From how far away did the robot fetch the medicine?

100. Bill can throw a ball vertically at a speed 1.5 times faster than Joe can. How many times higher will Bill's ball go than Joe's?

101. On an audio compact disc (CD), digital bits of information are encoded sequentially along a spiral path. Each bit occupies about 0.28 μm. A CD player's readout laser scans along the spiral's sequence of bits at a constant rate of about 1.2 m/s as the CD spins. (*a*) Determine the number N of digital bits that a CD player reads every second. (*b*) The audio information is sent to each of the two loudspeakers 44,100 times per second. Each of these samplings requires 16 bits, and so you might expect the required bit rate for a CD player to be

$$N_0 = 2\left(44{,}100\,\frac{\text{samplings}}{\text{s}}\right)\left(16\,\frac{\text{bits}}{\text{sampling}}\right) = 1.4\times10^6\,\frac{\text{bits}}{\text{s}},$$

where the 2 is for the 2 loudspeakers (the 2 stereo channels). Note that N_0 is less than the number N of bits actually read per second by a CD player. The excess number of bits $(= N - N_0)$ is needed for encoding and error correction. What percentage of the bits on a CD are dedicated to encoding and error correction?

102. Figure 2–55 shows the position vs. time graph for two bicycles, A and B. (*a*) Identify any instant at which the two bicycles have the same velocity. (*b*) Which bicycle has the larger acceleration? (*c*) At which instant(s) are the bicycles passing each other? Which bicycle is passing the other? (*d*) Which bicycle has the larger instantaneous velocity? (*e*) Which bicycle has the larger average velocity?

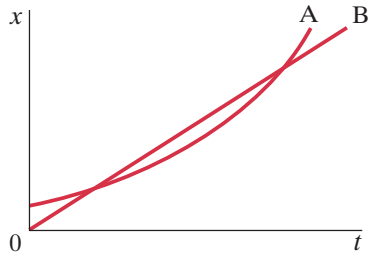


FIGURE 2–55
Problem 102.

103. You are traveling at a constant speed v_M, and there is a car in front of you traveling with a speed v_A. You realize that $v_M > v_A$, so you start slowing down with a constant acceleration a when the distance between you and the other car is x. What relationship between a and x determines whether or not you run into the car in front of you?

ANSWERS TO EXERCISES

A: (*a*) displacement $= -30$ cm; (*b*) total distance $= 50$ cm.
B: (*b*).
C: (*b*).
D: (*a*) +; (*b*) −; (*c*) −; (*d*) +.
E: (*c*).
F: (*b*).
G: (*e*).
H: (*c*).

This snowboarder flying through the air is an example of motion in two dimensions. In the absence of air resistance, the path would be a perfect parabola. The gold arrow represents the downward acceleration of gravity, $\vec{\mathbf{g}}$.

Galileo analyzed the motion of objects in two dimensions under the action of gravity near the Earth's surface (now called "projectile motion") by separating its horizontal and vertical components.

We will discuss how to manipulate vectors and how to add them. Besides analyzing projectile motion, we will also discuss unit vectors and vector kinematics, plus see how to work with relative velocity.



CHAPTER 3

Kinematics in Two or Three Dimensions; Vectors

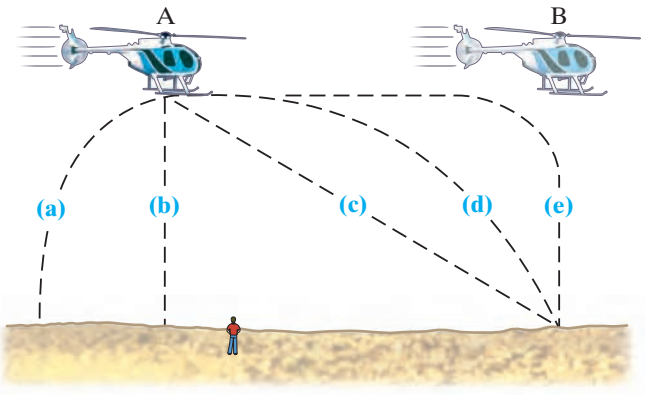
CONTENTS

3–1 Vectors and Scalars
3–2 Addition of Vectors—Graphical Methods
3–3 Subtraction of Vectors, and Multiplication of a Vector by a Scalar
3–4 Adding Vectors by Components
3–5 Unit Vectors
3–6 Vector Kinematics
3–7 Projectile Motion
3–8 Solving Problems Involving Projectile Motion
3–9 Relative Velocity

CHAPTER-OPENING QUESTION—Guess now!

[*Don't worry about getting the right answer now—you will get another chance later in the Chapter. See also page 1 of Chapter 1 for more explanation.*]

A small heavy box of emergency supplies is dropped from a moving helicopter at point A as it flies at constant speed in a horizontal direction. Which path in the drawing below best describes the path of the box (neglecting wind and air resistance) as seen by a person standing on the ground?



In Chapter 2 we dealt with motion along a straight line. We now consider the motion of objects that move in paths in two (or three) dimensions. In particular, we discuss an important type of motion known as *projectile motion*: objects projected outward near the Earth's surface, such as struck baseballs and golf balls, kicked footballs, and other projectiles. Before beginning our discussion of motion in two dimensions, we will need to discuss vectors, and how to add them.

3–1 Vectors and Scalars

We mentioned in Chapter 2 that the term *velocity* refers not only to how fast an object is moving but also to its direction. A quantity such as velocity, which has *direction* as well as *magnitude*, is a **vector** quantity. Other quantities that are also vectors are displacement, force, and momentum. However, many quantities have no direction associated with them, such as mass, time, and temperature. They are specified completely by a number and units. Such quantities are called **scalar** quantities.

Drawing a diagram of a particular physical situation is always helpful in physics, and this is especially true when dealing with vectors. On a diagram, each vector is represented by an arrow. The arrow is always drawn so that it points in the direction of the vector quantity it represents. The *length* of the arrow is drawn proportional to the magnitude of the vector quantity. For example, in Fig. 3–1, green arrows have been drawn representing the velocity of a car at various places as it rounds a curve. The magnitude of the velocity at each point can be read off Fig. 3–1 by measuring the length of the corresponding arrow and using the scale shown (1 cm = 90 km/h).

When we write the symbol for a vector, we will always use boldface type, with a tiny arrow over the symbol. Thus for velocity we write $\vec{\mathbf{v}}$. If we are concerned only with the magnitude of the vector, we will write simply v, in italics, as we do for other symbols.

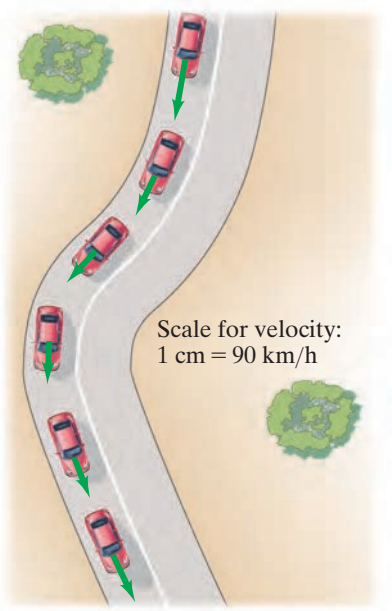


FIGURE 3–1 Car traveling on a road, slowing down to round the curve. The green arrows represent the velocity vector at each position.

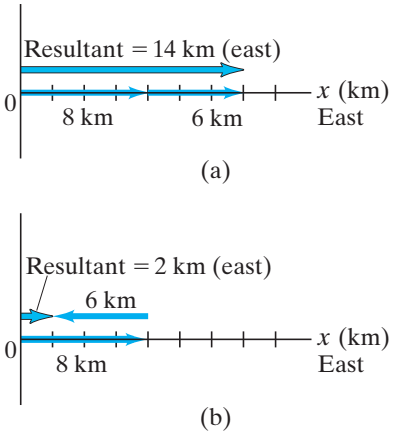
3–2 Addition of Vectors—Graphical Methods

Because vectors are quantities that have direction as well as magnitude, they must be added in a special way. In this Chapter, we will deal mainly with displacement vectors, for which we now use the symbol $\vec{\mathbf{D}}$, and velocity vectors, $\vec{\mathbf{v}}$. But the results apply for acceleration and other vectors we will encounter later.

We use simple arithmetic for adding scalars. Simple arithmetic can also be used for adding vectors if they are in the same direction. For example, if a person walks 8 km east one day, and 6 km east the next day, the person will be 8 km + 6 km = 14 km east of the point of origin. We say that the *net* or *resultant* displacement is 14 km to the east (Fig. 3–2a). If, on the other hand, the person walks 8 km east on the first day, and 6 km west (in the reverse direction) on the second day, then the person will end up 2 km from the origin (Fig. 3–2b), so the resultant displacement is 2 km to the east. In this case, the resultant displacement is obtained by subtraction: 8 km − 6 km = 2 km.

But simple arithmetic cannot be used if the two vectors are not along the same line. For example, suppose a person walks 10.0 km east and then walks 5.0 km north. These displacements can be represented on a graph in which the positive y axis points north and the positive x axis points east, Fig. 3–3. On this graph, we draw an arrow, labeled $\vec{\mathbf{D}}_1$, to represent the 10.0-km displacement to the east. Then we draw a second arrow, $\vec{\mathbf{D}}_2$, to represent the 5.0-km displacement to the north. Both vectors are drawn to scale, as shown in Fig. 3–3.

FIGURE 3–2 Combining vectors in one dimension.



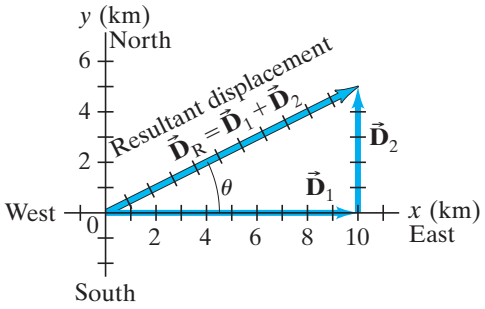


FIGURE 3–3 A person walks 10.0 km east and then 5.0 km north. These two displacements are represented by the vectors $\vec{\mathbf{D}}_1$ and $\vec{\mathbf{D}}_2$, which are shown as arrows. Also shown is the resultant displacement vector, $\vec{\mathbf{D}}_R$, which is the vector sum of $\vec{\mathbf{D}}_1$ and $\vec{\mathbf{D}}_2$. Measurement on the graph with ruler and protractor shows that $\vec{\mathbf{D}}_R$ has a magnitude of 11.2 km and points at an angle $\theta = 27°$ north of east.

After taking this walk, the person is now 10.0 km east and 5.0 km north of the point of origin. The **resultant displacement** is represented by the arrow labeled $\vec{\mathbf{D}}_R$ in Fig. 3–3. (The subscript R stands for resultant.) Using a ruler and a protractor, you can measure on this diagram (Fig. 3–3 on previous page) that the person is 11.2 km from the origin at an angle $\theta = 27°$ north of east. In other words, the resultant displacement vector has a magnitude of 11.2 km and makes an angle $\theta = 27°$ with the positive x axis. The magnitude (length) of $\vec{\mathbf{D}}_R$ can also be obtained using the theorem of Pythagoras in this case, because D_1, D_2, and D_R form a right triangle with D_R as the hypotenuse. Thus

$$D_R = \sqrt{D_1^2 + D_2^2} = \sqrt{(10.0\,\text{km})^2 + (5.0\,\text{km})^2}$$
$$= \sqrt{125\,\text{km}^2} = 11.2\,\text{km}.$$

You can use the Pythagorean theorem only when the vectors are *perpendicular* to each other.

The resultant displacement vector, $\vec{\mathbf{D}}_R$, is the sum of the vectors $\vec{\mathbf{D}}_1$ and $\vec{\mathbf{D}}_2$. That is,

$$\vec{\mathbf{D}}_R = \vec{\mathbf{D}}_1 + \vec{\mathbf{D}}_2.$$

This is a *vector* equation. An important feature of adding two vectors that are not along the same line is that the magnitude of the resultant vector is not equal to the sum of the magnitudes of the two separate vectors, but is smaller than their sum. That is,

$$D_R \leq (D_1 + D_2),$$

where the equals sign applies only if the two vectors point in the same direction. In our example (Fig. 3–3), $D_R = 11.2$ km, whereas $D_1 + D_2$ equals 15 km, which is the total distance traveled. Note also that we cannot set $\vec{\mathbf{D}}_R$ equal to 11.2 km, because we have a vector equation and 11.2 km is only a part of the resultant vector, its magnitude. We could write something like this, though: $\vec{\mathbf{D}}_R = \vec{\mathbf{D}}_1 + \vec{\mathbf{D}}_2 = (11.2\,\text{km}, 27°\ \text{N of E})$.

Figure 3–3 illustrates the general rules for graphically adding two vectors together, no matter what angles they make, to get their sum. The rules are as follows:

1. On a diagram, draw one of the vectors—call it $\vec{\mathbf{D}}_1$—to scale.
2. Next draw the second vector, $\vec{\mathbf{D}}_2$, to scale, placing its tail at the tip of the first vector and being sure its direction is correct.
3. The arrow drawn from the tail of the first vector to the tip of the second vector represents the *sum*, or **resultant**, of the two vectors.

The length of the resultant vector represents its magnitude. Note that vectors can be moved parallel to themselves on paper (maintaining the same length and angle) to accomplish these manipulations. The length of the resultant can be measured with a ruler and compared to the scale. Angles can be measured with a protractor. This method is known as the **tail-to-tip method of adding vectors**.

The resultant is not affected by the order in which the vectors are added. For example, a displacement of 5.0 km north, to which is added a displacement of 10.0 km east, yields a resultant of 11.2 km and angle $\theta = 27°$ (see Fig. 3–4), the same as when they were added in reverse order (Fig. 3–3). Thus we can write, using $\vec{\mathbf{V}}$ to represent any type of vector,

$$\vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2 = \vec{\mathbf{V}}_2 + \vec{\mathbf{V}}_1, \qquad \text{[commutative property]} \quad \textbf{(3–1a)}$$

which is known as the *commutative* property of vector addition.

The tail-to-tip method of adding vectors can be extended to three or more vectors. The resultant is drawn from the tail of the first vector to the tip of the last

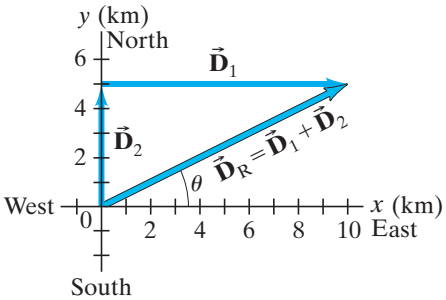


FIGURE 3–4 If the vectors are added in reverse order, the resultant is the same. (Compare to Fig. 3–3.)

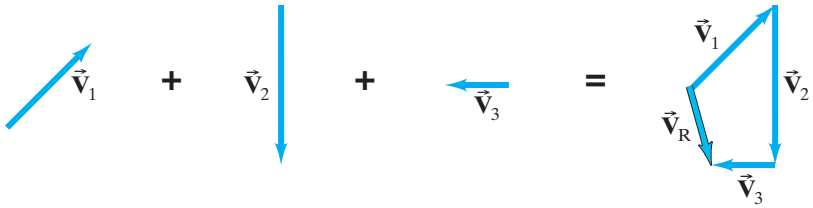


FIGURE 3–5 The resultant of three vectors: $\vec{V}_R = \vec{V}_1 + \vec{V}_2 + \vec{V}_3$.

one added. An example is shown in Fig. 3–5; the three vectors could represent displacements (northeast, south, west) or perhaps three forces. Check for yourself that you get the same resultant no matter in which order you add the three vectors; that is,

$$(\vec{V}_1 + \vec{V}_2) + \vec{V}_3 = \vec{V}_1 + (\vec{V}_2 + \vec{V}_3), \quad \text{[associative property]} \quad \textbf{(3–1b)}$$

which is known as the *associative* property of vector addition.

A second way to add two vectors is the **parallelogram method**. It is fully equivalent to the tail-to-tip method. In this method, the two vectors are drawn starting from a common origin, and a parallelogram is constructed using these two vectors as adjacent sides as shown in Fig. 3–6b. The resultant is the diagonal drawn from the common origin. In Fig. 3–6a, the tail-to-tip method is shown, and we can see that both methods yield the same result.

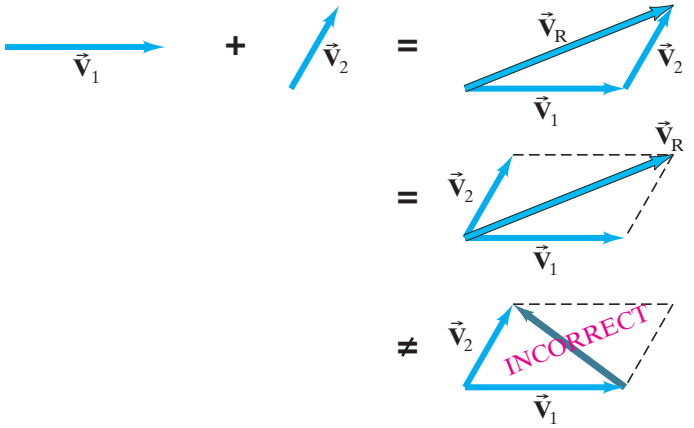


FIGURE 3–6 Vector addition by two different methods, (a) and (b). Part (c) is incorrect.

It is a common error to draw the sum vector as the diagonal running between the tips of the two vectors, as in Fig. 3–6c. *This is incorrect*: it does not represent the sum of the two vectors. (In fact, it represents their difference, $\vec{V}_2 - \vec{V}_1$, as we will see in the next Section.)

CAUTION

Be sure to use the correct diagonal on the parallelogram to get the resultant

CONCEPTUAL EXAMPLE 3–1 **Range of vector lengths.** Suppose two vectors each have length 3.0 units. What is the range of possible lengths for the vector representing the sum of the two?

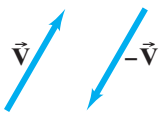
RESPONSE The sum can take on any value from 6.0 $(= 3.0 + 3.0)$ where the vectors point in the same direction, to 0 $(= 3.0 - 3.0)$ when the vectors are antiparallel. Magnitudes between 0 and 6.0 occur when the two vectors are at an angle other than 0° and 180°.

EXERCISE A If the two vectors of Example 3–1 are perpendicular to each other, what is the resultant vector length?

3–3 Subtraction of Vectors, and Multiplication of a Vector by a Scalar

Given a vector $\vec{V}$, we define the *negative* of this vector $(-\vec{V})$ to be a vector with the same magnitude as $\vec{V}$ but opposite in direction, Fig. 3–7. Note, however, that no vector is ever negative in the sense of its magnitude: the magnitude of every vector is positive. Rather, a minus sign tells us about its direction.

FIGURE 3–7 The negative of a vector is a vector having the same length but opposite direction.



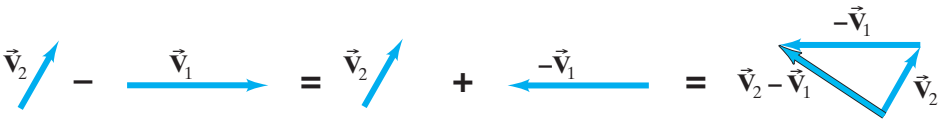


FIGURE 3–8 Subtracting two vectors: $\vec{\mathbf{V}}_2 - \vec{\mathbf{V}}_1$.

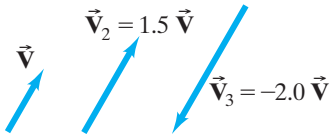
We can now define the subtraction of one vector from another: the difference between two vectors $\vec{\mathbf{V}}_2 - \vec{\mathbf{V}}_1$ is defined as

$$\vec{\mathbf{V}}_2 - \vec{\mathbf{V}}_1 = \vec{\mathbf{V}}_2 + (-\vec{\mathbf{V}}_1).$$

That is, the difference between two vectors is equal to the sum of the first plus the negative of the second. Thus our rules for addition of vectors can be applied as shown in Fig. 3–8 using the tail-to-tip method.

A vector $\vec{\mathbf{V}}$ can be multiplied by a scalar c. We define their product so that $c\vec{\mathbf{V}}$ has the same direction as $\vec{\mathbf{V}}$ and has magnitude cV. That is, multiplication of a vector by a positive scalar c changes the magnitude of the vector by a factor c but doesn't alter the direction. If c is a negative scalar, the magnitude of the product $c\vec{\mathbf{V}}$ is still $|c|V$ (where $|c|$ means the magnitude of c), but the direction is precisely opposite to that of $\vec{\mathbf{V}}$. See Fig. 3–9.

FIGURE 3–9 Multiplying a vector $\vec{\mathbf{V}}$ by a scalar c gives a vector whose magnitude is c times greater and in the same direction as $\vec{\mathbf{V}}$ (or opposite direction if c is negative).



EXERCISE B What does the "incorrect" vector in Fig. 3–6c represent? (*a*) $\vec{\mathbf{V}}_2 - \vec{\mathbf{V}}_1$, (*b*) $\vec{\mathbf{V}}_1 - \vec{\mathbf{V}}_2$, (*c*) something else (specify).

3–4 Adding Vectors by Components

Adding vectors graphically using a ruler and protractor is often not sufficiently accurate and is not useful for vectors in three dimensions. We discuss now a more powerful and precise method for adding vectors. But do not forget graphical methods—they are useful for visualizing, for checking your math, and thus for getting the correct result.

Consider first a vector $\vec{\mathbf{V}}$ that lies in a particular plane, and we have chosen an x and a y axis on this plane. Then $\vec{\mathbf{V}}$ can be expressed as the sum of two other vectors, called the **components** of the original vector, which are usually chosen to be along the x and y axes. The process of finding the components is known as **resolving the vector into its components**. An example is shown in Fig. 3–10; the vector $\vec{\mathbf{V}}$ could be a displacement vector that points at an angle $\theta = 30°$ north of east, where we have chosen the positive x axis to be to the east and the positive y axis north. This vector $\vec{\mathbf{V}}$ is resolved into its x and y components by drawing dashed lines out from the tip (A) of the vector (lines AB and AC), making them perpendicular to the x and y axes. Then the lines 0B and 0C represent the x and y components of $\vec{\mathbf{V}}$, respectively, as shown in Fig. 3–10b. These *vector components* are written $\vec{\mathbf{V}}_x$ and $\vec{\mathbf{V}}_y$. In this book we usually show vector components as arrows, like vectors, but dashed. The *scalar components*, V_x and V_y, are the magnitudes of the vector components, with units, accompanied by a positive or negative sign depending on whether they point along the positive or negative x or y axis. As can be seen in Fig. 3–10, $\vec{\mathbf{V}}_x + \vec{\mathbf{V}}_y = \vec{\mathbf{V}}$ by the parallelogram method of adding vectors.

Space is made up of three dimensions, and sometimes it is necessary to resolve a vector into components along three mutually perpendicular directions. In rectangular coordinates the components are $\vec{\mathbf{V}}_x$, $\vec{\mathbf{V}}_y$, and $\vec{\mathbf{V}}_z$.

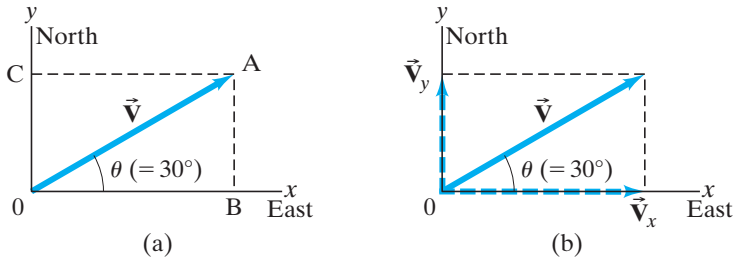


FIGURE 3–10 Resolving a vector $\vec{\mathbf{V}}$ into its components along a chosen set of x and y axes. The components, once found, themselves represent the vector. That is, the components contain as much information as the vector itself.

The use of trigonometric functions for finding the components of a vector is illustrated in Fig. 3–11, where a vector and its two components are thought of as making up a right triangle. (See also Appendix A for other details on trigonometric functions and identities.) We then see that the sine, cosine, and tangent are as given in Fig. 3–11, where θ is the angle $\vec{\mathbf{V}}$ makes with the $+x$ axis, measured positive counterclockwise. If we multiply the definition of $\sin\theta = V_y/V$ by V on both sides, we get

$$V_y = V\sin\theta. \tag{3–2a}$$

Similarly, from the definition of $\cos\theta$, we obtain

$$V_x = V\cos\theta. \tag{3–2b}$$

Keep in mind that θ is chosen (by convention) to be the angle that the vector makes with the positive x axis, measured positive counterclockwise.

The components of a given vector will be different for different choices of coordinate axes. It is therefore crucial to specify the choice of coordinate system when giving the components.

There are two ways to specify a vector in a given coordinate system:

1. We can give its components, V_x and V_y.
2. We can give its magnitude V and the angle θ it makes with the positive x axis.

We can shift from one description to the other using Eqs. 3–2, and, for the reverse, by using the theorem of Pythagoras[†] and the definition of tangent:

$$V = \sqrt{V_x^2 + V_y^2} \tag{3–3a}$$

$$\tan\theta = \frac{V_y}{V_x} \tag{3–3b}$$

as can be seen in Fig. 3–11.

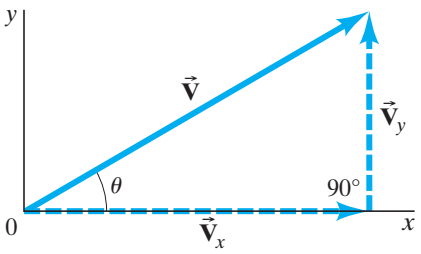
We can now discuss how to add vectors using components. The first step is to resolve each vector into its components, Eqs. 3–2. Next we can see, using Fig. 3–12, that the addition of any two vectors $\vec{\mathbf{V}}_1$ and $\vec{\mathbf{V}}_2$ to give a resultant $\vec{\mathbf{V}}_R = \vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2$, implies that

$$\begin{aligned} V_{Rx} &= V_{1x} + V_{2x} \\ V_{Ry} &= V_{1y} + V_{2y}. \end{aligned} \tag{3–4}$$

That is, the sum of the x components equals the x component of the resultant vector, and the sum of the y components equals the y component of the resultant vector, as can be verified by a careful examination of Fig. 3–12. Note that we do *not* add x components to y components.

If the magnitude and direction of the resultant vector are desired, they can be obtained using Eqs. 3–3.

[†]In three dimensions, the theorem of Pythagoras becomes $V = \sqrt{V_x^2 + V_y^2 + V_z^2}$, where V_z is the component along the third, or z, axis.



$$\sin\theta = \frac{V_y}{V}$$
$$\cos\theta = \frac{V_x}{V}$$
$$\tan\theta = \frac{V_y}{V_x}$$
$$V^2 = V_x^2 + V_y^2$$

FIGURE 3–11 Finding the components of a vector using trigonometric functions. The equations are valid only if θ is the angle $\vec{\mathbf{V}}$ makes with the positive x axis.

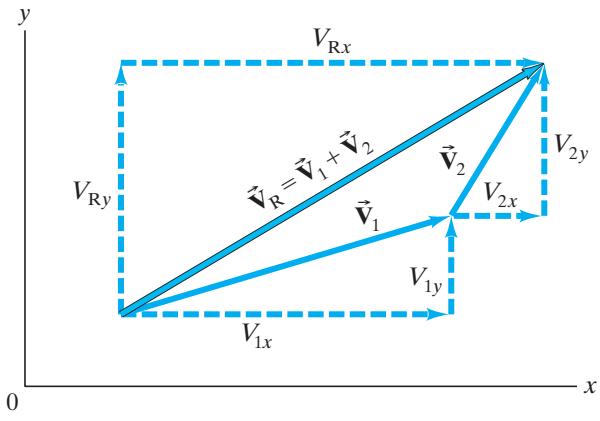


FIGURE 3–12 The components of $\vec{\mathbf{V}}_R = \vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2$ are
$V_{Rx} = V_{1x} + V_{2x}$
$V_{Ry} = V_{1y} + V_{2y}$.

The components of a given vector depend on the choice of coordinate axes. You can often reduce the work involved in adding vectors by a good choice of axes—for example, by choosing one of the axes to be in the same direction as one of the vectors. Then that vector will have only one nonzero component.

EXAMPLE 3–2 **Mail carrier's displacement.** A rural mail carrier leaves the post office and drives 22.0 km in a northerly direction. She then drives in a direction 60.0° south of east for 47.0 km (Fig. 3–13a). What is her displacement from the post office?

APPROACH We choose the positive x axis to be east and the positive y axis to be north, since those are the compass directions used on most maps. The origin of the xy coordinate system is at the post office. We resolve each vector into its x and y components. We add the x components together, and then the y components together, giving us the x and y components of the resultant.

SOLUTION Resolve each displacement vector into its components, as shown in Fig. 3–13b. Since $\vec{\mathbf{D}}_1$ has magnitude 22.0 km and points north, it has only a y component:

$$D_{1x} = 0, \qquad D_{1y} = 22.0\,\text{km}.$$

$\vec{\mathbf{D}}_2$ has both x and y components:

$$D_{2x} = +(47.0\,\text{km})(\cos 60°) = +(47.0\,\text{km})(0.500) = +23.5\,\text{km}$$
$$D_{2y} = -(47.0\,\text{km})(\sin 60°) = -(47.0\,\text{km})(0.866) = -40.7\,\text{km}.$$

Notice that D_{2y} is negative because this vector component points along the negative y axis. The resultant vector, $\vec{\mathbf{D}}_\text{R}$, has components:

$$D_{\text{R}x} = D_{1x} + D_{2x} = 0\,\text{km} + 23.5\,\text{km} = +23.5\,\text{km}$$
$$D_{\text{R}y} = D_{1y} + D_{2y} = 22.0\,\text{km} + (-40.7\,\text{km}) = -18.7\,\text{km}.$$

This specifies the resultant vector completely:

$$D_{\text{R}x} = 23.5\,\text{km}, \qquad D_{\text{R}y} = -18.7\,\text{km}.$$

We can also specify the resultant vector by giving its magnitude and angle using Eqs. 3–3:

$$D_\text{R} = \sqrt{D_{\text{R}x}^2 + D_{\text{R}y}^2} = \sqrt{(23.5\,\text{km})^2 + (-18.7\,\text{km})^2} = 30.0\,\text{km}$$

$$\tan\theta = \frac{D_{\text{R}y}}{D_{\text{R}x}} = \frac{-18.7\,\text{km}}{23.5\,\text{km}} = -0.796.$$

A calculator with a key labeled INV TAN, or ARC TAN, or TAN^{-1} gives $\theta = \tan^{-1}(-0.796) = -38.5°$. The negative sign means $\theta = 38.5°$ below the x axis, Fig. 3–13c. So, the resultant displacement is 30.0 km directed at 38.5° in a southeasterly direction.

NOTE Always be attentive about the quadrant in which the resultant vector lies. An electronic calculator does not fully give this information, but a good diagram does.

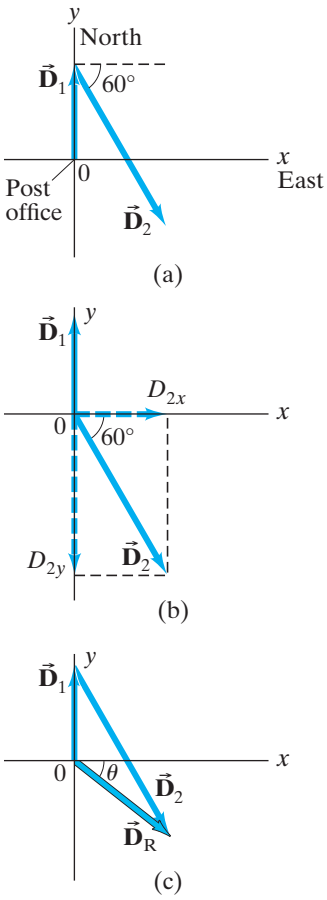


FIGURE 3–13 Example 3–2. (a) The two displacement vectors, $\vec{\mathbf{D}}_1$ and $\vec{\mathbf{D}}_2$. (b) $\vec{\mathbf{D}}_2$ is resolved into its components. (c) $\vec{\mathbf{D}}_1$ and $\vec{\mathbf{D}}_2$ are added graphically to obtain the resultant $\vec{\mathbf{D}}_\text{R}$. The component method of adding the vectors is explained in the Example.

PROBLEM SOLVING
Identify the correct quadrant by drawing a careful diagram

As we saw in Example 3–2, any component that points along the negative x or y axis gets a minus sign. The signs of trigonometric functions depend on which "quadrant" the angle falls in: for example, the tangent is positive in the first and third quadrants (from 0° to 90°, and 180° to 270°), but negative in the second and fourth quadrants; see Appendix A–9, Fig. A–6. The best way to keep track of angles, and to check any vector result, is always to draw a vector diagram, like Fig. 3–13. A vector diagram gives you something tangible to look at when analyzing a problem, and provides a check on the results.

The following Problem Solving Strategy should not be considered a prescription. Rather it is a summary of things to do to get you thinking and involved in the problem at hand.

PROBLEM SOLVING

Adding Vectors

Here is a brief summary of how to add two or more vectors using components:

1. **Draw a diagram**, adding the vectors graphically by either the parallelogram or tail-to-tip method.
2. **Choose** x and y **axes**. Choose them in a way, if possible, that will make your work easier. (For example, choose one axis along the direction of one of the vectors, which then will have only one component.)
3. **Resolve** each vector into its x and y **components**, showing each component along its appropriate (x or y) axis as a (dashed) arrow.
4. **Calculate each component** (when not given) using sines and cosines. If θ_1 is the angle that vector $\vec{\mathbf{V}}_1$ makes with the positive x axis, then:

$$V_{1x} = V_1 \cos\theta_1, \qquad V_{1y} = V_1 \sin\theta_1.$$

Pay careful attention to **signs**: any component that points along the negative x or y axis gets a minus sign.

5. **Add** the x **components** together to get the x component of the resultant. Similarly for y:

$$V_{Rx} = V_{1x} + V_{2x} + \text{any others}$$
$$V_{Ry} = V_{1y} + V_{2y} + \text{any others}.$$

This is the answer: the components of the resultant vector. Check signs to see if they fit the quadrant shown in your diagram (point 1 above).

6. If you want to know the **magnitude and direction** of the resultant vector, use Eqs. 3–3:

$$V_R = \sqrt{V_{Rx}^2 + V_{Ry}^2}, \qquad \tan\theta = \frac{V_{Ry}}{V_{Rx}}.$$

The vector diagram you already drew helps to obtain the correct position (quadrant) of the angle θ.

EXAMPLE 3–3 **Three short trips.** An airplane trip involves three legs, with two stopovers, as shown in Fig. 3–14a. The first leg is due east for 620 km; the second leg is southeast (45°) for 440 km; and the third leg is at 53° south of west, for 550 km, as shown. What is the plane's total displacement?

APPROACH We follow the steps in the Problem Solving Strategy above.

SOLUTION

1. **Draw a diagram** such as Fig. 3–14a, where $\vec{\mathbf{D}}_1$, $\vec{\mathbf{D}}_2$, and $\vec{\mathbf{D}}_3$ represent the three legs of the trip, and $\vec{\mathbf{D}}_R$ is the plane's total displacement.
2. **Choose axes**: Axes are also shown in Fig. 3–14a: x is east, y north.
3. **Resolve components**: It is imperative to draw a good diagram. The components are drawn in Fig. 3–14b. Instead of drawing all the vectors starting from a common origin, as we did in Fig. 3–13b, here we draw them "tail-to-tip" style, which is just as valid and may make it easier to see.
4. **Calculate the components**:

$$\vec{\mathbf{D}}_1: D_{1x} = +D_1\cos 0° = D_1 = 620\text{ km}$$
$$D_{1y} = +D_1\sin 0° = 0\text{ km}$$
$$\vec{\mathbf{D}}_2: D_{2x} = +D_2\cos 45° = +(440\text{ km})(0.707) = +311\text{ km}$$
$$D_{2y} = -D_2\sin 45° = -(440\text{ km})(0.707) = -311\text{ km}$$
$$\vec{\mathbf{D}}_3: D_{3x} = -D_3\cos 53° = -(550\text{ km})(0.602) = -331\text{ km}$$
$$D_{3y} = -D_3\sin 53° = -(550\text{ km})(0.799) = -439\text{ km}.$$

We have given a minus sign to each component that in Fig. 3–14b points in the $-x$ or $-y$ direction. The components are shown in the Table in the margin.

5. **Add the components**: We add the x components together, and we add the y components together to obtain the x and y components of the resultant:

$$D_{Rx} = D_{1x} + D_{2x} + D_{3x} = 620\text{ km} + 311\text{ km} - 331\text{ km} = 600\text{ km}$$
$$D_{Ry} = D_{1y} + D_{2y} + D_{3y} = 0\text{ km} - 311\text{ km} - 439\text{ km} = -750\text{ km}.$$

The x and y components of the resultant are 600 km and −750 km, and point respectively to the east and south. This is one way to give the answer.

6. **Magnitude and direction**: We can also give the answer as

$$D_R = \sqrt{D_{Rx}^2 + D_{Ry}^2} = \sqrt{(600)^2 + (-750)^2}\,\text{km} = 960\text{ km}$$
$$\tan\theta = \frac{D_{Ry}}{D_{Rx}} = \frac{-750\text{ km}}{600\text{ km}} = -1.25, \qquad \text{so } \theta = -51°.$$

Thus, the total displacement has magnitude 960 km and points 51° below the x axis (south of east), as was shown in our original sketch, Fig. 3–14a.

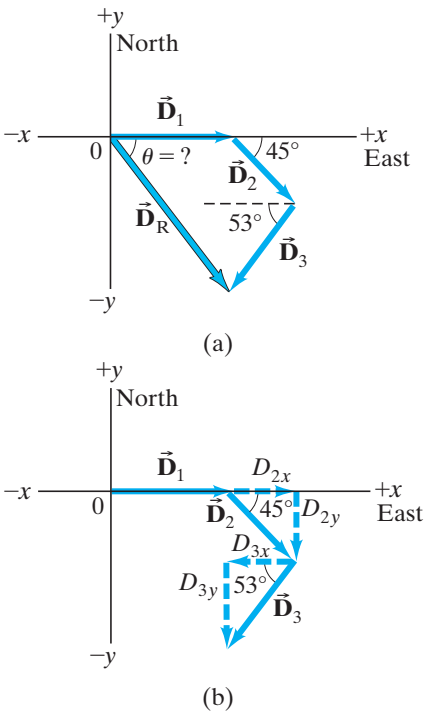


FIGURE 3–14 Example 3–3.

Vector	Components x (km)	y (km)
$\vec{\mathbf{D}}_1$	620	0
$\vec{\mathbf{D}}_2$	311	−311
$\vec{\mathbf{D}}_3$	−331	−439
$\vec{\mathbf{D}}_R$	600	−750

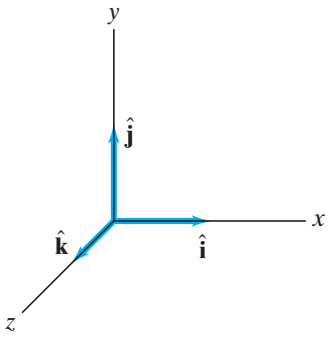


FIGURE 3–15 Unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ along the x, y, and z axes.

3–5 Unit Vectors

Vectors can be conveniently written in terms of *unit vectors*. A **unit vector** is defined to have a magnitude exactly equal to one (1). (The word "unit" comes from the Latin, unus, meaning "one".) It is useful to define unit vectors that point along coordinate axes, and in an x, y, z rectangular coordinate system these unit vectors are called $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$. They point, respectively, along the positive x, y, and z axes as shown in Fig. 3–15. Like other vectors, $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ do not have to be placed at the origin, but can be placed elsewhere as long as the direction and unit length remain unchanged. It is common to write unit vectors with a "hat": $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ (and we will do so in this book) as a reminder that each has magnitude of exactly one unit.

Because of the definition of multiplication of a vector by a scalar (Section 3–3), the components of a vector $\vec{\mathbf{V}}$ can be written $\vec{\mathbf{V}}_x = V_x\hat{\mathbf{i}}$, $\vec{\mathbf{V}}_y = V_y\hat{\mathbf{j}}$, and $\vec{\mathbf{V}}_z = V_z\hat{\mathbf{k}}$. Hence any vector $\vec{\mathbf{V}}$ can be written in terms of its components as

$$\vec{\mathbf{V}} = V_x\hat{\mathbf{i}} + V_y\hat{\mathbf{j}} + V_z\hat{\mathbf{k}}. \qquad \textbf{(3–5)}$$

Unit vectors are helpful when adding vectors analytically by components. For example, Eq. 3–4 can be seen to be true by using unit vector notation for each vector (which we write for the two-dimensional case, with the extension to three dimensions being straightforward):

$$\begin{aligned}\vec{\mathbf{V}}_R = (V_{Rx})\hat{\mathbf{i}} + (V_{Ry})\hat{\mathbf{j}} &= \vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2 \\ &= (V_{1x}\hat{\mathbf{i}} + V_{1y}\hat{\mathbf{j}}) + (V_{2x}\hat{\mathbf{i}} + V_{2y}\hat{\mathbf{j}}) \\ &= (V_{1x} + V_{2x})\hat{\mathbf{i}} + (V_{1y} + V_{2y})\hat{\mathbf{j}}.\end{aligned}$$

Comparing the first line to the third line, we get Eqs. 3–4.

EXAMPLE 3–4 **Using unit vectors.** Write the vectors of Example 3–2 in unit vector notation, and perform the addition.

APPROACH We use the components we found in Example 3–2, $D_{1x} = 0$, $D_{1y} = 22.0\,\text{km}$, and $D_{2x} = 23.5\,\text{km}$, $D_{2y} = -40.7\,\text{km}$, and we now write them in the form of Eq. 3–5.

SOLUTION We have

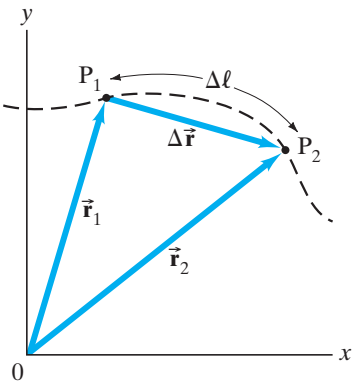
$$\begin{aligned}\vec{\mathbf{D}}_1 &= 0\hat{\mathbf{i}} + 22.0\,\text{km}\,\hat{\mathbf{j}} \\ \vec{\mathbf{D}}_2 &= 23.5\,\text{km}\,\hat{\mathbf{i}} - 40.7\,\text{km}\,\hat{\mathbf{j}}.\end{aligned}$$

Then the resultant displacement is

$$\begin{aligned}\vec{\mathbf{D}}_R = \vec{\mathbf{D}}_1 + \vec{\mathbf{D}}_2 &= (0 + 23.5)\,\text{km}\,\hat{\mathbf{i}} + (22.0 - 40.7)\,\text{km}\,\hat{\mathbf{j}} \\ &= 23.5\,\text{km}\,\hat{\mathbf{i}} - 18.7\,\text{km}\,\hat{\mathbf{j}}.\end{aligned}$$

The components of the resultant displacement, $\vec{\mathbf{D}}_R$, are $D_x = 23.5\,\text{km}$ and $D_y = -18.7\,\text{km}$. The magnitude of $\vec{\mathbf{D}}_R$ is $D_R = \sqrt{(23.5\,\text{km})^2 + (18.7\,\text{km})^2} = 30.0\,\text{km}$, just as in Example 3–2.

FIGURE 3–16 Path of a particle in the xy plane. At time t_1 the particle is at point P_1 given by the position vector $\vec{\mathbf{r}}_1$; at t_2 the particle is at point P_2 given by the position vector $\vec{\mathbf{r}}_2$. The displacement vector for the time interval $t_2 - t_1$ is $\Delta\vec{\mathbf{r}} = \vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1$. The actual distance traveled along the path between P_1 and P_2 is $\Delta\ell$.



3–6 Vector Kinematics

We can now extend our definitions of velocity and acceleration in a formal way to two- and three-dimensional motion. Suppose a particle follows a path in the xy plane as shown in Fig. 3–16. At time t_1, the particle is at point P_1, and at time t_2, it is at point P_2. The vector $\vec{\mathbf{r}}_1$ is the position vector of the particle at time t_1 (it represents the displacement of the particle from the origin of the coordinate system). And $\vec{\mathbf{r}}_2$ is the position vector at time t_2.

In one dimension, we defined displacement as the *change in position* of the particle. In the more general case of two or three dimensions, the **displacement vector** is defined as the vector representing change in position. We call it $\Delta\vec{\mathbf{r}}$,[†] where

$$\Delta\vec{\mathbf{r}} = \vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1.$$

This represents the displacement during the time interval $\Delta t = t_2 - t_1$.

[†]We used $\vec{\mathbf{D}}$ for the displacement vector earlier in the Chapter for illustrating vector addition. The new notation here, $\Delta\vec{\mathbf{r}}$, emphasizes that it is the difference between two position vectors.

In unit vector notation, we can write

$$\vec{\mathbf{r}}_1 = x_1\hat{\mathbf{i}} + y_1\hat{\mathbf{j}} + z_1\hat{\mathbf{k}}, \tag{3–6a}$$

where x_1, y_1, and z_1 are the coordinates of point P_1. Similarly,

$$\vec{\mathbf{r}}_2 = x_2\hat{\mathbf{i}} + y_2\hat{\mathbf{j}} + z_2\hat{\mathbf{k}}.$$

Hence

$$\Delta\vec{\mathbf{r}} = (x_2 - x_1)\hat{\mathbf{i}} + (y_2 - y_1)\hat{\mathbf{j}} + (z_2 - z_1)\hat{\mathbf{k}}. \tag{3–6b}$$

If the motion is along the x axis only, then $y_2 - y_1 = 0$, $z_2 - z_1 = 0$, and the magnitude of the displacement is $\Delta r = x_2 - x_1$, which is consistent with our earlier one-dimensional equation (Section 2–1). Even in one dimension, displacement is a vector, as are velocity and acceleration.

The **average velocity vector** over the time interval $\Delta t = t_2 - t_1$ is defined as

$$\text{average velocity} = \frac{\Delta\vec{\mathbf{r}}}{\Delta t}. \tag{3–7}$$

Now let us consider shorter and shorter time intervals—that is, we let Δt approach zero so that the distance between points P_2 and P_1 also approaches zero, Fig. 3–17a. We define the **instantaneous velocity vector** as the limit of the average velocity as Δt approaches zero:

$$\vec{\mathbf{v}} = \lim_{\Delta t \to 0} \frac{\Delta\vec{\mathbf{r}}}{\Delta t} = \frac{d\vec{\mathbf{r}}}{dt}. \tag{3–8}$$

The direction of $\vec{\mathbf{v}}$ at any moment is along the line tangent to the path at that moment (Fig. 3–17b).

Note that the magnitude of the average velocity in Fig. 3–16 is not equal to the average speed, which is the actual distance traveled along the path, $\Delta\ell$, divided by Δt. In some special cases, the average speed and average velocity are equal in magnitude (such as motion along a straight line in one direction), but in general they are not. However, in the limit $\Delta t \to 0$, Δr always approaches $\Delta\ell$, so the instantaneous speed *always* equals the magnitude of the instantaneous velocity at any time.

The instantaneous velocity (Eq. 3–8) is equal to the derivative of the position vector with respect to time. Equation 3–8 can be written in terms of components starting with Eq. 3–6a as:

$$\vec{\mathbf{v}} = \frac{d\vec{\mathbf{r}}}{dt} = \frac{dx}{dt}\hat{\mathbf{i}} + \frac{dy}{dt}\hat{\mathbf{j}} + \frac{dz}{dt}\hat{\mathbf{k}} = v_x\hat{\mathbf{i}} + v_y\hat{\mathbf{j}} + v_z\hat{\mathbf{k}}, \tag{3–9}$$

where $v_x = dx/dt$, $v_y = dy/dt$, $v_z = dz/dt$ are the x, y, and z components of the velocity. Note that $d\hat{\mathbf{i}}/dt = d\hat{\mathbf{j}}/dt = d\hat{\mathbf{k}}/dt = 0$ since these unit vectors are constant in both magnitude and direction.

Acceleration in two or three dimensions is treated in a similar way. The **average acceleration vector**, over a time interval $\Delta t = t_2 - t_1$ is defined as

$$\text{average acceleration} = \frac{\Delta\vec{\mathbf{v}}}{\Delta t} = \frac{\vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_1}{t_2 - t_1}, \tag{3–10}$$

where $\Delta\vec{\mathbf{v}}$ is the change in the instantaneous velocity vector during that time interval: $\Delta\vec{\mathbf{v}} = \vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_1$. Note that $\vec{\mathbf{v}}_2$ in many cases, such as in Fig. 3–18a, may not be in the same direction as $\vec{\mathbf{v}}_1$. Hence the average acceleration vector may be in a different direction from either $\vec{\mathbf{v}}_1$ or $\vec{\mathbf{v}}_2$ (Fig. 3–18b). Furthermore, $\vec{\mathbf{v}}_2$ and $\vec{\mathbf{v}}_1$ may have the same magnitude but different directions, and the difference of two such vectors will not be zero. Hence acceleration can result from either a change in the magnitude of the velocity, or from a change in direction of the velocity, or from a change in both.

The **instantaneous acceleration vector** is defined as the limit of the average acceleration vector as the time interval Δt is allowed to approach zero:

$$\vec{\mathbf{a}} = \lim_{\Delta t \to 0} \frac{\Delta\vec{\mathbf{v}}}{\Delta t} = \frac{d\vec{\mathbf{v}}}{dt}, \tag{3–11}$$

and is thus the derivative of $\vec{\mathbf{v}}$ with respect to t.

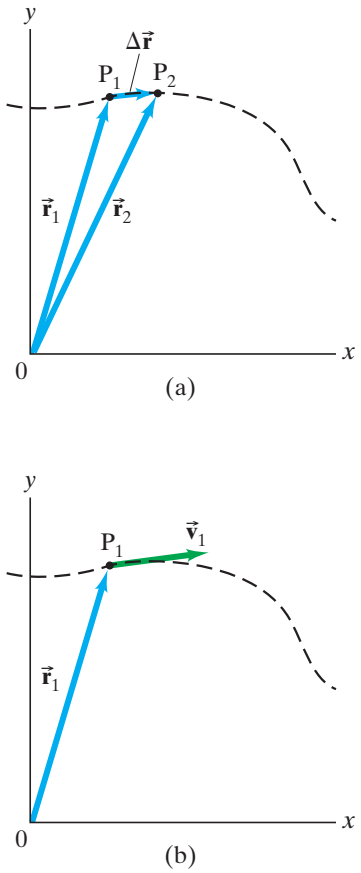


FIGURE 3–17 (a) As we take Δt and $\Delta\vec{\mathbf{r}}$ smaller and smaller [compare to Fig. 3–16] we see that the direction of $\Delta\vec{\mathbf{r}}$ and of the instantaneous velocity ($\Delta\vec{\mathbf{r}}/\Delta t$, where $\Delta t \to 0$) is (b) tangent to the curve at P_1.

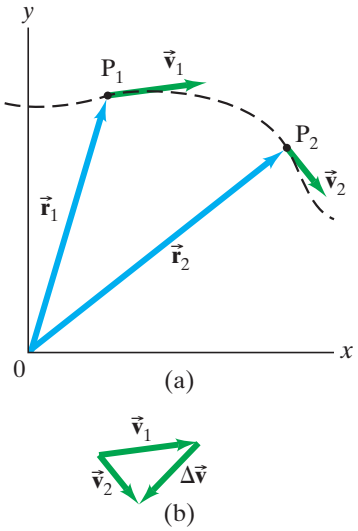


FIGURE 3–18 (a) Velocity vectors $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$ at instants t_1 and t_2 for a particle at points P_1 and P_2, as in Fig. 3–16. (b) The direction of the average acceleration is in the direction of $\Delta\vec{\mathbf{v}} = \vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_1$.

We can write $\vec{\mathbf{a}}$ using components:

$$\vec{\mathbf{a}} = \frac{d\vec{\mathbf{v}}}{dt} = \frac{dv_x}{dt}\hat{\mathbf{i}} + \frac{dv_y}{dt}\hat{\mathbf{j}} + \frac{dv_z}{dt}\hat{\mathbf{k}}$$
$$= a_x\hat{\mathbf{i}} + a_y\hat{\mathbf{j}} + a_z\hat{\mathbf{k}}, \qquad \textbf{(3–12a)}$$

where $a_x = dv_x/dt$, etc. Because $v_x = dx/dt$, then $a_x = dv_x/dt = d^2x/dt^2$, as we saw in Section 2–4. Thus we can also write the acceleration as

$$\vec{\mathbf{a}} = \frac{d^2x}{dt^2}\hat{\mathbf{i}} + \frac{d^2y}{dt^2}\hat{\mathbf{j}} + \frac{d^2z}{dt^2}\hat{\mathbf{k}}. \qquad \textbf{(3–12b)}$$

The instantaneous acceleration will be nonzero not only when the magnitude of the velocity changes, but also if its direction changes. For example, a person riding in a car traveling at constant speed around a curve, or a child riding on a merry-go-round, will both experience an acceleration because of a change in the direction of the velocity, even though the speed may be constant. (More on this in Chapter 5.)

In general, we will use the terms "velocity" and "acceleration" to mean the instantaneous values. If we want to discuss average values, we will use the word "average."

EXAMPLE 3–5 **Position given as a function of time.** The position of a particle as a function of time is given by

$$\vec{\mathbf{r}} = \left[(5.0\,\text{m/s})t + (6.0\,\text{m/s}^2)t^2\right]\hat{\mathbf{i}} + \left[(7.0\,\text{m}) - (3.0\,\text{m/s}^3)t^3\right]\hat{\mathbf{j}},$$

where r is in meters and t is in seconds. (*a*) What is the particle's displacement between $t_1 = 2.0\,\text{s}$ and $t_2 = 3.0\,\text{s}$? (*b*) Determine the particle's instantaneous velocity and acceleration as a function of time. (*c*) Evaluate $\vec{\mathbf{v}}$ and $\vec{\mathbf{a}}$ at $t = 3.0\,\text{s}$.

APPROACH For (*a*), we find $\Delta\vec{\mathbf{r}} = \vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1$, inserting $t_1 = 2.0\,\text{s}$ for finding $\vec{\mathbf{r}}_1$, and $t_2 = 3.0\,\text{s}$ for $\vec{\mathbf{r}}_2$. For (*b*), we take derivatives (Eqs. 3–9 and 3–12), and for (*c*) we substitute $t = 3.0\,\text{s}$ into our results in (*b*).

SOLUTION (*a*) We insert $t_1 = 2.0\,\text{s}$ into the given equation for $\vec{\mathbf{r}}$:

$$\vec{\mathbf{r}}_1 = \left[(5.0\,\text{m/s})(2.0\,\text{s}) + (6.0\,\text{m/s}^2)(2.0\,\text{s})^2\right]\hat{\mathbf{i}} + \left[(7.0\,\text{m}) - (3.0\,\text{m/s}^3)(2.0\,\text{s})^3\right]\hat{\mathbf{j}}$$
$$= (34\,\text{m})\hat{\mathbf{i}} - (17\,\text{m})\hat{\mathbf{j}}.$$

Similarly, at $t_2 = 3.0\,\text{s}$,

$$\vec{\mathbf{r}}_2 = (15\,\text{m} + 54\,\text{m})\hat{\mathbf{i}} + (7.0\,\text{m} - 81\,\text{m})\hat{\mathbf{j}} = (69\,\text{m})\hat{\mathbf{i}} - (74\,\text{m})\hat{\mathbf{j}}.$$

Thus

$$\Delta\vec{\mathbf{r}} = \vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1 = (69\,\text{m} - 34\,\text{m})\hat{\mathbf{i}} + (-74\,\text{m} + 17\,\text{m})\hat{\mathbf{j}} = (35\,\text{m})\hat{\mathbf{i}} - (57\,\text{m})\hat{\mathbf{j}}.$$

That is, $\Delta x = 35\,\text{m}$, and $\Delta y = -57\,\text{m}$.

(*b*) To find velocity, we take the derivative of the given $\vec{\mathbf{r}}$ with respect to time, noting (Appendix B–2) that $d(t^2)/dt = 2t$, and $d(t^3)/dt = 3t^2$:

$$\vec{\mathbf{v}} = \frac{d\vec{\mathbf{r}}}{dt} = \left[5.0\,\text{m/s} + (12\,\text{m/s}^2)t\right]\hat{\mathbf{i}} + \left[0 - (9.0\,\text{m/s}^3)t^2\right]\hat{\mathbf{j}}.$$

The acceleration is (keeping only two significant figures):

$$\vec{\mathbf{a}} = \frac{d\vec{\mathbf{v}}}{dt} = (12\,\text{m/s}^2)\hat{\mathbf{i}} - (18\,\text{m/s}^3)t\,\hat{\mathbf{j}}.$$

Thus $a_x = 12\,\text{m/s}^2$ is constant; but $a_y = -(18\,\text{m/s}^3)t$ depends linearly on time, increasing in magnitude with time in the negative y direction.

(*c*) We substitute $t = 3.0\,\text{s}$ into the equations we just derived for $\vec{\mathbf{v}}$ and $\vec{\mathbf{a}}$:

$$\vec{\mathbf{v}} = (5.0\,\text{m/s} + 36\,\text{m/s})\hat{\mathbf{i}} - (81\,\text{m/s})\hat{\mathbf{j}} = (41\,\text{m/s})\hat{\mathbf{i}} - (81\,\text{m/s})\hat{\mathbf{j}}$$
$$\vec{\mathbf{a}} = (12\,\text{m/s}^2)\hat{\mathbf{i}} - (54\,\text{m/s}^2)\hat{\mathbf{j}}.$$

Their magnitudes at $t = 3.0\,\text{s}$ are $v = \sqrt{(41\,\text{m/s})^2 + (81\,\text{m/s})^2} = 91\,\text{m/s}$, and $a = \sqrt{(12\,\text{m/s}^2)^2 + (54\,\text{m/s}^2)^2} = 55\,\text{m/s}^2$.

Constant Acceleration

In Chapter 2 we studied the important case of one-dimensional motion for which the acceleration is constant. In two or three dimensions, if the acceleration vector, $\vec{\mathbf{a}}$, is constant in magnitude and direction, then a_x = constant, a_y = constant, a_z = constant. The average acceleration in this case is equal to the instantaneous acceleration at any moment. The equations we derived in Chapter 2 for one dimension, Eqs. 2–12a, b, and c, apply separately to each perpendicular component of two- or three-dimensional motion.

In two dimensions we let $\vec{\mathbf{v}}_0 = v_{x0}\hat{\mathbf{i}} + v_{y0}\hat{\mathbf{j}}$ be the initial velocity, and we apply Eqs. 3–6a, 3–9, and 3–12b for the position vector $\vec{\mathbf{r}}$, velocity $\vec{\mathbf{v}}$, and acceleration $\vec{\mathbf{a}}$. We can then write Eqs. 2–12a, b, and c for two dimensions as shown in Table 3–1.

TABLE 3–1 Kinematic Equations for Constant Acceleration in 2 Dimensions

x component (horizontal)		*y* component (vertical)
$v_x = v_{x0} + a_x t$	(Eq. 2–12a)	$v_y = v_{y0} + a_y t$
$x = x_0 + v_{x0}t + \frac{1}{2}a_x t^2$	(Eq. 2–12b)	$y = y_0 + v_{y0}t + \frac{1}{2}a_y t^2$
$v_x^2 = v_{x0}^2 + 2a_x(x - x_0)$	(Eq. 2–12c)	$v_y^2 = v_{y0}^2 + 2a_y(y - y_0)$

The first two of the equations in Table 3–1 can be written more formally in vector notation.

$$\vec{\mathbf{v}} = \vec{\mathbf{v}}_0 + \vec{\mathbf{a}}t \qquad [\vec{\mathbf{a}} = \text{constant}] \quad \textbf{(3–13a)}$$

$$\vec{\mathbf{r}} = \vec{\mathbf{r}}_0 + \vec{\mathbf{v}}_0 t + \tfrac{1}{2}\vec{\mathbf{a}}t^2. \qquad [\vec{\mathbf{a}} = \text{constant}] \quad \textbf{(3–13b)}$$

Here, $\vec{\mathbf{r}}$ is the position vector at any time, and $\vec{\mathbf{r}}_0$ is the position vector at $t = 0$. These equations are the vector equivalent of Eqs. 2–12a and b. In practical situations, we usually use the component form given in Table 3–1.

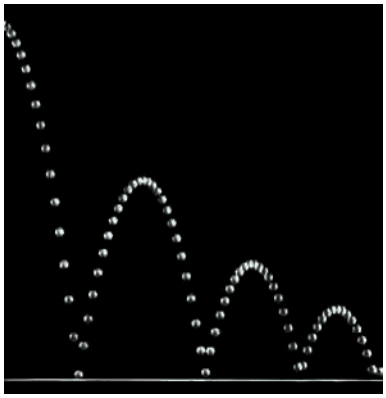
3–7 Projectile Motion

In Chapter 2, we studied one-dimensional motion of an object in terms of displacement, velocity, and acceleration, including purely vertical motion of a falling object undergoing acceleration due to gravity. Now we examine the more general translational motion of objects moving through the air in two dimensions near the Earth's surface, such as a golf ball, a thrown or batted baseball, kicked footballs, and speeding bullets. These are all examples of **projectile motion** (see Fig. 3–19), which we can describe as taking place in two dimensions if there is no wind.

Although air resistance is often important, in many cases its effect can be ignored, and we will ignore it in the following analysis. We will not be concerned now with the process by which the object is thrown or projected. We consider only its motion *after* it has been projected, and *before* it lands or is caught—that is, we analyze our projected object only when it is moving freely through the air under the action of gravity alone. Then the acceleration of the object is that due to gravity, which acts downward with magnitude $g = 9.80\,\text{m/s}^2$, and we assume it is constant.[†]

Galileo was the first to describe projectile motion accurately. He showed that it could be understood by analyzing the horizontal and vertical components of the motion separately. For convenience, we assume that the motion begins at time $t = 0$ at the origin of an xy coordinate system (so $x_0 = y_0 = 0$).

[†]This restricts us to objects whose distance traveled and maximum height above the Earth are small compared to the Earth's radius (6400 km).



(a)



(b)

FIGURE 3–19 Photographs of (a) a bouncing ball and (b) a thrown basketball, each showing the characteristic "parabolic" path of projectile motion.

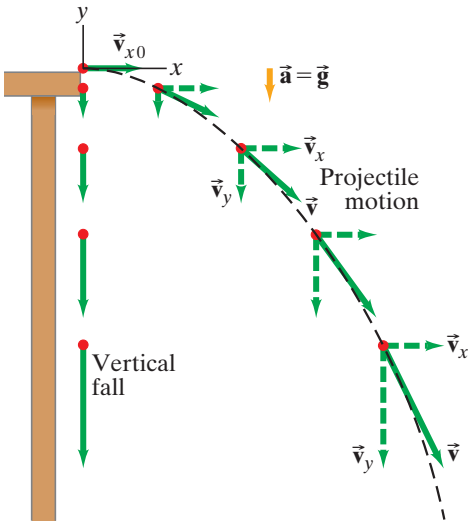


FIGURE 3–20 Projectile motion of a small ball projected horizontally with initial velocity $\vec{\mathbf{v}} = \vec{\mathbf{v}}_{x0}$. The dashed black line represents the path of the object. The velocity vector $\vec{\mathbf{v}}$ is in the direction of motion at each point, and thus is tangent to the path. The velocity vectors are green arrows, and velocity components are dashed. (A vertically falling object starting from rest at the same place and time is shown at the left for comparison; v_y is the same at each instant for the falling object and the projectile.)

Let us look at a (tiny) ball rolling off the end of a horizontal table with an initial velocity in the horizontal (x) direction, $\vec{\mathbf{v}}_{x0}$. See Fig. 3–20, where an object falling vertically is also shown for comparison. The velocity vector $\vec{\mathbf{v}}$ at each instant points in the direction of the ball's motion at that instant and is thus always tangent to the path. Like Galileo, we treat the horizontal and vertical components of the velocity and acceleration separately, and we apply the kinematic equations (Eqs. 2–12) to the x and y components of the motion.

First we examine the vertical (y) component of the motion. At the instant the ball leaves the table's top ($t = 0$), it has only an x component of velocity. Once the ball leaves the table (at $t = 0$), it experiences a vertically downward acceleration g, the acceleration due to gravity. Thus v_y is initially zero ($v_{y0} = 0$) but increases continually in the downward direction (until the ball hits the ground). Let us take y to be positive upward. Then the acceleration due to gravity is in the $-y$ direction, so $a_y = -g$. From Eq. 2–12a (using y in place of x) we can write $v_y = v_{y0} + a_y t = -gt$ since we set $v_{y0} = 0$. The vertical displacement is given by Eq. 2–12b written in terms of y: $y = y_0 + v_{y0} + \frac{1}{2}a_y t^2$. Given $y_0 = 0$, $v_{y0} = 0$, and $a_y = -g$, then $y = -\frac{1}{2}gt^2$.

In the horizontal direction, the acceleration is zero (ignoring air resistance). With $a_x = 0$, the horizontal component of velocity, v_x, remains constant, equal to its initial value, v_{x0}, and thus has the same magnitude at each point on the path. The horizontal displacement (with $a_x = 0$) is given by $x = v_{x0}t + \frac{1}{2}a_x t^2 = v_{x0}t$.

The two vector components, $\vec{\mathbf{v}}_x$ and $\vec{\mathbf{v}}_y$, can be added vectorially at any instant to obtain the velocity $\vec{\mathbf{v}}$ at that time (each point on the path), as shown in Fig. 3–20.

One result of this analysis, which Galileo himself predicted, is that *an object projected horizontally will reach the ground in the same time as an object dropped vertically*. The vertical motions are the same in both cases, as shown in Fig. 3–20. Figure 3–21 is a multiple-exposure photograph of an experiment that confirms this.

If an object is projected at an upward angle, as in Fig. 3–22, the analysis is similar, but now there is an initial vertical component of velocity, v_{y0}. Because of the downward acceleration of gravity, the upward component of velocity v_y gradually decreases with time until the object reaches the highest point on its path, at which point $v_y = 0$. The object then moves downward (Fig. 3–22) and v_y increases in the downward direction (becoming more negative). As before, v_x remains constant.

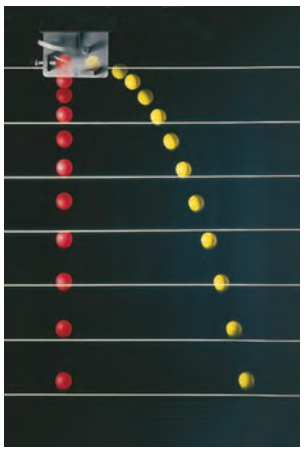


FIGURE 3–21 Multiple-exposure photograph showing positions of two balls at equal time intervals. One ball was dropped from rest at the same time the other was projected horizontally outward. The vertical position of each ball is seen to be the same at each instant.

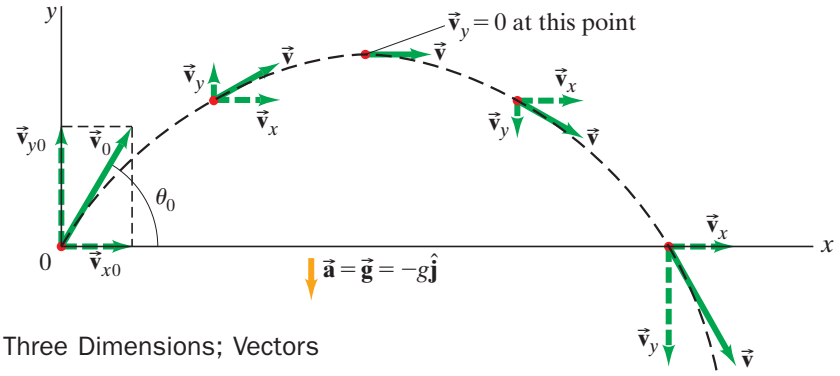
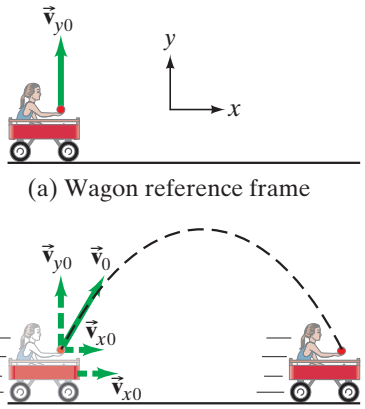


FIGURE 3–22 Path of a projectile fired with initial velocity $\vec{\mathbf{v}}_0$ at angle θ_0 to the horizontal. Path is shown dashed in black, the velocity vectors are green arrows, and velocity components are dashed. The acceleration $\vec{\mathbf{a}} = d\vec{\mathbf{v}}/dt$ is downward. That is, $\vec{\mathbf{a}} = \vec{\mathbf{g}} = -g\hat{\mathbf{j}}$ where $\hat{\mathbf{j}}$ is the unit vector in the positive y direction. Not shown is where the projectile hits the ground (at that point projectile motion ceases).

CONCEPTUAL EXAMPLE 3–6 **Where does the apple land?** A child sits upright in a wagon that moves to the right at constant speed, Fig. 3–23. The child extends her hand and throws an apple straight up (from her own point of view, Fig. 3–23a), as the wagon moves forward at constant speed. Neglecting air resistance, will the apple land (*a*) behind the wagon, (*b*) in the wagon, or (*c*) in front of the wagon?

RESPONSE The child throws the apple straight up in her own reference frame with initial velocity $\vec{\mathbf{v}}_{y0}$ (Fig. 3–23a). But when viewed by someone on the ground, the apple also has an initial horizontal component of velocity equal to the velocity of the wagon, $\vec{\mathbf{v}}_{x0}$. Thus, to a person on the ground, the apple follows the path of a projectile as shown in Fig. 3–23b. The apple experiences no horizontal acceleration, so $\vec{\mathbf{v}}_{x0}$ stays constant, equal to the speed of the wagon. As the apple follows its arc, the wagon will be directly under the apple at all times because they have the same horizontal velocity. When the apple comes down, it will drop right into the outstretched hand of the child. The answer is (*b*).



(a) Wagon reference frame

(b) Ground reference frame

FIGURE 3–23 Example 3–6.

EXERCISE C Return to the Chapter-Opening Question, page 54, and answer it again now. Try to explain why you may have answered differently the first time. Describe the role of the helicopter in this example of projectile motion.

3–8 Solving Problems Involving Projectile Motion

We can simplify Eqs. 2–12 (Table 3–1) for the case of projectile motion because we can set $a_x = 0$. See Table 3–2, which assumes y is positive upward, so $a_y = -g = -9.80\,\text{m/s}^2$. If θ is chosen relative to the $+x$ axis, as in Fig. 3–22, then

$$v_{x0} = v_0 \cos\theta_0, \quad \text{and} \quad v_{y0} = v_0 \sin\theta_0.$$

In doing Problems involving projectile motion, we must consider a time interval for which our chosen object is in the air, influenced only by gravity. We do not consider the throwing (or projecting) process, nor the time after the object lands or is caught, because then other influences act on the object, and we can no longer set $\vec{\mathbf{a}} = \vec{\mathbf{g}}$.

PROBLEM SOLVING
Choice of time interval

TABLE 3–2 Kinematic Equations for Projectile Motion
(y positive upward; $a_x = 0$, $a_y = -g = -9.80\,\text{m/s}^2$)

Horizontal Motion ($a_x = 0$, v_x = constant)		Vertical Motion† ($a_y = -g$ = constant)
$v_x = v_{x0}$	(Eq. 2–12a)	$v_y = v_{y0} - gt$
$x = x_0 + v_{x0}t$	(Eq. 2–12b)	$y = y_0 + v_{y0}t - \frac{1}{2}gt^2$
	(Eq. 2–12c)	$v_y^2 = v_{y0}^2 - 2g(y - y_0)$

†If y is taken positive downward, the minus $(-)$ signs in front of g become plus $(+)$ signs.

PROBLEM SOLVING

Projectile Motion

Our approach to solving Problems in Section 2–6 also applies here. Solving projectile motion Problems can require creativity, and cannot be done just by following rules. You must avoid just plugging numbers into equations that seem to "work."

1. As always, **read** carefully; **choose** the **object** (or objects) you are going to analyze.
2. **Draw** a careful **diagram** showing what is happening.
3. **Choose** an origin and an xy **coordinate system**.
4. Decide on the **time interval**, which for projectile motion can only include motion under the effect of gravity alone, not throwing or landing. The time interval must be the same for the x and y analyses. The x and y motions are connected by the common time, t.
5. **Examine** the horizontal (x) and vertical (y) **motions** separately. If you are given the initial velocity, you may want to resolve it into its x and y components.
6. List the **known** and **unknown** quantities, choosing $a_x = 0$ and $a_y = -g$ or $+g$, where $g = 9.80\,\text{m/s}^2$, depending on choice of y positive up or down. Recall: v_x never changes throughout the trajectory, and $v_y = 0$ at the highest point of any trajectory that returns downward. The velocity just before landing is generally not zero.
7. Think for a minute before jumping into the equations. **Apply** the **relevant equations** (Table 3–2), combining them if necessary. You may need to combine components of a vector to get magnitude and direction (Eqs. 3–3).

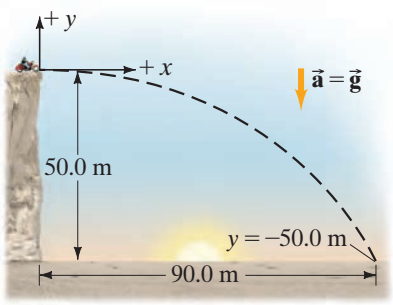


FIGURE 3–24 Example 3–7.

EXAMPLE 3–7 **Driving off a cliff.** A movie stunt driver on a motorcycle speeds horizontally off a 50.0-m-high cliff. How fast must the motorcycle leave the clifftop to land on level ground below, 90.0 m from the base of the cliff where the cameras are? Ignore air resistance.

APPROACH We explicitly follow the steps of the Problem Solving Strategy on the previous page.

SOLUTION

1. and **2. Read, choose the object, and draw a diagram**. Our object is the motorcycle and driver, taken as a single unit. The diagram is shown in Fig. 3–24.
3. **Choose a coordinate system**. We choose the y direction to be positive upward, with the top of the cliff as $y_0 = 0$. The x direction is horizontal with $x_0 = 0$ at the point where the motorcycle leaves the cliff.
4. **Choose a time interval**. We choose our time interval to begin ($t = 0$) just as the motorcycle leaves the clifftop at position $x_0 = 0$, $y_0 = 0$. Our time interval ends just before the motorcycle touches the ground below.
5. **Examine *x* and *y* motions**. In the horizontal (x) direction, the acceleration $a_x = 0$, so the velocity is constant. The value of x when the motorcycle reaches the ground is $x = +90.0\,\text{m}$. In the vertical direction, the acceleration is the acceleration due to gravity, $a_y = -g = -9.80\,\text{m/s}^2$. The value of y when the motorcycle reaches the ground is $y = -50.0\,\text{m}$. The initial velocity is horizontal and is our unknown, v_{x0}; the initial vertical velocity is zero, $v_{y0} = 0$.
6. **List knowns and unknowns**. See the Table in the margin. Note that in addition to not knowing the initial horizontal velocity v_{x0} (which stays constant until landing), we also do not know the time t when the motorcycle reaches the ground.
7. **Apply relevant equations**. The motorcycle maintains constant v_x as long as it is in the air. The time it stays in the air is determined by the y motion—when it reaches the ground. So we first find the time using the y motion, and then use this time value in the x equations. To find out how long it takes the motorcycle to reach the ground below, we use Eq. 2–12b (Table 3–2) for the vertical (y) direction with $y_0 = 0$ and $v_{y0} = 0$:

$$\begin{aligned} y &= y_0 + v_{y0}t + \tfrac{1}{2}a_y t^2 \\ &= 0 + 0 + \tfrac{1}{2}(-g)t^2 \end{aligned}$$

or

$$y = -\tfrac{1}{2}gt^2.$$

We solve for t and set $y = -50.0\,\text{m}$:

$$t = \sqrt{\frac{2y}{-g}} = \sqrt{\frac{2(-50.0\,\text{m})}{-9.80\,\text{m/s}^2}} = 3.19\,\text{s}.$$

To calculate the needed initial velocity, v_{x0}, we again use Eq. 2–12b, but this time for the horizontal (x) direction, with $a_x = 0$ and $x_0 = 0$:

$$\begin{aligned} x &= x_0 + v_{x0}t + \tfrac{1}{2}a_x t^2 \\ &= 0 + v_{x0}t + 0 \end{aligned}$$

or

$$x = v_{x0}t.$$

So the motorcycle needs to leave the clifftop with a speed

$$v_{x0} = \frac{x}{t} = \frac{90.0\,\text{m}}{3.19\,\text{s}} = 28.2\,\text{m/s},$$

which is about 100 km/h (roughly 60 mi/h).

Known	Unknown
$x_0 = y_0 = 0$	v_{x0}
$x = 90.0\,\text{m}$	t
$y = -50.0\,\text{m}$	
$a_x = 0$	
$a_y = -g = -9.80\,\text{m/s}^2$	
$v_{y0} = 0$	

NOTE In the time interval of the projectile motion, the only acceleration is g in the negative y direction. The acceleration in the x direction is zero.

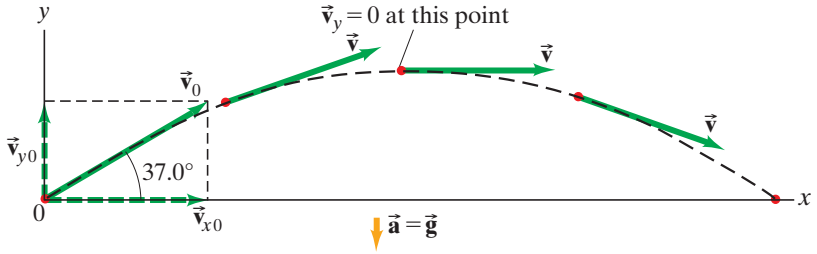


FIGURE 3–25 Example 3–8.

EXAMPLE 3–8 **A kicked football.** A football is kicked at an angle $\theta_0 = 37.0°$ with a velocity of 20.0 m/s, as shown in Fig. 3–25. Calculate (*a*) the maximum height, (*b*) the time of travel before the football hits the ground, and (*c*) how far away it hits the ground. Assume the ball leaves the foot at ground level, and ignore air resistance, wind, and rotation of the ball.

PHYSICS APPLIED
Sports

APPROACH This may seem difficult at first because there are so many questions. But we can deal with them one at a time. We take the y direction as positive upward, and treat the x and y motions separately. The total time in the air is again determined by the y motion. The x motion occurs at constant velocity. The y component of velocity varies, being positive (upward) initially, decreasing to zero at the highest point, and then becoming negative as the football falls.

SOLUTION We resolve the initial velocity into its components (Fig. 3–25):

$$v_{x0} = v_0 \cos 37.0° = (20.0\,\text{m/s})(0.799) = 16.0\,\text{m/s}$$
$$v_{y0} = v_0 \sin 37.0° = (20.0\,\text{m/s})(0.602) = 12.0\,\text{m/s}.$$

(*a*) To find the maximum height, we consider a time interval that begins just after the football loses contact with the foot until the ball reaches its maximum height. During this time interval, the acceleration is g downward. At the maximum height, the velocity is horizontal (Fig. 3–25), so $v_y = 0$; and this occurs at a time given by $v_y = v_{y0} - gt$ with $v_y = 0$ (see Eq. 2–12a in Table 3–2). Thus

$$t = \frac{v_{y0}}{g} = \frac{(12.0\,\text{m/s})}{(9.80\,\text{m/s}^2)} = 1.224\,\text{s} \approx 1.22\,\text{s}.$$

From Eq. 2–12b, with $y_0 = 0$, we can solve for y at this time $t = v_{y0}/g$:

$$y = v_{y0}t - \tfrac{1}{2}gt^2 = \frac{v_{y0}^2}{g} - \frac{1}{2}\frac{v_{y0}^2}{g} = \frac{v_{y0}^2}{2g} = \frac{(12.0\,\text{m/s})^2}{2(9.80\,\text{m/s}^2)} = 7.35\,\text{m}.$$

The maximum height is 7.35 m. [Solving Eq. 2–12c for y gives the same result.]

(*b*) To find the time it takes for the ball to return to the ground, we consider a different time interval, starting at the moment the ball leaves the foot ($t = 0$, $y_0 = 0$) and ending just before the ball touches the ground ($y = 0$ again). We can use Eq. 2–12b with $y_0 = 0$ and also set $y = 0$ (ground level):

$$y = y_0 + v_{y0}t - \tfrac{1}{2}gt^2$$
$$0 = 0 + v_{y0}t - \tfrac{1}{2}gt^2.$$

This equation can be easily factored:

$$t(\tfrac{1}{2}gt - v_{y0}) = 0.$$

There are two solutions, $t = 0$ (which corresponds to the initial point, y_0), and

$$t = \frac{2v_{y0}}{g} = \frac{2(12.0\,\text{m/s})}{(9.80\,\text{m/s}^2)} = 2.45\,\text{s},$$

which is the total travel time of the football.

(*c*) The total distance traveled in the x direction is found by applying Eq. 2–12b with $x_0 = 0$, $a_x = 0$, $v_{x0} = 16.0\,\text{m/s}$, and $t = 2.45\,\text{s}$:

$$x = v_{x0}t = (16.0\,\text{m/s})(2.45\,\text{s}) = 39.2\,\text{m}.$$

NOTE In (*b*) the time needed for the whole trip, $t = 2v_{y0}/g = 2.45\,\text{s}$, is double the time to reach the highest point, calculated in (*a*). That is, the time to go up equals the time to come back down to the same level (ignoring air resistance).

EXERCISE D In Example 3–8, what is (*a*) the velocity vector at the maximum height, and (*b*) the acceleration vector at maximum height?

In Example 3–8 we treated the football as if it were a particle, ignoring its rotation. We also ignored air resistance. Because air resistance is significant on a football, our results are only estimates (mainly overestimates). We also ignored any wind. If there happened to be a cross-wind, we would need a third axis for such 3-dimensional motion.

EXERCISE E Two balls are thrown in the air at different angles, but each reaches the same height. Which ball remains in the air longer: the one thrown at the steeper angle or the one thrown at a shallower angle?

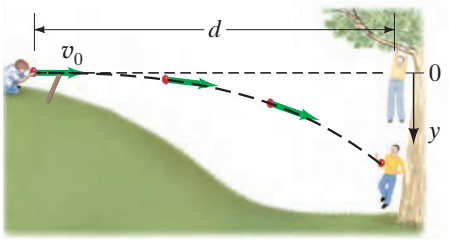


FIGURE 3–26 Example 3–9.

CONCEPTUAL EXAMPLE 3–9 **The wrong strategy.** A boy on a small hill aims his water-balloon slingshot horizontally, straight at a second boy hanging from a tree branch a distance *d* away, Fig. 3–26. At the instant the water balloon is released, the second boy lets go and falls from the tree, hoping to avoid being hit. Show that he made the wrong move. (He hadn't studied physics yet.) Ignore air resistance.

RESPONSE Both the water balloon and the boy in the tree start falling at the same instant, and in a time t they each fall the same vertical distance $y = \frac{1}{2}gt^2$, much like Fig. 3–21. In the time it takes the water balloon to travel the horizontal distance d, the balloon will have the same y position as the falling boy. Splat. If the boy had stayed in the tree, he would have avoided the humiliation.

Level Horizontal Range

The total distance the football traveled in Example 3–8 is called the horizontal **range** R. We now derive a formula for the range, which applies to a projectile that lands at the same level it started ($= y_0$): that is, $y(\text{final}) = y_0$ (see Fig. 3–27a). Looking back at Example 3–8 part (*c*), we see that $x = R = v_{x0}t$ where (from part *b*) $t = 2v_{y0}/g$. Thus

$$R = v_{x0}t = v_{x0}\left(\frac{2v_{y0}}{g}\right) = \frac{2v_{x0}v_{y0}}{g} = \frac{2v_0^2 \sin\theta_0 \cos\theta_0}{g}, \quad [y = y_0]$$

where $v_{x0} = v_0 \cos\theta_0$ and $v_{y0} = v_0 \sin\theta_0$. This can be rewritten, using the trigonometric identity $2\sin\theta\cos\theta = \sin 2\theta$ (Appendix A or inside the rear cover):

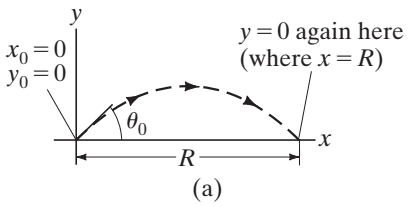
$$R = \frac{v_0^2 \sin 2\theta_0}{g}. \qquad [\text{only if } y(\text{final}) = y_0]$$

Note that the *maximum* range, for a given initial velocity v_0, is obtained when $\sin 2\theta$ takes on its maximum value of 1.0, which occurs for $2\theta_0 = 90°$; so

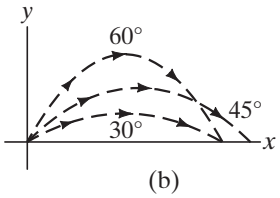
$$\theta_0 = 45° \text{ for maximum range, and } R_{\max} = v_0^2/g.$$

The maximum range increases by the square of v_0, so doubling the initial velocity of a projectile increases its maximum range by a factor of 4.

When air resistance is important, the range is less for a given v_0, and the maximum range is obtained at an angle smaller than 45°.



(a)



(b)

FIGURE 3–27 (a) The range R of a projectile. (b) There are generally two angles θ_0 that will give the same range. If one angle is θ_{01}, the other is $\theta_{02} = 90° - \theta_{01}$. Also see Example 3–10.

EXAMPLE 3–10 **Range of a cannon ball.** Suppose one of Napoleon's cannons had a muzzle speed, v_0, of 60.0 m/s. At what angle should it have been aimed (ignore air resistance) to strike a target 320 m away?

APPROACH We use the equation just derived for the range, $R = v_0^2 \sin 2\theta_0/g$, with $R = 320$ m.

SOLUTION We solve for $\sin 2\theta_0$ in the range formula:

$$\sin 2\theta_0 = \frac{Rg}{v_0^2} = \frac{(320\text{ m})(9.80\text{ m/s}^2)}{(60.0\text{ m/s})^2} = 0.871.$$

We want to solve for an angle θ_0 that is between 0° and 90°, which means $2\theta_0$ in this equation can be as large as 180°. Then $\sin^{-1}(0.871) = 2\theta_0 = 60.6°$ is a solution,

so the cannon should be aimed at $\theta_0 = 30.3°$. But $2\theta_0 = 180° - 60.6° = 119.4°$ is also a solution (see Appendix A–9), so θ_0 can also be $\theta_0 = 59.7°$. In general we have two solutions (see Fig. 3–27b), which in the present case are given by

$$\theta_0 = 30.3° \quad \text{or} \quad 59.7°.$$

Either angle gives the same range. Only when $\sin 2\theta_0 = 1$ (so $\theta_0 = 45°$) is there a single solution (that is, both solutions are the same).

The level range formula applies only if takeoff and landing are at the same height $(y = y_0)$. Example 3–11 below considers a case where they are not equal heights $(y \neq y_0)$.

EXAMPLE 3–11 **A punt.** Suppose the football in Example 3–8 was punted and left the punter's foot at a height of 1.00 m above the ground. How far did the football travel before hitting the ground? Set $x_0 = 0$, $y_0 = 0$.

PHYSICS APPLIED
Sports

PROBLEM SOLVING
Do not use any formula unless you are sure its range of validity fits the problem; the range formula does not apply here because $y \neq y_0$

APPROACH The x and y motions are again treated separately. But we cannot use the range formula as in Example 3–10 because the range formula is valid only if $y(\text{final}) = y_0$, which is not the case here. Now we have $y_0 = 0$, and the football hits the ground where $y = -1.00$ m (see Fig. 3–28). We choose our time interval to start when the ball leaves his foot $(t = 0,\ y_0 = 0,\ x_0 = 0)$ and end just before the ball hits the ground $(y = -1.00\text{ m})$. We can get x from Eq. 2–12b, $x = v_{x0}t$, and we saw that $v_{x0} = 16.0$ m/s in Example 3–8. But first we must find t, the time at which the ball hits the ground, which we obtain from the y motion.

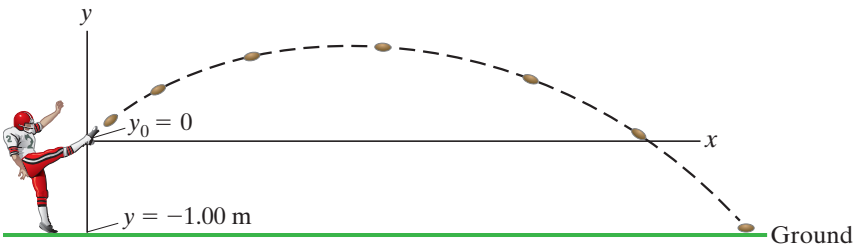


FIGURE 3–28 Example 3–11: the football leaves the punter's foot at $y = 0$, and reaches the ground where $y = -1.00$ m.

SOLUTION To find t with $y = -1.00$ m and $v_{y0} = 12.0$ m/s (see Example 3–8), we use the equation

$$y = y_0 + v_{y0}t - \tfrac{1}{2}gt^2,$$

and obtain

$$-1.00\text{ m} = 0 + (12.0\text{ m/s})t - (4.90\text{ m/s}^2)t^2.$$

We rearrange this equation into standard form $(ax^2 + bx + c = 0)$ so we can use the quadratic formula:

$$(4.90\text{ m/s}^2)t^2 - (12.0\text{ m/s})t - (1.00\text{ m}) = 0.$$

The quadratic formula (Appendix A–1) gives

$$t = \frac{12.0\text{ m/s} \pm \sqrt{(-12.0\text{ m/s})^2 - 4(4.90\text{ m/s}^2)(-1.00\text{ m})}}{2(4.90\text{ m/s}^2)}$$
$$= 2.53\text{ s} \quad \text{or} \quad -0.081\text{ s}.$$

The second solution would correspond to a time prior to our chosen time interval that begins at the kick $(t = 0)$, so it doesn't apply. With $t = 2.53$ s for the time at which the ball touches the ground, the horizontal distance the ball traveled is (using $v_{x0} = 16.0$ m/s from Example 3–8):

$$x = v_{x0}t = (16.0\text{ m/s})(2.53\text{ s}) = 40.5\text{ m}.$$

Our assumption in Example 3–8 that the ball leaves the foot at ground level gives a result (39.2 m) that is an underestimate of about 1.3 m in the distance our punt traveled. (But Example 3–8 would apply for a kickoff or field goal in American football.)

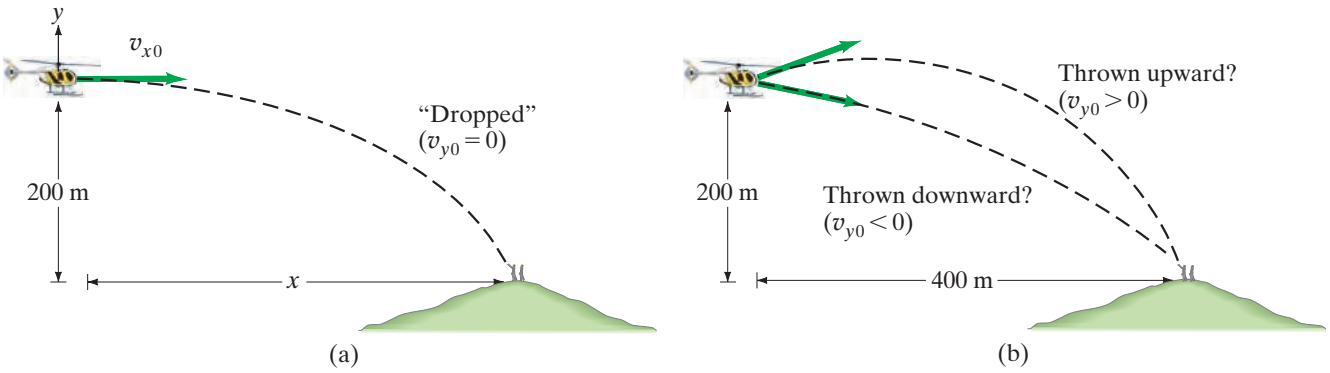


FIGURE 3–29 Example 3–12.

PHYSICS APPLIED
Reaching a target from a moving helicopter

EXAMPLE 3–12 **Rescue helicopter drops supplies.** A rescue helicopter wants to drop a package of supplies to isolated mountain climbers on a rocky ridge 200 m below. If the helicopter is traveling horizontally with a speed of 70 m/s (250 km/h), (*a*) how far in advance of the recipients (horizontal distance) must the package be dropped (Fig. 3–29a)? (*b*) Suppose, instead, that the helicopter releases the package a horizontal distance of 400 m in advance of the mountain climbers. What vertical velocity should the package be given (up or down) so that it arrives precisely at the climbers' position (Fig. 3–29b)? (*c*) With what speed does the package land in the latter case?

APPROACH We choose the origin of our xy coordinate system at the initial position of the helicopter, taking $+y$ upward, and use the kinematic equations (Table 3–2).

SOLUTION (*a*) We can find the time to reach the climbers using the vertical distance of 200 m. The package is "dropped" so initially it has the velocity of the helicopter, $v_{x0} = 70\ \text{m/s}$, $v_{y0} = 0$. Then, since $y = -\frac{1}{2}gt^2$, we have

$$t = \sqrt{\frac{-2y}{g}} = \sqrt{\frac{-2(-200\ \text{m})}{9.80\ \text{m/s}^2}} = 6.39\ \text{s}.$$

The horizontal motion of the falling package is at constant speed of 70 m/s. So

$$x = v_{x0}t = (70\ \text{m/s})(6.39\ \text{s}) = 447\ \text{m} \approx 450\ \text{m},$$

assuming the given numbers were good to two significant figures.

(*b*) We are given $x = 400\ \text{m}$, $v_{x0} = 70\ \text{m/s}$, $y = -200\ \text{m}$, and we want to find v_{y0} (see Fig. 3–29b). Like most problems, this one can be approached in various ways. Instead of searching for a formula or two, let's try to reason it out in a simple way, based on what we did in part (*a*). If we know t, perhaps we can get v_{y0}. Since the horizontal motion of the package is at constant speed (once it is released we don't care what the helicopter does), we have $x = v_{x0}t$, so

$$t = \frac{x}{v_{x0}} = \frac{400\ \text{m}}{70\ \text{m/s}} = 5.71\ \text{s}.$$

Now let's try to use the vertical motion to get v_{y0}: $y = y_0 + v_{y0}t - \frac{1}{2}gt^2$. Since $y_0 = 0$ and $y = -200\ \text{m}$, we can solve for v_{y0}:

$$v_{y0} = \frac{y + \frac{1}{2}gt^2}{t} = \frac{-200\ \text{m} + \frac{1}{2}(9.80\ \text{m/s}^2)(5.71\ \text{s})^2}{5.71\ \text{s}} = -7.0\ \text{m/s}.$$

Thus, in order to arrive at precisely the mountain climbers' position, the package must be thrown *downward* from the helicopter with a speed of 7.0 m/s.

(*c*) We want to know v of the package at $t = 5.71\ \text{s}$. The components are:

$$v_x = v_{x0} = 70\ \text{m/s}$$
$$v_y = v_{y0} - gt = -7.0\ \text{m/s} - (9.80\ \text{m/s}^2)(5.71\ \text{s}) = -63\ \text{m/s}.$$

So $v = \sqrt{(70\ \text{m/s})^2 + (-63\ \text{m/s})^2} = 94\ \text{m/s}$. (Better not to release the package from such an altitude, or use a parachute.)

Projectile Motion Is Parabolic

We now show that the path followed by any projectile is a *parabola*, if we can ignore air resistance and can assume that only gravity is acting with $\vec{\mathbf{g}}$ constant. To do so, we need to find y as a function of x by eliminating t between the two equations for horizontal and vertical motion (Eq. 2–12b in Table 3–2), and for simplicity we set $x_0 = y_0 = 0$:

$$x = v_{x0}t$$
$$y = v_{y0}t - \tfrac{1}{2}gt^2.$$

From the first equation, we have $t = x/v_{x0}$, and we substitute this into the second equation to obtain

$$y = \left(\frac{v_{y0}}{v_{x0}}\right)x - \left(\frac{g}{2v_{x0}^2}\right)x^2. \qquad \textbf{(3–14)}$$

We see that y as a function of x has the form

$$y = Ax - Bx^2,$$

where A and B are constants for any specific projectile motion. This is the well-known equation for a parabola. See Figs. 3–19 and 3–30.

The idea that projectile motion is parabolic was at the forefront of physics research in Galileo's day. Today we discuss it in Chapter 3 of introductory physics!



(a)



(b)

FIGURE 3–30 Examples of projectile motion: (a) a boy leaping, (b) glowing lava from the volcano Stromboli.

3–9 Relative Velocity

We now consider how observations made in different frames of reference are related to each other. For example, consider two trains approaching one another, each with a speed of 80 km/h with respect to the Earth. Observers on the Earth beside the train tracks will measure 80 km/h for the speed of each of the trains. Observers on either one of the trains (different frames of reference) will measure a speed of 160 km/h for the other train approaching them.

Similarly, when one car traveling 90 km/h passes a second car traveling in the same direction at 75 km/h, the first car has a speed relative to the second car of 90 km/h − 75 km/h = 15 km/h.

When the velocities are along the same line, simple addition or subtraction is sufficient to obtain the **relative velocity**. But if they are not along the same line, we must make use of vector addition. We emphasize, as mentioned in Section 2–1, that when specifying a velocity, it is important to specify what the reference frame is.

When determining relative velocity, it is easy to make a mistake by adding or subtracting the wrong velocities. It is important, therefore, to draw a diagram and use a careful labeling process. Each velocity is labeled by *two subscripts: the first refers to the object, the second to the reference frame in which it has this velocity*. For example, suppose a boat heads directly across a river, as shown in Fig. 3–31. We let $\vec{\mathbf{v}}_{BW}$ be the velocity of the **B**oat with respect to the **W**ater. (This is also what the boat's velocity would be relative to the shore if the water were still.) Similarly, $\vec{\mathbf{v}}_{BS}$ is the velocity of the **B**oat with respect to the **S**hore, and $\vec{\mathbf{v}}_{WS}$ is the velocity of the **W**ater with respect to the **S**hore (this is the river current). Note that $\vec{\mathbf{v}}_{BW}$ is what the boat's motor produces (against the water), whereas $\vec{\mathbf{v}}_{BS}$ is equal to $\vec{\mathbf{v}}_{BW}$ plus the effect of the current, $\vec{\mathbf{v}}_{WS}$. Therefore, the velocity of the boat relative to the shore is (see vector diagram, Fig. 3–31)

$$\vec{\mathbf{v}}_{BS} = \vec{\mathbf{v}}_{BW} + \vec{\mathbf{v}}_{WS}. \qquad \textbf{(3–15)}$$

By writing the subscripts using this convention, we see that the inner subscripts (the two W's) on the right-hand side of Eq. 3–15 are the same; also, the outer subscripts on the right of Eq. 3–15 (the B and the S) are the same as the two subscripts for the sum vector on the left, $\vec{\mathbf{v}}_{BS}$.

FIGURE 3–31 A boat heads north directly across a river which flows west. Velocity vectors are shown as green arrows:
$\vec{\mathbf{v}}_{BS}$ = velocity of **B**oat with respect to the **S**hore,
$\vec{\mathbf{v}}_{BW}$ = velocity of **B**oat with respect to the **W**ater,
$\vec{\mathbf{v}}_{WS}$ = velocity of **W**ater with respect to the **S**hore (river current).
As it crosses the river, the boat is dragged downstream by the current $(\vec{\mathbf{v}}_{WS})$.

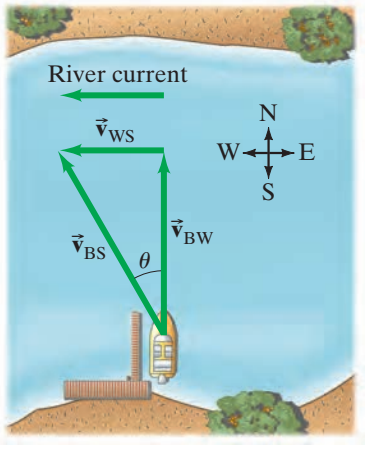


FIGURE 3–32 Derivation of relative velocity equation (Eq. 3–15), in this case for a person walking along the corridor in a train. We are looking down on the train and two reference frames are shown: xy on the Earth and $x'y'$ fixed on the train. We have:

$\vec{\mathbf{r}}_{PT}$ = position vector of person (P) relative to train (T),

$\vec{\mathbf{r}}_{PE}$ = position vector of person (P) relative to Earth (E),

$\vec{\mathbf{r}}_{TE}$ = position vector of train's coordinate system (T) relative to Earth (E).

From the diagram we see that

$$\vec{\mathbf{r}}_{PE} = \vec{\mathbf{r}}_{PT} + \vec{\mathbf{r}}_{TE}.$$

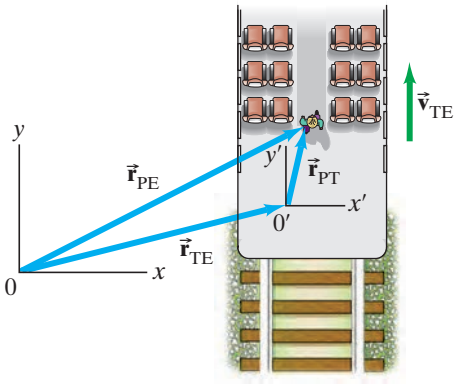
We take the derivative with respect to time to obtain

$$\frac{d}{dt}(\vec{\mathbf{r}}_{PE}) = \frac{d}{dt}(\vec{\mathbf{r}}_{PT}) + \frac{d}{dt}(\vec{\mathbf{r}}_{TE}),$$

or, since $d\vec{\mathbf{r}}/dt = \vec{\mathbf{v}}$,

$$\vec{\mathbf{v}}_{PE} = \vec{\mathbf{v}}_{PT} + \vec{\mathbf{v}}_{TE}.$$

This is the equivalent of Eq. 3–15 for the present situation (check the subscripts!).



By following this convention (first subscript for the object, second for the reference frame), you can write down the correct equation relating velocities in different reference frames.†

Figure 3–32 gives a derivation of Eq. 3–15, $\vec{\mathbf{v}}_{BS} = \vec{\mathbf{v}}_{BW} + \vec{\mathbf{v}}_{WS}$, but using different subscripts.

Equation 3–15 is valid in general and can be extended to three or more velocities. For example, if a fisherman on a boat walks with a velocity $\vec{\mathbf{v}}_{FB}$ relative to the boat, his velocity relative to the shore is $\vec{\mathbf{v}}_{FS} = \vec{\mathbf{v}}_{FB} + \vec{\mathbf{v}}_{BW} + \vec{\mathbf{v}}_{WS}$. The equations involving relative velocity will be correct when adjacent inner subscripts are identical and when the outermost ones correspond exactly to the two on the velocity on the left of the equation. But this works only with plus signs (on the right), not minus signs.

It is often useful to remember that for any two objects or reference frames, A and B, the velocity of A relative to B has the same magnitude, but opposite direction, as the velocity of B relative to A:

$$\vec{\mathbf{v}}_{BA} = -\vec{\mathbf{v}}_{AB}. \qquad \textbf{(3–16)}$$

For example, if a train is traveling 100 km/h relative to the Earth in a certain direction, objects on the Earth (such as trees) appear to an observer on the train to be traveling 100 km/h in the opposite direction.

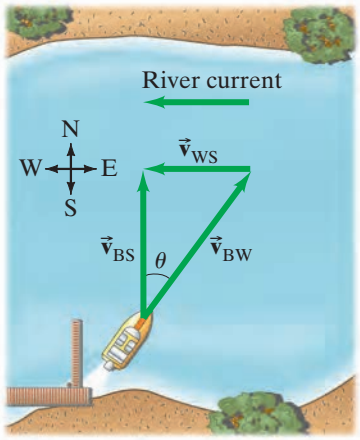


FIGURE 3–33 Example 3–13. A boat, in order to go directly across a moving river current, must head upstream.

EXAMPLE 3–13 **Heading upstream.** A boat's speed in still water is $v_{BW} = 1.85$ m/s. If the boat is to travel north directly across a river whose westward current has speed $v_{WS} = 1.20$ m/s, at what upstream angle must the boat head? See Fig. 3–33.

APPROACH If the boat heads straight across the river, the current will drag the boat downstream (westward). To overcome the river's current, the boat must have an upstream (eastward) component of velocity as well as a cross-stream (northward) component. Figure 3–33 has been drawn with $\vec{\mathbf{v}}_{BS}$, the velocity of the **B**oat relative to the **S**hore, pointing directly across the river because this is where the boat is supposed to go. Note that $\vec{\mathbf{v}}_{BS} = \vec{\mathbf{v}}_{BW} + \vec{\mathbf{v}}_{WS}$.

SOLUTION Vector $\vec{\mathbf{v}}_{BW}$ points upstream at angle θ as shown. From the diagram,

$$\sin\theta = \frac{v_{WS}}{v_{BW}} = \frac{1.20\text{ m/s}}{1.85\text{ m/s}} = 0.6486.$$

Thus $\theta = 40.4°$, so the boat must head upstream at a 40.4° angle.

†We thus can see, for example, that the equation $\vec{\mathbf{v}}_{BW} = \vec{\mathbf{v}}_{BS} + \vec{\mathbf{v}}_{WS}$ is wrong: the inner subscripts are not the same, and the outer ones on the right do not correspond to the subscripts on the left.

EXAMPLE 3–14 **Heading across the river.** The same boat ($v_{BW} = 1.85\,\text{m/s}$) now heads directly across the river whose current is still 1.20 m/s. (*a*) What is the velocity (magnitude and direction) of the boat relative to the shore? (*b*) If the river is 110 m wide, how long will it take to cross and how far downstream will the boat be then?

APPROACH The boat now heads directly across the river and is pulled downstream by the current, as shown in Fig. 3–34. The boat's velocity with respect to the shore, $\vec{\mathbf{v}}_{BS}$, is the sum of the boat's velocity with respect to the water, $\vec{\mathbf{v}}_{BW}$, plus the velocity of the water with respect to the shore, $\vec{\mathbf{v}}_{WS}$ (= the river's current). Just as before,

$$\vec{\mathbf{v}}_{BS} = \vec{\mathbf{v}}_{BW} + \vec{\mathbf{v}}_{WS}.$$

SOLUTION (*a*) Since $\vec{\mathbf{v}}_{BW}$ is perpendicular to $\vec{\mathbf{v}}_{WS}$, we can get v_{BS} using the theorem of Pythagoras:

$$v_{BS} = \sqrt{v_{BW}^2 + v_{WS}^2} = \sqrt{(1.85\,\text{m/s})^2 + (1.20\,\text{m/s})^2} = 2.21\,\text{m/s}.$$

We can obtain the angle (note how θ is defined in Fig. 3–34) from:

$$\tan\theta = v_{WS}/v_{BW} = (1.20\,\text{m/s})/(1.85\,\text{m/s}) = 0.6486.$$

A calculator with a key such as INV TAN or ARC TAN or TAN^{-1} gives $\theta = \tan^{-1}(0.6486) = 33.0°$. Note that this angle is not equal to the angle calculated in Example 3–13.

(*b*) To find the travel time for the boat to cross the river, recall the river's width $D = 110\,\text{m}$, and use the velocity component that points directly across the river, $v_{BW} = D/t$. Solving for t, we get $t = 110\,\text{m}/(1.85\,\text{m/s}) = 59.5\,\text{s}$. Also, the boat will have been carried downstream, in this time, a distance

$$d = v_{WS}t = (1.20\,\text{m/s})(59.5\,\text{s}) = 71.4\,\text{m} \approx 71\,\text{m}.$$

NOTE There is no acceleration in this Example, so the motion involves only constant velocities (of the boat or of the river).

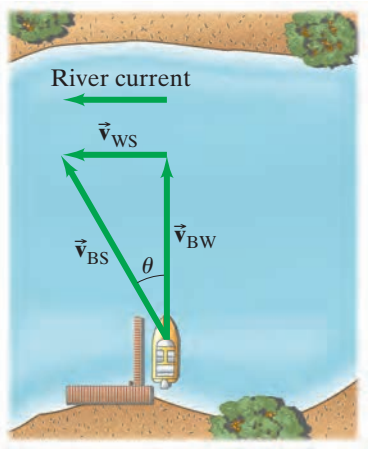


FIGURE 3–34 Example 3–14. A boat heading directly across a river whose current moves at 1.20 m/s.

EXAMPLE 3–15 **Car velocities at 90°.** Two cars approach a street corner at right angles to each other with the same speed of 40.0 km/h (= 11.11 m/s) relative to the ground, as shown in Fig. 3–35a. What is the relative velocity of car 1 as seen by car 2?

APPROACH Figure 3–35a shows the situation in a reference frame fixed to the Earth. But we want to view the situation from a reference frame in which car 2 is at rest, and this is shown in Fig. 3–35b. In this reference frame (the world as seen by the driver of car 2), the Earth moves toward car 2 with velocity $\vec{\mathbf{v}}_{E2}$ (speed of 40.0 km/h), which is of course equal and opposite to $\vec{\mathbf{v}}_{2E}$, the velocity of car 2 with respect to the Earth (Eq. 3–16):

$$\vec{\mathbf{v}}_{2E} = -\vec{\mathbf{v}}_{E2}.$$

Then the velocity of car 1 as seen by car 2 is (see Eq. 3–15)

$$\vec{\mathbf{v}}_{12} = \vec{\mathbf{v}}_{1E} + \vec{\mathbf{v}}_{E2}.$$

SOLUTION Because $\vec{\mathbf{v}}_{E2} = -\vec{\mathbf{v}}_{2E}$, then

$$\vec{\mathbf{v}}_{12} = \vec{\mathbf{v}}_{1E} - \vec{\mathbf{v}}_{2E}.$$

That is, the velocity of car 1 as seen by car 2 is the difference of their velocities, $\vec{\mathbf{v}}_{1E} - \vec{\mathbf{v}}_{2E}$, both measured relative to the Earth; see Fig. 3–35c. Since the magnitudes of $\vec{\mathbf{v}}_{1E}$, $\vec{\mathbf{v}}_{2E}$, and $\vec{\mathbf{v}}_{E2}$ are equal (40.0 km/h = 11.11 m/s), we see (Fig. 3–35b) that $\vec{\mathbf{v}}_{12}$ points at a 45° angle toward car 2; the speed is

$$v_{12} = \sqrt{(11.11\,\text{m/s})^2 + (11.11\,\text{m/s})^2} = 15.7\,\text{m/s}\ (= 56.6\,\text{km/h}).$$

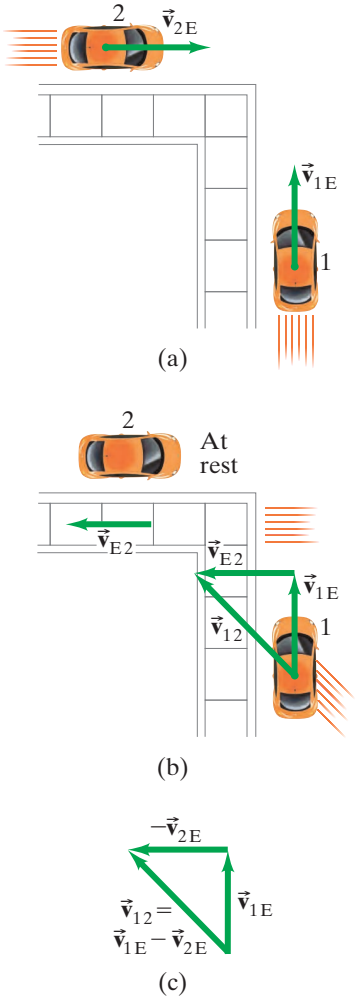


FIGURE 3–35 Example 3–15.

Summary

A quantity such as velocity, that has both a magnitude and a direction, is called a **vector**. A quantity such as mass, that has only a magnitude, is called a **scalar**. On diagrams, vectors are represented by arrows.

Addition of vectors can be done graphically by placing the tail of each successive arrow at the tip of the previous one. The sum, or **resultant vector**, is the arrow drawn from the tail of the first vector to the tip of the last vector. Two vectors can also be added using the *parallelogram* method.

Vectors can be added more accurately by adding their **components** along chosen axes with the aid of trigonometric functions. A vector of magnitude V making an angle θ with the $+x$ axis has components

$$V_x = V\cos\theta, \qquad V_y = V\sin\theta. \tag{3–2}$$

Given the components, we can find a vector's magnitude and direction from

$$V = \sqrt{V_x^2 + V_y^2}, \qquad \tan\theta = \frac{V_y}{V_x}. \tag{3–3}$$

It is often helpful to express a vector in terms of its components along chosen axes using **unit vectors**, which are vectors of unit length along the chosen coordinate axes; for Cartesian coordinates the unit vectors along the x, y, and z axes are called $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$.

The general definitions for the **instantaneous velocity**, $\vec{\mathbf{v}}$, and **acceleration**, $\vec{\mathbf{a}}$, of a particle (in one, two, or three dimensions) are

$$\vec{\mathbf{v}} = \frac{d\vec{\mathbf{r}}}{dt} \tag{3–8}$$

$$\vec{\mathbf{a}} = \frac{d\vec{\mathbf{v}}}{dt}, \tag{3–11}$$

where $\vec{\mathbf{r}}$ is the position vector of the particle. The kinematic equations for motion with constant acceleration can be written for each of the x, y, and z components of the motion and have the same form as for one-dimensional motion (Eqs. 2–12). Or they can be written in the more general vector form:

$$\begin{aligned}\vec{\mathbf{v}} &= \vec{\mathbf{v}}_0 + \vec{\mathbf{a}}t \\ \vec{\mathbf{r}} &= \vec{\mathbf{r}}_0 + \vec{\mathbf{v}}_0 t + \tfrac{1}{2}\vec{\mathbf{a}}t^2.\end{aligned} \tag{3–13}$$

Projectile motion is the motion of an object in the air near the Earth's surface under the effect of gravity alone. It can be analyzed as two separate motions if air resistance can be ignored. The horizontal component of motion is at constant velocity, whereas the vertical component is at constant acceleration, $\vec{\mathbf{g}}$, just as for an object falling vertically under the action of gravity.

The velocity of an object relative to one frame of reference can be found by vector addition if its velocity relative to a second frame of reference, and the **relative velocity** of the two reference frames, are known.

Questions

1. One car travels due east at 40 km/h, and a second car travels north at 40 km/h. Are their velocities equal? Explain.
2. Can you conclude that a car is not accelerating if its speedometer indicates a steady 60 km/h? Explain.
3. Give several examples of an object's motion in which a great distance is traveled but the displacement is zero.
4. Can the displacement vector for a particle moving in two dimensions be longer than the length of path traveled by the particle over the same time interval? Can it be less? Discuss.
5. During baseball practice, a player hits a very high fly ball and then runs in a straight line and catches it. Which had the greater displacement, the player or the ball? Explain.
6. If $\vec{\mathbf{V}} = \vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2$, is V necessarily greater than V_1 and/or V_2? Discuss.
7. Two vectors have lengths $V_1 = 4.5$ km and $V_2 = 5.0$ km. What are the maximum and minimum magnitudes of their vector sum?
8. Can two vectors, of unequal magnitude, add up to give the zero vector? Can *three* unequal vectors? Under what conditions?
9. Can the magnitude of a vector ever (*a*) equal, or (*b*) be less than, one of its components?
10. Does the odometer of a car measure a scalar or a vector quantity? What about the speedometer?
11. How could you determine the speed a slingshot imparts to a rock, using only a meter stick, a rock, and the slingshot?
12. In archery, should the arrow be aimed directly at the target? How should your angle of aim depend on the distance to the target?
13. Where in Fig. 3–22 is (*a*) $\vec{\mathbf{v}} = 0$, (*b*) $v_y = 0$, (*c*) $v_x = 0$?
14. It was reported in World War I that a pilot flying at an altitude of 2 km caught in his bare hands a bullet fired at the plane! Using the fact that a bullet slows down considerably due to air resistance, explain how this incident occurred.
15. You are on the street trying to hit a friend in his dorm window with a water balloon. He has a similar idea and is aiming at you with *his* water balloon. You aim straight at each other and throw at the same instant. Do the water balloons hit each other? Explain why or why not.
16. A projectile is launched at an upward angle of 30° to the horizontal with a speed of 30 m/s. How does the horizontal component of its velocity 1.0 s after launch compare with its horizontal component of velocity 2.0 s after launch, ignoring air resistance? Explain.
17. A projectile has the least speed at what point in its path?
18. Two cannonballs, A and B, are fired from the ground with identical initial speeds, but with θ_A larger than θ_B. (*a*) Which cannonball reaches a higher elevation? (*b*) Which stays longer in the air? (*c*) Which travels farther? Explain.
19. A person sitting in an enclosed train car, moving at constant velocity, throws a ball straight up into the air in her reference frame. (*a*) Where does the ball land? What is your answer if the car (*b*) accelerates, (*c*) decelerates, (*d*) rounds a curve, (*e*) moves with constant velocity but is open to the air?
20. If you are riding on a train that speeds past another train moving in the same direction on an adjacent track, it appears that the other train is moving backward. Why?
21. Two rowers, who can row at the same speed in still water, set off across a river at the same time. One heads straight across and is pulled downstream somewhat by the current. The other one heads upstream at an angle so as to arrive at a point opposite the starting point. Which rower reaches the opposite side first? Explain.
22. If you stand motionless under an umbrella in a rainstorm where the drops fall vertically, you remain relatively dry. However, if you start running, the rain begins to hit your legs even if they remain under the umbrella. Why?

MisConceptual Questions

1. You are adding vectors of length 20 and 40 units. Which of the following choices is a possible resultant magnitude?
(*a*) 0. (*b*) 18. (*c*) 37. (*d*) 64. (*e*) 100.

2. The magnitude of a component of a vector must be
(*a*) less than or equal to the magnitude of the vector.
(*b*) equal to the magnitude of the vector.
(*c*) greater than or equal to the magnitude of the vector.
(*d*) less than, equal to, or greater than the magnitude of the vector.

3. You are in the middle of a large field. You walk in a straight line for 100 m, then turn left and walk 100 m more in a straight line before stopping. When you stop, you are 100 m from your starting point. By how many degrees did you turn?
(*a*) 90°. (*b*) 120°. (*c*) 30°. (*d*) 180°.
(*e*) This is impossible. You cannot walk 200 m and be only 100 m away from where you started.

4. Which of the following equations correctly expresses the relation between vectors $\vec{\mathbf{A}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{C}}$, shown in Fig. 3–36?
(*a*) $\vec{\mathbf{A}} = \vec{\mathbf{B}} + \vec{\mathbf{C}}$.
(*b*) $\vec{\mathbf{B}} = \vec{\mathbf{A}} + \vec{\mathbf{C}}$.
(*c*) $\vec{\mathbf{C}} = \vec{\mathbf{A}} + \vec{\mathbf{B}}$.
(*d*) $\vec{\mathbf{A}} + \vec{\mathbf{B}} + \vec{\mathbf{C}} = 0$.

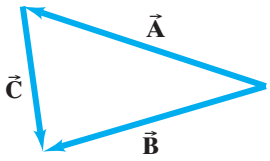
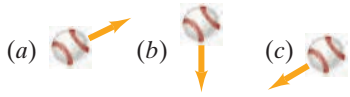


FIGURE 3–36
MisConceptual Question 4.

5. A car is driven at a constant speed of 10.0 m/s around a circle of radius 20.0 m. As it goes $\frac{1}{4}$ of the way around,
(*a*) the magnitude of the average velocity is 0.
(*b*) the magnitude of the average velocity is 10 m/s.
(*c*) the magnitude of the average velocity is between 0 and 10 m/s.
(*d*) the magnitude of the average velocity is greater than 10 m/s.

6. A bullet fired horizontally from a rifle begins to fall
(*a*) as soon as it leaves the barrel.
(*b*) after air friction reduces its speed.
(*c*) not at all if air resistance is ignored.

7. A baseball player hits a ball that soars high into the air. After the ball has left the bat, and while it is traveling upward (at point P in Fig. 3–37), what is the direction of acceleration? Ignore air resistance.



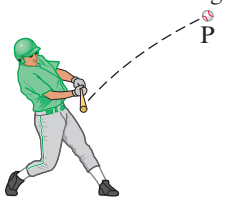


FIGURE 3–37
MisConceptual Question 7.

8. One ball is dropped vertically from a window. At the same instant, a second ball is thrown horizontally from the same window. Which ball has the greater speed at ground level?
(*a*) The dropped ball.
(*b*) The thrown ball.
(*c*) Neither—they both have the same speed on impact.
(*d*) It depends on how hard the ball was thrown.

9. Two balls having different speeds roll off the edge of a horizontal table at the same time. Which hits the floor sooner?
(*a*) The faster one.
(*b*) The slower one.
(*c*) Both the same.

10. You are riding in an enclosed train car moving at 90 km/h. If you throw a baseball straight up, where will the baseball land?
(*a*) In front of you.
(*b*) Behind you.
(*c*) In your hand.
(*d*) Can't decide from the given information.

11. Which of the three kicks in Fig. 3–38 is in the air for the longest time? They all reach the same maximum height *h*. Ignore air resistance.
(*a*), (*b*), (*c*), or
(*d*) all the same time.

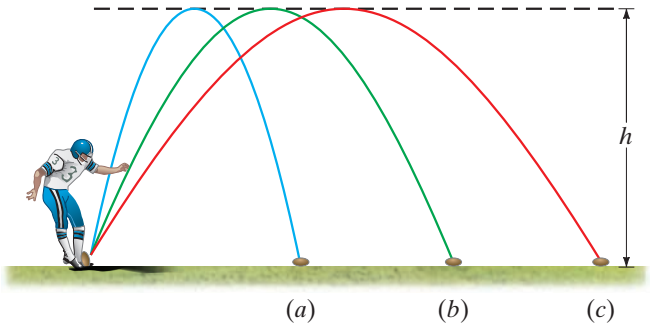


FIGURE 3–38 MisConceptual Question 11.

12. A baseball is hit high and far. Which of the following statements is true? At the highest point,
(*a*) the magnitude of the acceleration is zero.
(*b*) the magnitude of the velocity is zero.
(*c*) the magnitude of the velocity is the slowest.
(*d*) more than one of the above is true.
(*e*) none of the above are true.

13. A hunter is aiming horizontally at a monkey who is sitting in a tree. The monkey is so terrified when it sees the gun that it falls off the tree. At that very instant, the hunter pulls the trigger. What will happen?
(*a*) The bullet will miss the monkey because the monkey falls down while the bullet speeds straight forward.
(*b*) The bullet will hit the monkey because both the monkey and the bullet are falling downward at the same rate due to gravity.
(*c*) The bullet will miss the monkey because although both the monkey and the bullet are falling downward due to gravity, the monkey is falling faster.
(*d*) It depends on how far the hunter is from the monkey.

14. Which statements are *not* valid for a projectile? Take up as positive and ignore air resistance.
(*a*) The projectile has the same *x* velocity at any point on its path.
(*b*) The acceleration of the projectile is positive and decreasing when the projectile is moving upwards, zero at the top, and increasingly negative as the projectile descends.
(*c*) The acceleration of the projectile has a constant negative value.
(*d*) The *y* component of the velocity of the projectile is zero at the highest point of the projectile's path.
(*e*) The velocity at the highest point is zero.

15. A car travels 10 m/s east. Another car travels 10 m/s north. The relative speed of the first car with respect to the second is
(*a*) less than 20 m/s.
(*b*) exactly 20 m/s.
(*c*) more than 20 m/s.

Problems

3–2 to 3–5 Vector Addition; Unit Vectors

1. (I) A car is driven 245 km west and then 118 km southwest (45.0°). What is the displacement of the car from the point of origin (magnitude and direction)? Draw a diagram.
2. (I) A delivery truck travels 21 blocks north, 16 blocks east, and 26 blocks south. What is its final displacement from the origin? Assume the blocks are equal length.
3. (I) If $V_x = 9.40$ units and $V_y = -6.80$ units, determine the magnitude and direction of $\vec{\mathbf{V}}$.
4. (II) Graphically determine the resultant of the following three vector displacements: (1) 24 m, 36° north of east; (2) 18 m, 37° east of north; and (3) 26 m, 33° west of south.
5. (II) $\vec{\mathbf{V}}$ is a vector 21.8 units in magnitude and points at an angle of 23.4° above the negative x axis. (*a*) Sketch this vector. (*b*) Calculate V_x and V_y. (*c*) Use V_x and V_y to obtain (again) the magnitude and direction of $\vec{\mathbf{V}}$. [*Note*: Part (*c*) is a good way to check if you've resolved your vector correctly.]
6. (II) Vector $\vec{\mathbf{V}}_1$ is 6.2 units long and points along the negative x axis. Vector $\vec{\mathbf{V}}_2$ is 8.1 units long and points at +55° to the positive x axis. (*a*) What are the x and y components of each vector? (*b*) Determine the sum $\vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2$ (magnitude and angle).
7. (II) Figure 3–39 shows two vectors, $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$, whose magnitudes are $A = 6.8$ units and $B = 5.5$ units. Determine $\vec{\mathbf{C}}$ if (*a*) $\vec{\mathbf{C}} = \vec{\mathbf{A}} + \vec{\mathbf{B}}$, (*b*) $\vec{\mathbf{C}} = \vec{\mathbf{A}} - \vec{\mathbf{B}}$, (*c*) $\vec{\mathbf{C}} = \vec{\mathbf{B}} - \vec{\mathbf{A}}$. Give the magnitude and direction for each.

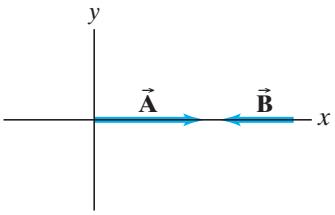


FIGURE 3–39 Problem 7.

8. (II) An airplane is traveling 815 km/h in a direction 41.5° west of north (Fig. 3–40). (*a*) Find the components of the velocity vector in the northerly and westerly directions. (*b*) How far north and how far west has the plane traveled after 1.75 h?

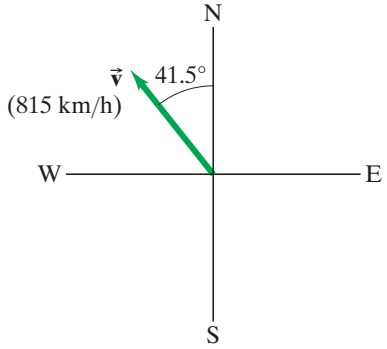


FIGURE 3–40 Problem 8.

9. (II) The summit of a mountain, 2450 m above base camp, is measured on a map to be 4580 m horizontally from the camp in a direction 38.4° west of north. What are the components of the displacement vector from camp to summit? What is its magnitude? Choose the x axis east, y axis north, and z axis up.
10. (II) Three vectors are shown in Fig. 3–41. Their magnitudes are given in arbitrary units. Determine the sum of the three vectors. Give the resultant in terms of (*a*) components, (*b*) magnitude and angle with the $+x$ axis.

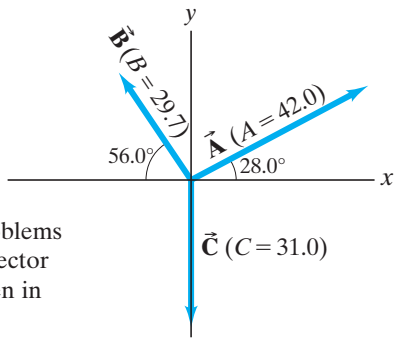


FIGURE 3–41 Problems 10, 11, 12, and 13. Vector magnitudes are given in arbitrary units.

11. (II) (*a*) Given the vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ shown in Fig. 3–41, determine $\vec{\mathbf{B}} - \vec{\mathbf{A}}$. (*b*) Determine $\vec{\mathbf{A}} - \vec{\mathbf{B}}$ without using your answer in (*a*). (*c*) Compare your results and see if they are opposite.
12. (II) Determine the vector $\vec{\mathbf{A}} - \vec{\mathbf{C}}$, given the vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{C}}$ in Fig. 3–41.
13. (II) For the vectors shown in Fig. 3–41, determine (*a*) $\vec{\mathbf{B}} - 3\vec{\mathbf{A}}$, (*b*) $2\vec{\mathbf{A}} - 3\vec{\mathbf{B}} + 2\vec{\mathbf{C}}$, and (*c*) $\vec{\mathbf{C}} - \vec{\mathbf{A}} - \vec{\mathbf{B}}$.
14. (II) Let $\vec{\mathbf{V}}_1 = -6.0\,\hat{\mathbf{i}} + 8.0\,\hat{\mathbf{j}}$ and $\vec{\mathbf{V}}_2 = -4.5\,\hat{\mathbf{i}} - 5.0\,\hat{\mathbf{j}}$. Determine the magnitude and direction of (*a*) $\vec{\mathbf{V}}_1$, (*b*) $\vec{\mathbf{V}}_2$, (*c*) $\vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2$, and (*d*) $\vec{\mathbf{V}}_1 - \vec{\mathbf{V}}_2$.
15. (II) (*a*) Determine the magnitude and direction of the sum of the three vectors $\vec{\mathbf{V}}_1 = 4.0\,\hat{\mathbf{i}} - 8.0\,\hat{\mathbf{j}}$, $\vec{\mathbf{V}}_2 = \hat{\mathbf{i}} + \hat{\mathbf{j}}$, and $\vec{\mathbf{V}}_3 = -2.0\,\hat{\mathbf{i}} + 4.0\,\hat{\mathbf{j}}$. (*b*) Determine $\vec{\mathbf{V}}_1 - \vec{\mathbf{V}}_2 + \vec{\mathbf{V}}_3$.
16. (II) Suppose a vector $\vec{\mathbf{V}}$ makes an angle ϕ with respect to the y axis. What could be the x and y components of the vector $\vec{\mathbf{V}}$?
17. (II) Two vectors, $\vec{\mathbf{V}}_1$ and $\vec{\mathbf{V}}_2$, add to a resultant $\vec{\mathbf{V}}_R = \vec{\mathbf{V}}_1 + \vec{\mathbf{V}}_2$. Describe $\vec{\mathbf{V}}_1$ and $\vec{\mathbf{V}}_2$ if (*a*) $V_R = V_1 + V_2$, (*b*) $V_R^2 = V_1^2 + V_2^2$, (*c*) $V_1 + V_2 = V_1 - V_2$.
18. (III) You are given a vector in the xy plane that has a magnitude of 95.0 units and a y component of −60.0 units. (*a*) What are the two possibilities for its x component? (*b*) Assuming the x component is known to be positive, specify the vector which, if you add it to the original one, would give a resultant vector that is 80.0 units long and points entirely in the $-x$ direction.

3–6 Vector Kinematics

19. (I) The position of a particular particle as a function of time is given by $\vec{\mathbf{r}} = (9.60t\,\hat{\mathbf{i}} + 6.45\,\hat{\mathbf{j}} - 1.50t^2\,\hat{\mathbf{k}})$ m. Determine the particle's velocity and acceleration as a function of time.
20. (I) What was the average velocity of the particle in Problem 19 between $t = 1.00$ s and $t = 3.00$ s? What is the magnitude of the instantaneous velocity at $t = 2.00$ s?
21. (II) A car is moving with speed 16.0 m/s due south at one moment and 25.7 m/s due east 8.00 s later. Over this time interval, determine the magnitude and direction of (*a*) its average velocity, (*b*) its average acceleration. (*c*) What is its average speed? [*Hint*: Can you determine all these from the information given?]

22. (II) At $t = 0$, a particle starts from rest at $x = 0$, $y = 0$, and moves in the xy plane with an acceleration $\vec{\mathbf{a}} = (4.0\,\hat{\mathbf{i}} + 3.0\,\hat{\mathbf{j}})\,\text{m/s}^2$. Determine (*a*) the x and y components of velocity, (*b*) the speed of the particle, and (*c*) the position of the particle, all as a function of time. (*d*) Evaluate all the above at $t = 2.0\,\text{s}$.

23. (II) (*a*) A skier is accelerating down a 30.0° hill at $1.80\,\text{m/s}^2$ (Fig. 3–42). What is the vertical component of her acceleration? (*b*) How long will it take her to reach the bottom of the hill, assuming she starts from rest and accelerates uniformly, if the elevation change is 125 m?

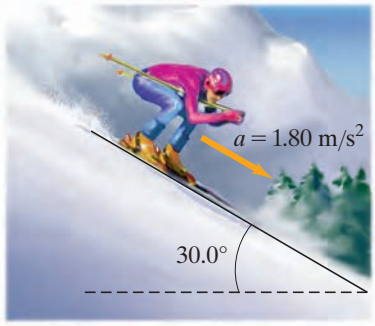


FIGURE 3–42 Problem 23.

24. (II) A hiker follows a winding trail for 5.5 hours while climbing a mountain. The distance along the trail is 11.5 km and the summit is 850 m above and 8.0 km due north of the starting point. What are the average speed and the magnitude and direction of the average velocity vector?

25. (II) An ant walks on a piece of graph paper straight along the x axis a distance of 10.0 cm in 2.40 s. It then turns left 40.0° and walks in a straight line another 10.0 cm in 1.80 s. Finally, it turns another 70.0° to the left and walks another 10.0 cm in 1.55 s. Determine (*a*) the x and y components of the ant's average velocity, and (*b*) its magnitude and direction.

26. (II) Suppose the position of an object is given by $\vec{\mathbf{r}} = (3.0t^2\,\hat{\mathbf{i}} - 6.0t^3\,\hat{\mathbf{j}})\,\text{m}$. (*a*) Determine its velocity $\vec{\mathbf{v}}$ and acceleration $\vec{\mathbf{a}}$, as a function of time. (*b*) Determine $\vec{\mathbf{r}}$ and $\vec{\mathbf{v}}$ at time $t = 3.5\,\text{s}$.

27. (II) A particle's position as a function of time t is given by $\vec{\mathbf{r}} = (5.0t + 6.0t^2)\,\text{m}\,\hat{\mathbf{i}} + (7.0 - 3.0t^3)\,\text{m}\,\hat{\mathbf{j}}$. At $t = 5.0\,\text{s}$, find the magnitude and direction of the particle's displacement vector $\Delta\vec{\mathbf{r}}$ relative to the point $\vec{\mathbf{r}}_0 = (2.0\,\hat{\mathbf{i}} + 7.0\,\hat{\mathbf{j}})\,\text{m}$.

28. (II) On mountainous downhill roads, an escape lane is sometimes placed to the side of the road for trucks whose brakes might fail. Assuming a constant upward slope of 26°, calculate the horizontal and vertical components of the acceleration of a truck that slowed from 110 km/h to rest in 7.0 s. See Fig. 3–43.

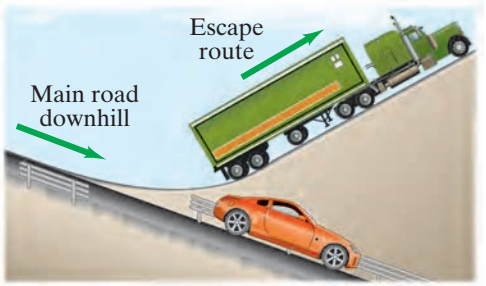


FIGURE 3–43 Problem 28.

29. (II) A light plane is headed due south with a speed relative to still air of 185 km/h. After 1.25 h, the pilot notices that they have covered only 135 km and their direction is not south but 15.0° east of south. What is the wind velocity?

30. (III) An object, which is at the origin at time $t = 0$, has initial velocity $\vec{\mathbf{v}}_0 = (-14.0\,\hat{\mathbf{i}} - 7.0\,\hat{\mathbf{j}})\,\text{m/s}$ and constant acceleration $\vec{\mathbf{a}} = (6.0\,\hat{\mathbf{i}} + 3.0\,\hat{\mathbf{j}})\,\text{m/s}^2$. Find the position $\vec{\mathbf{r}}$ where the object comes to rest (momentarily).

31. (III) A particle starts from the origin at $t = 0$ with an initial velocity of 5.0 m/s along the positive x axis. If the acceleration is $(-3.0\,\hat{\mathbf{i}} + 4.5\,\hat{\mathbf{j}})\,\text{m/s}^2$, determine the velocity and position of the particle at the moment it reaches its maximum x coordinate.

3–7 and 3–8 Projectile Motion (neglect air resistance)

32. (I) A tiger leaps horizontally from a 7.5-m-high rock with a speed of 3.0 m/s. How far from the base of the rock will she land?

33. (I) A diver running 2.5 m/s dives out horizontally from the edge of a vertical cliff and 3.5 s later reaches the water below. How high was the cliff and how far from its base did the diver hit the water?

34. (II) A ball is thrown horizontally from the roof of a building 7.5 m tall and lands 9.5 m from the base. What was the ball's initial speed?

35. (II) A ball thrown horizontally at 10.8 m/s from the roof of a building lands 21.0 m from the base of the building. How high is the building?

36. (II) A football is kicked at ground level with a speed of 18.0 m/s at an angle of 31.0° to the horizontal. How much later does it hit the ground?

37. (II) A fire hose held near the ground shoots water at a speed of 6.5 m/s. At what angle(s) should the nozzle point in order that the water land 2.5 m away (Fig. 3–44)? Why are there two different angles? Sketch the two trajectories.

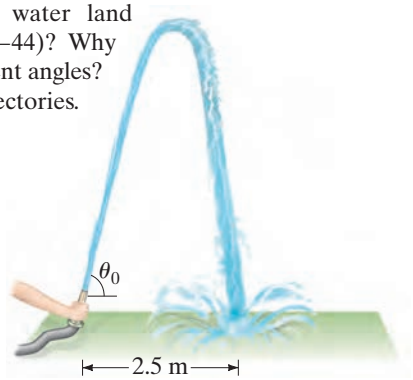


FIGURE 3–44 Problem 37.

38. (II) You buy a plastic dart gun, and being a clever physics student you decide to do a quick calculation to find its maximum horizontal range. You shoot the gun straight up, and it takes 3.4 s for the dart to land back at the barrel. What is the maximum horizontal range of your gun?

39. (II) A projectile is fired with an initial speed of 38.8 m/s at an angle of 42.2° above the horizontal on a long flat firing range. Determine (*a*) the maximum height reached by the projectile, (*b*) the total time in the air, (*c*) the total horizontal distance covered (that is, the range), and (*d*) the speed of the projectile 1.50 s after firing.

40. (II) A diver leaves the end of a diving board that is 2.0 m above the surface of the water. Her dive takes her 1.2 m above the board, and then into the water a horizontal distance of 2.2 m from the end of the board. At what speed and angle did she leave the board?

41. (II) The maximum range of a projectile is found to be 95 m. If the projectile strikes the ground a distance of 68 m away, what was the angle of launch?

42. (II) A projectile is shot from the edge of a cliff 125 m above ground level with an initial speed of 62.0 m/s at an angle of 35.0° with the horizontal, as shown in Fig. 3–45. (*a*) Determine the time taken by the projectile to hit point P at ground level. (*b*) Determine the distance X of point P from the base of the vertical cliff. At the instant just before the projectile hits point P, find (*c*) the horizontal and the vertical components of its velocity, (*d*) the magnitude of the velocity, and (*e*) the angle made by the velocity vector with the horizontal. (*f*) Find the maximum height above the cliff top reached by the projectile.

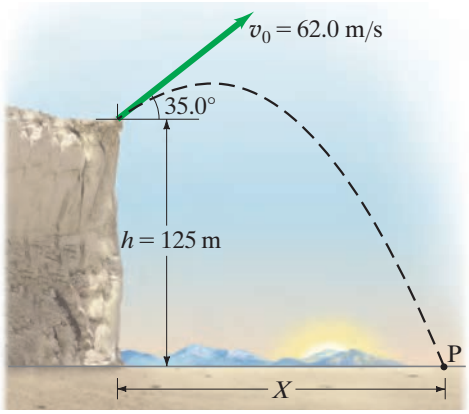


FIGURE 3–45 Problem 42.

43. (II) An athlete performing a long jump leaves the ground at a 27.0° angle and lands 7.80 m away. (*a*) What was the takeoff speed? (*b*) If this speed were increased by just 5.0%, how much longer would the jump be?

44. (II) In Example 3–11 we chose the x axis to the right and y axis up. Redo this problem by defining the x axis to the left and y axis down, and show that the conclusion remains the same—the football lands on the ground 40.5 m to the right of where it departed the punter's foot.

45. (II) A stunt driver wants to make his car jump over 8 cars parked side by side below a horizontal ramp (Fig. 3–46). (*a*) With what minimum speed must he drive off the horizontal ramp? The vertical height of the ramp is 1.5 m above the car roofs and the horizontal distance he must clear is 22 m. (*b*) If the ramp is now tilted upward, so that "takeoff angle" is 9.0° above the horizontal, what is the new minimum speed?

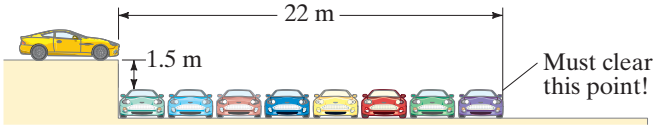


FIGURE 3–46 Problem 45.

46. (II) A baseball is hit with a speed of 27.0 m/s at an angle of 45.0°. It lands on the flat roof of a 12.5-m-tall nearby building. If the ball was hit when it was 1.2 m above the ground, what horizontal distance does it travel before it lands on the building?

47. (II) A rescue plane wants to drop supplies to isolated mountain climbers on a rocky ridge 265 m below. If the plane is traveling horizontally with a speed of 125 km/h, how far in advance of the recipients (horizontal distance) must the goods be dropped?

48. (II) Show that the time required for a projectile to reach its highest point is equal to the time for it to return to its original height if air resistance is negligible.

49. (II) Exactly 3.0 s after a projectile is fired into the air from the ground, it is observed to have a velocity $\vec{\mathbf{v}} = (7.8\,\hat{\mathbf{i}} + 5.2\,\hat{\mathbf{j}})$ m/s, where the x axis is horizontal and the y axis is positive upward. Determine (*a*) the horizontal range of the projectile, (*b*) its maximum height above the ground, and (*c*) its speed and angle of motion just before it strikes the ground.

50. (II) At what projection angle will the range of a projectile equal its maximum height?

51. (II) A ball is thrown horizontally from the top of a cliff with initial speed v_0 (at $t = 0$). At any moment, its direction of motion makes an angle θ to the horizontal (Fig. 3–47). Derive a formula for θ as a function of time, t, as the ball follows a projectile's path.

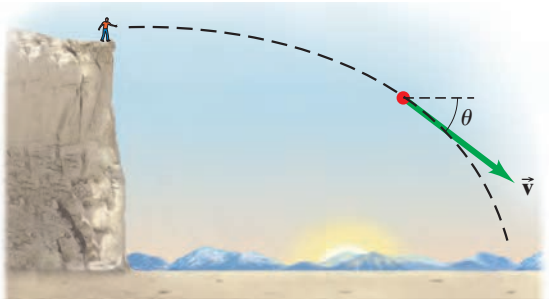


FIGURE 3–47 Problem 51.

52. (II) Romeo is throwing pebbles gently up to Juliet's window, and he wants the pebbles to hit the window with only a horizontal component of velocity. He is standing at the edge of a rose garden 8.0 m below her window and 8.5 m from the base of the wall (Fig. 3–48). How fast are the pebbles going when they hit her window?

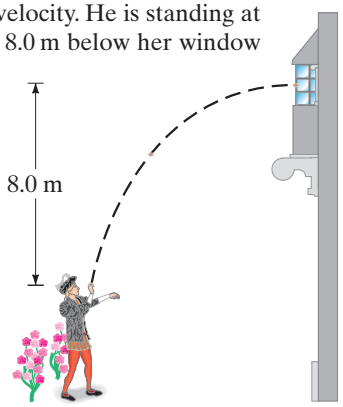


FIGURE 3–48 Problem 52.

53. (II) (*a*) A long jumper leaves the ground at 45° above the horizontal and lands 8.0 m away. What is her "takeoff" speed v_0? (*b*) Now she is out on a hike and comes to the left bank of a river. There is no bridge and the right bank is 10.0 m away horizontally and 2.5 m vertically below. If she jumps from the edge of the left bank at 45° with the speed calculated in (*a*), how long, or short, of the opposite bank will she land (Fig. 3–49)?

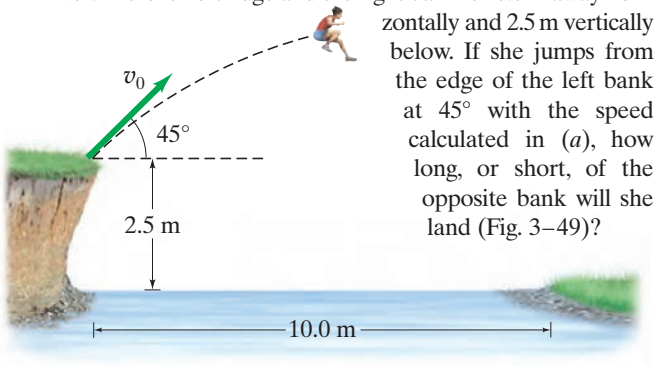


FIGURE 3–49 Problem 53.

54. (II) Show that the speed with which a projectile leaves the ground is equal to its speed just before it strikes the ground at the end of its journey, assuming the firing level equals the landing level.

55. (II) If a ball is kicked from ground level at 15.0 m/s, there are two launch angles that will make the ball land 20.0 m away. (*a*) What are the two angles? (*b*) What maximum height does the ball reach in each case? (*c*) How long is the ball in the air for each case? Ignore air resistance.

56. (II) At serve, a tennis player aims to hit the ball horizontally. What minimum speed is required for the ball to clear the 0.90-m-high net about 15.0 m from the server if the ball is "launched" from a height of 2.30 m? Where will the ball land if it just clears the net (and will it be "good" in the sense that it lands within 7.0 m of the net)? How long will it be in the air? See Fig. 3–50.

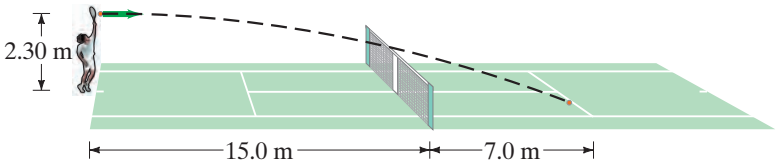


FIGURE 3–50 Problem 56.

57. (II) An Olympic long jumper is capable of jumping 8.0 m. Assuming his horizontal speed is 9.1 m/s as he leaves the ground, how long is he in the air and how high does he go? Assume that he lands standing upright—that is, the same way he left the ground.

58. (III) Revisit Example 3–9, and assume that the boy with the slingshot is *below* the boy in the tree (Fig. 3–51) and so aims *upward*, directly at the boy in the tree. Show that again the boy in the tree makes the wrong move by letting go at the moment the water balloon is shot.

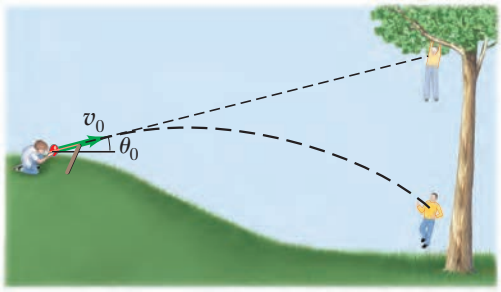
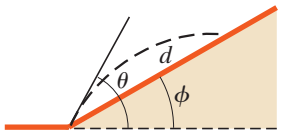


FIGURE 3–51 Problem 58.

59. (III) Suppose the kick in Example 3–8 is attempted 36.0 m from the goalposts, whose crossbar is 3.05 m above the ground. If the football is directed perfectly between the goalposts, will it pass over the bar and be a field goal? Show why or why not. If not, from what horizontal distance must this kick be made if it is to score?

60. (III) A person stands at the base of a hill that is a straight incline making an angle ϕ with the horizontal (Fig. 3–52). For a given initial speed v_0, at what angle θ (to the horizontal) should objects be thrown so that the distance d they land up the hill is as large as possible?

FIGURE 3–52 Problem 60. Given ϕ and v_0, determine θ to make d maximum.



61. (III) Derive a formula for the horizontal range R of a projectile when it lands at a height h above its initial point. (For $h < 0$, it lands a distance $-h$ below the starting point.) Assume it is projected at an angle θ_0 with initial speed v_0.

3–9 Relative Velocity

62. (I) A person going for a morning jog on the deck of a cruise ship is running toward the bow (front) of the ship at 2.5 m/s while the ship is moving ahead at 8.8 m/s. What is the velocity of the jogger relative to the water? Later, the jogger is moving toward the stern (rear) of the ship. What is the jogger's velocity relative to the water now?

63. (I) Huck Finn walks at a speed of 0.70 m/s across his raft (that is, he walks perpendicular to the raft's motion relative to the shore). The heavy raft is traveling down the Mississippi River at a speed of 1.50 m/s relative to the river bank (Fig. 3–53). What is Huck's velocity (speed and direction) relative to the river bank?

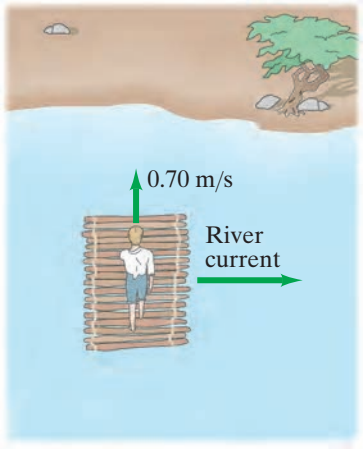


FIGURE 3–53 Problem 63.

64. (II) Determine the speed of the boat with respect to the shore in Example 3–13.

65. (II) A motorboat whose speed in still water is 4.30 m/s must aim upstream at an angle of 23.5° (with respect to a line perpendicular to the shore) in order to travel directly across the stream. (*a*) What is the speed of the current? (*b*) What is the resultant speed of the boat with respect to the shore? (See Fig. 3–33.)

66. (II) A passenger on a boat moving at 1.70 m/s on a still lake walks up a flight of stairs at a speed of 0.60 m/s, Fig. 3–54. The stairs are angled at 45° pointing in the direction of motion as shown. Write the vector velocity of the passenger relative to the water.

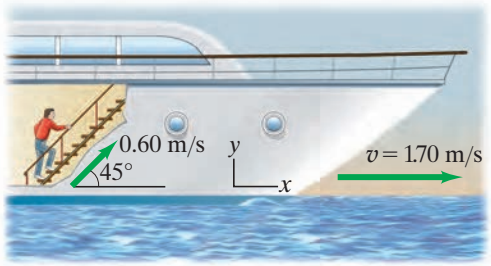


FIGURE 3–54 Problem 66.

67. (II) An airplane is heading due south at a speed of 688 km/h. If a wind begins blowing from the southwest at a speed of 85.0 km/h (average), calculate (*a*) the velocity (magnitude and direction) of the plane, relative to the ground, and (*b*) how far from its intended position it will be after 11.0 min if the pilot takes no corrective action. [*Hint*: First draw a diagram.]

68. (II) In what direction should the pilot aim the plane in Problem 67 so that it will fly due south?

69. (II) Raindrops make an angle θ with the vertical when viewed through a moving train window (Fig. 3–55). If the speed of the train is v_T, what is the speed of the raindrops in the reference frame of the Earth in which they are assumed to fall vertically?



FIGURE 3–55
Problem 69.

70. (II) A boat, whose speed in still water is 2.80 m/s, must cross a 265-m-wide river and arrive at a point 118 m upstream from where it starts (Fig. 3–56). To do so, the pilot must head the boat at a 45.0° upstream angle. What is the speed of the river's current?

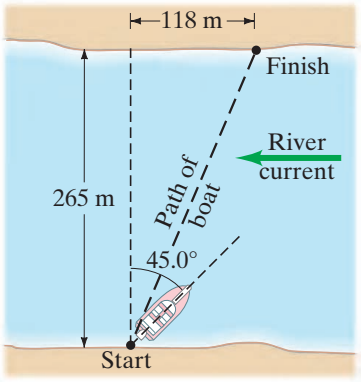


FIGURE 3–56
Problem 70.

71. (II) A swimmer is capable of swimming 0.60 m/s in still water. (*a*) If she aims her body directly across a 55-m-wide river whose current is 0.50 m/s, how far downstream (from a point opposite her starting point) will she land? (*b*) How long will it take her to reach the other side?

72. (II) (*a*) At what upstream angle must the swimmer in Problem 71 aim, if she is to arrive at a point directly across the stream? (*b*) How long will it take her?

73. (II) Two cars approach a street corner at right angles to each other (Fig. 3–57). Car 1 travels at a speed relative to Earth $v_{1E} = 35$ km/h, and car 2 at $v_{2E} = 55$ km/h. What is the relative velocity of car 1 as seen by car 2? What is the velocity of car 2 relative to car 1?

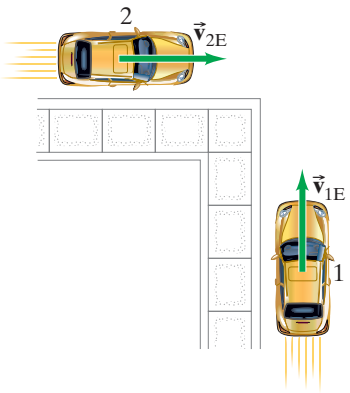


FIGURE 3–57 Problem 73.

74. (III) An airplane, whose air speed is 560 km/h, is supposed to fly in a straight path 38.0° N of E. But a steady 82 km/h wind is blowing from the north. In what direction should the plane head? [*Hint*: Use the law of sines, Appendix A–9.]

75. (III) Point B is located across the river from point A and at a 42.0° angle upstream from A. A boat travels from point A to point B, sailing at 18.0 km/h relative to the water. The current in the river flows at 5.60 km/h. The river is 1.45 km wide. (*a*) At what angle should the boat head? (*b*) What will be the boat's speed relative to the ground? (*c*) How much time does the trip from point A to point B take?

General Problems

76. A plumber steps out of his truck, walks 55 m east and 38 m south, and then takes an elevator 12 m into the subbasement of a building where a bad leak is occurring. What is the displacement of the plumber relative to his truck? Give your answer in components. Also give the magnitude and angles, with respect to the x axis, in the vertical and horizontal planes. Assume x is east, y is north, and z is up.

77. *Apollo* astronauts took a "nine iron" to the Moon and hit a golf ball about 180 m. Assuming that the swing, launch angle, and so on, were the same as on Earth where the same astronaut could hit it only 32 m, estimate the acceleration due to gravity on the surface of the Moon. (We neglect air resistance in both cases, but on the Moon there is none.)

78. A car moving at 95 km/h passes a 1.40-km-long train traveling in the same direction on a track that is parallel to the road. If the speed of the train is 75 km/h, how long does it take the car to pass the train, and how far will the car have traveled in this time? What are the results if the car and train are instead traveling in opposite directions?

79. When Babe Ruth hit a homer over the 8.0-m-high right-field fence 98 m from home plate, roughly what was the minimum speed of the *baseball* when it left the bat? Assume the ball was hit 1.0 m above the ground and its path initially made a 36° angle with the ground.

80. A child runs down a 12° hill and then suddenly jumps upward at a 15° angle above horizontal and lands 1.3 m down the hill as measured along the hill. What was the child's initial speed at the jump?

81. Here is something to try at a sporting event. Show that the maximum height h attained by an object projected into the air, such as a baseball, football, or soccer ball, is approximately given by

$$h \approx 1.2t^2 \text{ m},$$

where t is the total time of flight for the object in seconds. Assume that the object returns to the same level as that from which it was launched, as in Fig. 3–58. For example, if you count seconds and find that a baseball was in the air for $t = 5.0$ s, the maximum height attained was $h = 1.2 \times (5.0)^2 = 30$ m. The fun of this relation is that h can be determined without knowledge of the launch speed v_0 or launch angle θ_0. Why is that exactly? See Section 3–8.

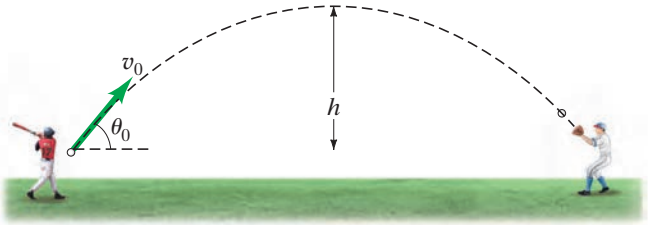


FIGURE 3–58 Problem 81.

82. The cliff divers of Acapulco push off horizontally from rock platforms about 35 m above the water, but they must clear rocky outcrops at water level that extend out into the water 5.0 m from the base of the cliff directly under their launch point. See Fig. 3–59. What minimum pushoff speed is necessary to clear the rocks? How long are they in the air?

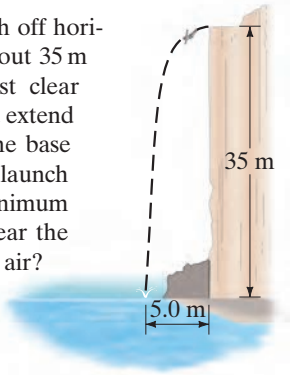


FIGURE 3–59
Problem 82.

83. A grasshopper hops along a level road. On each hop, the grasshopper launches itself at angle $\theta_0 = 45°$ and achieves a range $R = 0.80$ m. What is the average horizontal speed of the grasshopper as it hops along the road? Assume that the time spent on the ground between hops is negligible.

84. Spymaster Andrea, flying a constant 208 km/h horizontally in a low-flying helicopter, wants to drop secret documents into her contact's open car which is traveling 135 km/h on a level highway 78.0 m below. At what angle (with the horizontal) should the car be in her sights when the packet is released (Fig. 3–60)?

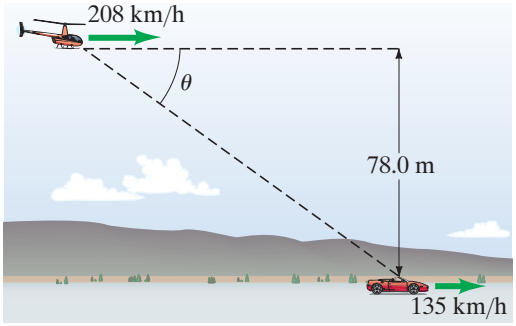


FIGURE 3–60 Problem 84.

85. A basketball leaves a player's hands at a height of 2.10 m above the floor. The basket is 3.05 m above the floor. The player likes to shoot the ball at a 38.0° angle. If the shot is made from a horizontal distance of 11.00 m and must be accurate to ± 0.22 m (horizontally), what is the range of initial speeds allowed to make the basket?

86. Extreme-sports enthusiasts have been known to jump off the top of El Capitan, a sheer granite cliff of height 910 m in Yosemite National Park. Assume a jumper runs horizontally off the top of El Capitan with speed 4.0 m/s and enjoys a free fall until she is 150 m above the valley floor, at which point she opens her parachute (Fig. 3–61). (*a*) How long is the jumper in free fall? Ignore air resistance. (*b*) It is important to be as far away from the cliff as possible before opening the parachute. How far from the cliff is this jumper when she opens her chute?

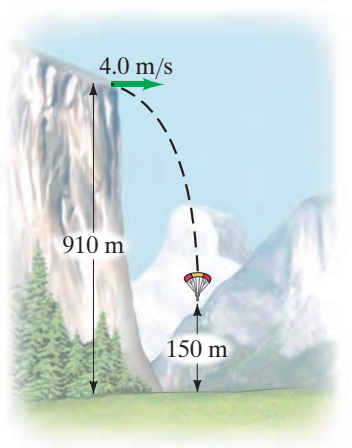


FIGURE 3–61
Problem 86.

87. If a baseball pitch leaves the pitcher's hand horizontally at a velocity of 130 km/h, by what % will the pull of gravity change the magnitude of the velocity when the ball reaches the batter, 18 m away? For this estimate, ignore air resistance and spin on the ball.

88. A projectile is launched from ground level to the top of a cliff which is 195 m away and 135 m high (see Fig. 3–62). If the projectile lands on top of the cliff 6.6 s after it is fired, find the initial velocity of the projectile (magnitude and direction). Neglect air resistance.

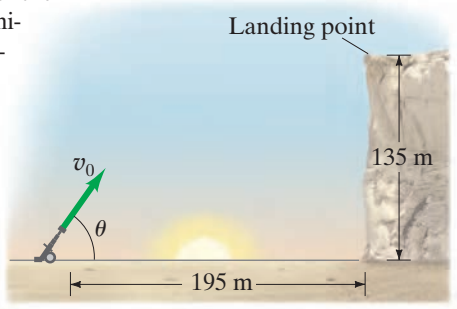


FIGURE 3–62
Problem 88.

89. A diver leaves the end of a 3.5-m-high diving board and strikes the water 1.3 s later, 3.0 m beyond the end of the board. Considering the diver as a particle, determine: (*a*) her initial velocity, $\vec{\mathbf{v}}_0$; (*b*) the maximum height reached; and (*c*) the velocity $\vec{\mathbf{v}}_f$ with which she enters the water.

90. A basketball is shot from an initial height of 2.40 m (Fig. 3–63) with an initial speed $v_0 = 12$ m/s directed at an angle $\theta_0 = 35°$ above the horizontal. (*a*) How far from the basket was the player if he made a basket? (*b*) At what angle to the horizontal did the ball enter the basket?

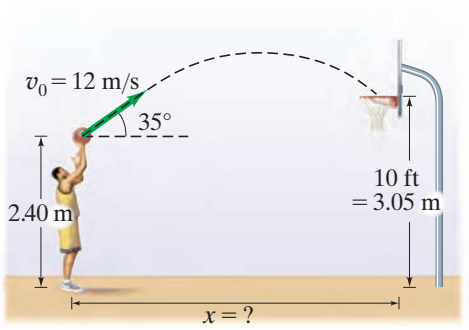


FIGURE 3–63
Problem 90.

91. A hunter aims directly at a target (on the same level) 38.0 m away. (*a*) If the arrow leaves the bow at a speed of 21.3 m/s, by how much will it miss the target? (*b*) At what angle should the bow be aimed so the target will be hit?

92. A person in the passenger basket of a hot-air balloon throws a ball horizontally outward from the basket with speed 12.0 m/s (Fig. 3–64). What initial velocity (magnitude and direction) does the ball have relative to a person standing on the ground (*a*) if the hot-air balloon is rising at 3.0 m/s relative to the ground during this throw, and (*b*) if the hot-air balloon is descending at 3.0 m/s relative to the ground?



FIGURE 3–64
Problem 92.

93. A rock is kicked horizontally at 15 m/s from a hill with a 45° slope (Fig. 3–65). How long does it take for the rock to hit the ground?

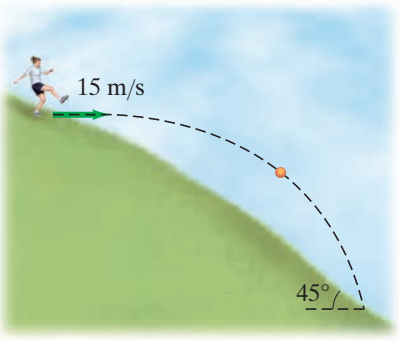


FIGURE 3–65 Problem 93.

94. A batter hits a fly ball which leaves the bat 0.90 m above the ground at an angle of 64° with an initial speed of 28 m/s heading toward centerfield. Ignore air resistance. (*a*) How far from home plate would the ball land if not caught? (*b*) The ball is caught by the centerfielder who, starting at a distance of 105 m from home plate just as the ball was hit, runs straight toward home plate at a constant speed and makes the catch at ground level. Find his speed.

95. A ball is shot from the top of a building with an initial velocity of 24 m/s at an angle $\theta = 42°$ above the horizontal. (*a*) What are the horizontal and vertical components of the initial velocity? (*b*) If a nearby building is the same height and 65 m away, how far below the top of the building will the ball strike the nearby building?

96. The speed of a boat in still water is v. The boat is to make a round trip in a river whose current travels at speed u. Derive a formula for the time needed to make a round trip of total distance D if the boat makes the round trip by moving (*a*) upstream and back downstream, and (*b*) directly across the river and back. We must assume $u < v$; why?

97. At $t = 0$ a batter hits a baseball with an initial speed of 28 m/s at a 55° angle to the horizontal. An outfielder is 85 m from the batter at $t = 0$ and, as seen from home plate, the line of sight to the outfielder makes a horizontal angle of 22° with the plane in which the ball moves (see Fig. 3–66). What speed and direction must the fielder take to catch the ball at the same height from which it was struck? Give the angle with respect to the outfielder's line of sight to home plate.

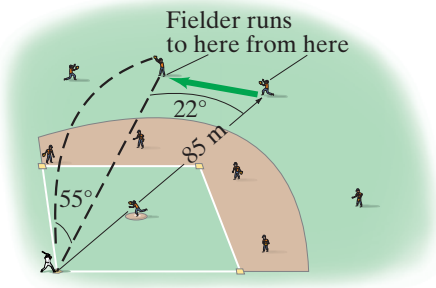


FIGURE 3–66
Problem 97.

98. A child, who is 45 m from the bank of a river, is being carried helplessly downstream by the river's swift current of 1.0 m/s. As the child passes a lifeguard on the river's bank, the lifeguard starts swimming in a straight line (Fig. 3–67) until she reaches the child at a point downstream. If the lifeguard can swim at a speed of 2.0 m/s relative to the water, how long does it take her to reach the child? How far downstream does the lifeguard intercept the child?

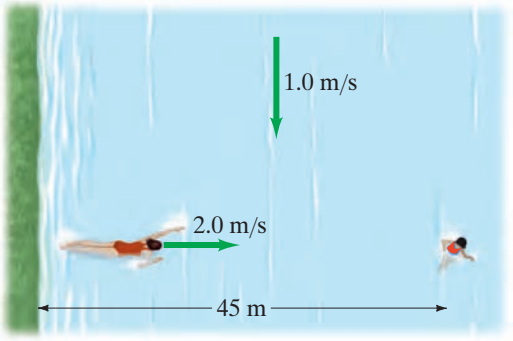


FIGURE 3–67 Problem 98.

99. A particle has a velocity of $\vec{\mathbf{v}} = (-3.0\,\hat{\mathbf{i}} + 4.5t\,\hat{\mathbf{j}})$ m/s. The particle starts at $\vec{\mathbf{r}} = (2.5\,\hat{\mathbf{i}} - 3.1\,\hat{\mathbf{j}})$ m at $t = 0$. Give the position and acceleration as a function of time. What is the shape of the resulting path?

100. In hot pursuit, Agent Logan of the FBI must get directly across a 1300-m-wide river in minimum time. The river's current is 0.80 m/s, he can row a boat at 1.60 m/s, and he can run 3.00 m/s. Describe the path he should take (rowing plus running along the shore) for the minimum crossing time, and determine the minimum time.

101. You are driving south on a highway at 12 m/s (approximately 25 mi/h) in a snowstorm. When you last stopped, you noticed that the snow was coming down vertically, but it is passing the windows of the moving car at an angle of 9.0° to the horizontal. Estimate the speed of the vertically falling snowflakes relative to the ground. [*Hint*: Construct a relative velocity diagram similar to Fig. 3–33 or 3–34. Be careful about which angle is the angle given.]

102. A boat is traveling where there is a current of 0.20 m/s east (Fig. 3–68). To avoid some offshore rocks, the boat must clear a buoy that is NNE (22.5°) and 2.8 km away. The boat's speed through still water is 2.1 m/s. If the boat wants to pass the right side of the buoy by 0.15 km, at what angle should the boat head?

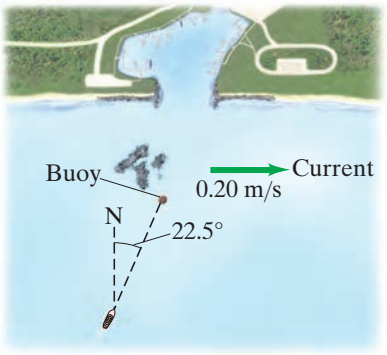
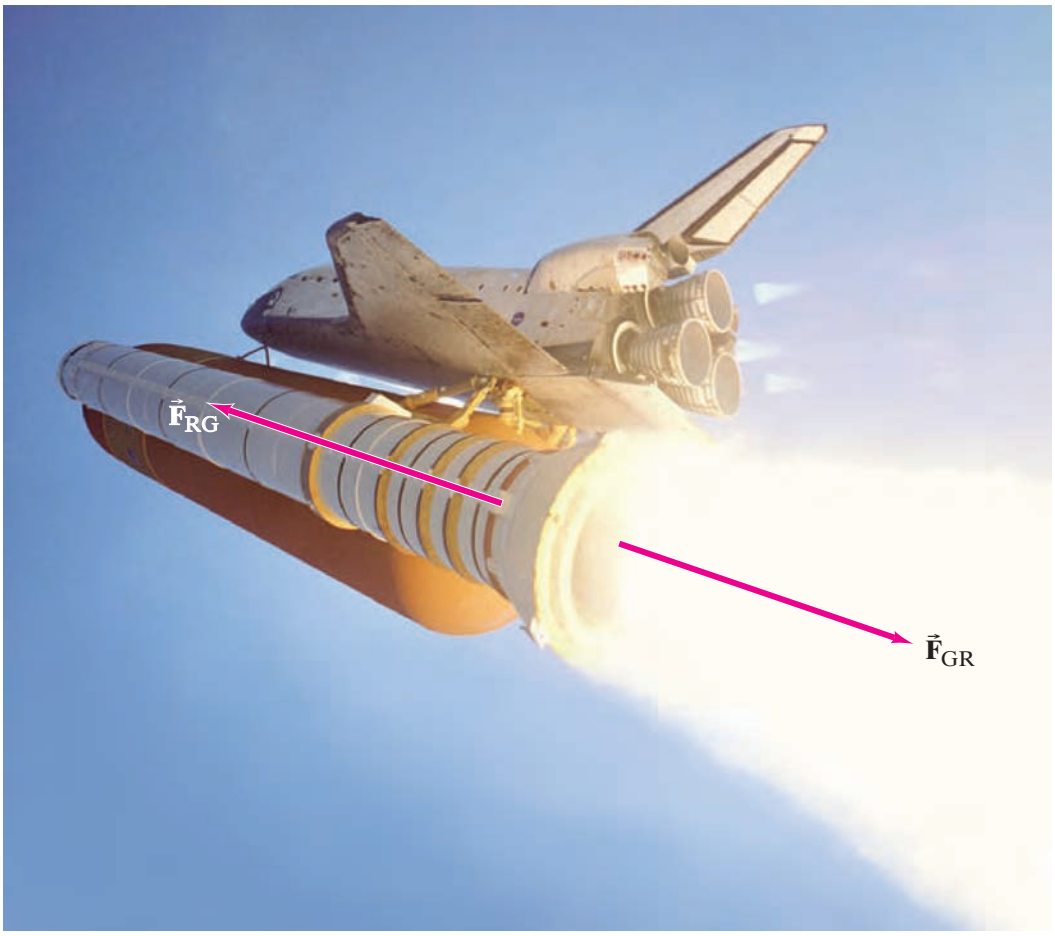


FIGURE 3–68 Problem 102.

ANSWERS TO EXERCISES

A: $3.0\sqrt{2} \approx 4.2$ units.
B: (*a*).
C: (*d*). Helicopter provides the initial velocity of the box.
D: (*a*) $v = v_{x0} = 16.0$ m/s, horizontal; (*b*) 9.80 m/s^2 down.
E: Both balls reach the same height, so are in the air for the same length of time.



A space shuttle is carried out into space by powerful rockets. They are accelerating, increasing in speed rapidly. To do so, a force must be exerted on them according to Newton's second law, $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$. What exerts this force? The **r**ocket engines exert a force on the **g**ases they push out (expel) from the rear of the rockets (labeled $\vec{\mathbf{F}}_{GR}$). According to Newton's third law, these ejected gases exert an equal and opposite force on the rockets in the forward direction. This "reaction" force exerted *on* the **r**ockets *by* the **g**ases, labeled $\vec{\mathbf{F}}_{RG}$, is the *net* force on the rockets and accelerates the rockets. (Any other forces, such as gravity, are assumed small in comparison.)

CHAPTER 4

Dynamics: Newton's Laws of Motion

CHAPTER-OPENING QUESTIONS—Guess now!

1. A 150-kg football player collides head-on with a 75-kg running back. During the collision, the heavier player exerts a force of magnitude F_A on the smaller player. If the smaller player exerts a force F_B back on the heavier player, which response is most accurate?

(a) $F_B = F_A$.
(b) $F_B < F_A$.
(c) $F_B > F_A$.
(d) $F_B = 0$.
(e) We need more information.

2. A line by the poet T. S. Eliot (from *Murder in the Cathedral*) has the women of Canterbury say "the earth presses up against our feet." What force is this?

(a) Gravity.
(b) The normal force.
(c) A friction force.
(d) Centrifugal force.
(e) No force—they are being poetic.

CONTENTS

4–1 Force
4–2 Newton's First Law of Motion
4–3 Mass
4–4 Newton's Second Law of Motion
4–5 Newton's Third Law of Motion
4–6 Weight—the Force of Gravity; and the Normal Force
4–7 Solving Problems with Newton's Laws: Free-Body Diagrams
4–8 Problem Solving—A General Approach

We have discussed how motion is *described* in terms of velocity and acceleration. Now we deal with the question of *why* objects move as they do: What makes an object at rest begin to move? What causes an object to accelerate or decelerate? What is involved when an object moves in a curved path? We can answer in each case that a force is required. In this Chapter,[†] we will investigate the connection between force and motion, which is the subject called **dynamics**.

4–1 Force

Intuitively, we experience **force** as any kind of a push or a pull on an object. When you push a stalled car or a grocery cart (Fig. 4–1), you are exerting a force on it. When a motor lifts an elevator, or a hammer hits a nail, or the wind blows the leaves of a tree, a force is being exerted. We often call these *contact forces* because the force is exerted when one object comes in contact with another object. On the other hand, we say that an object falls because of the *force of gravity* (which is not a contact force).



FIGURE 4–1 A force exerted on a grocery cart—in this case exerted by a person.

If an object is at rest, to start it moving requires force—that is, a force is needed to accelerate an object from zero velocity to a nonzero velocity. For an object already moving, if you want to change its velocity—either in direction or in magnitude—a force is required. In other words, to accelerate an object, a force is always required. In Section 4–4 we discuss the precise relation between acceleration and net force, which is Newton's second law.

One way to measure the magnitude (or strength) of a force is to use a spring scale (Fig. 4–2). Normally, such a spring scale is used to find the weight of an object; by weight we mean the force of gravity acting on the object (Section 4–6). The spring scale, once calibrated, can be used to measure other kinds of forces as well, such as the pulling force shown in Fig. 4–2.

A force exerted in a different direction has a different effect. Force has direction as well as magnitude, and is indeed a vector that follows the rules of vector addition discussed in Chapter 3. We can represent any force on a diagram by an arrow, just as we did with velocity. The direction of the arrow is the direction of the push or pull, and its length is drawn proportional to the magnitude of the force. Force vectors in this book are drawn as red in color, velocity is green.

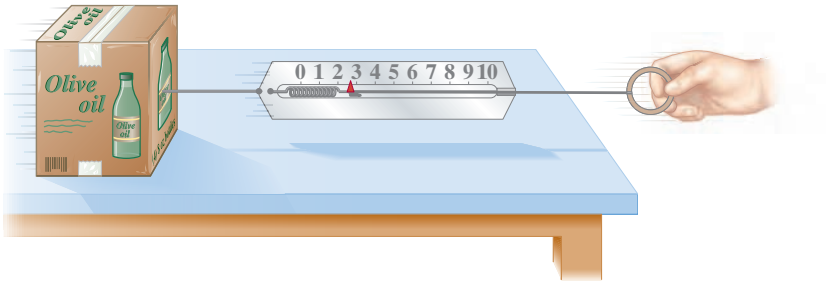


FIGURE 4–2 A spring scale used to measure a force.

4–2 Newton's First Law of Motion

What is the relationship between force and motion? The ancient Greeks, including Aristotle (384–322 B.C.), believed that a force was required to keep an object moving along a horizontal plane. To Aristotle, the natural state of an object was at rest, and a force was thought necessary to keep an object in motion. Aristotle also argued that the greater the force on the object, the greater its speed.

Some 2000 years later, in the early 1600s, Galileo disagreed. Galileo maintained that it is just as natural for an object to be in motion with a constant velocity as it is for it to be at rest. See Section 1–1, page 2.

[†]We treat everyday objects in motion here. When velocities are extremely high, close to the speed of light $(3.0 \times 10^8 \text{ m/s})$, we use the theory of relativity (Chapter 36), and in the submicroscopic world of atoms and molecules we use quantum theory (Chapter 37 *ff*).

To understand Galileo's idea, consider the following observations involving motion along a horizontal plane. To push an object with a rough surface along a tabletop at constant speed requires a certain amount of force. To push an equally heavy object with a very smooth surface across the table at the same speed will require less force. If a layer of oil or other lubricant is placed between the surface of the object and the table, then almost no force is required to keep the object moving. Notice that in each successive step, less force is required. As the next step, we imagine there is no friction at all, that the object does not rub against the table—or there is a perfect lubricant between the object and the table—and theorize that once started, the object would move across the table at constant speed with *no* force applied. A steel ball bearing rolling on a hard horizontal surface approaches this situation. So does a puck on an air table, in which a thin layer of air reduces friction almost to zero.

It was Galileo's genius to imagine such an idealized world—in this case, one where there is no friction—and to see that it could lead to a more accurate and richer understanding of the real world. This idealization led him to his remarkable conclusion that if no force is applied to a moving object, it will continue to move with constant speed in a straight line. An object slows down only if a force is exerted on it. Galileo thus interpreted friction as a force akin to ordinary pushes and pulls.

To push an object across a table at constant speed requires a force from your hand that can balance the force of friction (Fig. 4–3). When the object moves at constant speed, your pushing force is equal in magnitude to the friction force; but these two forces are in opposite directions, so the *net* force on the object (the vector sum of the two forces) is zero. This is consistent with Galileo's viewpoint: the object moves with constant velocity when no *net* force is exerted on it.

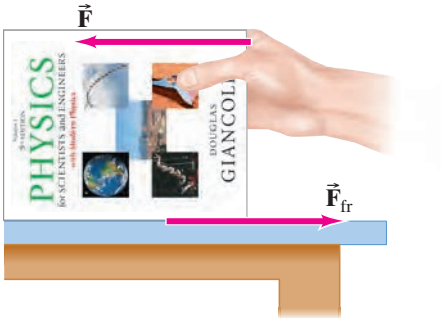


FIGURE 4–3 $\vec{\mathbf{F}}$ represents the force applied by the person and $\vec{\mathbf{F}}_{\text{fr}}$ represents the force of friction.

Upon this foundation laid by Galileo, Isaac Newton (Fig. 4–4) built his great theory of motion. Newton's analysis of motion is summarized in his famous "three laws of motion." In his great work, the *Principia* (published in 1687), Newton readily acknowledged his debt to Galileo. In fact, **Newton's first law of motion** is close to Galileo's conclusions. It states that

NEWTON'S FIRST LAW OF MOTION

> **Every object continues in its state of rest, or of uniform velocity in a straight line, as long as no net force acts on it.**

The tendency of an object to maintain its state of rest or of uniform velocity in a straight line is called **inertia**. Newton's first law is thus often called the **law of inertia**.

CONCEPTUAL EXAMPLE 4–1 **Newton's first law.** A school bus comes to a sudden stop, and all of the backpacks on the floor start to slide forward. What force causes them to do that?

RESPONSE It isn't "force" that does it. By Newton's first law, the backpacks continue their state of motion, maintaining their velocity. The backpacks slow down if a force is applied, such as friction with the floor.

FIGURE 4–4
Isaac Newton (1642–1727). Besides developing mechanics, including his three great laws of motion and the law of universal gravitation, he also tried to understand the nature of light.



Inertial Reference Frames

Newton's first law does not hold in every reference frame. For example, if your reference frame is an accelerating car, an object such as a cup resting on the dashboard may begin to move toward you (it stayed at rest as long as the car's velocity remained constant). The cup accelerated toward you, but neither you nor anything else exerted a force on it in that direction. Similarly, in the reference frame of the decelerating bus in Example 4–1, there was no force pushing the backpacks forward. In accelerating reference frames, Newton's first law does not hold. Physics is easier in reference frames in which Newton's first law *does* hold, and they are called **inertial reference frames** (the law of inertia is valid in them). For most purposes, we usually make the approximation that a reference frame fixed on the Earth is an inertial frame. This is not precisely true, due to the Earth's rotation, but usually it is close enough.

Any reference frame that moves with constant velocity (say, a car or an airplane) relative to an inertial frame is also an inertial reference frame. Reference frames where the law of inertia does *not* hold, such as the accelerating reference frames discussed above, are called **noninertial** reference frames. How can we be sure a reference frame is inertial or not? By checking to see if Newton's first law holds. Thus Newton's first law serves as the definition of inertial reference frames.

4–3 Mass

Newton's second law, which we come to in the next Section, makes use of the concept of mass. Newton used the term *mass* as a synonym for "quantity of matter." This intuitive notion of the mass of an object is not very precise because the concept "quantity of matter" is not very well defined. More precisely, we can say that **mass** is a *measure of the inertia* of an object. The more mass an object has, the greater the force needed to give it a particular acceleration. It is harder to start it moving from rest, or to stop it when it is moving, or to change its velocity sideways out of a straight-line path. A truck has much more inertia than a baseball moving at the same speed, and a much greater force is needed to change the truck's velocity at the same rate as the ball's. The truck therefore has much more mass.

To quantify the concept of mass, we must define a standard. In SI units, the unit of mass is the **kilogram** (kg) as we discussed in Chapter 1, Section 1–4.

⚠ **CAUTION**
Distinguish mass from weight

The terms *mass* and *weight* are often confused with one another, but it is important to distinguish between them. Mass is a property of an object itself (a measure of an object's inertia, or its "quantity of matter"). Weight, on the other hand, is a force, the pull of gravity acting on an object. To see the difference, suppose we take an object to the Moon. The object will weigh only about one-sixth as much as it did on Earth, since the force of gravity is weaker. But its mass will be the same. It will have the same amount of matter as on Earth, and will have just as much inertia. It will be just as hard to start it moving on the Moon as on Earth. (More on weight in Section 4–6.)

4–4 Newton's Second Law of Motion

Newton's first law states that if no net force is acting on an object at rest, the object remains at rest; or if the object is moving, it continues moving with constant speed in a straight line. But what happens if a net force *is* exerted on an object? Newton perceived that the object's velocity will change (Fig. 4–5). A net force exerted on an object may make its velocity increase. Or, if the net force is in a direction opposite to the motion, that force will reduce the object's velocity. If the net force acts sideways on a moving object, the *direction* of the object's velocity changes. That change in the *direction* of the velocity is also an acceleration. So a sideways net force on an object also causes acceleration. In general, we can say that *a net force causes acceleration*.

FIGURE 4–5 The bobsled accelerates because the team exerts a force.



What precisely is the relationship between acceleration and force? Everyday experience can suggest an answer. Consider the force required to push a cart when friction is small enough to ignore. (If there is friction, consider the *net* force, which is the force you exert minus the force of friction.) If you push the cart horizontally with a gentle but constant force for a certain period of time, you will make the cart accelerate from rest up to some speed, say 3 km/h. If you push with twice the force, the cart will reach 3 km/h in half the time. The acceleration will be twice as great. If you triple the force, the acceleration is tripled, and so on. Thus, the acceleration of an object is directly proportional to the net applied force. But the acceleration depends on the mass of the object as well. If you push an empty grocery cart with the same force as you push one that is filled with groceries, you will find that the full cart has a smaller acceleration. The greater the mass, the less the acceleration for the same net force. The mathematical relation, as Newton argued, is that the

acceleration of an object is inversely proportional to its mass. These relationships are found to hold in general and can be summarized as follows:

The acceleration of an object is directly proportional to the net force acting on it, and is inversely proportional to the object's mass. The direction of the acceleration is in the direction of the net force acting on the object.

NEWTON'S SECOND LAW OF MOTION

This is **Newton's second law of motion**.

Newton's second law can be written as an equation:

$$\vec{\mathbf{a}} = \frac{\Sigma\vec{\mathbf{F}}}{m},$$

where $\vec{\mathbf{a}}$ stands for acceleration, m for the mass, and $\Sigma\vec{\mathbf{F}}$ for the *net force* on the object. The symbol Σ (Greek "sigma") stands for "sum of"; $\vec{\mathbf{F}}$ stands for force, so $\Sigma\vec{\mathbf{F}}$ means the *vector sum of all forces* acting on the object, which we define as the **net force**.

We rearrange this equation to obtain the familiar statement of Newton's second law:

$$\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}. \qquad \textbf{(4–1a)}$$

NEWTON'S SECOND LAW OF MOTION

Newton's second law relates the description of motion (acceleration) to the cause of motion (force). It is one of the most fundamental relationships in physics. From Newton's second law we can make a more precise definition of **force** as *an action capable of accelerating an object*.

Every force $\vec{\mathbf{F}}$ is a vector, with magnitude and direction. Equation 4–1a is a vector equation valid in any inertial reference frame. It can be written in component form in rectangular coordinates as

$$\Sigma F_x = ma_x, \qquad \Sigma F_y = ma_y, \qquad \Sigma F_z = ma_z, \qquad \textbf{(4–1b)}$$

where

$$\vec{\mathbf{F}} = F_x\hat{\mathbf{i}} + F_y\hat{\mathbf{j}} + F_z\hat{\mathbf{k}}.$$

The component of acceleration in each direction is affected only by the component of the net force in that direction. (See Section 3–5 for unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$.)

If the motion is all along a line (one-dimensional), we can leave out the subscripts and simply write $\Sigma F = ma$. Again, a is the acceleration of an object of mass m, and ΣF includes all the forces acting on that object, and *only* forces acting on that object. Sometimes the net force ΣF is written as F_{net}, so $F_{\text{net}} = ma$.

In SI units, with the mass in kilograms, the unit of force is called the **newton** (N). One newton is the force required to impart an acceleration of $1\ \text{m/s}^2$ to a mass of 1 kg. Thus $1\ \text{N} = 1\ \text{kg}\cdot\text{m/s}^2$.

In cgs units, the unit of mass is the **gram** (g), which is $\frac{1}{1000}$ of a kilogram.[†] The unit of force is the **dyne**, which is defined as the net force needed to impart an acceleration of $1\ \text{cm/s}^2$ to a mass of 1 g. Thus $1\ \text{dyne} = 1\ \text{g}\cdot\text{cm/s}^2$. Because $1\ \text{g} = 10^{-3}\ \text{kg}$ and $1\ \text{cm} = 10^{-2}\ \text{m}$, then $1\ \text{dyne} = 10^{-5}\ \text{N}$.

In the British system, which we rarely use, the unit of force is the **pound** (abbreviated lb), where $1\ \text{lb} = 4.44822\ \text{N} \approx 4.45\ \text{N}$. The unit of mass is the **slug**, which is defined as that mass which will undergo an acceleration of $1\ \text{ft/s}^2$ when a force of 1 lb is applied to it. Thus $1\ \text{lb} = 1\ \text{slug}\cdot\text{ft/s}^2$. Table 4–1 summarizes the units in the different systems.

TABLE 4–1
Units for Mass and Force

System	Mass	Force
SI	kilogram (kg)	newton (N) $(= \text{kg}\cdot\text{m/s}^2)$
cgs	gram (g)	dyne $(= \text{g}\cdot\text{cm/s}^2)$
British	slug	pound (lb)

Conversion factors: $1\ \text{dyne} = 10^{-5}\ \text{N}$; $1\ \text{lb} \approx 4.45\ \text{N}$; $1\ \text{slug} \approx 14.6\ \text{kg}$.

PROBLEM SOLVING
Use a consistent set of units

It is very important that only one set of units be used in a given calculation or problem, with the SI being what we almost always use. If the force is given in, say, newtons, and the mass in grams, then before attempting to solve for the acceleration in SI units, we must change the mass to kilograms. For example, if the force is given as 2.0 N along the x axis and the mass is 500 g, we change the latter to 0.50 kg, and the acceleration will then automatically come out in m/s^2 when Newton's second law is used:

$$a_x = \frac{\Sigma F_x}{m} = \frac{2.0\ \text{N}}{0.50\ \text{kg}} = \frac{2.0\ \text{kg}\cdot\text{m/s}^2}{0.50\ \text{kg}} = 4.0\ \text{m/s}^2,$$

where we set $1\ \text{N} = 1\ \text{kg}\cdot\text{m/s}^2$.

[†]Be careful not to confuse g for gram with g for the acceleration due to gravity. The latter is always italicized (or boldface when shown as a vector).

EXAMPLE 4–2 **ESTIMATE** **Force to accelerate a fast car.** Estimate the net force needed to accelerate (*a*) a 1000-kg car at $\frac{1}{2}g$; (*b*) a 200-gram apple at the same rate.

APPROACH We use Newton's second law to find the net force needed for each object; we are given the mass and the acceleration. This is an estimate (the $\frac{1}{2}$ is not said to be precise) so we round off to one significant figure.

SOLUTION (*a*) The car's acceleration is $a = \frac{1}{2}g = \frac{1}{2}(9.8\,\text{m/s}^2) \approx 5\,\text{m/s}^2$. We use Newton's second law to get the net force needed to achieve this acceleration:

$$\Sigma F = ma \approx (1000\,\text{kg})(5\,\text{m/s}^2) = 5000\,\text{N}.$$

(If you are used to British units, to get an idea of what a 5000-N force is, you can divide by 4.45 N/lb and get a force of about 1000 lb.)

(*b*) For the apple, $m = 200\,\text{g} = 0.2\,\text{kg}$, so

$$\Sigma F = ma \approx (0.2\,\text{kg})(5\,\text{m/s}^2) = 1\,\text{N}.$$

EXAMPLE 4–3 **Force to stop a car.** What average net force is required to bring a 1500-kg car to rest from a speed of 100 km/h within a distance of 55 m?

APPROACH We use Newton's second law, $\Sigma F = ma$, to determine the force, but first we need to calculate the acceleration *a*. We assume the acceleration is constant so that we can use the kinematic equations, Eqs. 2–12, to calculate it.



FIGURE 4–6 Example 4–3.

SOLUTION We assume the motion is along the $+x$ axis (Fig. 4–6). We are given the initial velocity $v_0 = 100\,\text{km/h} = 27.8\,\text{m/s}$ (Section 1–5), the final velocity $v = 0$, and the distance traveled $x - x_0 = 55\,\text{m}$. From Eq. 2–12c, we have

$$v^2 = v_0^2 + 2a(x - x_0),$$

so

$$a = \frac{v^2 - v_0^2}{2(x - x_0)} = \frac{0 - (27.8\,\text{m/s})^2}{2(55\,\text{m})} = -7.0\,\text{m/s}^2.$$

The net force required is then

$$\Sigma F = ma = (1500\,\text{kg})(-7.0\,\text{m/s}^2) = -1.1 \times 10^4\,\text{N},$$

or 11,000 N. The force must be exerted in the direction *opposite* to the initial velocity, which is what the negative sign means.

NOTE If the acceleration is not precisely constant, then we are determining an "average" acceleration and we obtain an "average" net force.

Newton's second law, like the first law, is valid only in inertial reference frames (Section 4–2). In the noninertial reference frame of a car that begins accelerating, a cup on the dashboard starts sliding—it accelerates—even though the net force on it is zero. Thus $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ does not work in such an accelerating reference frame $(\Sigma\vec{\mathbf{F}} = 0$, but $\vec{\mathbf{a}} \neq 0$ in this noninertial frame$)$.

EXERCISE A Suppose you watch a cup slide on the (smooth) dashboard of an accelerating car as we just discussed, but this time from an inertial reference frame outside the car, on the street. From your inertial frame, Newton's laws are valid. What force pushes the cup off the dashboard?

*Precise Definition of Mass

As mentioned in Section 4–3, we can quantify the concept of mass using its definition as a measure of inertia. How to do this is evident from Eq. 4–1a, where we see that the acceleration of an object is inversely proportional to its mass.

If the same net force ΣF acts to accelerate each of two masses, m_1 and m_2, then the ratio of their masses can be defined as the inverse ratio of their accelerations:

$$\frac{m_2}{m_1} = \frac{a_1}{a_2}.$$

If one of the masses is known (it could be the standard kilogram) and the two accelerations are precisely measured, then the unknown mass is obtained from this definition. For example, if $m_1 = 1.00\,\text{kg}$, and for a particular force $a_1 = 3.00\,\text{m/s}^2$ and $a_2 = 2.00\,\text{m/s}^2$, then $m_2 = 1.50\,\text{kg}$.

4–5 Newton's Third Law of Motion

Newton's second law of motion describes quantitatively how forces affect motion. But where do forces come from? Observations suggest that a force exerted on any object is always exerted *by another object*. A horse pulls a wagon, a person pushes a grocery cart, a hammer pushes on a nail, a magnet attracts a paper clip. In each of these examples, a force is exerted *on* one object, and that force is exerted *by* another object. The force exerted *on* the nail is exerted *by* the hammer.

But Newton realized that things are not so one-sided. True, the hammer exerts a force on the nail (Fig. 4–7). But the nail evidently exerts a force back on the hammer as well, for the hammer's speed is rapidly reduced to zero upon contact. Only a strong force could cause such a rapid deceleration of the hammer. Thus, said Newton, the two objects must be treated on an equal basis. The hammer exerts a force on the nail, and the nail exerts a force back on the hammer. This is the essence of **Newton's third law of motion**:

Whenever one object exerts a force on a second object, the second object exerts an equal force in the opposite direction on the first.

This law is sometimes stated as "to every action there is an equal and opposite reaction." But to avoid confusion, it is very important to remember that the "action" force and the "reaction" force are acting on *different* objects.

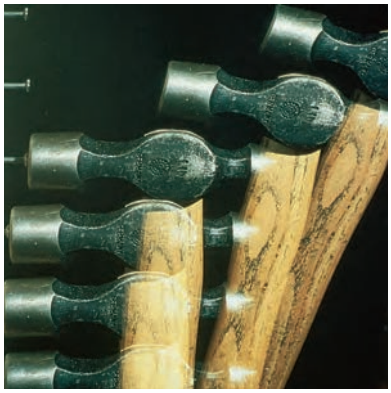


FIGURE 4–7 A hammer striking a nail. The hammer exerts a force on the nail and the nail exerts a force back on the hammer. The latter force decelerates the hammer and brings it to rest.

NEWTON'S THIRD LAW OF MOTION

CAUTION

Action and reaction forces act on different objects

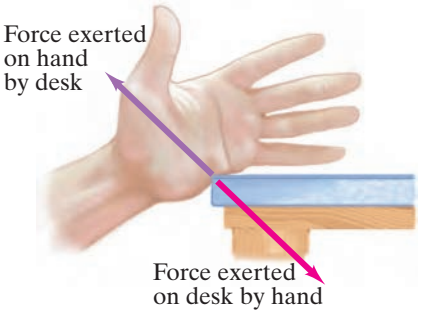


FIGURE 4–8 If your hand pushes against the edge of a desk (the force vector is shown in red), the desk pushes back against your hand (this force vector is shown in a different color, violet, to remind us that this force acts on a different object).

As evidence for the validity of Newton's third law, look at your hand when you push against the edge of a desk, Fig. 4–8. Your hand's shape is distorted, clear evidence that a force is being exerted on it. You can *see* the edge of the desk pressing into your hand. You can even *feel* the desk exerting a force on your hand; it hurts! The harder you push against the desk, the harder the desk pushes back on your hand. (You only feel forces exerted *on* you; when you exert a force on another object, what you feel is that object pushing back on you.)

The force the desk exerts on your hand has the same magnitude as the force your hand exerts on the desk. This is true not only if the desk is at rest but is true even if the desk is accelerating due to the force your hand exerts.

As another demonstration of Newton's third law, consider the ice skater in Fig. 4–9. There is very little friction between her skates and the ice, so she will move freely if a force is exerted on her. She pushes against the wall; and then *she* starts moving backward. The force she exerts on the wall cannot make *her* start moving, because that force acts on the wall. Something had to exert a force *on her* to start her moving, and that force could only have been exerted by the wall. The wall pushes on her with a force, by Newton's third law, equal and opposite to the force she exerts on the wall.

FIGURE 4–9 An example of Newton's third law: when an ice skater pushes against the wall, the wall pushes back and this force causes her to accelerate away.

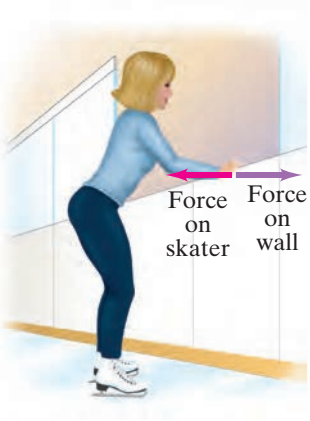
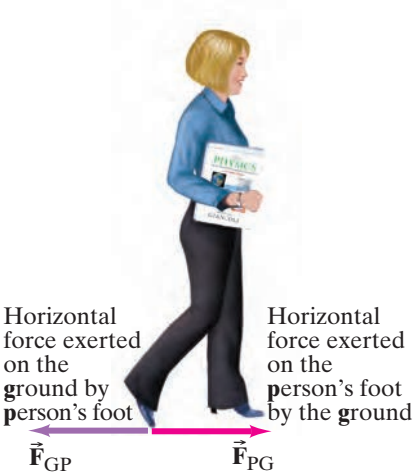




FIGURE 4–10 Another example of Newton's third law: the launch of a rocket. The rocket engine pushes the gases downward, and the gases exert an equal and opposite force upward on the rocket, accelerating it upward. (A rocket does *not* accelerate as a result of its expelled gases pushing against the ground.)

FIGURE 4–11 We can walk forward because, when one foot pushes backward against the ground, the ground pushes forward on that foot (Newton's third law). The two forces shown *act on different objects*.



⚠ CAUTION

*Distinguish **on** what object a force is exerted, and **by** what object*

NEWTON'S THIRD LAW OF MOTION

When a person throws a package out of a small boat (initially at rest), the boat starts moving in the opposite direction. The person exerts a force on the package. The package exerts an equal and opposite force back on the person, and this force propels the person (and the boat) backward slightly.

Rocket propulsion also is explained using Newton's third law (Fig. 4–10). A common misconception is that rockets accelerate because the gases rushing out the back of the engine push against the ground or the atmosphere. Not true. What happens, instead, is that a rocket exerts a strong force on the gases, expelling them; and the gases exert an equal and opposite force *on the rocket*. It is this latter force that propels the rocket forward—the force exerted *on* the rocket *by* the gases (see Chapter-Opening Photo, page 85). Thus, a space vehicle is maneuvered in empty space by firing its rockets in the direction opposite to that in which it needs to accelerate. When the rocket pushes on the gases in one direction, the gases push back on the rocket in the opposite direction. Jet aircraft too accelerate because the gases they thrust out backwards exert a forward force on the engines (Newton's third law).

Consider how we walk. A person begins walking by pushing with the foot backward against the ground. The ground then exerts an equal and opposite force forward on the person (Fig. 4–11), and it is this force, *on* the person, that moves the person forward. (If you doubt this, try walking normally where there is no friction, such as on very smooth slippery ice.) In a similar way, a bird flies forward by exerting a backward force on the air, but it is the air pushing forward (Newton's third law) on the bird's wings that propels the bird forward.

CONCEPTUAL EXAMPLE 4–4 **What exerts the force to move a car?**

What makes a car go forward?

RESPONSE A common answer is that the engine makes the car move forward. But it is not so simple. The engine makes the wheels and tires go around. But if the tires are on slick ice or wet mud, they just spin. Friction is needed. On firm ground, the tires push backward against the ground because of friction between the tires and the ground. By Newton's third law, the ground pushes on the tires in the opposite direction, accelerating the car forward.

We tend to associate forces with active objects such as humans, animals, engines, or a moving object like a hammer. It is often difficult to see how an inanimate object at rest, such as a wall or a desk (Fig. 4–8), or the wall of an ice rink (Fig. 4–9), can exert a force. The explanation is that every material, no matter how hard, is elastic (springy) at least to some degree. A stretched rubber band can exert a force on a wad of paper and accelerate it to fly across the room. Other materials may not stretch as readily as rubber, but they do stretch or compress when a force is applied to them. And just as a stretched rubber band exerts a force, so does a stretched (or compressed) wall, desk, or car fender.

From the examples discussed above, we can see how important it is to remember *on* what object a given force is exerted and *by* what object that force is exerted. A force influences the motion of an object only when it is applied *on* that object. A force exerted *by* an object does not influence that same object; it only influences the other object *on* which it is exerted. Thus, to avoid confusion, the two prepositions *on* and *by* must always be used—and used with care.

One way to keep clear which force acts on which object is to use double subscripts. For example, the force exerted *on* the **P**erson *by* the **G**round as the person walks in Fig. 4–11 can be labeled $\vec{\mathbf{F}}_{PG}$. And the force exerted on the ground by the person is $\vec{\mathbf{F}}_{GP}$. By Newton's third law

$$\vec{\mathbf{F}}_{GP} = -\vec{\mathbf{F}}_{PG}. \qquad (4\text{–}2)$$

$\vec{\mathbf{F}}_{GP}$ and $\vec{\mathbf{F}}_{PG}$ have the same magnitude (Newton's third law), and the minus sign reminds us that these two forces are in opposite directions.

Note carefully that the two forces shown in Fig. 4–11 act on different objects—to emphasize this we used slightly different colors for the vector arrows representing these forces. These two forces would never appear together in a sum of forces in Newton's second law, $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$. Why not? Because they act on different objects: $\vec{\mathbf{a}}$ is the acceleration of one particular object, and $\Sigma\vec{\mathbf{F}}$ must include *only* the forces on that *one* object.

⚠ CAUTION

The 2 forces in Newton's third law act on different bodies. Only one can be included in $\Sigma F = ma$ for an object

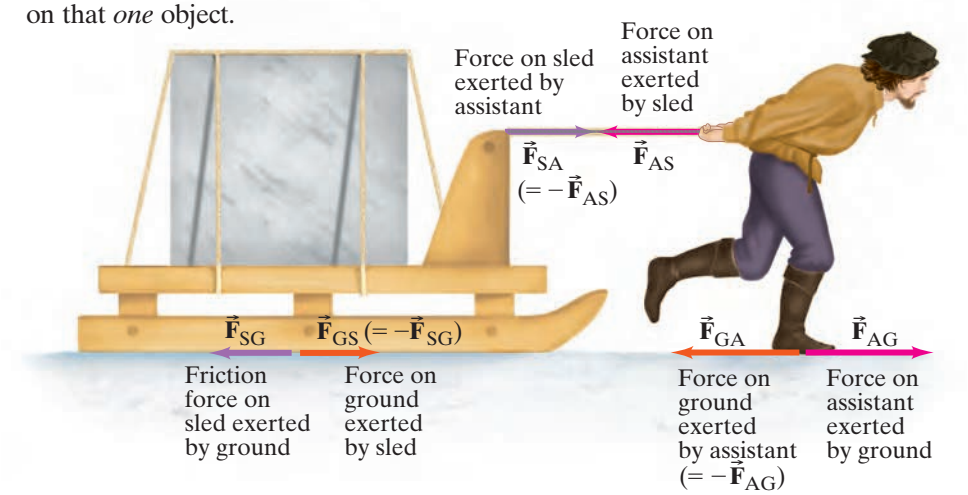


FIGURE 4–12 Example 4–5, showing only horizontal forces. Michelangelo has selected a fine block of marble for his next sculpture. Shown here is his assistant pulling it on a sled away from the quarry. Forces on the assistant are shown as red (magenta) arrows. Forces on the sled are purple arrows. Forces acting on the ground are orange arrows. Action–reaction forces that are equal and opposite are labeled by the same subscripts but reversed (such as $\vec{\mathbf{F}}_{GA}$ and $\vec{\mathbf{F}}_{AG}$) and are of different colors because they act on different objects.

CONCEPTUAL EXAMPLE 4–5 **Third law clarification.** Michelangelo's assistant has been assigned the task of moving a block of marble using a sled (Fig. 4–12). He says to his boss, "When I exert a forward force on the sled, the sled exerts an equal and opposite force backward. So how can I ever start it moving? No matter how hard I pull, the backward reaction force always equals my forward force, so the net force must be zero. I'll never be able to move this load." Is he correct?

RESPONSE No. Although it is true that the action and reaction forces are equal in magnitude, the assistant has forgotten that they are exerted on different objects. The forward ("action") force is exerted by the assistant on the sled (Fig. 4–12), whereas the backward "reaction" force is exerted by the sled on the assistant. To determine if the *assistant* moves or not, we must consider only the forces *on the assistant* and then apply $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$, where $\Sigma\vec{\mathbf{F}}$ is the net force *on the assistant*, $\vec{\mathbf{a}}$ is the acceleration of the assistant, and m is the assistant's mass. There are two forces on the assistant that affect his forward motion; they are shown as bright red (magenta) arrows in Figs. 4–12 and 4–13: they are (1) the horizontal force $\vec{\mathbf{F}}_{AG}$ exerted on the assistant by the ground (the harder he pushes backward against the ground, the harder the ground pushes forward on him—Newton's third law), and (2) the force $\vec{\mathbf{F}}_{AS}$ exerted on the assistant by the sled, pulling backward on him; see Fig. 4–13. If he pushes hard enough on the ground, the force on him exerted by the ground, $\vec{\mathbf{F}}_{AG}$, will be larger than the sled pulling back, $\vec{\mathbf{F}}_{AS}$, and the assistant accelerates forward (Newton's second law). The sled, on the other hand, accelerates forward when the force on *it* exerted by the assistant is greater than the frictional force exerted backward on it by the ground (that is, when $\vec{\mathbf{F}}_{SA}$ has greater magnitude than $\vec{\mathbf{F}}_{SG}$ in Fig. 4–12).

PROBLEM SOLVING

A study of Newton's second and third laws

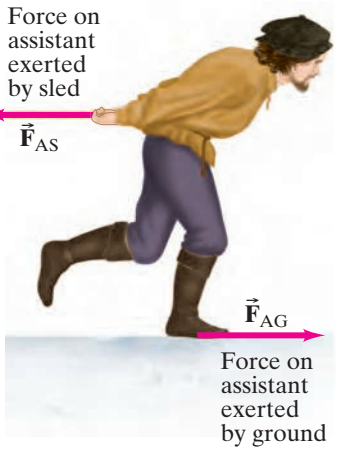


FIGURE 4–13 Example 4–5. The horizontal forces on the assistant.

Using double subscripts to clarify Newton's third law can become cumbersome, and we won't usually use them in this way. We will usually use a single subscript referring to what exerts the force on the object being discussed. Nevertheless, if there is any confusion in your mind about a given force, go ahead and use two subscripts to identify *on* what object and *by* what object the force is exerted.

EXERCISE B A tennis ball collides head-on with a more massive baseball. (i) Which ball experiences the greater force of impact? (ii) Which experiences the greater acceleration during the impact? (iii) Which of Newton's laws are useful to obtain the correct answers?

EXERCISE C If you push on a heavy desk, does it always push back on you? (*a*) No. (*b*) Yes. (*c*) Not unless someone else also pushes on it. (*d*) Yes, if it is out in space. (*e*) A desk never pushes.

EXERCISE D Return to the first Chapter-Opening Question, page 85, and answer it again now. Try to explain why you may have answered differently the first time.

4–6 Weight—the Force of Gravity; and the Normal Force

As we saw in Chapter 2, Galileo claimed that all objects dropped near the surface of the Earth would fall with the same acceleration, $\vec{\mathbf{g}}$, if air resistance was negligible. The force that causes this acceleration is called the *force of gravity* or *gravitational force*. What exerts the gravitational force on an object? It is the Earth, as we will discuss in Chapter 5, and the force acts vertically[†] downward, toward the center of the Earth. Let us apply Newton's second law to an object of mass m falling freely due to gravity. For the acceleration, $\vec{\mathbf{a}}$, we use the downward acceleration due to gravity, $\vec{\mathbf{g}}$. Thus, the **gravitational force** on an object, $\vec{\mathbf{F}}_G$, can be written as

$$\vec{\mathbf{F}}_G = m\vec{\mathbf{g}}. \qquad \textbf{(4–3)}$$

The direction of this force is down toward the center of the Earth. The magnitude of the force of gravity on an object, mg, is commonly called the object's **weight**.

In SI units, $g = 9.80\,\text{m/s}^2 = 9.80\,\text{N/kg}$,[‡] so the weight of a 1.00-kg mass on Earth is $1.00\,\text{kg} \times 9.80\,\text{m/s}^2 = 9.80\,\text{N}$. We will mainly be concerned with the weight of objects on Earth, but we note that on the Moon, on other planets, or in space, the weight of a given mass will be different than it is on Earth. For example, on the Moon the acceleration due to gravity is about one-sixth what it is on Earth, and a 1.0-kg mass weighs only 1.6 N. Although we will not use British units, we note that for practical purposes on the Earth, a mass of 1.0 kg weighs about 2.2 lb. (On the Moon, 1 kg weighs only about 0.4 lb.)

The force of gravity acts on an object when it is falling. When an object is at rest on the Earth, the gravitational force on it does not disappear, as we know if we weigh it on a spring scale. The same force, given by Eq. 4–3, continues to act. Why, then, doesn't the object move? From Newton's second law, the net force on an object that remains at rest is zero. There must be another force on the object to balance the gravitational force. For an object resting on a table, the table exerts this upward force; see Fig. 4–14a. The table is compressed slightly beneath the object, and due to its elasticity, it pushes up on the object as shown. The force exerted by the table is often called a **contact force**, since it occurs when two objects are in contact. (The force of your hand pushing on a cart is also a contact force.) When a contact force acts *perpendicular* to the common surface of contact, it is referred to as the **normal force** ("normal" means perpendicular); hence it is labeled $\vec{\mathbf{F}}_N$ in Fig. 4–14a.

The two forces shown in Fig. 4–14a are both acting on the statue, which remains at rest, so the vector sum of these two forces must be zero (Newton's second law). Hence $\vec{\mathbf{F}}_G$ and $\vec{\mathbf{F}}_N$ must be of equal magnitude and in opposite directions. But they are *not* the equal and opposite forces spoken of in Newton's third law. The action and reaction forces of Newton's third law act on *different objects*, whereas the two forces shown in Fig. 4–14a act on the *same* object. For each of the forces shown in Fig. 4–14a, we can ask, "What is the reaction force?" The upward force $\vec{\mathbf{F}}_N$ on the statue is exerted by the table. The reaction to this force is a force exerted by the statue downward on the table. It is shown in Fig. 4–14b, where it is labeled $\vec{\mathbf{F}}'_N$. This force, $\vec{\mathbf{F}}'_N$, exerted on the table by the statue, is the reaction force to $\vec{\mathbf{F}}_N$ in accord with Newton's third law. What about the other force on the statue, the force of gravity $\vec{\mathbf{F}}_G$ exerted by the Earth? Can you guess what the reaction is to this force? We will see in Chapter 5 that the reaction force is also a gravitational force, exerted on the Earth by the statue.

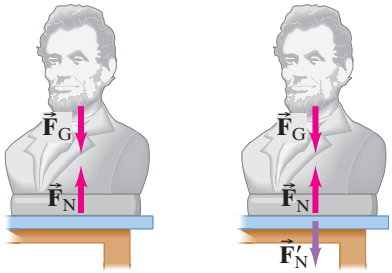


FIGURE 4–14 (a) The net force on an object at rest is zero according to Newton's second law. Therefore the downward force of gravity $(\vec{\mathbf{F}}_G)$ on an object at rest must be balanced by an upward force (the normal force $\vec{\mathbf{F}}_N$) exerted by the table in this case. (b) $\vec{\mathbf{F}}'_N$ is the force exerted on the table by the statue and is the reaction force to $\vec{\mathbf{F}}_N$ by Newton's third law. ($\vec{\mathbf{F}}'_N$ is shown in a different color to remind us it acts on a different object.) The reaction force to $\vec{\mathbf{F}}_G$ is not shown.

⚠ CAUTION

Weight and normal force are **not** *action–reaction pairs*

[†]The concept of "vertical" is tied to gravity. The best definition of *vertical* is that it is the direction in which objects fall. A surface that is "horizontal," on the other hand, is a surface on which a round object won't start rolling: gravity has no effect. Horizontal is perpendicular to vertical.

[‡]Since $1\,\text{N} = 1\,\text{kg}\cdot\text{m/s}^2$ (Section 4–4), then $1\,\text{m/s}^2 = 1\,\text{N/kg}$.

EXERCISE E Return to the second Chapter-Opening Question, page 85, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 4–6 Weight, normal force, and a box. A friend has given you a special gift, a box of mass 10.0 kg with a mystery surprise inside. The box is resting on the smooth (frictionless) horizontal surface of a table (Fig. 4–15a). (*a*) Determine the weight of the box and the normal force exerted on it by the table. (*b*) Now your friend pushes down on the box with a force of 40.0 N, as in Fig. 4–15b. Again determine the normal force exerted on the box by the table. (*c*) If your friend instead pulls upward on the box with a force of 40.0 N (Fig. 4–15c), what now is the normal force exerted on the box by the table?

APPROACH The box is at rest on the table, so the net force on the box in each case is zero (Newton's first or second law). The weight of the box has magnitude mg in all three cases.

SOLUTION (*a*) The weight of the box is $mg = (10.0\,\text{kg})(9.80\,\text{m/s}^2) = 98.0\,\text{N}$, and this force acts downward. The only other force on the box is the normal force $\vec{\mathbf{F}}_\text{N}$ exerted upward on it by the table, as shown in Fig. 4–15a. We choose the upward direction as the positive y direction; then the net force ΣF_y on the box is $\Sigma F_y = F_\text{N} - mg$; the minus sign means mg acts in the negative y direction (m and g are magnitudes). The box is at rest, so the net force on it must be zero (Newton's second law, $\Sigma F_y = ma_y$, and $a_y = 0$). Thus

$$\begin{aligned} \Sigma F_y &= ma_y \\ F_\text{N} - mg &= 0, \end{aligned}$$

so we have

$$F_\text{N} = mg.$$

The normal force on the box, exerted by the table, is 98.0 N upward, and has magnitude equal to the box's weight.

(*b*) Your friend is pushing down on the box with a force of 40.0 N. So instead of only two forces acting on the box, now there are three forces acting on the box, as shown in Fig. 4–15b. The weight of the box is still $mg = 98.0\,\text{N}$. The net force is $\Sigma F_y = F_\text{N} - mg - 40.0\,\text{N}$, and is equal to zero because the box remains at rest ($a = 0$). Newton's second law gives

$$\Sigma F_y = F_\text{N} - mg - 40.0\,\text{N} = 0.$$

We solve this equation for the normal force:

$$F_\text{N} = mg + 40.0\,\text{N} = 98.0\,\text{N} + 40.0\,\text{N} = 138.0\,\text{N},$$

which is greater than in (*a*). The table pushes back with more force when a person pushes down on the box. The normal force is not always equal to the weight!

(*c*) The box's weight is still 98.0 N and acts downward. The force exerted by your friend and the normal force both act upward (positive direction), as shown in Fig. 4–15c. The box doesn't move since your friend's upward force is less than the weight. The net force, again set to zero in Newton's second law because $a = 0$, is

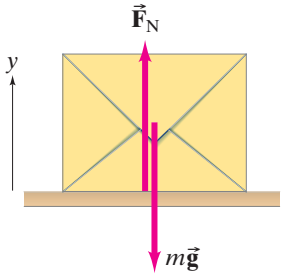
$$\Sigma F_y = F_\text{N} - mg + 40.0\,\text{N} = 0,$$

so

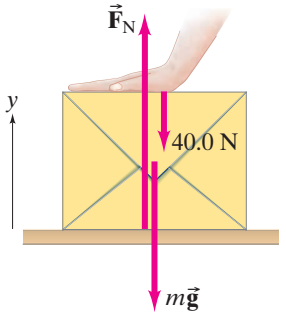
$$F_\text{N} = mg - 40.0\,\text{N} = 98.0\,\text{N} - 40.0\,\text{N} = 58.0\,\text{N}.$$

The table does not push against the full weight of the box because of the upward force exerted by your friend.

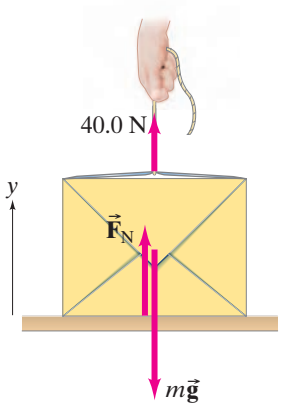
NOTE The weight of the box ($= mg$) does not change as a result of your friend's push or pull. Only the normal force is affected.



(a) $\Sigma F_y = F_\text{N} - mg = 0$



(b) $\Sigma F_y = F_\text{N} - mg - 40.0\,\text{N} = 0$



(c) $\Sigma F_y = F_\text{N} - mg + 40.0\,\text{N} = 0$

FIGURE 4–15 Example 4–6. (a) A 10-kg gift box is at rest on a table. (b) A person pushes down on the box with a force of 40.0 N. (c) A person pulls upward on the box with a force of 40.0 N. The forces are all assumed to act along a line; they are shown slightly displaced in order to be distinguishable. Only forces acting on the box are shown.

CAUTION

The normal force F_N is not always equal to the weight

CAUTION

The normal force, $\vec{\mathbf{F}}_N$, is not necessarily vertical

Recall that the normal force is elastic in origin (the table in Fig. 4–15 sags slightly under the weight of the box). The normal force in Example 4–6 is vertical, perpendicular to the horizontal table. The normal force is not always vertical, however. When you push against a wall, for example, the normal force with which the wall pushes back on you is horizontal (see Fig. 4–9). If an object is on a plane inclined at an angle to the horizontal, such as a skier or car on a hill, the normal force acts perpendicular to the plane and so is not vertical.

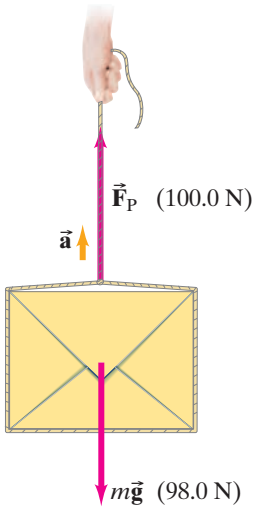


FIGURE 4–16 Example 4–7. The box accelerates upward because $F_P > mg$.

EXAMPLE 4–7 **Accelerating the box.** What happens when a person pulls upward on the box in Example 4–6*c* with a force equal to, or greater than, the box's weight? For example, let $F_P = 100.0\,\text{N}$ (Fig. 4–16) rather than the 40.0 N shown in Fig. 4–15c.

APPROACH We can start just as in Example 4–6, but be ready for a surprise.

SOLUTION The net force on the box is

$$\begin{aligned}\Sigma F_y &= F_N - mg + F_P \\ &= F_N - 98.0\,\text{N} + 100.0\,\text{N},\end{aligned}$$

and if we set this equal to zero (thinking the acceleration might be zero), we would get $F_N = -2.0\,\text{N}$. This is nonsense, since the negative sign implies F_N points downward, and the table surely cannot *pull* down on the box (unless there's glue on the table). The least F_N can be is zero, which it will be in this case, $F_N = 0$. What really happens here is that the box leaves the table and accelerates upward ($a \neq 0$, see gold vector in Fig. 4–16) because the net force is not zero. The net force (setting the normal force $F_N = 0$) is

$$\begin{aligned}\Sigma F_y &= F_P - mg = 100.0\,\text{N} - 98.0\,\text{N} \\ &= 2.0\,\text{N}\end{aligned}$$

upward. See Fig. 4–16. We apply Newton's second law and see that the box moves upward with an acceleration

$$a_y = \frac{\Sigma F_y}{m} = \frac{2.0\,\text{N}}{10.0\,\text{kg}} = 0.20\,\text{m/s}^2.$$

EXAMPLE 4–8 **Apparent weight loss.** A 65-kg woman descends in an elevator that briefly accelerates at $0.20\,g$ downward. She stands on a scale that reads in kg. (*a*) During this acceleration, what is her weight and what does the scale read? (*b*) What does the scale read when the elevator descends at a constant speed of 2.0 m/s?

APPROACH Figure 4–17 shows all the forces that act on the woman (and *only* those that act on her). The direction of the acceleration is downward, so we choose the positive direction as down (this is the opposite choice from Examples 4–6 and 4–7).

SOLUTION (*a*) From Newton's second law,

$$\begin{aligned}\Sigma F &= ma \\ mg - F_N &= m(0.20\,g).\end{aligned}$$

We solve for F_N:

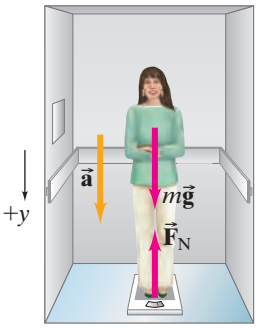
$$\begin{aligned}F_N &= mg - 0.20\,mg \\ &= 0.80\,mg,\end{aligned}$$

and it acts upward. The normal force $\vec{\mathbf{F}}_N$ is the force the scale exerts on the person, and it is equal and opposite to the force she exerts on the scale: $F'_N = 0.80\,mg$ downward. Her weight (force of gravity on her) is still $mg = (65\,\text{kg})(9.8\,\text{m/s}^2) = 640\,\text{N}$. But the scale, needing to exert a force of only $0.80\,mg$, will give a reading of $0.80\,m = 52\,\text{kg}$.

(*b*) Now there is no acceleration, $a = 0$, so by Newton's second law, $mg - F_N = 0$ and $F_N = mg$. The scale reads her true mass of 65 kg.

NOTE The scale in (*a*) gives a reading of 52 kg (as an "apparent mass"), but her mass doesn't change as a result of the acceleration: it stays at 65 kg.

FIGURE 4–17 Example 4–8. The acceleration vector is shown in gold to distinguish it from the red force vectors.



4–7 Solving Problems with Newton's Laws: Free-Body Diagrams

Newton's second law tells us that the acceleration of an object is proportional to the *net force* acting on the object. The **net force**, as mentioned earlier, is the *vector sum* of all forces acting on the object. Indeed, extensive experiments have shown that forces do add together as vectors precisely according to the rules we developed in Chapter 3. For example, in Fig. 4–18, two forces of equal magnitude (100 N each) are shown acting on an object at right angles to each other. Intuitively, or from *symmetry*, we can see that the object will start moving at a 45° angle and thus the net force acts at a 45° angle. This is just what the rules of vector addition give. From the theorem of Pythagoras, the magnitude of the resultant force is $F_R = \sqrt{(100\,\text{N})^2 + (100\,\text{N})^2} = 141\,\text{N}$.

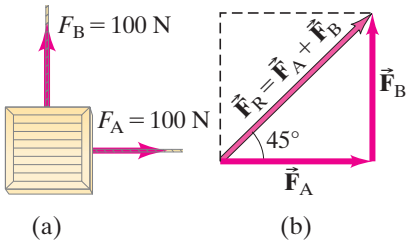


FIGURE 4–18 (a) Two horizontal forces, $\vec{F}_A$ and $\vec{F}_B$, exerted by workers A and B, act on a crate (we are looking down from above). (b) The sum, or resultant, of $\vec{F}_A$ and $\vec{F}_B$ is $\vec{F}_R$.

EXAMPLE 4–9 **Adding force vectors.** Calculate the sum of the two forces exerted on the boat by workers A and B in Fig. 4–19a.

APPROACH We add force vectors like any other vectors as described in Chapter 3. The first step is to choose an xy coordinate system (see Fig. 4–19a), and then resolve vectors into their components.

SOLUTION The two force vectors are shown resolved into components in Fig. 4–19b. We add the forces using the method of components. The components of $\vec{F}_A$ are

$$F_{Ax} = F_A \cos 45.0° = (40.0\,\text{N})(0.707) = 28.3\,\text{N},$$
$$F_{Ay} = F_A \sin 45.0° = (40.0\,\text{N})(0.707) = 28.3\,\text{N}.$$

The components of $\vec{F}_B$ are

$$F_{Bx} = +F_B \cos 37.0° = +(30.0\,\text{N})(0.799) = +24.0\,\text{N},$$
$$F_{By} = -F_B \sin 37.0° = -(30.0\,\text{N})(0.602) = -18.1\,\text{N}.$$

F_{By} is negative because it points along the negative y axis. The components of the resultant force are (see Fig. 4–19c)

$$F_{Rx} = F_{Ax} + F_{Bx} = 28.3\,\text{N} + 24.0\,\text{N} = 52.3\,\text{N},$$
$$F_{Ry} = F_{Ay} + F_{By} = 28.3\,\text{N} - 18.1\,\text{N} = 10.2\,\text{N}.$$

To find the magnitude of the resultant force, we use the Pythagorean theorem,

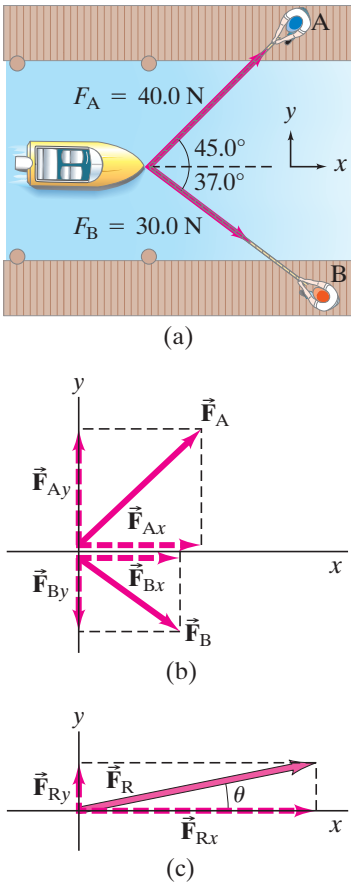
$$F_R = \sqrt{F_{Rx}^2 + F_{Ry}^2} = \sqrt{(52.3)^2 + (10.2)^2}\,\text{N} = 53.3\,\text{N}.$$

The only remaining question is the angle θ that the net force $\vec{F}_R$ makes with the x axis, Fig. 4–19c. We use:

$$\tan\theta = \frac{F_{Ry}}{F_{Rx}} = \frac{10.2\,\text{N}}{52.3\,\text{N}} = 0.195,$$

and $\tan^{-1}(0.195) = 11.0°$. The net force on the boat has magnitude 53.3 N and acts at an 11.0° angle to the x axis.

FIGURE 4–19 Example 4–9: Two force vectors act on a boat.



When solving problems involving Newton's laws and force, it is very important to draw a diagram showing all the forces acting *on* each object involved. Such a diagram is called a **free-body diagram**, or **force diagram**: choose one object, and draw an arrow to represent each force acting on it. Include *every* force acting on that object. Do not show forces that the chosen object exerts on *other* objects. To help you identify each and every force that is exerted on your chosen object, ask yourself what other objects could exert a force on it. If your problem involves more than one object, a separate free-body diagram is needed for each object. For now, the likely forces that could be acting are *gravity* and *contact forces* (one object pushing or pulling another, normal force, friction). Later we will consider other types of force such as buoyancy, fluid pressure, and electric and magnetic forces.

PROBLEM SOLVING
Free-body diagram

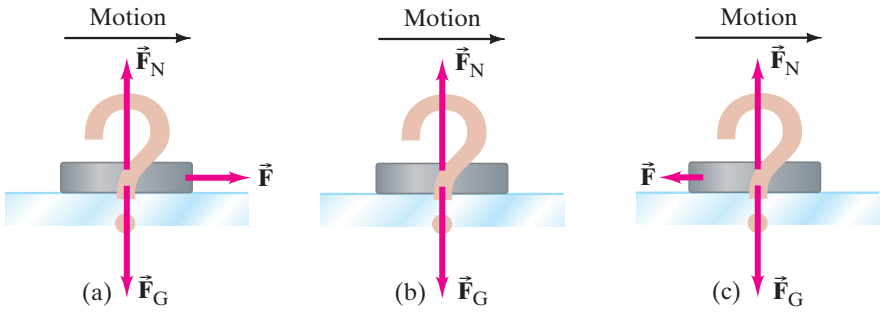


FIGURE 4–20 Example 4–10. Which is the correct free-body diagram for a hockey puck sliding across frictionless ice?

CONCEPTUAL EXAMPLE 4–10 **The hockey puck.** A hockey puck is sliding at constant velocity across a flat horizontal ice surface that is assumed to be frictionless. Which of the sketches in Fig. 4–20 is the correct free-body diagram for this puck? What would your answer be if the puck slowed down?

RESPONSE Did you choose (*a*)? If so, can you answer the question: what exerts the horizontal force labeled $\vec{\mathbf{F}}$ on the puck? If you say that it is the force needed to maintain the motion, ask yourself: what exerts this force? Remember that another object must exert any force—and there simply isn't any possibility here. Therefore, (*a*) is wrong. Besides, the force $\vec{\mathbf{F}}$ in Fig. 4–20a would give rise to an acceleration by Newton's second law. It is (*b*) that is correct. No net force acts on the puck, and the puck slides at constant velocity across the ice.

In the real world, where even smooth ice exerts at least a tiny friction force, (*c*) is the correct answer. The tiny friction force is in the direction opposite to the motion, and the puck's velocity decreases, even if very slowly.

PROBLEM SOLVING

Newton's Laws; Free-Body Diagrams

1. Draw a sketch of the situation, after carefully reading the problem at least twice.

2. Consider only one object (at a time), and draw a **free-body diagram** for that object, showing *all* the forces acting *on* that object. Include any unknown forces that you have to solve for. Do not show any forces that the chosen object exerts on other objects.

Draw the arrow for each force vector reasonably accurately for direction and magnitude. Label each force acting on the object, including forces you must solve for, according to its source (gravity, person, friction, and so on).

If several objects are involved, draw a free-body diagram for each object *separately*. For each object, show all the forces acting *on that object* (and *only* forces acting on that object). For each (and every) force, you must be clear about: *on* what object that force acts, and *by* what object that force is exerted. Only forces acting *on* a given object can be included in $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ for that object.

3. Newton's second law involves vectors, and it is usually important to **resolve vectors** into components. **Choose** x and y **axes** in a way that simplifies the calculation. For example, it often saves work if you choose one coordinate axis to be in the direction of the acceleration (if known).

4. For each object, **apply Newton's second law** to the x and y components separately. That is, the x component of the net force on that object is related to the x component of that object's acceleration: $\Sigma F_x = ma_x$, and similarly for the y direction.

5. Solve the equation or equations for the unknown(s). Put in numerical values only at the end, and keep track of units.

This Problem Solving Strategy should not be considered a prescription. Rather it is a summary of things to do that will start you thinking and getting involved in the problem at hand.

CAUTION
Treating an object as a particle

When we are concerned only about *translational motion*, all the forces on a given object can be drawn as acting at the center of the object, thus treating the object as a **point particle**. However, for problems involving rotation or statics, the place *where* each force acts is also important, as we shall see in Chapters 10, 11, and 12.

In the Examples in this Section, we assume that all surfaces are very smooth so that friction can be ignored. (Friction, and Examples using it, are discussed in Chapter 5.)

EXAMPLE 4–11 **Pulling the mystery box.** Suppose a friend asks to examine the 10.0-kg box you were given (Example 4–6, Fig. 4–15), hoping to guess what is inside; and you respond, "Sure, pull the box over to you." She then pulls the box by the attached cord, as shown in Fig. 4–21a, along the smooth surface of the table. The magnitude of the force exerted by the person is $F_P = 40.0\,N$, and it is exerted at a 30.0° angle as shown. Calculate (*a*) the acceleration of the box, and (*b*) the magnitude of the upward force F_N exerted by the table on the box. Assume that friction can be neglected.

APPROACH We follow the Problem Solving Strategy on the previous page.

SOLUTION

1. **Draw a sketch**: The situation is shown in Fig. 4–21a; it shows the box and the force applied by the person, F_P.
2. **Free-body diagram**: Figure 4–21b shows the free-body diagram of the box. To draw it correctly, we show *all* the forces acting on the box and *only* the forces acting on the box. They are: the force of gravity $m\vec{\mathbf{g}}$; the normal force exerted by the table $\vec{\mathbf{F}}_N$; and the force exerted by the person $\vec{\mathbf{F}}_P$. We are interested only in translational motion, so we can show the three forces acting at a point, Fig. 4–21c.
3. **Choose axes and resolve vectors**: We expect the motion to be horizontal, so we choose the x axis horizontal and the y axis vertical. The pull of 40.0 N has components
$$F_{Px} = (40.0\,N)(\cos 30.0°) = (40.0\,N)(0.866) = 34.6\,N,$$
$$F_{Py} = (40.0\,N)(\sin 30.0°) = (40.0\,N)(0.500) = 20.0\,N.$$
In the horizontal (x) direction, $\vec{\mathbf{F}}_N$ and $m\vec{\mathbf{g}}$ have zero components. Thus the horizontal component of the net force is F_{Px}.
4. (*a*) **Apply Newton's second law** to determine the x component of the acceleration:
$$F_{Px} = ma_x.$$
5. (*a*) **Solve**:
$$a_x = F_{Px}/m = 34.6\,N/10.0\,kg = 3.46\,m/s^2.$$
The acceleration of the box is 3.46 m/s^2 to the right.

(*b*) Next we want to find F_N.

4′. (*b*) **Apply Newton's second law** to the vertical (y) direction, with upward as positive:
$$\Sigma F_y = ma_y$$
$$F_N - mg + F_{Py} = ma_y.$$

5′. (*b*) **Solve**: We have $mg = (10.0\,kg)(9.80\,m/s^2) = 98.0\,N$ and, from step 3 above, $F_{Py} = 20.0\,N$. Furthermore, since $F_{Py} < mg$, the box does not move vertically, so $a_y = 0$. Thus
$$F_N - 98.0\,N + 20.0\,N = 0,$$
so $F_N = 78.0\,N$.

NOTE F_N is less than mg: the table does not push against the full weight of the box because part of the pull exerted by the person is in the upward direction.

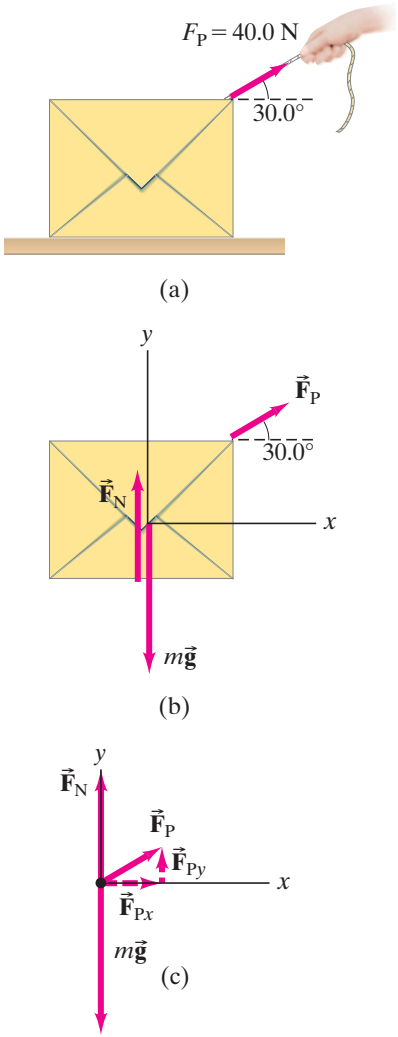


FIGURE 4–21 (a) Pulling the box, Example 4–11; (b) is the free-body diagram for the box, and (c) is the free-body diagram considering all the forces to act at a point (translational motion only, which is what we have here).

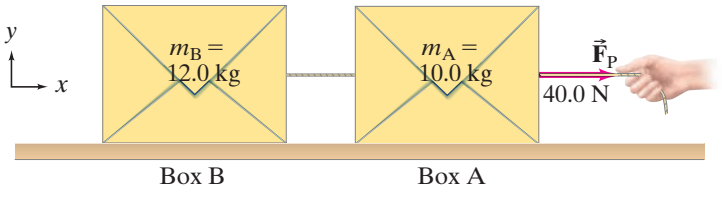
EXERCISE F If the force $\vec{\mathbf{F}}_P$ (Fig. 4–21) exerted by the person is doubled in magnitude, what now will be the normal force?

EXERCISE G A 10.0-kg box is dragged on a horizontal frictionless surface by a horizontal force of 10.0 N. If the applied force is doubled, the normal force on the box will (*a*) increase; (*b*) remain the same; (*c*) decrease.

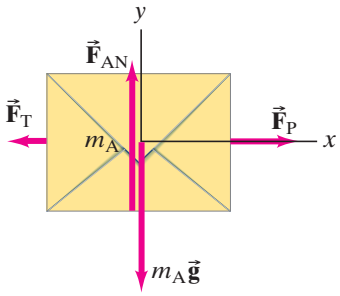
Tension in a Flexible Cord

When a flexible cord pulls on an object, the cord is said to be under **tension**, and the force it exerts on the object is the tension F_T. If the cord has negligible mass, the force exerted at one end is transmitted undiminished to each adjacent piece of cord along the entire length to the other end. Why? Because $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}} = 0$ for the cord if the cord's mass m is zero (or negligible) no matter what $\vec{\mathbf{a}}$ is. Hence the forces pulling on the cord at its two ends must add up to zero (F_T and $-F_T$). Note that flexible cords and strings can only pull. They can't push because they bend.

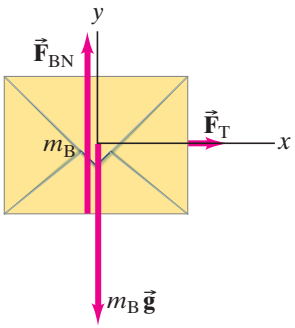
PROBLEM SOLVING
Cords can pull but can't push; tension exists throughout a taut cord



(a)



(b)



(c)

FIGURE 4–22 Example 4–12. (a) Two boxes, A and B, are connected by a cord. A person pulls horizontally on box A with force $F_P = 40.0\,\text{N}$. (b) Free-body diagram for box A. (c) Free-body diagram for box B.

Our next Example involves two boxes connected by a cord. We can refer to this group of objects as a system. A **system** is any group of one or more objects we choose to consider and study.

EXAMPLE 4–12 **Two boxes connected by a cord.** Two boxes, A and B, are connected by a lightweight cord and are resting on a smooth (frictionless) table. The boxes have masses of 12.0 kg and 10.0 kg. A horizontal force F_P of 40.0 N is applied to the 10.0-kg box by a person pulling on a cord, as shown in Fig. 4–22a. Find (*a*) the acceleration of each box, and (*b*) the tension in the cord connecting the boxes.

APPROACH We streamline our approach by not listing each step. We have two boxes so we draw a free-body diagram for each. To draw them correctly, we must consider the forces on *each* box by itself, so that Newton's second law can be applied to each. The person exerts a force F_P on box A. Box A exerts a force F_T on the connecting cord, and the cord exerts an opposite but equal magnitude force F_T back on box A (Newton's third law). The two horizontal forces on box A are shown in Fig. 4–22b, along with the force of gravity $m_A\vec{\mathbf{g}}$ downward and the normal force $\vec{\mathbf{F}}_{AN}$ exerted upward by the table. The cord is light, so we neglect its mass. The tension at each end of the cord is thus the same. Hence the cord exerts a force F_T on the second box. Figure 4–22c shows the forces on box B, which are $\vec{\mathbf{F}}_T$, $m_B\vec{\mathbf{g}}$, and the normal force $\vec{\mathbf{F}}_{BN}$. There will be only horizontal motion. We take the positive x axis to the right.

SOLUTION (*a*) We apply $\Sigma F_x = ma_x$ to box A:

$$\Sigma F_x = F_P - F_T = m_A a_A. \qquad \text{[box A]}$$

For box B, the only horizontal force is F_T, so

$$\Sigma F_x = F_T = m_B a_B. \qquad \text{[box B]}$$

The boxes are connected, and if the cord remains taut and doesn't stretch, then the two boxes will have the same acceleration a. Thus $a_A = a_B = a$. We now have 2 unknowns, F_T and a, and the two equations above to solve for them. We are given $m_A = 10.0\,\text{kg}$ and $m_B = 12.0\,\text{kg}$. We can add the two equations above to eliminate an unknown (F_T) and obtain

$$(m_A + m_B)\,a = F_P - F_T + F_T = F_P$$

or

$$a = \frac{F_P}{m_A + m_B} = \frac{40.0\,\text{N}}{22.0\,\text{kg}} = 1.82\,\text{m/s}^2.$$

This is what we sought.

(*b*) From the equation for box B above $(F_T = m_B a_B)$, the tension in the cord is

$$F_T = m_B a = (12.0\,\text{kg})(1.82\,\text{m/s}^2) = 21.8\,\text{N}.$$

Thus, $F_T < F_P\ (= 40.0\,\text{N})$, as we expect, since F_T acts to accelerate only m_B.

Alternate Solution to (a) We would have obtained the same result had we considered a single system, of mass $m_A + m_B$, acted on by a net horizontal force equal to F_P. (The tension forces F_T would then be considered internal to the system as a whole, and summed together would make zero contribution to the net force on the *whole* system.)

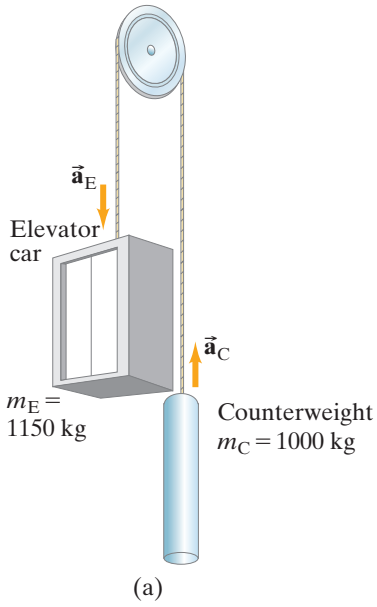
NOTE It might be tempting to say that the force the person exerts, F_P, acts not only on box A but also on box B. It doesn't. F_P acts only on box A. It affects box B via the tension in the cord, F_T, which acts on box B and accelerates it. (You could look at it this way: $F_T < F_P$ because F_P accelerates *both* boxes whereas F_T only accelerates box B.)

⚠ CAUTION

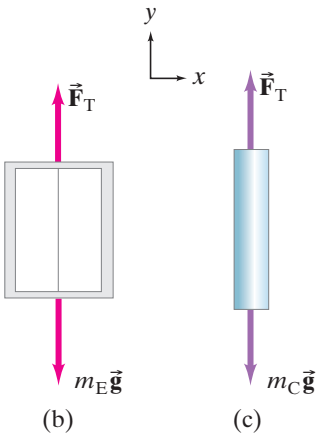
For any object, use only the forces on that object in calculating $\Sigma F = ma$

EXAMPLE 4–13 **Elevator and counterweight (Atwood machine).** A system of two objects suspended over a pulley by a flexible cable, as shown in Fig. 4–23a, is sometimes referred to as an *Atwood machine*. Consider the real-life application of an elevator (m_E) and its counterweight (m_C). To minimize the work done by the motor to raise and lower the elevator safely, m_E and m_C are made similar in mass. We leave the motor out of the system for this calculation, and assume that the cable's mass is negligible and that the mass of the pulley, as well as any friction, is small and ignorable. These assumptions ensure that the tension F_T in the cable has the same magnitude on both sides of the pulley. Let the mass of the counterweight be $m_C = 1000\,\text{kg}$. Assume the mass of the empty elevator is 850 kg, and its mass when carrying four passengers is $m_E = 1150\,\text{kg}$. For the latter case $(m_E = 1150\,\text{kg})$, calculate (*a*) the acceleration of the elevator and (*b*) the tension in the cable.

PHYSICS APPLIED
Elevator (as Atwood machine)



(a)



(b) (c)

FIGURE 4–23 Example 4–13. (a) Atwood machine in the form of an elevator–counterweight system. (b) and (c) Free-body diagrams for the two objects.

APPROACH Again we have two objects, and we will need to apply Newton's second law to each of them separately. Each mass has two forces acting on it: gravity downward and the cable tension pulling upward, $\vec{\mathbf{F}}_T$. Figures 4–23b and c show the free-body diagrams for the elevator (m_E) and for the counterweight (m_C). The elevator, being the heavier, will accelerate downward, whereas the counterweight will accelerate upward. The magnitudes of their accelerations will be equal (we assume the cable is massless and doesn't stretch). For the counterweight, $m_C g = (1000\,\text{kg})(9.80\,\text{m/s}^2) = 9800\,\text{N}$, so F_T must be greater than 9800 N (in order that m_C will accelerate upward). For the elevator, $m_E g = (1150\,\text{kg})(9.80\,\text{m/s}^2) = 11{,}300\,\text{N}$, which must have greater magnitude than F_T so that m_E accelerates downward. Thus our calculation must give F_T between 9800 N and 11,300 N.

SOLUTION (*a*) To find F_T as well as the acceleration a, we apply Newton's second law, $\Sigma F = ma$, to each object. We take upward as the positive y direction for both objects. With this choice of axes, $a_C = a$ because m_C accelerates upward, and $a_E = -a$ because m_E accelerates downward. Thus

$$F_T - m_E g = m_E a_E = -m_E a$$
$$F_T - m_C g = m_C a_C = +m_C a.$$

We can subtract the first equation from the second to get

$$(m_E - m_C)g = (m_E + m_C)a,$$

where a is now the only unknown. We solve this for a:

$$a = \frac{m_E - m_C}{m_E + m_C}g = \frac{1150\,\text{kg} - 1000\,\text{kg}}{1150\,\text{kg} + 1000\,\text{kg}}g = 0.070\,g = 0.68\,\text{m/s}^2.$$

The elevator (m_E) accelerates downward (and the counterweight m_C upward) at $a = 0.070\,g = 0.68\,\text{m/s}^2$.

(*b*) The tension in the cable F_T can be obtained from either of the two $\Sigma F = ma$ equations at the start of our solution, setting $a = 0.070\,g = 0.68\,\text{m/s}^2$:

$$F_T = m_E g - m_E a = m_E(g - a)$$
$$= 1150\,\text{kg}\,(9.80\,\text{m/s}^2 - 0.68\,\text{m/s}^2) = 10{,}500\,\text{N},$$

or

$$F_T = m_C g + m_C a = m_C(g + a)$$
$$= 1000\,\text{kg}\,(9.80\,\text{m/s}^2 + 0.68\,\text{m/s}^2) = 10{,}500\,\text{N},$$

which are consistent. As predicted, our result lies between 9800 N and 11,300 N.

NOTE We can check our equation for the acceleration a in this Example by noting that if the masses were equal $(m_E = m_C)$, then our equation above for a would give $a = 0$, as we should expect. Also, if one of the masses is zero (say, $m_C = 0$), then the other mass $(m_E \neq 0)$ would be predicted by our equation to accelerate at $a = g$, again as expected.

PROBLEM SOLVING
Check your result by seeing if it works in situations where the answer is easily guessed

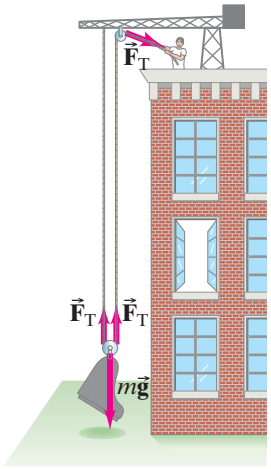


FIGURE 4–24 Example 4–14.

CONCEPTUAL EXAMPLE 4–14 **The advantage of a pulley.** A mover is trying to lift a piano (slowly) up to a second-story apartment (Fig. 4–24). He is using a rope looped over two pulleys as shown. What force must he exert on the rope to slowly lift the piano's 1600-N weight?

RESPONSE The magnitude of the tension force F_T within the rope is the same at any point along the rope if we assume we can ignore the mass of the rope and pulleys. First notice the forces acting on the lower pulley at the piano. The weight of the piano $(= mg)$ pulls down on the pulley. The tension in the rope, looped through this pulley, pulls up *twice*, once on each side of the pulley. Let us apply Newton's second law to the pulley–piano combination (of mass m), choosing the upward direction as positive:

$$2F_T - mg = ma.$$

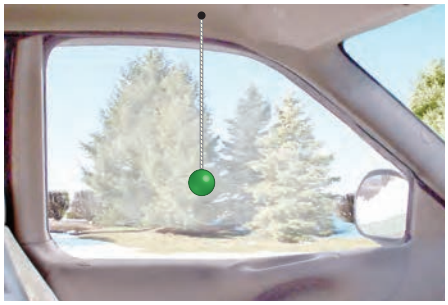
To move the piano with constant speed (set $a = 0$ in this equation) thus requires a tension in the rope, and hence a pull on the rope, of $F_T = mg/2$. The piano mover can exert a force equal to half the piano's weight.

NOTE We say the pulley has given a **mechanical advantage** of 2, since without the pulley the mover would have to exert twice the force.

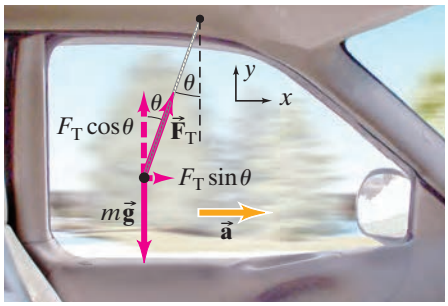
PHYSICS APPLIED
Accelerometer

EXAMPLE 4–15 **Accelerometer.** A small mass m hangs from a thin string and can swing like a pendulum. You attach it above the window of your car as shown in Fig. 4–25a. When the car is at rest, the string hangs vertically. What angle θ does the string make (a) when the car accelerates at a constant $a = 1.20\,\text{m/s}^2$, and (b) when the car moves at constant velocity, $v = 90\,\text{km/h}$?

FIGURE 4–25 Example 4–15.



(a)



(b)

APPROACH The free-body diagram of Fig. 4–25b shows the pendulum at some angle θ relative to the vertical, and the forces on it: $m\vec{g}$ downward, and the tension $\vec{F}_T$ in the cord (including its components). These forces do not add up to zero if $\theta \neq 0$; and since we have an acceleration a, we expect $\theta \neq 0$.

SOLUTION (a) The acceleration $a = 1.20\,\text{m/s}^2$ is horizontal $(= a_x)$, and the only horizontal force is the x component of $\vec{F}_T$: $F_T \sin\theta$ (Fig. 4–25b). Then from Newton's second law,

$$ma = F_T \sin\theta.$$

The vertical component of Newton's second law gives, since $a_y = 0$,

$$0 = F_T \cos\theta - mg.$$

So

$$mg = F_T \cos\theta.$$

Dividing the two equations involving F_T, we obtain

$$\tan\theta = \frac{F_T \sin\theta}{F_T \cos\theta} = \frac{ma}{mg} = \frac{a}{g}$$

or

$$\tan\theta = \frac{1.20\,\text{m/s}^2}{9.80\,\text{m/s}^2} = 0.122,$$

so

$$\theta = 7.0°.$$

(b) The velocity is constant, so $a = 0$ and therefore $\tan\theta = 0$. Hence the pendulum hangs vertically $(\theta = 0°)$.

NOTE This simple device is an **accelerometer**—it can be used to determine acceleration, by measuring the angle θ.

Inclines

Now we consider what happens when an object slides down an incline, such as a hill or ramp. Such problems are interesting because gravity is the accelerating force, yet the acceleration is not vertical. Solving problems is usually easier if we choose the xy coordinate system so the x axis points along the incline (the direction of motion) and the y axis is perpendicular to the incline, as shown in Fig. 4–26a. Note also that the normal force is not vertical, but is perpendicular to the sloping surface of the plane, along the y axis in Fig. 4–26b.

PROBLEM SOLVING
Good choice of coordinate system simplifies the calculation

EXAMPLE 4–16 **Box slides down an incline.** A box of mass m is placed on a smooth (frictionless) incline that makes an angle θ with the horizontal, as shown in Fig. 4–26. (*a*) Determine the normal force on the box. (*b*) Determine the box's acceleration. (*c*) Evaluate for a mass $m = 10\,\text{kg}$ and an incline of $\theta = 30°$.

APPROACH We expect the motion to be along the incline, so we choose the x axis along the slope, positive down the slope (the direction of motion). The y axis is perpendicular to the incline, positive upwards. The free-body diagram is shown in Fig. 4–26b. The forces on the box are its weight $m\vec{\mathbf{g}}$ vertically downward, which is shown resolved into its components parallel and perpendicular to the incline, and the normal force $\vec{\mathbf{F}}_\text{N}$ along the $+y$ axis. The components of $m\vec{\mathbf{g}}$ along the x and y axes are found using the definitions of sine ("side opposite") and cosine ("side adjacent"): $(mg)_x = mg\sin\theta$, and $(mg)_y = mg\cos\theta$. (The angle θ of the plane equals the angle between $m\vec{\mathbf{g}}$ and its y component because the left sides of the two angles are perpendicular to each other and so are the right sides—see Appendix A–6, Fig. A–2.)

SOLUTION (*a*) There is no motion in the y direction, so $a_y = 0$. Applying Newton's second law we have

$$\begin{aligned} F_y &= ma_y \\ F_\text{N} - mg\cos\theta &= 0, \end{aligned}$$

where F_N and the y component of gravity $(mg\cos\theta)$ are all the forces acting on the box in the y direction. Thus the normal force is given by

$$F_\text{N} = mg\cos\theta.$$

Note carefully that unless $\theta = 0°$, F_N has magnitude less than the weight mg.
(*b*) In the x direction the only force acting is the x component of $m\vec{\mathbf{g}}$, which is $mg\sin\theta$. The acceleration a is in the x direction so

$$\begin{aligned} F_x &= ma_x \\ mg\sin\theta &= ma, \end{aligned}$$

and we see that the acceleration down the plane is

$$a = g\sin\theta.$$

Thus the acceleration along an incline is always less than g, except at $\theta = 90°$, for which $\sin\theta = 1$ and $a = g$. This makes sense since $\theta = 90°$ is pure vertical fall. For $\theta = 0°$, $a = 0$, which makes sense because $\theta = 0°$ means the plane is horizontal so gravity causes no acceleration. Note too that the *acceleration does not depend on the mass* (m).
(*c*) For $\theta = 30°$, $\cos\theta = 0.866$ and $\sin\theta = 0.500$, so

$$F_\text{N} = 0.866\,mg = 85\,\text{N},$$

and

$$a = 0.500\,g = 4.9\,\text{m/s}^2.$$

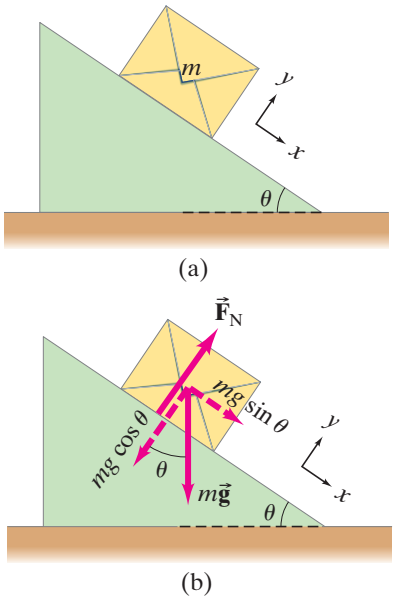


FIGURE 4–26 Example 4–16. (a) Box sliding on inclined plane. (b) Free-body diagram of box.

EXERCISE H Is the normal force always perpendicular to an inclined plane? Is it always vertical?

We will discuss more Examples of motion on an incline in the next Chapter, where friction will be included.

4–8 Problem Solving—A General Approach

A basic part of a physics course is solving problems effectively. The approach discussed here, though emphasizing Newton's laws, can be applied generally for other topics discussed throughout this book.

PROBLEM SOLVING

In General

1. Read and reread written problems carefully. A common error is to skip a word or two when reading, which can completely change the meaning of a problem.

2. Draw an accurate picture or diagram of the situation. (This is probably the most overlooked, yet most crucial, part of solving a problem.) Use arrows to represent vectors such as velocity or force, and label the vectors with appropriate symbols. When dealing with forces and applying Newton's laws, make sure to include all forces on a given object, including unknown ones, and make clear what forces act on what object (otherwise you may make an error in determining the *net force* on a particular object).

3. A separate **free-body diagram** needs to be drawn for each object involved, and it must show *all* the forces acting on a given object (and only on that object). Do not show forces that act on other objects.

4. Choose a convenient xy **coordinate system** (one that makes your calculations easier, such as one axis in the direction of the acceleration). Vectors are to be resolved into components along the coordinate axes. When using Newton's second law, apply $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ separately to x and y components, remembering that x direction forces are related to a_x, and similarly for y. If more than one object is involved, you can choose different (convenient) coordinate systems for each.

5. List the knowns and the unknowns (what you are trying to determine), and decide what you need in order to find the unknowns. For problems in the present Chapter, we use Newton's laws. More generally, it may help to see if one or more **relationships** (or **equations**) relate the unknowns to the knowns. But be sure each relationship is applicable in the given case. It is very important to know the limitations of each formula or relationship—when it is valid and when not. In this book, the more general equations have been given numbers, but even these can have a limited range of validity (often stated in brackets to the right of the equation).

6. Try to solve the problem approximately, to see if it is doable (to check if enough information has been given) and reasonable. Use your intuition, and make **rough calculations**—see "Order of Magnitude Estimating" in Section 1–6. A rough calculation, or a reasonable guess about what the range of final answers might be, is very useful. And a rough calculation can be checked against the final answer to catch errors in calculation, such as in a decimal point or the powers of 10.

7. Solve the problem, which may include algebraic manipulation of equations and/or numerical calculations. Recall the mathematical rule that you need as many independent equations as you have unknowns; if you have three unknowns, for example, then you need three independent equations. It is usually best to work out the algebra symbolically before putting in the numbers. Why? Because (*a*) you can then solve a whole class of similar problems with different numerical values; (*b*) you can check your result for cases already understood (say, $\theta = 0^\circ$ or 90°); (*c*) there may be cancellations or other simplifications; (*d*) there is usually less chance for numerical error; and (*e*) you may gain better insight into the problem.

8. Be sure to keep track of **units**, for they can serve as a check (they must balance on both sides of any equation).

9. Again consider if your answer is **reasonable**. The use of dimensional analysis, described in Section 1–7, can also serve as a check for many problems.

Summary

Newton's three laws of motion are the basic classical laws describing motion.

Newton's first law (the **law of inertia**) states that if the net force on an object is zero, an object originally at rest remains at rest, and an object in motion remains in motion in a straight line with constant velocity.

Newton's second law states that the acceleration of an object is directly proportional to the net force acting on it, and inversely proportional to its mass:

$$\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}. \qquad \textbf{(4–1a)}$$

Newton's second law is one of the most important and fundamental laws in classical physics.

Newton's third law states that whenever one object exerts a force on a second object, the second object always exerts a force on the first object which is equal in magnitude but opposite in direction:

$$\vec{\mathbf{F}}_{AB} = -\vec{\mathbf{F}}_{BA}, \qquad \textbf{(4–2)}$$

where $\vec{\mathbf{F}}_{BA}$ is the force on object B exerted by object A. This is true even if objects are moving and accelerating, and/or have different masses.

The tendency of an object to resist a change in its motion is called **inertia**. **Mass** is a measure of the inertia of an object.

Weight refers to the **gravitational force** on an object, and is equal to the product of the object's mass m and the acceleration of gravity $\vec{\mathbf{g}}$:

$$\vec{\mathbf{F}}_G = m\vec{\mathbf{g}}. \qquad \textbf{(4–3)}$$

Force, which is a vector, can be considered as a push or pull; or, from Newton's second law, force can be defined as an action capable of giving rise to acceleration. The **net force** on an object is the vector sum of all forces acting on that object.

For solving problems involving the forces on one or more objects, it is essential to draw a **free-body diagram** for *each* object, showing all the forces acting on only that object. Newton's second law can be applied to the vector components for each object.

Questions

1. Why does a child in a wagon seem to fall backward when you give the wagon a sharp pull forward?
2. A box rests on a frictionless part of a truck bed. The truck driver starts the truck and accelerates forward. The box immediately starts to slide toward the rear of the truck bed. Discuss the motion of the box, in terms of Newton's laws, as seen (*a*) by Francesca standing on the ground beside the truck, and (*b*) by Phil who is riding on the truck (Fig. 4–27).

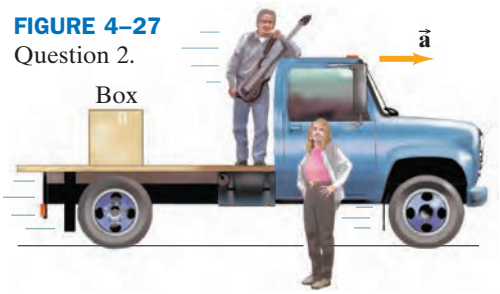


FIGURE 4–27
Question 2.

3. If an object is moving, is it possible for the net force acting on it to be zero? Explain.
4. If the acceleration of an object is zero, are no forces acting on it? Explain.
5. Only one force acts on an object. Can the object have zero acceleration? Can it have zero velocity? Explain.
6. When a golf ball is dropped to the pavement, it bounces back up. (*a*) Is a force needed to make it bounce back up? (*b*) If so, what exerts the force?
7. If you walk on a log floating on a lake, why does the log move in the opposite direction?
8. (*a*) Why do you push down harder on the pedals of a bicycle when first starting out than when moving at constant speed? (*b*) Why do you need to pedal at all when cycling at constant speed?
9. A stone hangs by a fine thread from the ceiling, and a section of the same thread dangles from the bottom of the stone (Fig. 4–28). If a person gives a sharp pull on the dangling thread, where is the thread likely to break: below the stone or above it? What if the person gives a slow and steady pull? Explain your answers.



FIGURE 4–28
Question 9.

10. The force of gravity on a 2-kg rock is twice as great as that on a 1-kg rock. Why then doesn't the heavier rock fall faster?
11. You pull a box with a constant force across a frictionless table using an attached rope held horizontally. If you now pull the rope with the same force at an angle to the horizontal (with the box remaining flat on the table), does the acceleration of the box increase, decrease, or remain the same? Explain.
12. When an object falls freely under the influence of gravity there is a net force mg exerted on it by the Earth. Yet by Newton's third law the object exerts an equal and opposite force on the Earth. Does the Earth move? Explain.
13. Compare the effort (or force) needed to lift a 10-kg object when you are on the Moon with the force needed to lift it on Earth. Compare the force needed to throw a 2-kg object horizontally with a given speed on the Moon and on Earth.

14. According to Newton's third law, each team in a tug of war (Fig. 4–29) pulls with equal force on the other team. What, then, determines which team will win?



FIGURE 4–29 Question 14. A tug of war. Describe the forces on each of the teams and on the rope.

15. When you stand still on the ground, how large a force does the ground exert on you? Why doesn't this force make you rise up into the air?
16. Whiplash sometimes results from an automobile accident when the victim's car is struck violently from the rear. Explain why the head of the victim seems to be thrown backward in this situation. Is it really?
17. Mary exerts an upward force of 40 N to hold a bag of groceries. Describe the "reaction" force (Newton's third law) by stating (*a*) its magnitude, (*b*) its direction, (*c*) *on* what object it is exerted, and (*d*) *by* what object it is exerted.
18. A father and his young daughter are ice skating. They face each other at rest and then push each other so they begin moving in opposite directions. Which one has the greater speed? Explain.
19. What would your bathroom scale read if you weighed yourself on an inclined plane? Assume the mechanism functions properly, even at an angle.
20. Which of the following objects weighs about 1 N: (*a*) an apple, (*b*) a mosquito, (*c*) this book, (*d*) you?
21. Why might your foot hurt if you kick a heavy desk or a wall?
22. When you are running and want to stop quickly, you must decelerate quickly. (*a*) What is the origin of the force that causes you to stop? (*b*) Estimate (using your own experience) the maximum rate of deceleration of a person running at top speed to come to rest.
23. Suppose that you are standing on top of a cardboard carton that just barely supports you. What would happen to it if you jumped up into the air? It would: (*a*) collapse; (*b*) be unaffected; (*c*) spring upward a bit; (*d*) move sideways. Explain your answer.

MisConceptual Questions

1. A truck is traveling horizontally to the right (Fig. 4–30). When the truck starts to slow down, the crate on the frictionless truck bed starts to slide. In what direction could the net force be on the crate?
   (*a*) No direction. The net force is zero.
   (*b*) Straight down (gravity).
   (*c*) Straight up (the normal force).
   (*d*) Horizontal and to the right.
   (*e*) Horizontal and to the left.

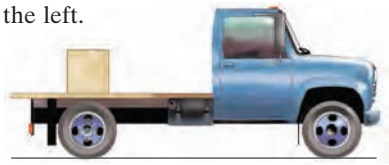


FIGURE 4–30 MisConceptual Question 1.

2. George, in the foreground of Fig. 4–31, is able to move the large truck because
   (*a*) he is stronger than the truck.
   (*b*) he is heavier in some respects than the truck.
   (*c*) the truck offers no resistance because its brakes are off.
   (*d*) the ground exerts a greater friction force on George than it does on the truck.
   (*e*) he exerts a greater force on the truck than the truck exerts back on him.



FIGURE 4–31 MisConceptual Question 2.

3. A bear sling, Fig. 4–32, is used in some national parks for placing backpackers' food out of the reach of bears. As the backpacker raises the pack by pulling down on the rope, the force F needed:
   (*a*) decreases as the pack rises until the rope is straight across.
   (*b*) doesn't change.
   (*c*) increases until the rope is straight.
   (*d*) increases but the rope always sags where the pack hangs.

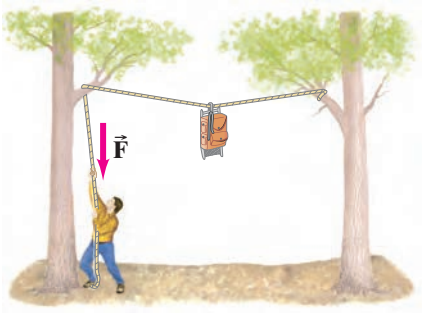


FIGURE 4–32 MisConceptual Question 3.

4. What causes the boat in Fig. 4–33 to move forward?
   (*a*) The force the man exerts on the paddle.
   (*b*) The force the paddle exerts on the water.
   (*c*) The force the water exerts on the paddle.
   (*d*) The motion of the water itself.



FIGURE 4–33 MisConceptual Question 4.

5. You are trying to push your stalled car. Although you apply a horizontal force of 400 N on the car, it doesn't budge, and neither do you. What other force(s) must also have a magnitude of 400 N?
(*a*) The force exerted by the car on you.
(*b*) The friction force exerted by the car on the road.
(*c*) The normal force exerted by the road on you.
(*d*) The friction force exerted by the road on you.

6. When a skier skis down a hill, the normal force exerted on the skier by the hill is
(*a*) equal to the weight of the skier.
(*b*) greater than the weight of the skier.
(*c*) less than the weight of the skier.

7. A golf ball is hit with a golf club. While the ball flies through the air, which forces act on the ball? Neglect air resistance.
(*a*) The force of the golf club acting on the ball.
(*b*) The force of gravity acting on the ball.
(*c*) The force of the ball moving forward through the air.
(*d*) All of the above.
(*e*) Both (*a*) and (*b*).

8. Suppose an object is accelerated by a force of 100 N. Suddenly a second force of 100 N in the opposite direction is exerted on the object, so that the forces cancel. The object
(*a*) quickly stops.
(*b*) decelerates gradually to rest.
(*c*) continues at the velocity it had before the second force was applied.
(*d*) is gradually brought to rest and then accelerates in the direction of the second force.

9. You are pushing a heavy box across a rough floor. When you are initially pushing the box and it is accelerating,
(*a*) you exert a force on the box, but the box does not exert a force on you.
(*b*) the box is so heavy it exerts a force on you, but you do not exert a force on the box.
(*c*) the force you exert on the box is greater than the force of the box pushing back on you.
(*d*) the force you exert on the box is equal to the force of the box pushing back on you.
(*e*) the force that the box exerts on you is greater than the force you exert on the box.

10. Two 5-newton boxes are attached to opposite ends of a spring scale and suspended from pulleys as shown in Fig. 4–34. What is the reading on the scale?
(*a*) 0 N.
(*b*) Between 0 and 5 N.
(*c*) 5 N.
(*d*) Between 5 and 10 N.
(*e*) 10 N.

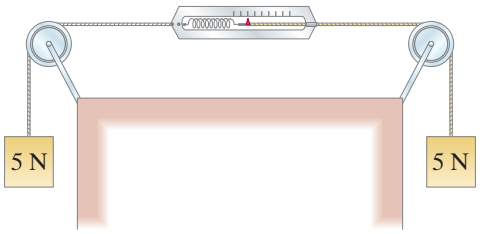


FIGURE 4–34 MisConceptual Question 10.

11. Two tug of war opponents each pull with a force of 500 N on opposite ends of a rope. Assume the rope is massless. What is the tension in the rope?
(*a*) 0 N.
(*b*) 250 N.
(*c*) 500 N.
(*d*) 1000 N.
(*e*) Impossible to tell.

12. The normal force on an extreme skier descending a very steep slope (Fig. 4–35) can be zero if
(*a*) his speed is great enough.
(*b*) he leaves the slope (no longer touches the snow).
(*c*) the slope is greater than 75°.
(*d*) the slope is vertical (90°).

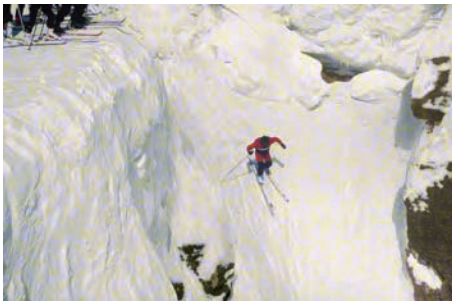


FIGURE 4–35 MisConceptual Question 12.

13. To pull an old stump out of the ground, you and a friend tie two ropes to the stump. You pull on it with a force of 500 N to the north while your friend pulls with a force of 450 N to the northwest. The total force exerted by the two ropes is
(*a*) less than 950 N.
(*b*) exactly 950 N.
(*c*) more than 950 N.

Problems

4–4 to 4–6 Newton's Laws, Gravitational Force, Normal Force [Assume no friction.]

1. (I) What net force is needed to accelerate a 45-kg sled at 1.4 m/s^2 on horizontal frictionless ice?

2. (I) What is the weight of a 74-kg astronaut (*a*) on Earth, (*b*) on the Moon ($g = 1.7\ m/s^2$), (*c*) on Mars ($g = 3.7\ m/s^2$), (*d*) in outer space traveling with constant velocity?

3. (I) How much tension must a rope withstand if it is used to accelerate a 1210-kg car horizontally along a frictionless surface at 1.35 m/s^2?

4. (I) A net force of 215 N accelerates a bike and rider at 2.30 m/s^2. What is the mass of the bike and rider together?

5. (II) What average force is required to stop a 950-kg car in 8.0 s if the car is traveling at 95 km/h?

6. (II) According to a simplified model of a *mammalian heart*, at each pulse approximately 20 g of blood is accelerated from 0.25 m/s to 0.35 m/s during a period of 0.10 s. What is the magnitude of the force exerted by the heart muscle?

7. (II) Superman must stop a 120-km/h train in 150 m to keep it from hitting a stalled car on the tracks. If the train's mass is 3.6×10^5 kg, how much force must he exert? Compare to the weight of the train (give as %). How much force does the train exert on Superman?

8. (II) A person has a reasonable chance of surviving an *automobile crash* if the deceleration is no more than 30 g's. Calculate the force on a 65-kg person accelerating at this rate. What distance is traveled if brought to rest at this rate from 85 km/h?

9. (II) Estimate the average force exerted by a shot-putter on a 7.0-kg shot if the shot is moved through a distance of 2.8 m and is released with a speed of 13 m/s.

10. (II) A 0.140-kg baseball traveling 35.0 m/s strikes the catcher's mitt, which, in bringing the ball to rest, recoils backward 11.0 cm. What was the average force applied by the ball on the glove?

11. (II) A fisherman yanks a fish vertically out of the water with an acceleration of 2.5 m/s^2 using very light fishing line that has a breaking strength of 18 N ($\approx$ 4 lb). The fisherman unfortunately loses the fish as the line snaps. What can you say about the mass of the fish?

12. (II) How much tension must a cable withstand if it is used to accelerate a 1400-kg car vertically upward at 0.70 m/s^2?

13. (II) A 20.0-kg box rests on a table. (*a*) What is the weight of the box and the normal force acting on it? (*b*) A 10.0-kg box is placed on top of the 20.0-kg box, as shown in Fig. 4–36. Determine the normal force that the table exerts on the 20.0-kg box and the normal force that the 20.0-kg box exerts on the 10.0-kg box.

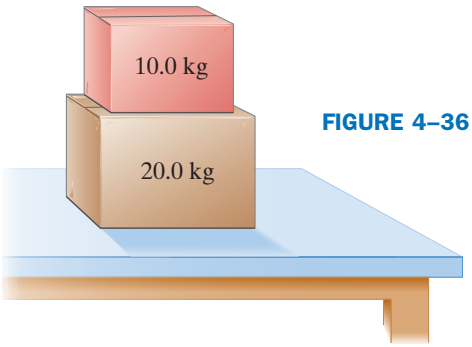


FIGURE 4–36 Problem 13.

14. (II) A particular race car can cover a quarter-mile track (402 m) in 6.40 s starting from a standstill. Assuming the acceleration is constant, how many horizontal "g's" does the driver experience? If the combined mass of the driver and race car is 535 kg, what horizontal force must the road exert on the tires? Ignore air resistance.

15. (II) A 14.0-kg bucket is lowered vertically by a rope in which there is 132 N of tension at a given instant. What is the acceleration of the bucket? Is it up or down?

16. (II) A 75-kg petty thief wants to escape from a third-story jail window. Unfortunately, a makeshift rope made of sheets tied together can support a mass of only 62 kg. How might the thief use this "rope" to escape? Give a quantitative answer.

17. (II) A woman stands on a bathroom scale in a motionless elevator. When the elevator begins to move, the scale briefly reads only 0.75 of her regular weight. Calculate the acceleration of the elevator, and find the direction of acceleration.

18. (II) Can cars "stop on a dime"? Calculate the acceleration of a 1400-kg car if it can stop from 35 km/h on a dime (diameter = 1.7 cm). How many g's is this? What is the force felt by the 68-kg occupant of the car?

19. (II) The cable supporting a 2375-kg elevator has a maximum strength of 24,950 N. What maximum upward acceleration can it give the elevator without breaking?

20. (II) Using focused laser light, *optical tweezers* can apply a force of about 10 pN (piconewtons) to a 1.0-μm-diameter polystyrene bead, which has a density about equal to that of water: a volume of 1.0 cm^3 has a mass of about 1.0 g. Estimate the bead's acceleration in g's.

21. (II) A *rocket* with a mass of 2.45×10^6 kg is launched by exerting a vertical force of 3.55×10^7 N on the gases it expels. Determine (*a*) the acceleration of the rocket, (*b*) its velocity after 8.0 s, and (*c*) how long it takes to reach an altitude of 9500 m. Assume g remains constant, and ignore the mass of gas expelled (not realistic).

22. (II) (*a*) What is the acceleration of two falling *sky divers* (total mass = 148 kg including parachute) when the upward force of air resistance is equal to one-fourth of their weight? (*b*) After opening the parachute, the divers descend leisurely to the ground at constant speed. What now is the force of air resistance on the sky divers and their parachute?

23. (II) An elevator (mass 4850 kg, with people) is to be designed so that the maximum acceleration is 0.0640 g. What are the maximum and minimum forces the motor should exert on the supporting cable?

24. (II) An exceptional standing jump would raise a person 0.80 m off the ground. To do this, what force must a 68-kg person exert against the ground? Assume the person crouches a distance of 0.20 m prior to jumping, and thus the upward force has this distance to act over before he leaves the ground.

25. (II) High-speed *elevators* function under two limitations: (1) the maximum magnitude of vertical acceleration that a typical human body can experience without discomfort is about 1.2 m/s^2, and (2) the typical maximum speed attainable is about 9.0 m/s. You board an elevator on a skyscraper's ground floor and are transported 180 m above the ground level in three steps: acceleration of magnitude 1.2 m/s^2 from rest to 9.0 m/s, followed by constant upward velocity of 9.0 m/s, then deceleration of magnitude 1.2 m/s^2 from 9.0 m/s to rest. (*a*) Determine the elapsed time for each of these 3 stages. (*b*) Determine the change in the magnitude of the normal force, expressed as a % of your normal weight during each stage. (*c*) What fraction of the total transport time does the normal force not equal the person's weight?

26. (III) The 100-m dash can be run by the best sprinters in 10.0 s. A 66-kg sprinter accelerates uniformly for the first 45 m to reach top speed, which he maintains for the remaining 55 m. (*a*) What is the average horizontal component of force exerted on his feet by the ground during acceleration? (*b*) What is the speed of the sprinter over the last 55 m of the race (that is, his top speed)?

27. (III) A person jumps from the roof of a house 2.8 m high. When he strikes the ground below, he bends his knees so that his torso decelerates over an approximate distance of 0.70 m. If the mass of his torso (excluding legs) is 42 kg, find (*a*) his velocity just before his feet strike the ground, and (*b*) the average force exerted on his torso by his legs during deceleration.

4–7 Newton's Laws and Vectors [Ignore friction.]

28. (I) A box weighing 66.0 N rests on a table. A rope tied to the box runs vertically upward over a pulley and a weight is hung from the other end (Fig. 4–37). Determine the force that the table exerts on the box if the weight hanging on the other side of the pulley weighs (*a*) 30.0 N, (*b*) 60.0 N, and (*c*) 90.0 N.

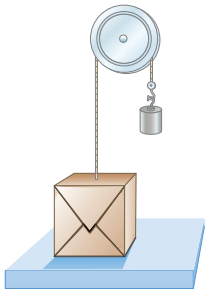


FIGURE 4–37
Problem 28.

29. (I) Draw the free-body diagram for a basketball player (*a*) just before leaving the ground on a jump, and (*b*) while in the air. See Fig. 4–38.



FIGURE 4–38
Problem 29.

30. (I) A 650-N force acts in a northwesterly direction. A second 650-N force must be exerted in what direction so that the resultant of the two forces points westward? Illustrate your answer with a vector diagram.

31. (I) Sketch the free-body diagram of a baseball (*a*) at the moment it is hit by the bat, and again (*b*) after it has left the bat and is flying toward the outfield. Ignore air resistance.

32. (II) Arlene is to walk across a "high wire" strung horizontally between two buildings 10.0 m apart. The sag in the rope when she is at the midpoint is 10.0°, as shown in Fig. 4–39. If her mass is 50.0 kg, what is the tension in the rope at this point?

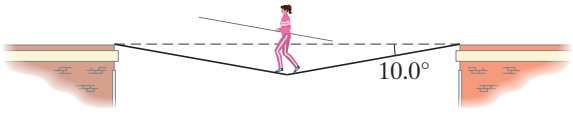


FIGURE 4–39 Problem 32.

33. (II) A window washer pulls herself upward using the bucket–pulley apparatus shown in Fig. 4–40. (*a*) How hard must she pull downward to raise herself slowly at constant speed? (*b*) If she increases this force by 15%, what will her acceleration be? The mass of the person plus the bucket is 78 kg.



FIGURE 4–40
Problem 33.

34. (II) One 3.2-kg paint bucket is hanging by a massless cord from another 3.2-kg paint bucket, also hanging by a massless cord, as shown in Fig. 4–41. (*a*) If the buckets are at rest, what is the tension in each cord? (*b*) If the two buckets are pulled upward with an acceleration of 1.45 m/s^2 by the upper cord, calculate the tension in each cord.

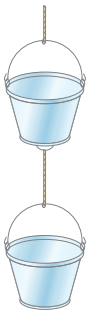


FIGURE 4–41
Problems 34 and 35.

35. (II) The cords accelerating the buckets in Problem 34b, Fig. 4–41, each have a weight of 2.0 N. Determine the tension in each cord at the three points of attachment.

36. (II) Two large snowcats are towing a housing unit north, as shown in Fig. 4–42. The sum of the forces $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ exerted on the unit by the horizontal cables is north, parallel to the line L, and $F_A = 4200$ N. Determine F_B and the magnitude of $\vec{\mathbf{F}}_A + \vec{\mathbf{F}}_B$.

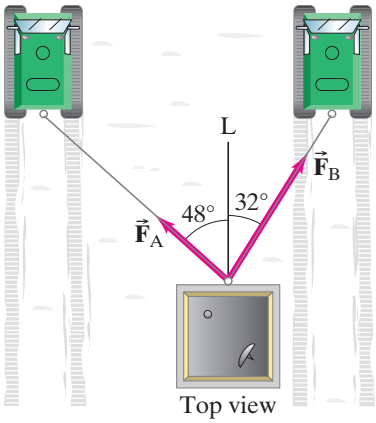


FIGURE 4–42
Problem 36.

37. (II) A train locomotive is pulling two cars of the same mass behind it, Fig. 4–43. Determine the ratio of the tension in the coupling (think of it as a cord) between the locomotive and the first car (F_{T1}) to that between the first car and the second car (F_{T2}), for any nonzero acceleration of the train.

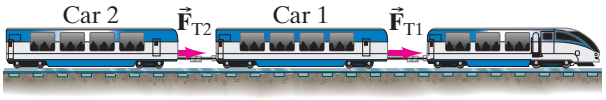


FIGURE 4–43 Problem 37.

38. (II) The two forces $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ shown in Fig. 4–44a and b (looking down) act on an 18.5-kg object on a frictionless tabletop. If $F_1 = 10.4$ N and $F_2 = 16.2$ N, find the net force on the object and its acceleration for (a) and (b).

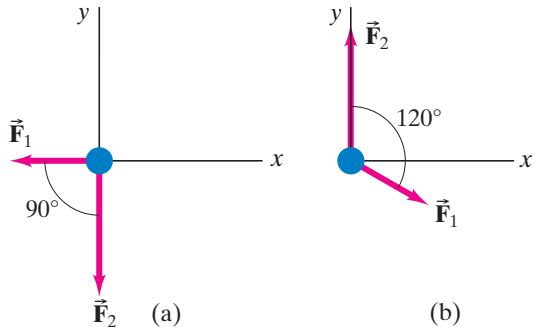


FIGURE 4–44 Problem 38.

39. (II) A skateboarder, with an initial speed of 2.0 m/s, rolls virtually friction free down a straight incline of length 18 m in 3.3 s. At what angle θ is the incline oriented above the horizontal?

40. (II) At the instant a race began, a 65-kg sprinter exerted a force of 720 N on the starting block at a 22° angle with respect to the ground. (*a*) What was the horizontal acceleration of the sprinter? (*b*) If the force was exerted for 0.32 s, with what speed did the sprinter leave the starting block?

41. (II) A mass m is at rest on a horizontal frictionless surface at $t = 0$. Then a constant force F_0 acts on it for a time t_0. Suddenly the force doubles to $2F_0$ and remains constant until $t = 2t_0$. Determine the total distance traveled from $t = 0$ to $t = 2t_0$.

42. (II) A 3.0-kg object has the following two forces acting on it:

$$\vec{\mathbf{F}}_1 = (16\hat{\mathbf{i}} + 12\hat{\mathbf{j}})\,\text{N}$$
$$\vec{\mathbf{F}}_2 = (-10\hat{\mathbf{i}} + 22\hat{\mathbf{j}})\,\text{N}$$

If the object is initially at rest, determine its velocity $\vec{\mathbf{v}}$ at $t = 4.0$ s.

43. (II) A 27-kg chandelier hangs from a ceiling on a vertical 3.4-m-long wire. (*a*) What horizontal force would be necessary to displace its position 0.15 m to one side? (*b*) What will be the tension in the wire?

44. (II) Redo Example 4–13 but (*a*) set up the equations so that the direction of the acceleration $\vec{\mathbf{a}}$ of each object is in the direction of motion of that object. (In Example 4–13, we took $\vec{\mathbf{a}}$ as positive upward for both masses.) (*b*) Solve the equations to obtain the same answers as in Example 4–13.

45. (II) The block shown in Fig. 4–45 has mass $m = 7.0$ kg and lies on a fixed smooth frictionless plane tilted at an angle $\theta = 22.0°$ to the horizontal. (*a*) Determine the acceleration of the block as it slides down the plane. (*b*) If the block starts from rest 12.0 m up the plane from its base, what will be the block's speed when it reaches the bottom of the incline?

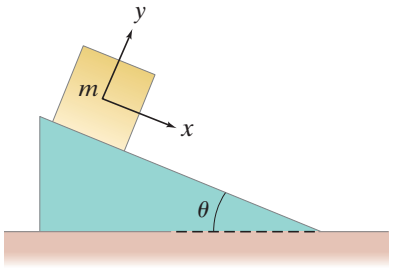


FIGURE 4–45 Block on inclined plane. Problems 45 and 46.

46. (II) A block is given an initial speed of 5.2 m/s up the 22.0° plane shown in Fig. 4–45. (*a*) How far up the plane will it go? (*b*) How much time elapses before it returns to its starting point? Ignore friction.

47. (II) An object is hanging by a string from your rearview mirror. While you are accelerating at a constant rate from rest to 28 m/s in 5.0 s, what angle θ does the string make with the vertical? See Fig. 4–46.

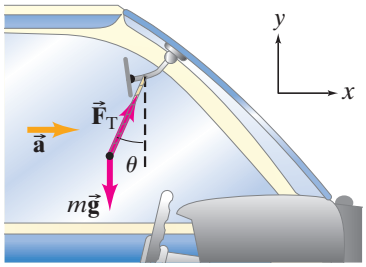


FIGURE 4–46 Problem 47.

48. (II) A 2.0-kg purse is dropped from the top of the Leaning Tower of Pisa and falls 55 m before reaching the ground with a speed of 27 m/s. What was the average force of air resistance?

49. (II) Bob traverses a chasm by stringing a rope between a tree on one side of the chasm and a tree on the opposite side, 25 m away, Fig. 4–47. Assume the rope can provide a tension force of up to 31 kN before breaking, and use a "safety factor" of 10 (that is, the rope should only be required to undergo a tension force of 3.1 kN). (*a*) If Bob's mass is 72.0 kg, determine the distance x that the rope must sag at a point halfway across if it is to be within its recommended safety range. (*b*) If the rope sags by only one-fourth the distance found in (*a*), determine the tension force in the rope. Will the rope break?

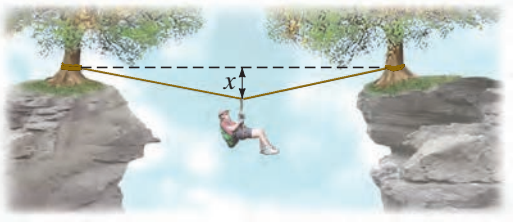


FIGURE 4–47 Problem 49.

50. (II) As shown in Fig. 4–48, five balls (masses 2.00, 2.05, 2.10, 2.15, 2.20 kg) hang from a crossbar. Each mass is supported by "5-lb test" fishing line which will break when its tension force exceeds 22.2 N (= 5.00 lb). When this device is placed in an elevator, which accelerates upward, only the lines attached to the 2.05 and 2.00 kg masses do not break. Within what range is the elevator's acceleration?

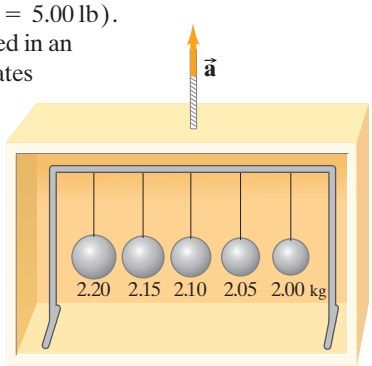


FIGURE 4–48 Problem 50.

51. (II) A high-speed 14-car Italian train has a mass of 640 metric tons (640,000 kg). It can exert a maximum force of 400 kN horizontally against the tracks, whereas at maximum constant velocity (300 km/h), it exerts a force of about 150 kN. Calculate (*a*) its maximum acceleration, and (*b*) estimate the total force of friction and air resistance at top speed.

52. (II) A 450-kg piano is being unloaded from a truck by rolling it down a ramp inclined at 15°. There is negligible friction and the ramp is 4.0 m long. Two workers slow the rate at which the piano moves by pushing with a combined force of 1020 N parallel to the ramp. If the piano starts from rest, how fast is it moving at the bottom?

53. (II) Uphill escape ramps are sometimes provided to the side of steep downhill highways for trucks with overheated brakes. For a simple 11° upward ramp, what length would be needed to stop a runaway truck traveling 140 km/h? Note the large size of your calculated length. (If sand is used for the bed of the ramp, its length can be reduced by a factor of about 2.)

54. (II) A child on a sled reaches the bottom of a hill with a velocity of 10.0 m/s and travels 25.0 m along a horizontal straightaway to a stop. If the child and sled together have a mass of 60.0 kg, what is the average retarding force on the sled on the horizontal straightaway?

55. (II) Figure 4–49 shows a block (mass m_A) on a smooth horizontal surface, connected by a thin cord that passes over a pulley to a second block (m_B), which hangs vertically. (*a*) Draw a free-body diagram for each block, showing the force of gravity on each, the force (tension) exerted by the cord, and any normal force. (*b*) Apply Newton's second law to find formulas for the acceleration of the system and for the tension in the cord. Ignore friction and the masses of the pulley and cord.

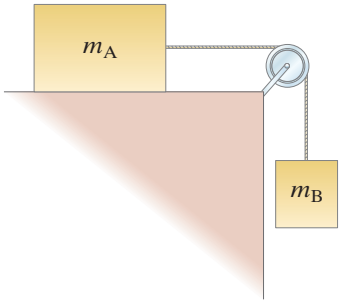


FIGURE 4–49 Problems 55, 56, and 57. Mass m_A rests on a smooth horizontal surface; m_B hangs vertically.

56. (II) (*a*) If $m_A = 14.0\,\text{kg}$ and $m_B = 5.0\,\text{kg}$ in Fig. 4–49, determine the acceleration of each block. (*b*) If initially m_A is at rest 1.250 m from the edge of the table, how long does it take to reach the edge of the table if the system is allowed to move freely? (*c*) If $m_B = 1.0\,\text{kg}$, how large must m_A be if the acceleration of the system is to be kept at $\frac{1}{100}g$?

57. (III) Determine a formula for the acceleration of the system shown in Fig. 4–49 (see Problem 55) if the cord has a non-negligible mass m_C. Specify in terms of ℓ_A and ℓ_B, the lengths of cord from the respective masses to the pulley. (The total cord length is $\ell = \ell_A + \ell_B$.)

58. (III) Suppose the pulley in Fig. 4–50 is suspended by a cord C. Determine the tension in this cord after the masses are released and before one hits the ground. Ignore the mass of the pulley and cords.

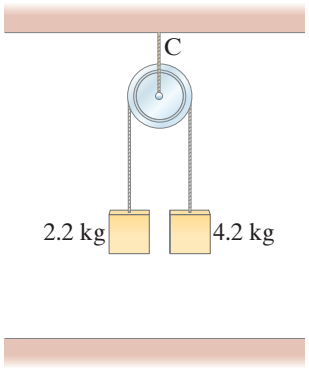


FIGURE 4–50 Problem 58.

59. (III) Suppose two boxes on a frictionless table are connected by a heavy cord of mass 1.0 kg. Calculate the acceleration of each box and the tension at each end of the cord, using the free-body diagrams shown in Fig. 4–51. Assume $F_P = 40.0\,\text{N}$, and ignore sagging of the cord. Compare your results to Example 4–12 and Fig. 4–22.

60. (III) Three blocks on a frictionless horizontal surface are in contact with each other as shown in Fig. 4–52. A force $\vec{\mathbf{F}}$ is applied to block A (mass m_A). (*a*) Draw a free-body diagram for each block. Determine (*b*) the acceleration of the system (in terms of m_A, m_B, and m_C), (*c*) the net force on each block, and (*d*) the force of contact that each block exerts on its neighbor. (*e*) If $m_A = m_B = m_C = 10.0\,\text{kg}$ and $F = 96.0\,\text{N}$, give numerical answers to (*b*), (*c*), and (*d*). Explain how your answers make sense intuitively.

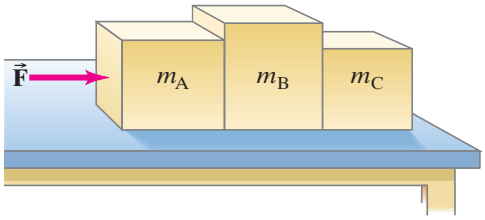


FIGURE 4–52 Problem 60.

61. (III) A 2.5-kg block is placed on a frictionless table. The block is connected by massless ropes over massless pulleys to a 5.0-kg block on the right, and a 3.0-kg block on the left, as shown in Fig. 4–53. Find the acceleration of the block on the table.

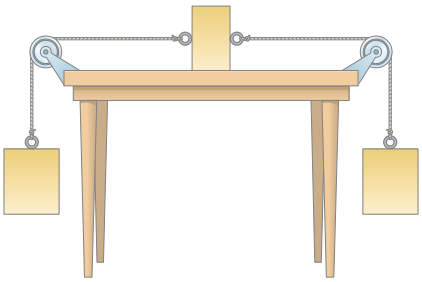


FIGURE 4–53 Problem 61.

62. (III) A small block of mass m rests on the sloping side of a triangular block of mass M which itself rests on a horizontal table as shown in Fig. 4–54. Assuming all surfaces are frictionless, determine the magnitude of the force $\vec{\mathbf{F}}$ that must be applied to M so that m remains in a fixed position relative to M (that is, m doesn't move on the incline). [*Hint*: Take x and y axes horizontal and vertical.]

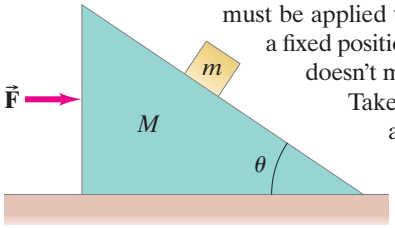
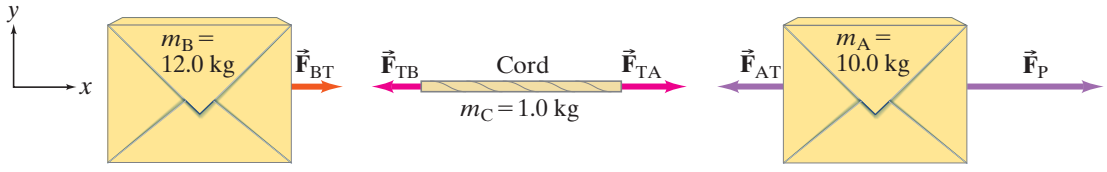


FIGURE 4–54 Problem 62.

FIGURE 4–51 Problem 59. Free-body diagrams for each of the objects of the system shown in Fig. 4–22a. Vertical forces, $\vec{\mathbf{F}}_N$ and $\vec{\mathbf{F}}_G$, are not shown.



(a) (b) (c)

63. (III) The double Atwood machine shown in Fig. 4–55 has frictionless, massless pulleys and cords. Determine (*a*) the acceleration of masses m_A, m_B, and m_C, and (*b*) the tensions F_{TA} and F_{TC} in the cords.

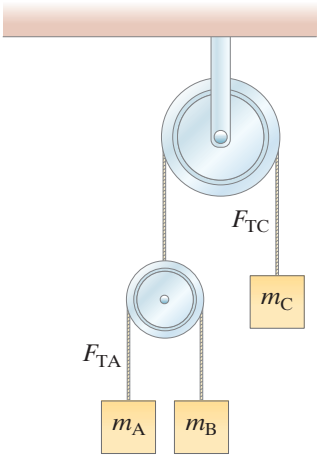


FIGURE 4–55 Problem 63.

64. (III) Determine a formula for the magnitude of the force $\vec{\mathbf{F}}$ exerted on the large block (m_C) in Fig. 4–56 so that the mass m_A does not move relative to m_C. Ignore all friction. Assume m_B does not make contact with m_C.

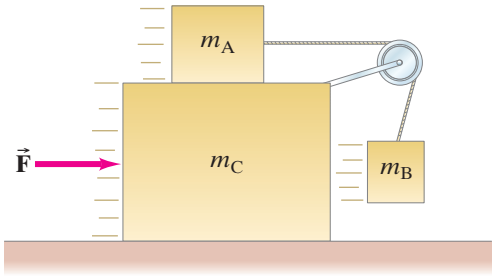


FIGURE 4–56 Problem 64.

65. (III) The two masses shown in Fig. 4–57 are each initially 1.8 m above the ground, and the massless frictionless pulley is 4.8 m above the ground. What maximum height does the lighter object reach after the system is released? [*Hint*: First determine the acceleration of the lighter mass and then its velocity at the moment the heavier one hits the ground. This is its "launch" speed. Assume the mass doesn't hit the pulley or the ceiling. Ignore the mass of the cord.]

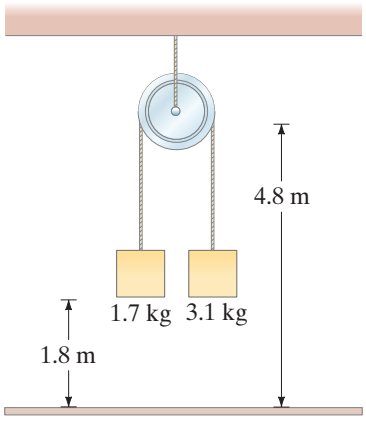


FIGURE 4–57 Problem 65.

66. (III) A particle of mass m, initially at rest at $x = 0$, is accelerated by a force that increases in time as $F = Ct^2$. Determine its velocity v and position x as a function of time.

General Problems

67. A crane's trolley at point P in Fig. 4–58 moves for a few seconds to the right with constant acceleration, and the 780-kg load hangs on a light cable at a 5.0° angle to the vertical as shown. What is the acceleration of the trolley and load?

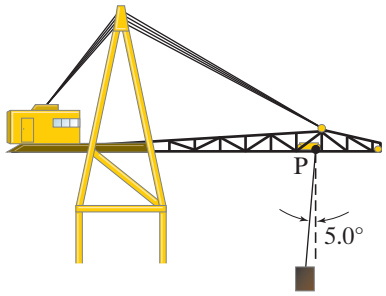


FIGURE 4–58 Problem 67.

68. A 75.0-kg person stands on a scale in an elevator. What does the scale read (in N and in kg) when (*a*) the elevator is at rest, (*b*) the elevator is climbing at a constant speed of 2.0 m/s, (*c*) the elevator is descending at 2.0 m/s, (*d*) the elevator is accelerating upward at 2.0 m/s^2, (*e*) the elevator is accelerating downward at 2.0 m/s^2?

69. A *city planner* is working on the redesign of a hilly portion of a city. An important consideration is how steep the roads can be so that even low-powered cars can get up the hills without slowing down. A particular small car, with a mass of 920 kg, can accelerate on a level road from rest to 21 m/s (75 km/h) in 14.5 s. Using these data, calculate the maximum steepness of a hill.

70. If a bicyclist of mass 65 kg (including the *bicycle*) can coast down a 6.5° hill at a steady speed of 6.0 km/h because of air resistance, how much force must be applied to climb the hill at the same speed (and the same air resistance)?

71. Tom's *hang glider* supports his weight using the six ropes shown in Fig. 4–59. Each rope is designed to support an equal fraction of Tom's weight. Tom's mass is 78.0 kg. What is the tension in each of the support ropes?

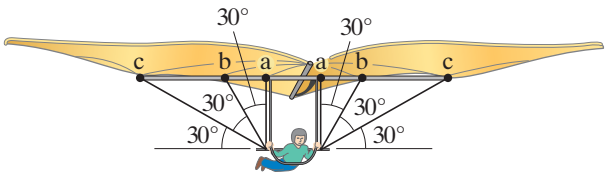


FIGURE 4–59 Problem 71.

72. A wet bar of soap $(m = 150\,\text{g})$ slides freely down a ramp 3.0 m long inclined at 8.5°. How long does it take to reach the bottom? How would this change if the soap's mass were 250 g?

73. A parent pulls his child in a wagon across a floor. The child and wagon have a combined mass of 35 kg. The wagon handle is inclined at 55° to the horizontal, and the child and wagon (whose wheels are nearly frictionless) are accelerating at 0.42 m/s^2. With what force is the parent pulling on the wagon handle?

74. Two equal masses in contact on a frictionless surface are acted on by the forces F_1 and F_2 as shown in Fig. 4–60. Determine the magnitude of the contact force exerted on each mass by the other when (*a*) $F_2 = F_1$; (*b*) $F_2 = \frac{1}{2}F_1$; (*c*) $F_2 = 0$. Express your answer in terms of m and F_1.

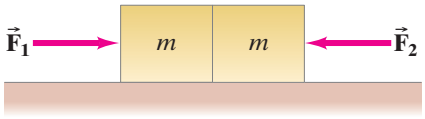


FIGURE 4–60 Problem 74.

75. Andrea dangles her watch from a thin piece of string while the jetliner she is in accelerates for takeoff, which takes about 21 s. Estimate the takeoff speed of the aircraft if the string makes an angle of 24° with respect to the vertical, Fig. 4–61.

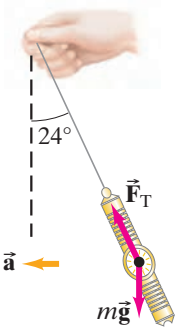


FIGURE 4–61
Problem 75.

76. A block (mass m_A) lying on a fixed frictionless inclined plane is connected to a mass m_B by a cord passing over a pulley, as shown in Fig. 4–62. (*a*) Determine a formula for the acceleration of the system in terms of m_A, m_B, θ, and g. (*b*) What conditions apply to masses m_A and m_B for the acceleration to be in one direction (say, m_A down the plane), or in the opposite direction? Ignore the mass of the cord and pulley.

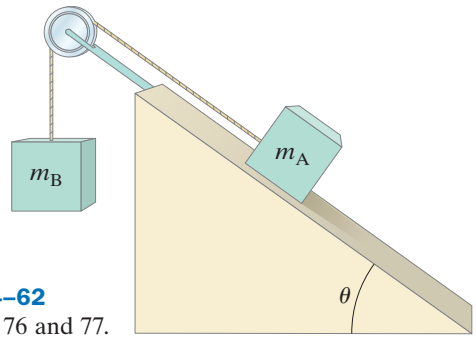


FIGURE 4–62
Problems 76 and 77.

77. (*a*) In Fig. 4–62, if $m_A = m_B = 1.00\,\text{kg}$ and $\theta = 38.0°$, what will be the acceleration of the system? (*b*) If $m_A = 1.00\,\text{kg}$ and the system remains at rest, what must the mass m_B be? (*c*) Calculate the tension in the cord for (*a*) and (*b*).

78. The masses m_A and m_B slide on the smooth (frictionless) inclines fixed as shown in Fig. 4–63. (*a*) Determine a formula for the acceleration of the system in terms of m_A, m_B, θ_A, θ_B, and g. (*b*) If $\theta_A = 34°$, $\theta_B = 21°$, and $m_A = 5.0\,\text{kg}$, what value of m_B would keep the system at rest? What would be the tension in the cord (negligible mass) in this case? (*c*) What ratio, m_A/m_B, would allow the masses to move at constant speed along their ramps in either direction?

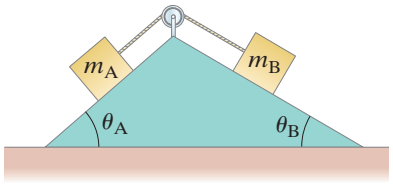


FIGURE 4–63
Problem 78.

79. (*a*) What minimum force F is needed to lift a piano (mass M) using the pulley apparatus shown in Fig. 4–64? (*b*) Determine the tension in each section of rope: F_{T1}, F_{T2}, F_{T3}, and F_{T4}. Assume pulleys are massless and frictionless, and that ropes are massless.

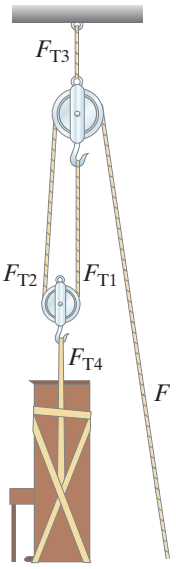


FIGURE 4–64
Problem 79.

80. In the *design of a supermarket*, there are to be several ramps connecting different parts of the store. Customers will have to push grocery carts up the ramps and it is desirable that this not be too difficult. The engineer has done a survey and found that almost no one complains if the force required is no more than 23 N. Ignoring friction, at what maximum angle θ should the ramps be built, assuming a full 25-kg cart?

81. A jet aircraft is accelerating at $3.8\,\text{m/s}^2$ as it climbs at an angle of 18° above the horizontal (Fig. 4–65). What is the total force that the cockpit seat exerts on the 75-kg pilot?

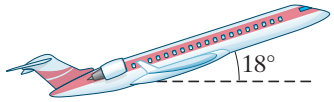


FIGURE 4–65 Problem 81.

82. A 6750-kg helicopter accelerates upward at $0.80\,\text{m/s}^2$ while lifting a 1080-kg frame at a construction site, Fig. 4–66. (*a*) What is the lift force exerted by the air on the helicopter rotors? (*b*) What is the tension in the cable (ignore its mass) which connects the frame to the helicopter? (*c*) What force does the cable exert on the helicopter?

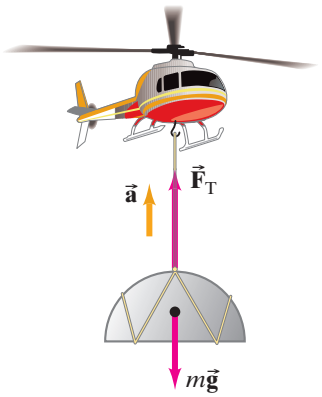


FIGURE 4–66 Problem 82.

83. An elevator in a tall building is allowed to reach a maximum speed of 3.5 m/s going down. What must the tension be in the cable to stop this elevator over a distance of 2.6 m if the elevator has a mass of 1650 kg including occupants?

84. A fisherman in a boat is using a "10-lb test" fishing line. This means that the line can exert a force of 45 N without breaking (1 lb = 4.45 N). (*a*) How heavy a fish can the fisherman land if he pulls the fish up vertically at constant speed? (*b*) If he accelerates the fish upward at 2.0 m/s^2, what maximum weight fish can he land? (*c*) Is it possible to land a 15-lb trout on 10-lb test line? Why or why not?

85. A "doomsday" *asteroid* with a mass of 2.0×10^{10} kg is hurtling through space. Unless the asteroid's speed is changed by about 0.20 cm/s, it will collide with Earth and cause tremendous damage. Researchers suggest that a small "space tug" sent to the asteroid's surface could exert a gentle constant force of 2.5 N. For how long must this force act?

86. Three mountain climbers who are roped together in a line are ascending an icefield inclined at 29° to the horizontal (Fig. 4–67). The last climber slips, pulling the second climber off his feet. The first climber is able to hold them both. If each climber has a mass of 75 kg, calculate the tension in each of the two sections of rope between the three climbers. Ignore friction between the ice and the fallen climbers.

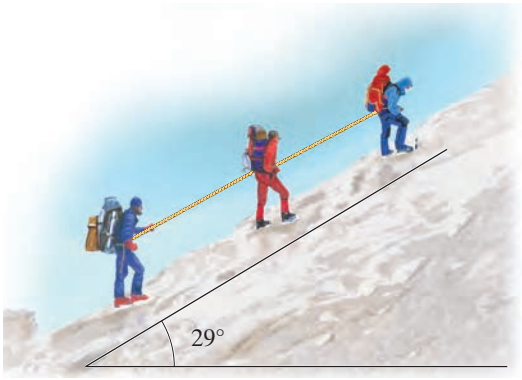


FIGURE 4–67 Problem 86.

87. A *bicyclist* can coast down a 5.0° hill at a constant speed of 6.0 km/h. If the force of air resistance is proportional to the speed v so that $F_{air} = cv$, calculate (*a*) the value of the constant c, and (*b*) the average force that must be applied (by the rider) in order to descend the hill at 18.0 km/h. The mass of the cyclist plus bicycle is 72.0 kg.

88. Consider the system shown in Fig. 4–68 with $m_A = 8.2$ kg and $m_B = 11.5$ kg. The angles $\theta_A = 59°$ and $\theta_B = 32°$. (*a*) In the absence of friction, what force $\vec{\mathbf{F}}$ would be required to pull the masses at a constant velocity up the fixed inclines? (*b*) The force $\vec{\mathbf{F}}$ is now removed. What are the magnitude and direction of the acceleration of the two blocks? (*c*) In the absence of $\vec{\mathbf{F}}$, what is the tension in the string?

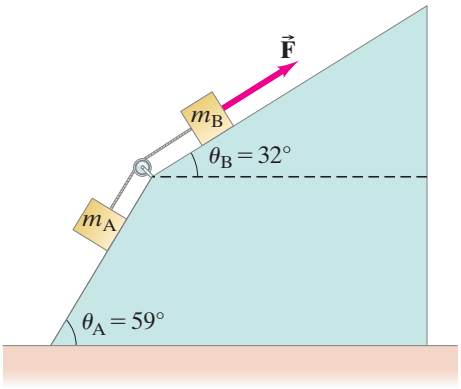


FIGURE 4–68
Problem 88.

89. A car starts rolling down a 1-in-4 hill (1-in-4 means that for each 4 m traveled along the sloping road, the elevation change is 1 m). How fast is it going when it reaches the bottom after traveling 55 m? Ignore friction.

90. An 18-kg child is riding in a child-restraint chair, securely fastened to the seat of a car (Fig. 4–69). Assume the car has speed 45 km/h when it hits a tree and is brought to rest in 0.20 s. Assuming constant deceleration during the collision, estimate the net horizontal force F that the straps of the restraint chair exert on the child to hold her in the chair.



FIGURE 4–69 Problem 90.

91. A 1.5-kg block rests on top of a 7.5-kg block (Fig. 4–70). The cord and pulley have negligible mass, and there is no significant friction anywhere. (*a*) What force F must be applied to the bottom block so the top block accelerates to the right at 2.2 m/s^2? (*b*) What is the tension in the connecting cord?

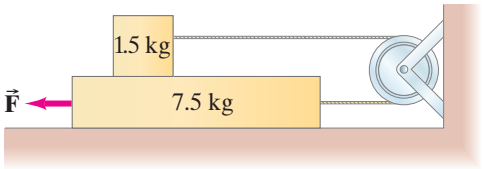


FIGURE 4–70 Problem 91.

92. You are driving home in your 860-kg car at 15 m/s. At a point 45 m from the beginning of an intersection, you see a green traffic light change to yellow, which you expect will last 4.0 s, and the distance to the far side of the intersection is 65 m (Fig. 4–71). (*a*) If you choose to accelerate, your car's engine will furnish a forward force of 1200 N. Will you make it completely through the intersection before the light turns red? (*b*) If you decide to panic stop, your brakes will provide a force of 2300 N. Will you stop before entering the intersection? Assume your car is 4 m long.

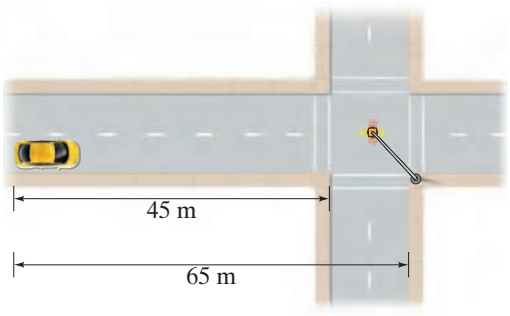


FIGURE 4–71 Problem 92.

93. A person jumps from a height of 5.0 m into a swimming pool entering feet first with her hands at her sides. She maintains this position upon entering the water and experiences a constant upward force equal to 350% of her weight due to the water itself. How far down does she go?

94. Two rock climbers, Paul and Jeanne, use safety ropes of similar length. Jeanne's rope is more elastic, called a *dynamic rope* by climbers. Paul has a *static rope*, not recommended for safety reasons. (*a*) Jeanne (Fig. 4–72) falls freely about 2.0 m and then the rope stops her over a distance of 1.0 m. Estimate how large a force (assume constant) she will feel from the rope. (Express the result in multiples of her weight.) (*b*) In a similar fall, Paul's rope stretches by only 30 cm. How many times his weight will the rope pull on him? Which climber is more likely to be hurt?



FIGURE 4–72
Problem 94.

95. (*a*) Finding her car stuck in the mud, a bright graduate of a good physics course ties a strong rope to the back bumper of the car, and the other end to a boulder, as shown in Fig. 4–73a. She pushes at the midpoint of the rope with her maximum effort, which she estimates to be a force $F_P \approx 300$ N. The car just begins to budge with the rope at an angle θ, which she estimates to be 5°. With what force is the rope pulling on the car? Neglect the mass of the rope. (*b*) What is the "mechanical advantage" of this technique (see Example 4–14)? (*c*) At what angle θ would this technique become counterproductive? [*Hint*: Consider the forces on a small segment of rope where $\vec{\mathbf{F}}_P$ acts, Fig. 4–73b.]

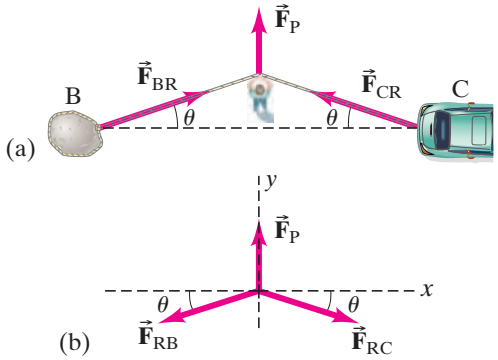
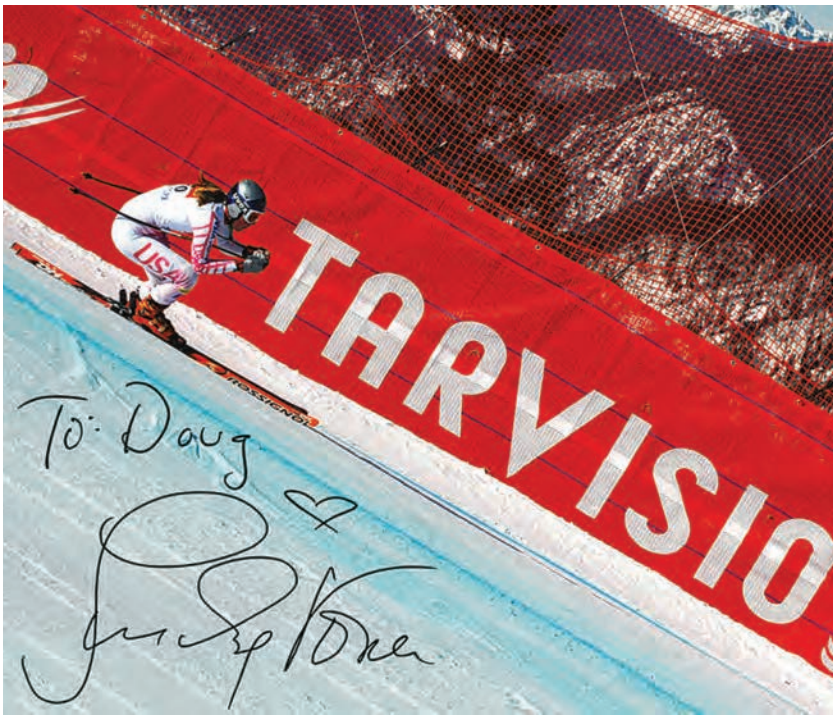


FIGURE 4–73 Problem 95. (a) Getting a car out of the mud, showing the forces on the boulder, on the car, and exerted by the person. (b) The free-body diagram: forces on a small segment of rope.

ANSWERS TO EXERCISES

A: No force is needed. The car accelerates out from under the cup, which tends to remain at rest as seen from the reference frame of the street. Think of Newton's first law (see Example 4–1).
B: (i) The same; (ii) the tennis ball; (iii) Newton's third law for part (i), second law for part (ii).
C: (*b*).
D: (*a*).
E: (*b*).
F: 58.0 N.
G: (*b*).
H: Yes; no.



Newton's laws are fundamental in physics. These photos show two situations of using Newton's laws which involve some new elements in addition to those discussed in the previous Chapter. The downhill skier illustrates *friction* on an incline; she is also retarded by air resistance, a velocity-dependent force.

The people on the rotating amusement park ride below illustrate the dynamics of circular motion.



CHAPTER 5

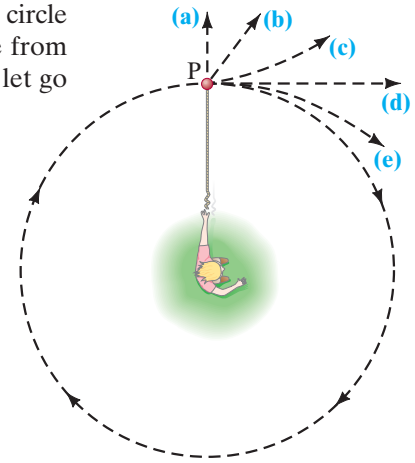
Using Newton's Laws: Friction, Circular Motion, Drag Forces

CONTENTS

5–1 Using Newton's Laws with Friction
5–2 Uniform Circular Motion—Kinematics
5–3 Dynamics of Uniform Circular Motion
5–4 Highway Curves: Banked and Unbanked
5–5 Nonuniform Circular Motion
*5–6 Velocity-Dependent Forces: Drag and Terminal Velocity

CHAPTER-OPENING QUESTION—Guess now!

You revolve a ball around you in a horizontal circle at constant speed on a string, as shown here from above. Which path will the ball follow if you let go of the string when the ball is at point P?



This chapter continues our study of Newton's laws and emphasizes their fundamental importance in physics. We will see how to apply Newton's laws to understand important situations including friction, and in circular motion, as well as with drag forces which are velocity-dependent (an optional-advanced Section).

5–1 Using Newton's Laws with Friction

Until now we have mostly ignored friction, but it must be taken into account in most practical situations. Friction exists between two solid surfaces because even the smoothest looking surface is quite rough on a microscopic scale, Fig. 5–1. When we try to slide an object across a surface, these microscopic bumps impede the motion. Exactly what is happening at the microscopic level is not fully understood. One possibility is that the atoms on a bump of one surface may come so close to the atoms of the other surface that the atoms form a sort of "bond" or brief tiny weld between the two surfaces. Sliding an object across a surface is often jerky, perhaps due to the making and breaking of these bonds.

Even when a round object rolls across a surface, there is still some friction, called *rolling friction*, although it is generally much less than when an object slides across a surface.

We focus our attention now on sliding friction, which is usually called **kinetic friction** (*kinetic* is from the Greek for "moving"). When an object slides along a rough surface, the force of kinetic friction acts opposite to the direction of the object's velocity. The magnitude of the force of kinetic friction depends on the nature of the two sliding surfaces. For given surfaces, experiments show that the friction force is approximately proportional to the *normal force* between the two surfaces, which is the force that either object exerts on the other and is perpendicular to their common surface of contact (see Fig. 5–2). The force of friction in many cases depends very little on the total surface area of contact; that is, the friction force on a book is roughly the same whether it is being slid across a table on its wide face or on its spine, assuming the surfaces have similar smoothness. We consider a simple model of friction in which we make this assumption that the friction force is independent of area. Then we write the proportionality between the magnitudes of the friction force F_{fr} and the normal force F_N as an equation by inserting a constant of proportionality, μ_k:

$$F_{fr} = \mu_k F_N. \qquad \text{[kinetic friction]}$$

This relation is not a fundamental law. It is an experimental relation between the magnitude of the friction force F_{fr}, which acts parallel to the two surfaces, and the magnitude of the normal force F_N, which acts perpendicular to the surfaces. It is *not* a vector equation since the two forces have different directions, perpendicular to one another. The term μ_k is called the *coefficient of kinetic friction*, and its value depends on the roughness of the two surfaces. Measured values for a variety of surfaces are given in Table 5–1. These are only approximate, however, since μ depends on whether the surfaces are wet or dry, on how much they have been sanded or rubbed, if any burrs remain, and other such factors. But μ_k (which has no units) is roughly independent of the sliding speed, as well as the area in contact.

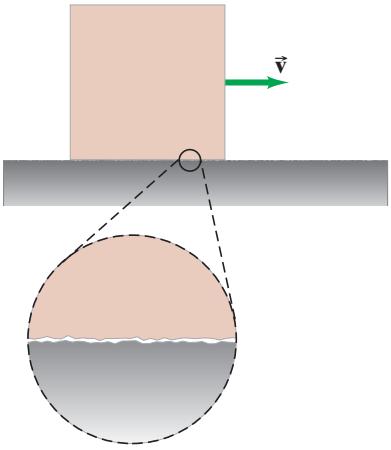


FIGURE 5–1 An object moving to the right on a table or floor. The two surfaces in contact are rough, at least on a microscopic scale.

FIGURE 5–2 When an object is pulled along a surface by an applied force $(\vec{\mathbf{F}}_A)$, the force of friction $\vec{\mathbf{F}}_{fr}$ opposes the motion. The magnitude of $\vec{\mathbf{F}}_{fr}$ is proportional to the magnitude of the normal force (F_N).

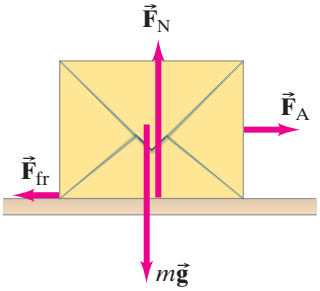


TABLE 5–1 Coefficients of Friction†

Surfaces	Coefficient of Kinetic Friction, μ_k	Coefficient of Static Friction, μ_s
Wood on wood	0.2	0.4
Ice on ice	0.03	$\lesssim$0.1
Metal on metal (lubricated)	0.07	0.1
Steel on steel (unlubricated)	0.6	0.7
Rubber on dry concrete	0.8	0.8
Rubber on wet concrete	0.5	0.7
Rubber on other solid surfaces	1	1–4
Teflon® on Teflon in air	0.04	0.04
Teflon on steel in air	0.04	0.04
Lubricated ball bearings	<0.01	<0.01
Synovial joints (in human limbs)	<0.01	<0.01

†Values are approximate and intended only as a guide.

What we have been discussing up to now is *kinetic friction*, when one object slides over another. There is also **static friction**, which refers to a force parallel to the two surfaces that can arise even when they are not moving. Suppose an object such as a desk is resting on a horizontal floor. If no horizontal force is exerted on the desk, there also is no friction force. But now suppose you try to push the desk, and it doesn't move. You are exerting a horizontal force, but the desk isn't moving, so there must be another force on the desk, in the opposite direction, keeping it from moving (the net force is zero on an object at rest). This is the force of *static friction* exerted by the floor on the desk. If you push with a greater force without moving the desk, the force of static friction in the opposite direction also has increased. If you push hard enough, the desk will eventually start to move, and kinetic friction takes over. At this point, you have exceeded the maximum force of static friction, which is given by $(F_{fr})_{max} = \mu_s F_N$, where μ_s is the *coefficient of static friction* (Table 5–1). Because the force of static friction can vary from zero to this maximum value, we write

$$F_{fr} \leq \mu_s F_N. \qquad \text{[static friction]}$$

You may have noticed that it is often easier to keep a heavy object sliding than it is to start it sliding in the first place. This is consistent with μ_s generally being greater than μ_k (see Table 5–1).

It seems that this simple relation for friction, $F_{fr} \leq \mu F_N$, was first established by the great "Renaissance man" Leonardo da Vinci (1452–1519).

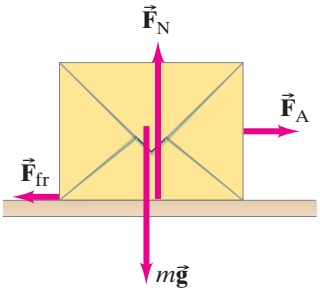
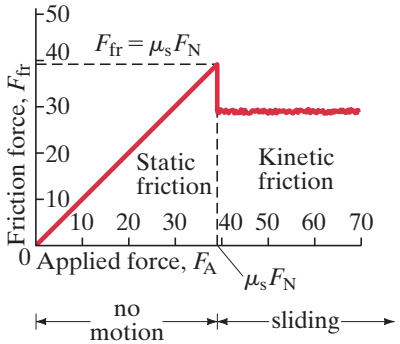


FIGURE 5–2 Repeated for Example 5-1.

FIGURE 5–3 Example 5–1. Magnitude of the force of friction as a function of the external force applied to an object initially at rest. As the applied force is increased in magnitude, the force of static friction increases in proportion until the applied force equals $\mu_s F_N$. If the applied force increases further, the object will begin to move, and the friction force drops to a roughly constant value characteristic of kinetic friction.



EXAMPLE 5–1 **Friction: static and kinetic.** A 10.0-kg box rests on a horizontal floor. The coefficient of static friction is $\mu_s = 0.40$ and the coefficient of kinetic friction is $\mu_k = 0.30$. Determine the force of friction, F_{fr}, acting on the box if a horizontal external applied force F_A is exerted on it of magnitude: (*a*) 0, (*b*) 10 N, (*c*) 20 N, (*d*) 38 N, and (*e*) 40 N.

APPROACH We don't know, right off, if we are dealing with static friction or kinetic friction, nor if the box remains at rest or accelerates. We need to draw a free-body diagram, and then determine in each case whether or not the box will move: the box starts moving if F_A is greater than the maximum static friction force. The forces on the box are shown in Fig. 5–2: gravity $m\vec{\mathbf{g}}$; the normal force exerted upward by the floor $\vec{\mathbf{F}}_N$; the horizontal applied force $\vec{\mathbf{F}}_A$; and the friction force $\vec{\mathbf{F}}_{fr}$.

SOLUTION The free-body diagram of the box is shown in Fig. 5–2. In the vertical direction there is no motion, so Newton's second law in the vertical direction gives $\Sigma F_y = ma_y = 0$, which tells us $F_N - mg = 0$. Hence the normal force is

$$F_N = mg = (10.0\,\text{kg})(9.80\,\text{m/s}^2) = 98.0\,\text{N}.$$

(*a*) Because $F_A = 0$ in this first case, the box doesn't move, and $F_{fr} = 0$.

(*b*) The force of static friction will oppose any applied force up to a maximum of

$$\mu_s F_N = (0.40)(98.0\,\text{N}) = 39\,\text{N}.$$

When the applied force is $F_A = 10\,\text{N}$, the box will not move. Newton's second law gives $\Sigma F_x = F_A - F_{fr} = 0$, so $F_{fr} = 10\,\text{N}$.

(*c*) An applied force of 20 N is also not sufficient to move the box. Thus $F_{fr} = 20\,\text{N}$ to balance the applied force.

(*d*) The applied force of 38 N is still not quite large enough to move the box. So the friction force has now increased to 38 N to keep the box at rest.

(*e*) A force of 40 N will start the box moving since it exceeds the maximum force of static friction, $\mu_s F_N = (0.40)(98\,\text{N}) = 39\,\text{N}$. Instead of static friction, we now have kinetic friction, and its magnitude is

$$F_{fr} = \mu_k F_N = (0.30)(98.0\,\text{N}) = 29\,\text{N}.$$

There is now a net (horizontal) force on the box of magnitude $F = 40\,\text{N} - 29\,\text{N} = 11\,\text{N}$, so the box will accelerate at a rate

$$a_x = \frac{\Sigma F}{m} = \frac{11\,\text{N}}{10.0\,\text{kg}} = 1.1\,\text{m/s}^2$$

as long as the applied force is 40 N.

NOTE Figure 5–3 shows a graph that summarizes this Example.

Friction can be a hindrance. It slows down moving objects and causes heating and binding of moving parts in machinery. Friction can be reduced by using lubricants such as oil. More effective in reducing friction between two surfaces is to maintain a layer of air or other gas between them. Devices using this concept, which is not practical for most situations, include air tracks and air tables in which the layer of air is maintained by forcing air through many tiny holes. Another technique to maintain the air layer is to suspend objects in air using magnetic fields ("magnetic levitation").

On the other hand, friction can be helpful. Our ability to walk (see Fig. 4–11) depends on friction between the soles of our shoes (or feet) and the ground. Walking involves static friction, not kinetic friction. The movement of a car, and also its stability, depend on friction. When friction is low, such as on ice, safe walking or driving becomes difficult.

CONCEPTUAL EXAMPLE 5–2 **A box against a wall.** You can hold a box against a rough wall (Fig. 5–4) and prevent it from slipping down by pressing hard horizontally. How does the application of a horizontal force keep an object from moving vertically?

RESPONSE This won't work well if the wall is slippery. You need friction. Even then, if you don't press hard enough, the box will slip. The horizontal force you apply produces a normal force on the box exerted by the wall (net force horizontally is zero since box doesn't move horizontally.) The force of gravity mg, acting downward on the box, can now be balanced by an upward static friction force whose maximum magnitude is proportional to the normal force. The harder you push, the greater F_N is and the greater F_{fr} can be. If you don't press hard enough, then $mg > \mu_s F_N$ and the box begins to slide down.

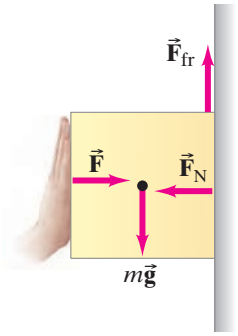


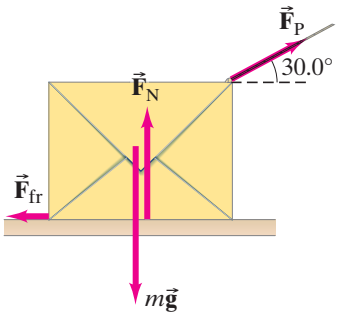
FIGURE 5–4 Example 5–2.

EXERCISE A If $\mu_s = 0.40$ and $mg = 20\,\text{N}$, what minimum force F will keep the box from falling: (*a*) 100 N; (*b*) 80 N; (*c*) 50 N; (*d*) 20 N; (*e*) 8 N?

EXAMPLE 5–3 **Pulling against friction.** A 10.0-kg box is being pulled along a horizontal surface by a force F_P of 40.0 N applied at a 30.0° angle above horizontal. This is like Example 4–11 except now there is friction, and we assume a coefficient of kinetic friction of 0.30. Calculate the acceleration.

APPROACH The free-body diagram is shown in Fig. 5–5. It is much like that in Fig. 4–21, but with one more force, that of friction.

FIGURE 5–5 Example 5–3.



SOLUTION The calculation for the vertical (y) direction is just the same as in Example 4–11, $mg = (10.0\,\text{kg})(9.80\,\text{m/s}^2) = 98.0\,\text{N}$ and $F_{Py} = (40.0\,\text{N})(\sin 30.0°) = 20.0\,\text{N}$. With y positive upward and $a_y = 0$, we have

$$F_N - mg + F_{Py} = ma_y$$
$$F_N - 98.0\,\text{N} + 20.0\,\text{N} = 0,$$

so the normal force is $F_N = 78.0\,\text{N}$. Now we apply Newton's second law for the horizontal (x) direction (positive to the right), and include the friction force:

$$F_{Px} - F_{fr} = ma_x.$$

The friction force is kinetic as long as $F_{fr} = \mu_k F_N$ is less than $F_{Px} = (40.0\,\text{N})\cos 30.0° = 34.6\,\text{N}$, which it is:

$$F_{fr} = \mu_k F_N = (0.30)(78.0\,\text{N}) = 23.4\,\text{N}.$$

Hence the box does accelerate:

$$a_x = \frac{F_{Px} - F_{fr}}{m} = \frac{34.6\,\text{N} - 23.4\,\text{N}}{10.0\,\text{kg}} = 1.1\,\text{m/s}^2.$$

In the absence of friction, as we saw in Example 4–11, the acceleration would be much greater than this.

NOTE Our final answer has only two significant figures because our least significant input value $(\mu_k = 0.30)$ has two.

EXERCISE B If $\mu_k F_N$ were greater than F_{Px}, what would you conclude?

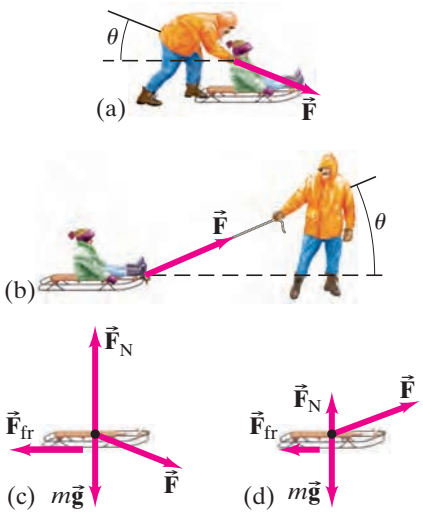
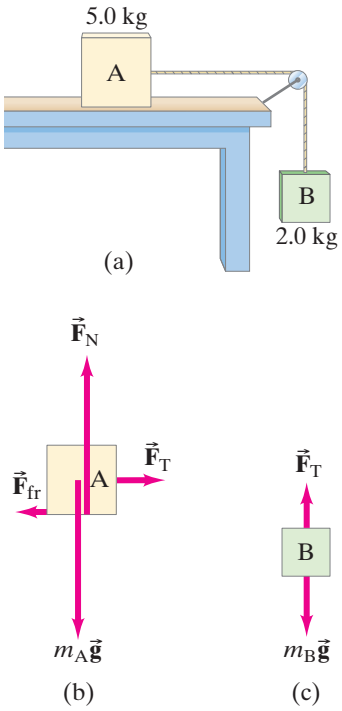


FIGURE 5–6 Example 5–4.

FIGURE 5–7 Example 5–5.



CONCEPTUAL EXAMPLE 5–4 **To push or to pull a sled?** Your little sister wants a ride on her sled. If you are on flat snow, will you exert less force if you push her or pull her? See Figs. 5–6a and b. Assume the same angle θ in each case.

RESPONSE Let us draw free-body diagrams for the sled–sister combination, as shown in Figs. 5–6c and d. They show, for the two cases: the force exerted by you, $\vec{\mathbf{F}}$ (an unknown); by the snow, $\vec{\mathbf{F}}_N$ and $\vec{\mathbf{F}}_{fr}$; and gravity $m\vec{\mathbf{g}}$. (*a*) If you push her, and $\theta > 0$, there is a vertically downward component to your force. Hence the upward normal force exerted by the ground (Fig. 5–6c) will be larger than mg (where m is the mass of sister plus sled). (*b*) If you pull her, your force has a vertically upward component, so the normal force F_N will be less than mg, Fig. 5–6d. Because the friction force is proportional to the normal force, F_{fr} will be less if you pull her. So you can exert less force if you pull her.

EXAMPLE 5–5 **Two boxes and a pulley.** In Fig. 5–7a, two boxes are connected by a cord running over a pulley. The coefficient of kinetic friction between box A and the table is 0.20. We ignore the mass of the cord and pulley and any friction in the pulley, which means we can assume that a force applied to one end of the cord will have the same magnitude at the other end. We wish to find the acceleration of the system, a, which will have the same magnitude for both boxes assuming the cord doesn't stretch. As box B moves down, box A moves to the right.

APPROACH The free-body diagrams for each box are shown in Figs. 5–7b and c. The forces on box A are the pulling force of the cord F_T, gravity $m_A g$, the normal force exerted by the table F_N, and a friction force exerted by the table F_{fr}. The forces on box B are gravity $m_B g$, and the cord pulling up, F_T.

SOLUTION Box A does not move vertically, so Newton's second law tells us the normal force just balances the weight,

$$F_N = m_A g = (5.0\,\text{kg})(9.8\,\text{m/s}^2) = 49\,\text{N}.$$

In the horizontal direction, there are two forces on box A (Fig. 5–7b): F_T, the tension in the cord (whose value we don't know), and the force of friction

$$F_{fr} = \mu_k F_N = (0.20)(49\,\text{N}) = 9.8\,\text{N}.$$

The horizontal acceleration of box A is what we wish to find; we use Newton's second law in the x direction, $\Sigma F_{Ax} = m_A a_x$, which becomes (taking the positive direction to the right and setting $a_{Ax} = a$):

$$\Sigma F_{Ax} = F_T - F_{fr} = m_A a. \qquad \text{[box A]}$$

Next consider box B. The force of gravity $m_B g = (2.0\,\text{kg})(9.8\,\text{m/s}^2) = 19.6\,\text{N}$ pulls downward; and the cord pulls upward with a force F_T. So we can write Newton's second law for box B (taking the downward direction as positive):

$$\Sigma F_{By} = m_B g - F_T = m_B a. \qquad \text{[box B]}$$

[Notice that if $a \neq 0$, then F_T is not equal to $m_B g$.]

We have two unknowns, a and F_T, and we also have two equations. We solve the box A equation for F_T:

$$F_T = F_{fr} + m_A a,$$

and substitute this into the box B equation:

$$m_B g - F_{fr} - m_A a = m_B a.$$

Now we solve for a and put in numerical values:

$$a = \frac{m_B g - F_{fr}}{m_A + m_B} = \frac{19.6\,\text{N} - 9.8\,\text{N}}{5.0\,\text{kg} + 2.0\,\text{kg}} = 1.4\,\text{m/s}^2,$$

which is the acceleration of box A to the right, and of box B down.

If we wish, we can calculate F_T using the third equation up from here:

$$F_T = F_{fr} + m_A a = 9.8\,\text{N} + (5.0\,\text{kg})(1.4\,\text{m/s}^2) = 17\,\text{N}.$$

NOTE Box B is not in free fall. It does not fall at $a = g$ because an additional force, F_T, is acting upward on it.

In Chapter 4 we examined motion on ramps and inclines, and saw that it is usually an advantage to choose the x axis along the plane, in the direction of acceleration. There we ignored friction, but now we take it into account.

EXAMPLE 5–6 **The skier.** The skier in Fig. 5–8a descends a 30° slope. If the coefficient of kinetic friction is 0.10, what is her acceleration when she is in contact with the snow?

APPROACH We choose the x axis along the slope, positive downslope in the direction of the skier's motion. The y axis is perpendicular to the surface. The forces acting on the skier (Fig. 5–8b) are gravity, $\vec{\mathbf{F}}_G = m\vec{\mathbf{g}}$, which points vertically downward (*not* perpendicular to the slope), and the two forces exerted on her skis by the snow—the normal force perpendicular to the snowy slope (*not* vertical), and the friction force parallel to the surface. These three forces are shown acting at one point in Fig. 5–8b, which is our free-body diagram for the skier.

SOLUTION We have to resolve only one vector into components, the weight $\vec{\mathbf{F}}_G$, and its components are shown as dashed lines in Fig. 5–8c. To be general, we use θ rather than 30° for now. We use the definitions of sine ("side opposite") and cosine ("side adjacent") to obtain the components:

$$\begin{aligned} F_{Gx} &= mg\sin\theta, \\ F_{Gy} &= -mg\cos\theta \end{aligned}$$

where F_{Gy} is in the negative y direction. To calculate the skier's acceleration down the hill, a_x, we apply Newton's second law to the x direction:

$$\begin{aligned} \Sigma F_x &= ma_x \\ mg\sin\theta - \mu_k F_N &= ma_x \end{aligned}$$

where the two forces are the x component of the gravity force ($+x$ direction) and the friction force ($-x$ direction). We want to find the value of a_x, but we don't yet know F_N in the last equation. Let's see if we can get F_N from the y component of Newton's second law:

$$\begin{aligned} \Sigma F_y &= ma_y \\ F_N - mg\cos\theta &= ma_y = 0 \end{aligned}$$

where we set $a_y = 0$ because there is no motion in the y direction (perpendicular to the slope) when the skier is in contact with the slope. Thus we can solve for F_N:

$$F_N = mg\cos\theta$$

and we can substitute this into our equation above for ma_x:

$$mg\sin\theta - \mu_k(mg\cos\theta) = ma_x.$$

There is an m in each term which can be canceled out. Thus (setting $\theta = 30°$ and $\mu_k = 0.10$):

$$\begin{aligned} a_x &= g\sin 30° - \mu_k g\cos 30° \\ &= 0.50g - (0.10)(0.866)g = 0.41g. \end{aligned}$$

The skier's acceleration is 0.41 times the acceleration of gravity, which in numbers† is $a = (0.41)(9.8\,\text{m/s}^2) = 4.0\,\text{m/s}^2$.

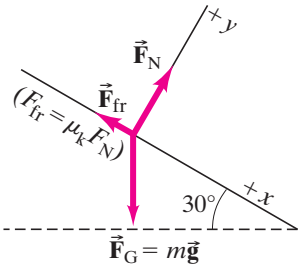
NOTE The mass canceled out, so we have the useful conclusion that *the acceleration doesn't depend on the mass*. That such a cancellation sometimes occurs, and thus may give a useful conclusion as well as saving calculation, is a big advantage of working with the algebraic equations and putting in the numbers only at the end.

NOTE The friction force on high-speed alpine skiers is very small—it seems the heat produced under the skis by friction melts the snow so the skis float on tiny balls of water. Fast skiers are slowed more by air resistance, which they can reduce by going into a tuck to reduce their surface area. When going airborne (as over the top of a hill) they are slowed more (can't hold that tuck, especially when landing) than when in contact with the snow.

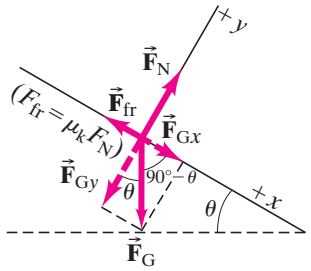
PHYSICS APPLIED
Skiing



(a)



(b)



(c)

FIGURE 5–8 Example 5–6. A skier descending a slope; $\vec{\mathbf{F}}_G = m\vec{\mathbf{g}}$ is the force of gravity (weight) on the skier.

PROBLEM SOLVING
It is often helpful to put in numbers only at the end

PHYSICS APPLIED
Skiers are faster on snow than when airborne

†We used values rounded off to 2 significant figures to obtain $a = 4.0\,\text{m/s}^2$. If we kept all the extra digits in our calculator, we would find $a = 0.4134\,g \approx 4.1\,\text{m/s}^2$. This difference is within the expected precision (number of significant figures, Section 1–3).

CAUTION

Directions of gravity and the normal force

In problems involving a slope or "inclined plane," avoid making errors in the directions of the normal force and gravity. The normal force is *not* vertical: it is perpendicular to the slope or plane. And gravity is *not* perpendicular to the slope—gravity acts vertically downward toward the center of the Earth.

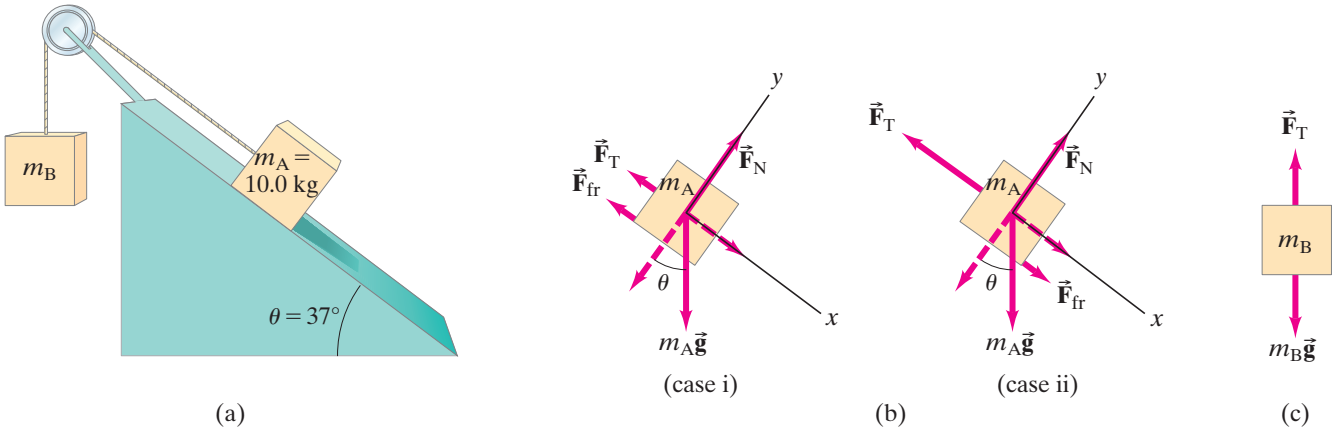


FIGURE 5–9 Example 5–7. Note choice of x and y axes.

EXAMPLE 5–7 **A ramp, a pulley, and two boxes.** A box of mass $m_A = 10.0\,\text{kg}$ rests on a surface inclined at $\theta = 37°$ to the horizontal. It is connected by a lightweight cord, which passes over a massless and frictionless pulley, to a second box of mass m_B, which hangs freely as shown in Fig. 5–9a. (*a*) If the coefficient of static friction is $\mu_s = 0.40$, determine what range of values for mass m_B will keep the system at rest. (*b*) If the coefficient of kinetic friction is $\mu_k = 0.30$, and $m_B = 10.0\,\text{kg}$, determine the acceleration of the system.

APPROACH Figure 5–9b shows two free-body diagrams for box m_A because the force of friction can be either up or down the slope, depending on which direction the box slides: (i) if $m_B = 0$ or is sufficiently small, m_A would tend to slide down the incline, so $\vec{\mathbf{F}}_{fr}$ would be directed up the incline; (ii) if m_B is large enough, m_A will tend to be pulled up the plane, so $\vec{\mathbf{F}}_{fr}$ would point down the plane. The tension force exerted by the cord is labeled $\vec{\mathbf{F}}_T$.

SOLUTION (*a*) For both cases (i) and (ii), Newton's second law for the y direction (perpendicular to the plane) is the same:

$$F_N - m_A g\cos\theta = m_A a_y = 0$$

since there is no y motion. So

$$F_N = m_A g\cos\theta.$$

Now for the x motion. We consider case (i) first for which $\Sigma F = ma$ gives

$$m_A g\sin\theta - F_T - F_{fr} = m_A a_x.$$

We want no acceleration, $a_x = 0$, and we solve for F_T since F_T is related to m_B (whose value we are seeking) by $F_T = m_B g$ (see Fig. 5–9c). Thus

$$m_A g\sin\theta - F_{fr} = F_T = m_B g.$$

We solve this for m_B and set F_{fr} at its maximum value $\mu_s F_N = \mu_s m_A g\cos\theta$ to find the minimum value that m_B can have to prevent motion $(a_x = 0)$:

$$\begin{aligned} m_B &= m_A\sin\theta - \mu_s m_A\cos\theta \\ &= (10.0\,\text{kg})(\sin 37° - 0.40\cos 37°) = 2.8\,\text{kg}. \end{aligned}$$

Thus if $m_B < 2.8\,\text{kg}$, then box A will slide down the incline.

Now for case (ii) in Fig. 5–9b, box A being tugged *up* the incline, but not actually quite accelerating. Newton's second law is

$$m_A g \sin\theta + F_{fr} - F_T = m_A a_x = 0.$$

Then the maximum value m_B can have without causing acceleration is given by

$$F_T = m_B g = m_A g \sin\theta + \mu_s m_A g \cos\theta$$

or

$$\begin{aligned} m_B &= m_A \sin\theta + \mu_s m_A \cos\theta \\ &= (10.0\,\text{kg})(\sin 37° + 0.40 \cos 37°) = 9.2\,\text{kg}. \end{aligned}$$

Thus, to prevent motion, we have the condition

$$2.8\,\text{kg} < m_B < 9.2\,\text{kg}.$$

(*b*) If $m_B = 10.0\,\text{kg}$ and $\mu_k = 0.30$, then m_B will fall and m_A will slide up the plane (case ii). To find their acceleration a, we use $\Sigma F = ma$ for box A:

$$m_A a = F_T - m_A g \sin\theta - \mu_k F_N.$$

Since m_B accelerates downward, Newton's second law for box B (Fig. 5–9c) tells us $m_B a = m_B g - F_T$, or $F_T = m_B g - m_B a$, and we substitute this into the equation above:

$$m_A a = m_B g - m_B a - m_A g \sin\theta - \mu_k F_N.$$

We solve for the acceleration a and substitute $F_N = m_A g \cos\theta$, and then $m_A = m_B = 10.0\,\text{kg}$, to find

$$\begin{aligned} a &= \frac{m_B g - m_A g \sin\theta - \mu_k m_A g \cos\theta}{m_A + m_B} \\ &= \frac{(10.0\,\text{kg})\left(9.80\,\text{m/s}^2\right)(1 - \sin 37° - 0.30 \cos 37°)}{20.0\,\text{kg}} \\ &= 0.78\,\text{m/s}^2. \end{aligned}$$

NOTE It is worth comparing this equation for acceleration a with that obtained in Example 5–5: if we set $\theta = 0$ in the equation just above, the plane is horizontal as in Example 5–5, and we obtain $a = (m_B g - \mu_k m_A g)/(m_A + m_B)$ just as in Example 5–5.

5–2 Uniform Circular Motion—Kinematics

An object moves in a straight line if the net force on it acts in the direction of motion, or the net force is zero. If the net force acts at an angle to the direction of motion at any moment, then the object moves in a curved path. An example of the latter is projectile motion, which we discussed in Chapter 3. Another important case is that of an object moving in a circle, such as a ball at the end of a string revolving around one's head, or the nearly circular motion of the Moon about the Earth.

An object that moves in a circle at constant speed v is said to experience **uniform circular motion**. The *magnitude* of the velocity remains constant in this case, but the *direction* of the velocity continuously changes as the object moves around the circle (Fig. 5–10). Because acceleration is defined as the rate of change of velocity, a change in direction of velocity is an acceleration, just as a change in its magnitude is. Thus, an object revolving in a circle is continuously accelerating, even when the speed remains constant ($v_1 = v_2 = v$ in Fig. 5–10). We now investigate this acceleration quantitatively.

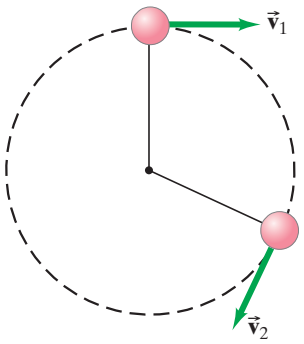


FIGURE 5–10 A small object moving in a circle, showing how the velocity changes. At each point, the instantaneous velocity is in a direction tangent to the circular path.

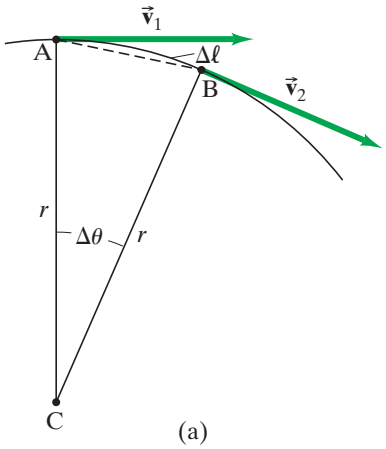
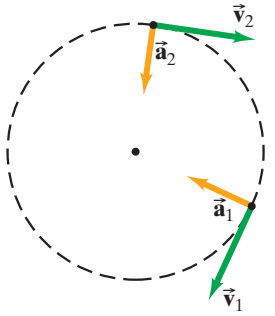


FIGURE 5–11 Determining the change in velocity, $\Delta\vec{\mathbf{v}}$, for a particle moving in a circle. The length $\Delta\ell$ is the distance along the arc, from A to B.

⚠ CAUTION

In uniform circular motion, the speed is constant, but the acceleration is not zero

FIGURE 5–12 For uniform circular motion, $\vec{\mathbf{a}}$ is always perpendicular to $\vec{\mathbf{v}}$.



Acceleration is defined as

$$\vec{\mathbf{a}} = \lim_{\Delta t \to 0} \frac{\Delta\vec{\mathbf{v}}}{\Delta t} = \frac{d\vec{\mathbf{v}}}{dt},$$

where $\Delta\vec{\mathbf{v}}$ is the change in velocity during the short time interval Δt. We will eventually consider the situation in which Δt approaches zero and thus obtain the instantaneous acceleration. But for purposes of making a clear drawing (Fig. 5–11), we consider a nonzero time interval. During the time interval Δt, the particle in Fig. 5–11a moves from point A to point B, covering a distance $\Delta\ell$ *along the arc* which subtends an angle $\Delta\theta$. The change in the velocity vector is $\vec{\mathbf{v}}_2 - \vec{\mathbf{v}}_1 = \Delta\vec{\mathbf{v}}$, and is shown in Fig. 5–11b.

Now we let Δt be very small, approaching zero. Then $\Delta\ell$ and $\Delta\theta$ are also very small, and $\vec{\mathbf{v}}_2$ will approach being parallel to $\vec{\mathbf{v}}_1$ (Fig. 5–11c); $\Delta\vec{\mathbf{v}}$ will be perpendicular to them (Fig. 5–11b). Thus $\Delta\vec{\mathbf{v}}$ points toward the center of the circle. Since $\vec{\mathbf{a}}$, by definition, is in the same direction as $\Delta\vec{\mathbf{v}}$ (equation above), it too must point toward the center of the circle. Therefore, this acceleration is called **centripetal acceleration** ("center-pointing" acceleration) or **radial acceleration** (since it is directed along the radius, toward the center of the circle), and we denote it by $\vec{\mathbf{a}}_R$.

Now that we have determined the direction, next we find the magnitude of the radial (centripetal) acceleration, a_R. Because the line CA in Fig. 5–11a is perpendicular to $\vec{\mathbf{v}}_1$, and line CB is perpendicular to $\vec{\mathbf{v}}_2$, then the angle $\Delta\theta$ between CA and CB is also the angle between $\vec{\mathbf{v}}_1$ and $\vec{\mathbf{v}}_2$. Hence the vectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, and $\Delta\vec{\mathbf{v}}$ in Fig. 5–11b form a triangle that is geometrically similar[†] to triangle ACB in Fig. 5–11a. If we take $\Delta\theta$ to be very small (letting Δt be very small) and setting $v = v_1 = v_2$ because the magnitude of the velocity is assumed not to change, we can write

$$\frac{\Delta v}{v} \approx \frac{\Delta\ell}{r},$$

or

$$\Delta v \approx \frac{v}{r}\Delta\ell.$$

This is an exact equality when Δt approaches zero, for then the arc length $\Delta\ell$ equals the chord length AB. We want to find the instantaneous acceleration, a_R, so we write the expression above as an equality, giving

$$a_R = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \lim_{\Delta t \to 0} \frac{v}{r}\frac{\Delta\ell}{\Delta t}.$$

Then, because

$$\lim_{\Delta t \to 0} \frac{\Delta\ell}{\Delta t}$$

is the linear speed, v, of the object, we have for the centripetal (radial) acceleration

$$a_R = \frac{v^2}{r}. \qquad \text{[centripetal (radial) acceleration]} \quad \textbf{(5–1)}$$

Equation 5–1 is valid at any instant in circular motion, and even when v is not constant. It is valid also for a portion of a circle, as for a car going around a curve, and r is referred to as the **radius of curvature** of the curve.

To summarize, *an object moving in a circle of radius r at constant speed v has an acceleration whose direction is toward the center of the circle and whose magnitude is* $a_R = v^2/r$. It makes sense that this acceleration depends on v and r: the greater the speed v, the faster the velocity changes direction; and the larger the radius, the less rapidly the velocity changes direction.

The acceleration vector points toward the center of the circle when v is constant. But the velocity vector always points in the direction of motion, which is tangential to the circle. Thus the velocity and acceleration vectors are perpendicular to each other at every point in the path for uniform circular motion (Fig. 5–12). Beware of thinking that acceleration and velocity are always in the same direction. For an object falling vertically, $\vec{\mathbf{a}}$ and $\vec{\mathbf{v}}$ are indeed parallel. But in uniform circular motion, $\vec{\mathbf{a}}$ and $\vec{\mathbf{v}}$ are perpendicular, not parallel (nor were they parallel in projectile motion, Section 3–7).

EXERCISE C Can Equations 2–12, the kinematic equations for constant acceleration, be used for uniform circular motion? For example, could Eq. 2–12b be used to calculate the time for the revolving ball in Fig. 5–12 to make one revolution?

[†]Appendix A contains a review of geometry.

Circular motion is often described in terms of the **frequency** f, the number of revolutions per second. The **period** T of an object revolving in a circle is the time required for one complete revolution. Period and frequency are related by

$$T = \frac{1}{f}. \qquad \textbf{(5–2)}$$

For example, if an object revolves at a frequency of 3 rev/s, then each revolution (rev) takes $\frac{1}{3}$ s. An object revolving in a circle (of circumference $2\pi r$) at constant speed v travels a distance $2\pi r$ in one revolution which takes a time T. Thus

$$v = \frac{\text{distance}}{\text{time}} = \frac{2\pi r}{T}.$$

EXAMPLE 5–8 **Acceleration of a revolving ball.** A 150-g ball at the end of a string is revolving uniformly in a horizontal circle of radius 0.600 m, as in Fig. 5–10 or 5–12. The ball makes 2.00 revolutions in a second. What is its centripetal acceleration?

APPROACH The centripetal acceleration is $a_R = v^2/r$. We are given r, and we can find the speed of the ball, v, from the given radius and frequency.

SOLUTION If the ball makes 2.00 complete revolutions per second, then the ball travels in a complete circle in a time interval equal to 0.500 s, which is its period T. The distance traveled in this time is the circumference of the circle, $2\pi r$, where r is the radius of the circle. Therefore, the ball has speed

$$v = \frac{2\pi r}{T} = \frac{2\pi(0.600\text{ m})}{(0.500\text{ s})} = 7.54\text{ m/s}.$$

The centripetal acceleration[†] is

$$a_R = \frac{v^2}{r} = \frac{(7.54\text{ m/s})^2}{(0.600\text{ m})} = 94.7\text{ m/s}^2.$$

EXERCISE D In Example 5–8, if the radius is doubled to 1.20 m, but the period stays the same, the centripetal acceleration will change by a factor of:
(*a*) 2; (*b*) 4; (*c*) $\frac{1}{2}$; (*d*) $\frac{1}{4}$; (*e*) none of these.

EXAMPLE 5–9 **Moon's centripetal acceleration.** The Moon's nearly circular orbit about the Earth has a radius of about 384,000 km and a period T of 27.3 days. Determine the acceleration of the Moon toward the Earth.

APPROACH Again we need to find the velocity v in order to find a_R. We will need to convert to SI units to get v in m/s.

SOLUTION In one orbit around the Earth, the Moon travels a distance $2\pi r$, where $r = 3.84 \times 10^8$ m is the radius of its circular path. The time required for one complete orbit is the Moon's period of 27.3 d, which in seconds is $T = (27.3\text{ d})(24.0\text{ h/d})(3600\text{ s/h}) = 2.36 \times 10^6$ s. The speed of the Moon in its orbit about the Earth is $v = 2\pi r/T$. Therefore,

$$a_R = \frac{v^2}{r} = \frac{(2\pi r)^2}{T^2 r} = \frac{4\pi^2 r}{T^2} = \frac{4\pi^2(3.84 \times 10^8\text{ m})}{(2.36 \times 10^6\text{ s})^2}$$
$$= 0.00272\text{ m/s}^2 = 2.72 \times 10^{-3}\text{ m/s}^2.$$

We can write this acceleration in terms of $g = 9.80\text{ m/s}^2$ (the acceleration of gravity at the Earth's surface) as

$$a = 2.72 \times 10^{-3}\text{ m/s}^2\left(\frac{g}{9.80\text{ m/s}^2}\right) = 2.78 \times 10^{-4} g.$$

NOTE The centripetal acceleration of the Moon, $a = 2.78 \times 10^{-4}\, g$, is *not* the acceleration of gravity for objects at the Moon's surface due to the Moon's gravity. Rather, it is the acceleration due to the *Earth's* gravity for any object (such as the Moon) that is 384,000 km from the Earth. Notice how small this acceleration is compared to the acceleration of objects near the Earth's surface.

⚠ CAUTION
Distinguish the Moon's gravity on objects at its surface, from the Earth's gravity acting on the Moon (this Example)

[†]Differences in the final digit can depend on whether you keep all digits in your calculator for v (which gives $a_R = 94.7\text{ m/s}^2$), or if you use $v = 7.54$ m/s (which gives $a_R = 94.8\text{ m/s}^2$). Both results are valid since our assumed accuracy is about $\pm 0.1\text{ m/s}^2$ (see Section 1–3).

PHYSICS APPLIED
Centrifuge

*Centrifugation

Centrifuges, and very high speed ultracentrifuges, are used to sediment materials quickly and to separate materials. Test tubes held in the centrifuge rotor are accelerated to very high rotational speeds: see Fig. 5–13, where one test tube is shown in two positions as the rotor turns. The small green dot represents a small particle, perhaps a macromolecule, in a fluid-filled test tube. At position A the particle has a tendency to move in a straight line, but the fluid resists the motion of the particles, exerting a centripetal force that keeps the particles moving nearly in a circle. The resistive force exerted by the fluid (liquid, gas, or gel, depending on the application) usually does not quite equal mv^2/r, and the particles move slowly toward the bottom of the tube. A centrifuge provides an "effective gravity" much larger than normal gravity because of the high rotational speeds, thus causing more rapid sedimentation.

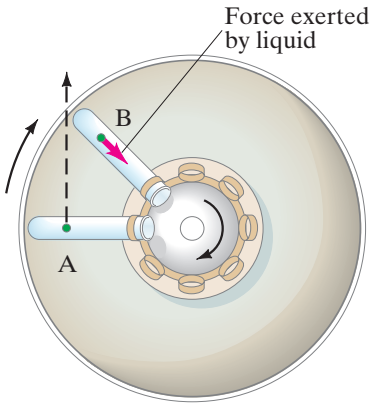


FIGURE 5–13 Two positions of a rotating test tube in a centrifuge (top view). At A, the green dot represents a macromolecule or other particle being sedimented. It would tend to follow the dashed line, heading toward the bottom of the tube, but the fluid resists this motion by exerting a force on the particle as shown at point B.

EXAMPLE 5–10 **Ultracentrifuge.** The rotor of an ultracentrifuge rotates at 50,000 rpm (revolutions per minute). A particle in a test tube (Fig. 5–13) is 6.00 cm from the rotation axis. Calculate its centripetal acceleration, in "g's."

APPROACH We calculate the centripetal acceleration from $a_R = v^2/r$.

SOLUTION The test tube makes 5.00×10^4 revolutions each minute, or, dividing by 60 s/min, 833 rev/s. The time to make one revolution, the period T, is

$$T = \frac{1}{f} = \frac{1}{(833\ \text{rev/s})} = 1.20 \times 10^{-3}\ \text{s/rev}.$$

The particle revolves in a circle of circumference $2\pi r = (2\pi)(0.0600\ \text{m}) = 0.377\ \text{m}$ per revolution. The speed of the particle is then

$$v = \frac{2\pi r}{T} = \left(\frac{0.377\ \text{m/rev}}{1.20 \times 10^{-3}\ \text{s/rev}}\right) = 3.14 \times 10^2\ \text{m/s}.$$

The centripetal acceleration is

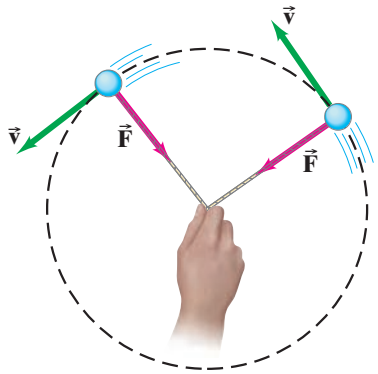
$$a_R = \frac{v^2}{r} = \frac{(3.14 \times 10^2\ \text{m/s})^2}{0.0600\ \text{m}} = 1.64 \times 10^6\ \text{m/s}^2,$$

which, dividing by $g = 9.80\ \text{m/s}^2$, is $1.67 \times 10^5\ g$'s $= 167{,}000\ g$'s.

Other uses of centrifugation include:
(1) *Astronauts* can undergo training in large centrifuges as experience for taking off from Earth at several g's.
(2) *Enrichment* of uranium: natural uranium is 99.3% uranium-238. The lighter form, U-235, is only 0.7%, but is the type (isotope) usable for nuclear fission in nuclear power plants and nuclear bombs. Complex centrifuges can partially separate the two isotopes, increasing (or "enriching") the proportion of U-235. Power or research reactors need 3% to 20%, although some use natural uranium; bombs need greater enrichment ($\gtrsim 80\%$).

PHYSICS APPLIED
Uranium enrichment by centrifugation

FIGURE 5–14 A force is required to keep an object moving in a circle. If the speed is constant, the force is directed toward the circle's center.



5–3 Dynamics of Uniform Circular Motion

According to Newton's second law $(\Sigma\vec{F} = m\vec{a})$, an object that is accelerating must have a net force acting on it. An object moving in a circle, such as a ball on the end of a string, must therefore have a force applied to it to keep it moving in that circle. That is, a net force is necessary to give it centripetal acceleration. The magnitude of the required force can be calculated using Newton's second law for the radial component, $\Sigma F_R = ma_R$, where a_R is the centripetal acceleration, $a_R = v^2/r$, and ΣF_R is the total (or net) force in the radial direction:

$$\Sigma F_R = ma_R = m\frac{v^2}{r}. \qquad \text{[circular motion]} \quad \textbf{(5–3)}$$

For uniform circular motion $(v = \text{constant})$, the total acceleration is a_R, which is directed toward the center of the circle at all times. Thus the *net force too must be directed toward the center of the circle* (Fig. 5–14). If no net force were exerted on the object, it would not move in a circle but in a straight line, as Newton's first law tells us.

The direction of the net force is continually changing so that it is always directed toward the center of the circle, perpendicular to $\vec{\mathbf{v}}$. This force is sometimes called a centripetal ("pointing toward the center") force. "Centripetal force" does not indicate some new kind of force; it only describes the *direction* of the net force needed to provide a circular path: the net force is directed toward the circle's center. The net force *must be applied by other objects*. For example, to swing a ball in a circle on the end of a string, you pull on the string and the string exerts the force on the ball. (Try it.) Here, the "centripetal force" that provides the centripetal acceleration is tension in the string. In other cases it can be gravity (on the Moon, for example); it could also be a normal force, a friction force, or even an electric force.

CAUTION

Centripetal force is not a new kind of force (Every force must be exerted by an object)

There is a common misconception that an object moving in a circle has an outward force acting on it, a so-called centrifugal ("center-fleeing") force. This is incorrect: *there is no outward force* on the revolving object. Consider, for example, a person swinging a ball on the end of a string around her head (Fig. 5–15). If you have ever done this yourself, you know that you feel a force pulling outward on your hand. The misconception arises when this pull is interpreted as an outward "centrifugal" force pulling on the ball that is transmitted along the string to your hand. This is not what is happening at all. To keep the ball moving in a circle, you pull *inwardly* on the string, and the string exerts this inward force on the ball. The ball exerts an equal and opposite force on the string (Newton's third law), and *this* is the outward force your hand feels (see Fig. 5–15).

CAUTION

There is no real "centrifugal force"

The force *on the ball* in Fig. 5–15 is the one exerted *inwardly* on it by you, via the string. To see even more convincing evidence that a "centrifugal force" does not act on the ball, consider what happens when you let go of the string. If a centrifugal force were acting, the ball would fly outward, as shown in Fig. 5–16a. But it doesn't; the ball flies off tangentially (Fig. 5–16b), in the direction of the velocity it had at the moment it was released, because the inward force no longer acts. Try it and see!

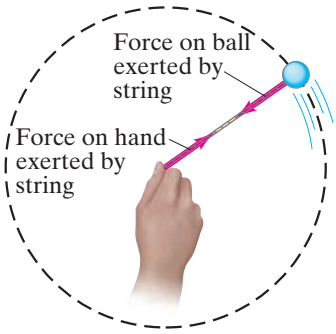
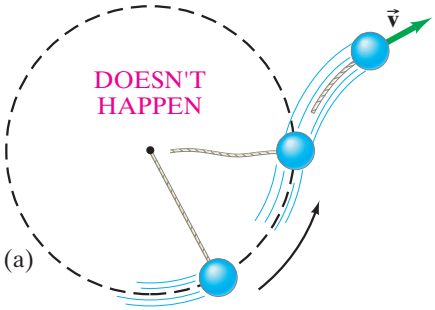
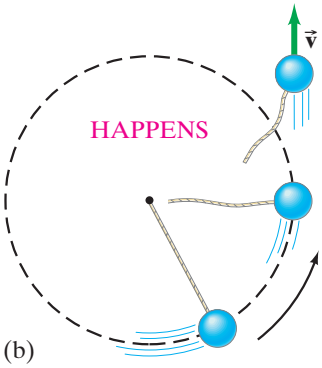


FIGURE 5–15 Swinging a ball on the end of a string (looking down from above).

EXERCISE E Return to the Chapter-Opening Question, page 116, and answer it again now. Try to explain why you may have answered differently the first time.






(c)

FIGURE 5–16 If centrifugal force existed, the revolving ball would fly outward as in (a) when released. In fact, it flies off tangentially as in (b). For example, in (c) sparks fly in straight lines tangentially from the edge of a rotating grinding wheel.

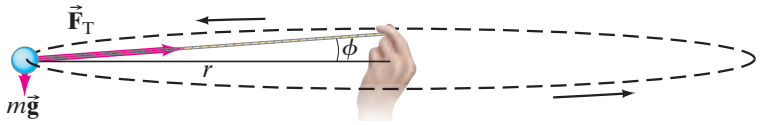


FIGURE 5–17
Example 5–11.

EXAMPLE 5–11 **ESTIMATE** **Force on revolving ball (horizontal).** Estimate the force a person must exert on a string attached to a 0.150-kg ball to make the ball revolve in a horizontal circle of radius 0.600 m. The ball makes 2.00 revolutions per second ($T = 0.500\,\text{s}$), as in Example 5–8. Ignore the string's mass.

APPROACH First we need to draw the free-body diagram for the ball. The forces acting on the ball are the force of gravity, $m\vec{\mathbf{g}}$ downward, and the tension force $\vec{\mathbf{F}}_T$ that the string exerts toward the hand at the center (which occurs because the person exerts that same force on the string). The free-body diagram for the ball is shown in Fig. 5–17. The ball's weight complicates matters and makes it impossible to revolve a ball with the cord perfectly horizontal. We estimate the force assuming the weight is small, and letting $\phi \approx 0$ in Fig. 5–17. Thus $\vec{\mathbf{F}}_T$ will act nearly horizontally and, in any case, provides the force necessary to give the ball its centripetal acceleration.

SOLUTION We apply Newton's second law to the radial direction, which we assume is horizontal:

$$(\Sigma F)_R = ma_R,$$

where $a_R = v^2/r$ and $v = 2\pi r/T = 2\pi(0.600\,\text{m})/(0.500\,\text{s}) = 7.54\,\text{m/s}$. Thus

$$F_T = m\frac{v^2}{r} = (0.150\,\text{kg})\frac{(7.54\,\text{m/s})^2}{(0.600\,\text{m})} \approx 14\,\text{N}.$$

NOTE We keep only two significant figures in the answer because we ignored the ball's weight; it is $mg = (0.150\,\text{kg})(9.80\,\text{m/s}^2) = 1.5\,\text{N}$, about $\frac{1}{10}$ of our result, which is small but maybe not so small as to justify stating a more precise answer for F_T.

NOTE To include the effect of $m\vec{\mathbf{g}}$, resolve $\vec{\mathbf{F}}_T$ in Fig. 5–17 into components, and set the horizontal component of $\vec{\mathbf{F}}_T$ equal to mv^2/r and its vertical component equal to mg.

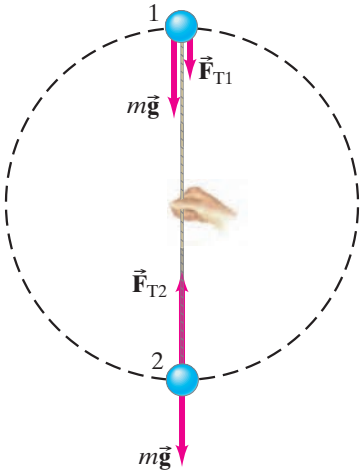


FIGURE 5–18 Example 5–12. Free-body diagrams for positions 1 and 2.

EXAMPLE 5–12 **Revolving ball (vertical circle).** A 0.150-kg ball on the end of a 1.10-m-long cord (negligible mass) is swung in a *vertical* circle. (*a*) Determine the minimum speed the ball must have at the top of its arc so that the ball continues moving in a circle. (*b*) Calculate the tension in the cord at the bottom of the arc, assuming the ball is moving at twice the speed of part (*a*).

APPROACH The ball moves in a vertical circle and is *not* undergoing uniform circular motion. The radius is assumed constant, but the speed v changes because of gravity. Nonetheless, Eq. 5–1 ($a_R = v^2/r$) is valid at each point along the circle, and we use it at the top and bottom points. The free-body diagram is shown in Fig. 5–18 for both positions.

SOLUTION (*a*) At the top (point 1), two forces act on the ball: $m\vec{\mathbf{g}}$, the force of gravity, and $\vec{\mathbf{F}}_{T1}$, the tension force the cord exerts at point 1. Both act downward, and their vector sum acts to give the ball its centripetal acceleration a_R. We apply Newton's second law, for the vertical direction, choosing downward as positive since the acceleration is downward (toward the center):

$$(\Sigma F)_R = ma_R$$

$$F_{T1} + mg = m\frac{v_1^2}{r}. \qquad \text{[at top]}$$

From this equation we can see that the tension force F_{T1} at point 1 will get larger if v_1 (ball's speed at top of circle) is made larger, as expected. But we are asked for the *minimum* speed to keep the ball moving in a circle. The cord will remain taut as long as there is tension in it. But if the tension disappears (because v_1 is too small) the cord can go limp, and the ball will fall out of its circular path. Thus, the minimum speed will occur if $F_{T1} = 0$, for which the equation above becomes

$$mg = m\frac{v_1^2}{r}. \qquad \text{[minimum speed at top]}$$

⚠ CAUTION
Circular motion only if cord is under tension

We solve for v_1, keeping an extra digit for use in (b):

$$v_1 = \sqrt{gr} = \sqrt{(9.80\,\text{m/s}^2)(1.10\,\text{m})} = 3.283\,\text{m/s} \approx 3.28\,\text{m/s}.$$

This is the minimum speed the ball must have at the top of the circle if it is to continue moving in a circular path.

(b) When the ball is at the bottom of the circle (point 2 in Fig. 5–18), the cord exerts its tension force F_{T2} upward, whereas the force of gravity, $m\vec{\mathbf{g}}$, still acts downward. Choosing *upward* as positive, Newton's second law gives:

$$(\Sigma F)_R = ma_R$$
$$F_{T2} - mg = m\frac{v_2^2}{r}. \qquad \text{[at bottom]}$$

The speed v_2 is given as twice that in (a), namely 6.566 m/s. We solve for F_{T2}:

$$F_{T2} = m\frac{v_2^2}{r} + mg$$
$$= (0.150\,\text{kg})\frac{(6.566\,\text{m/s})^2}{(1.10\,\text{m})} + (0.150\,\text{kg})(9.80\,\text{m/s}^2) = 7.35\,\text{N}.$$

EXERCISE F A rider on a Ferris wheel moves in a vertical circle of radius r at constant speed v (Fig. 5–19). Is the normal force that the seat exerts on the rider at the top of the wheel (a) less than, (b) more than, or (c) the same as, the force the seat exerts at the bottom of the wheel?

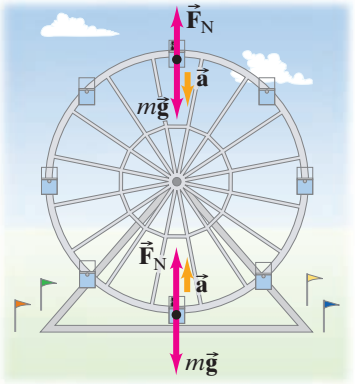


FIGURE 5–19 Exercise F.

EXAMPLE 5–13 **Conical pendulum.** A small ball of mass m, suspended by a cord of length ℓ, revolves in a horizontal circle of radius $r = \ell\sin\theta$, where θ is the angle the string makes with the vertical (Fig. 5–20). (a) In what direction is the acceleration of the ball, and what causes the acceleration? (b) Calculate the speed and period (time required for one revolution) of the ball in terms of ℓ, θ, g, and m.

APPROACH We can answer (a) by looking at Fig. 5–20, which shows the forces on the revolving ball at one instant: the acceleration points horizontally toward the center of the ball's circular path (not along the cord). The force responsible for the acceleration is the *net* force which here is the vector sum of the forces acting on the mass m: its weight $\vec{\mathbf{F}}_G$ (of magnitude $F_G = mg$), and the force exerted by the tension in the cord, $\vec{\mathbf{F}}_T$, which has horizontal and vertical components of magnitude $F_T\sin\theta$ and $F_T\cos\theta$.

SOLUTION (b) We apply Newton's second law to the horizontal and vertical directions. In the vertical direction, there is no motion, so the acceleration is zero and the net force in the vertical direction is zero:

$$F_T\cos\theta - mg = 0,$$

or

$$F_T = \frac{mg}{\cos\theta}.$$

In the horizontal direction there is only one force, of magnitude $F_T\sin\theta$, that acts toward the center of the circle and gives rise to the acceleration v^2/r. Newton's second law tells us:

$$F_T\sin\theta = m\frac{v^2}{r}.$$

We solve the second equation for v, and substitute for F_T from above (and use $r = \ell\sin\theta$):

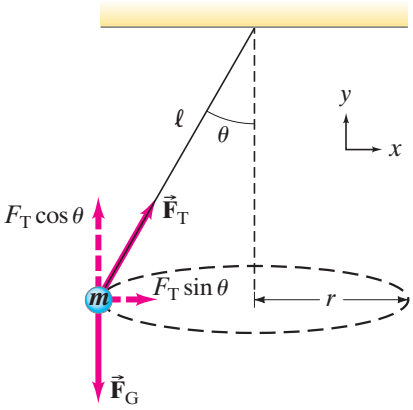
$$v = \sqrt{\frac{rF_T\sin\theta}{m}} = \sqrt{\frac{r}{m}\left(\frac{mg}{\cos\theta}\right)\sin\theta} = \sqrt{\frac{\ell g\sin^2\theta}{\cos\theta}}.$$

The period T is the time required to make one revolution, a distance of $2\pi r = 2\pi\ell\sin\theta$. The speed v can thus be written $v = 2\pi\ell\sin\theta/T$; then

$$T = \frac{2\pi\ell\sin\theta}{v} = \frac{2\pi\ell\sin\theta}{\sqrt{\ell g\sin^2\theta/\cos\theta}} = 2\pi\sqrt{\frac{\ell\cos\theta}{g}}.$$

NOTE The speed and period do not depend on the mass m of the ball. They do depend on ℓ and θ.

FIGURE 5–20 Example 5–13. Conical pendulum.



PROBLEM SOLVING

Uniform Circular Motion

1. **Draw a free-body diagram**, showing all the forces acting on each object under consideration. Be sure you can identify the source of each force (tension in a cord, Earth's gravity, friction, normal force, and so on). Don't put in something that doesn't belong (like a centrifugal force).
2. **Determine** which of the forces, or which of their components, act to provide the centripetal acceleration—that is, all the **forces or components that act radially**, toward or away from the center of the circular path. The sum of these forces (or components) provides the centripetal acceleration, $a_R = v^2/r$.
3. **Choose a convenient coordinate system**, preferably with one axis along the acceleration direction at a chosen instant.
4. **Apply Newton's second law** to the radial component:

$$\Sigma F_R = ma_R = m\frac{v^2}{r}. \qquad \text{[radial direction]}$$

5–4 Highway Curves: Banked and Unbanked

PHYSICS APPLIED
Driving around a curve

An example of circular dynamics occurs when an automobile rounds a curve, say to the left. In such a situation, you may feel that you are thrust outward toward the right side door. But there is no mysterious centrifugal force pulling on you.

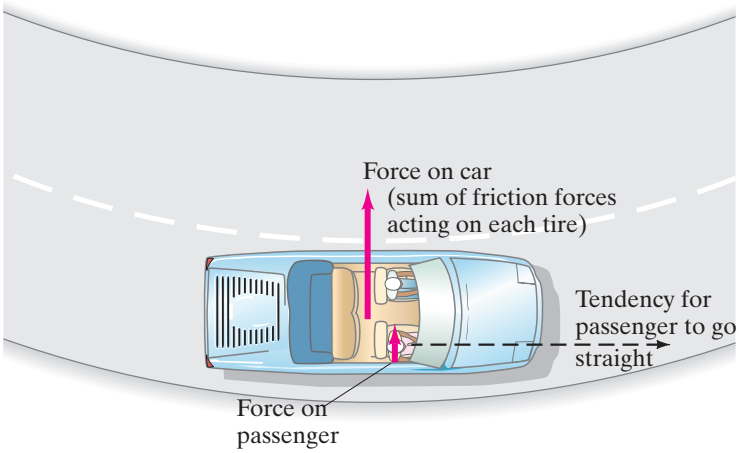


FIGURE 5–21 The road exerts a radially inward force (friction against the tires) on a car to make it move in a circle. The car exerts an inward force on the passenger.



FIGURE 5–22 Race car heading into a curve. From the tire marks we see that most cars experienced a sufficient friction force to give them the needed centripetal acceleration for rounding the curve safely. But, we also see tire tracks of cars on which there was not sufficient force—and which unfortunately followed more nearly straight-line paths.

What is happening is that you tend to move in a straight line, whereas the car has begun to follow a curved path. To make you go in the curved path, the seat (friction) or the door of the car (direct contact) exerts a force on you (Fig. 5–21). The car also must have a force exerted on it toward the center of the curve if it is to move in that curve. On a flat road, this force is supplied by friction between the tires and the pavement.

If the wheels and tires of the car are rolling normally without slipping or sliding, the bottom of the tire is at rest against the road at each instant. So the friction force the road exerts on the tires is static friction. But if static friction is not great enough, as under icy conditions or high speed, the static friction force is less than mv^2/r and the car will skid out of a circular path into a more nearly straight path. See Fig. 5–22. Once a car skids or slides, the friction force becomes kinetic friction, which is smaller than static friction.

EXAMPLE 5–14 **Skidding on a curve.** A 1000-kg car rounds a curve on a horizontal flat road of radius 50 m at a speed of 15 m/s (54 km/h). Will the car follow the curve, or will it skid? Assume: (*a*) the pavement is dry and the coefficient of static friction is $\mu_s = 0.60$; (*b*) the pavement is icy and $\mu_s = 0.25$.

APPROACH The forces on the car are gravity mg downward, the normal force F_N exerted upward by the road, and a horizontal friction force due to the level road. They are shown in Fig. 5–23, which is the free-body diagram for the car. The car will follow the curve if the maximum static friction force is greater than the mass times the centripetal acceleration.

SOLUTION In the vertical direction (y) there is no acceleration. Newton's second law tells us that the normal force F_N on the car is equal to the weight mg since the road is flat:

$$0 = \Sigma F_y = F_N - mg$$

so

$$F_N = mg = (1000\,\text{kg})(9.80\,\text{m/s}^2) = 9800\,\text{N}.$$

In the horizontal direction the only force is friction, and we must compare it to the force needed to produce the centripetal acceleration to see if it is sufficient. The net horizontal force required to keep the car moving in a circle around the curve is

$$(\Sigma F)_R = ma_R = m\frac{v^2}{r} = (1000\,\text{kg})\frac{(15\,\text{m/s})^2}{(50\,\text{m})} = 4500\,\text{N}.$$

Now we compute the maximum total static friction force (the sum of the friction forces acting on each of the four tires) to see if it can be large enough to provide a safe centripetal acceleration. For (*a*), $\mu_s = 0.60$, and the maximum friction force attainable (recall from Section 5–1 that $F_{fr} \leq \mu_s F_N$) is

$$(F_{fr})_{max} = \mu_s F_N = (0.60)(9800\,\text{N}) = 5880\,\text{N}.$$

Since a force of only 4500 N is needed, and that is, in fact, how much will be exerted by the road as a static friction force, the car can follow the curve. But in (*b*) the maximum static friction force possible is

$$(F_{fr})_{max} = \mu_s F_N = (0.25)(9800\,\text{N}) = 2450\,\text{N}.$$

The car will skid because the ground cannot exert sufficient force (4500 N is needed) to keep it moving in a curve of radius 50 m at a speed of 54 km/h.

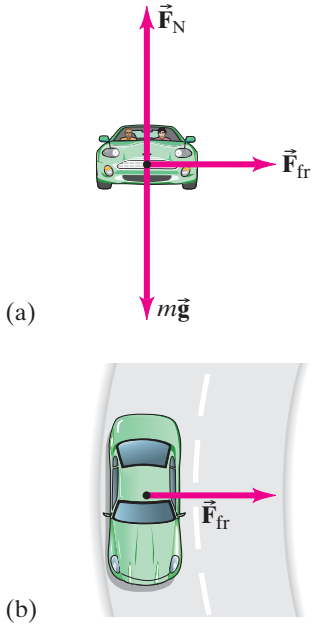


FIGURE 5–23 Example 5–14. Forces on a car rounding a curve on a level road. (a) Front view, (b) top view.

The possibility of skidding is worse if the wheels lock (stop rotating) when the brakes are applied too hard. When the tires are rolling, static friction exists. But if the wheels lock (stop rotating), the tires slide and the friction force, which is now kinetic friction, is less. More importantly, the *direction* of the friction force changes suddenly if the wheels lock. Static friction can point perpendicular to the velocity, as in Fig. 5–23b; but if the car slides, kinetic friction points *opposite* to the velocity. The force no longer points toward the center of the circle, and the car cannot continue in a curved path (see Fig. 5–22). Even worse, if the road is wet or icy, locking of the wheels occurs with less force on the brake pedal since there is less road friction to keep the wheels turning rather than sliding. **Antilock brakes (ABS)**, which all modern cars have, are designed to limit brake pressure just before the point where sliding would occur, by means of delicate sensors and a fast computer. But be careful if you drive an antique or older car.

PHYSICS APPLIED
Antilock brakes

EXERCISE G To negotiate a flat (unbanked) curve at a *faster* speed, a driver puts a couple of sand bags in his van aiming to increase the force of friction between the tires and the road. Will the sand bags help? Why or why not?

PHYSICS APPLIED

Banked curves

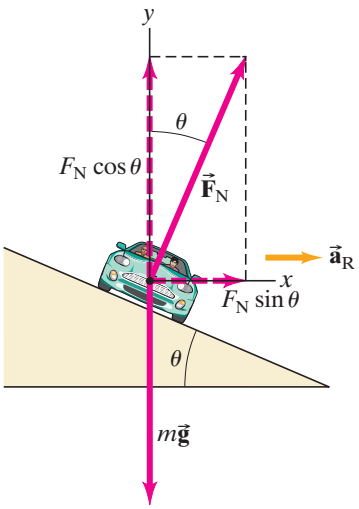


FIGURE 5–24 Normal force on a car rounding a banked curve, resolved into its horizontal and vertical components. The centripetal acceleration is horizontal (*not* parallel to the sloping road). The friction force on the tires, not shown, could point up or down along the slope, depending on the car's speed. The friction force will be zero for one particular speed. Note that here we choose the x axis horizontal (not along the plane) because the centripetal force is horizontal.

CAUTION

F_N is not always equal to mg

The *banking of curves* can reduce the chance of skidding. The normal force exerted by a banked road, acting perpendicular to the road, will have a component toward the center of the circle (Fig. 5–24), thus reducing the reliance on friction. For a given banking angle θ, there will be one speed for which no friction at all is required. This will be the case when the horizontal component of the normal force toward the center of the curve, $F_N \sin\theta$ (see Fig. 5–24), is just equal to the force required to give a vehicle its centripetal acceleration—that is, when

$$F_N \sin\theta = m\frac{v^2}{r}. \qquad \text{[no friction required]}$$

The banking angle of a road, θ, is chosen so that this condition holds for a particular speed, called the "design speed."

EXAMPLE 5–15 **Banking angle.** (*a*) For a car traveling with speed v around a curve of radius r, determine a formula for the angle at which a road should be banked so that no friction is required. (*b*) What is this angle for a road which has a curve of radius 50 m with a design speed of 50 km/h?

APPROACH Even though the road is banked, the car is still moving along a horizontal circle, so the centripetal acceleration needs to be horizontal. We choose our x and y axes as horizontal and vertical so that a_R, which is horizontal, is along the x axis at the moment shown in Fig. 5–24. The forces on the car are the Earth's gravity mg downward, and the normal force F_N exerted by the road perpendicular to its surface. See Fig. 5–24, where the components of F_N are also shown. We don't need to consider the friction of the road because we are designing a road to be banked so as to eliminate dependence on friction.

SOLUTION (*a*) Since there is no vertical motion, $a_y = 0$ and $\Sigma F_y = ma_y$ gives

$$F_N \cos\theta - mg = 0$$

or

$$F_N = \frac{mg}{\cos\theta}.$$

[Note in this case that $F_N \geq mg$ because $\cos\theta \leq 1$. Compare Example 5–14.] We substitute this relation for F_N into the equation for the horizontal motion,

$$F_N \sin\theta = m\frac{v^2}{r},$$

which becomes

$$\frac{mg}{\cos\theta}\sin\theta = m\frac{v^2}{r}$$

or

$$\tan\theta = \frac{v^2}{rg}.$$

This is the formula for the banking angle θ: no friction is needed at this speed v.

(*b*) For $r = 50\,\text{m}$ and $v = 50\,\text{km/h}$ $(= 14\,\text{m/s})$,

$$\tan\theta = \frac{(14\,\text{m/s})^2}{(50\,\text{m})(9.8\,\text{m/s}^2)} = 0.40,$$

so $\theta = \tan^{-1}(0.40) = 22°$.

NOTE Any friction force on the tires, not shown in Fig. 5–24, would point up or down along the banked slope (perpendicular to $\vec{F}_N$), depending on the car's speed.

EXERCISE H Can a heavy truck and a small car travel safely at the same speed around an icy banked-curve road? Why or why not?

5–5 Nonuniform Circular Motion

Circular motion at constant speed occurs when the net force on an object is exerted toward the center of the circle. If the net force is not directed toward the center but is at an angle, as shown in Fig. 5–25a, the force has two components. The component directed toward the center of the circle, $\vec{\mathbf{F}}_R$, gives rise to the centripetal acceleration, $\vec{\mathbf{a}}_R$, and keeps the object moving in a circle. The component tangent to the circle, $\vec{\mathbf{F}}_{tan}$, acts to increase (or decrease) the speed, and thus gives rise to a component of the acceleration tangent to the circle, $\vec{\mathbf{a}}_{tan}$. When the speed of the object is changing, a tangential component of force is acting.

When you first start revolving a ball on the end of a string around your head, you must give it tangential acceleration. You do this by pulling on the string with your hand displaced from the center of the circle. In athletics, a hammer thrower accelerates the hammer tangentially in a similar way so that it reaches a high speed before release.

The tangential component of the acceleration, a_{tan}, has magnitude equal to the rate of change of the *magnitude* of the object's velocity:

$$a_{tan} = \frac{dv}{dt}. \qquad \textbf{(5–4)}$$

The radial (centripetal) acceleration arises from the change in *direction* of the velocity and, as we have seen, has magnitude

$$a_R = \frac{v^2}{r}.$$

The tangential acceleration always points in a direction tangent to the circle, and is in the direction of motion (parallel to $\vec{\mathbf{v}}$, which is always tangent to the circle) if the speed is increasing, as shown in Fig. 5–25b. If the speed is decreasing, $\vec{\mathbf{a}}_{tan}$ points antiparallel to $\vec{\mathbf{v}}$. In either case, $\vec{\mathbf{a}}_{tan}$ and $\vec{\mathbf{a}}_R$ are always perpendicular to each other; and *their directions change* continually as the object moves along its circular path. The total vector acceleration $\vec{\mathbf{a}}$ is the sum of the two components:

$$\vec{\mathbf{a}} = \vec{\mathbf{a}}_{tan} + \vec{\mathbf{a}}_R. \qquad \textbf{(5–5)}$$

Since $\vec{\mathbf{a}}_R$ and $\vec{\mathbf{a}}_{tan}$ are always perpendicular to each other, the magnitude of $\vec{\mathbf{a}}$ at any moment is

$$a = \sqrt{a_{tan}^2 + a_R^2}.$$

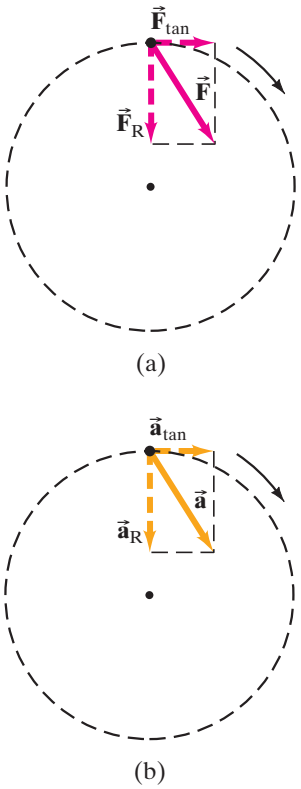


FIGURE 5–25 The speed of an object moving in a circle changes if the force on it has a tangential component, F_{tan}. Part (a) shows the force $\vec{\mathbf{F}}$ and its vector components; part (b) shows the acceleration vector and its vector components.

EXAMPLE 5–16 **Two components of acceleration.** A race car starts from rest and accelerates at a uniform rate to a speed of 35 m/s in 11 s, moving on a circular track of radius 250 m. Assuming constant tangential acceleration, find (*a*) the tangential acceleration, and (*b*) the radial acceleration, at the instant when the speed is $v = 15$ m/s.

APPROACH The tangential acceleration relates to the change in speed of the car, and can be calculated as $a_{tan} = \Delta v/\Delta t$. The centripetal acceleration relates to the change in the *direction* of the velocity vector and is calculated using $a_R = v^2/r$.

SOLUTION (*a*) During the 11-s time interval, we assume the tangential acceleration a_{tan} is constant. Its magnitude is

$$a_{tan} = \frac{\Delta v}{\Delta t} = \frac{(35\text{ m/s} - 0\text{ m/s})}{11\text{ s}} = 3.2\text{ m/s}^2.$$

(*b*) When $v = 15$ m/s, the radial acceleration is

$$a_R = \frac{v^2}{r} = \frac{(15\text{ m/s})^2}{(250\text{ m})} = 0.90\text{ m/s}^2.$$

NOTE The radial (centripetal) acceleration increases continually as the speed increases, whereas the tangential acceleration stays constant in this Example.

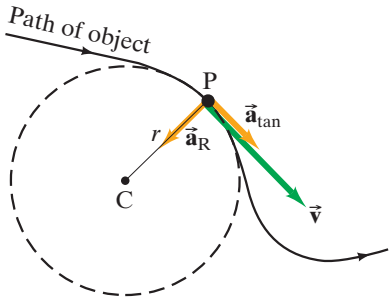


FIGURE 5–26 Object following a curved path (solid line). At point P the path has a radius of curvature r. The object has velocity $\vec{\mathbf{v}}$, tangential acceleration $\vec{\mathbf{a}}_{\tan}$ (the object is here increasing in speed), and radial (centripetal) acceleration $\vec{\mathbf{a}}_{\mathrm{R}}$ (magnitude $a_{\mathrm{R}} = v^2/r$) which points toward the center of curvature C.

EXERCISE I When the speed of the race car in Example 5–16 increases to 30 m/s, how are (*a*) $a_{\tan}$ and (*b*) a_{R} changed?

These concepts can be used for an object moving along any curved path, such as that shown in Fig. 5–26. We can treat any portion of a curve as an arc of a circle with a *radius of curvature r*. The velocity at any point is always tangent to the path. The acceleration can be written, in general, as a vector sum of two components: the tangential component $a_{\tan} = dv/dt$, and the radial (centripetal) component $a_{\mathrm{R}} = v^2/r$.

*5–6 Velocity-Dependent Forces: Drag and Terminal Velocity

When an object slides along a surface, the force of friction acting on the object is nearly independent of how fast the object is moving. But other types of resistive forces do depend on the object's velocity. An important example is for an object moving through a liquid or gas such as air resistance on a falling object. A fluid offers resistance to the motion of an object moving through it, and this resistive force, or **drag force**, depends on the velocity of the object.[†]

The way the drag force varies with velocity is complicated in general. But for small objects at very low speeds, a good approximation can often be made by assuming that the drag force, F_{D}, is directly proportional to the magnitude of the velocity, v:

$$F_{\mathrm{D}} = -bv. \qquad \textbf{(5–6)}$$

The minus sign is necessary because the drag force opposes the motion. Here b is a constant (approximately) that depends on the viscosity of the fluid and on the size and shape of the object. Equation 5–6 works well for small objects moving at low speed in a viscous liquid. It also works for very small objects moving in air at very low speeds, such as dust particles. For objects moving at high speeds, such as an airplane, a sky diver, a baseball, or an automobile, the force of air resistance can be better approximated as being proportional to v^2:

$$F_{\mathrm{D}} \propto v^2.$$

For accurate calculations, however, more complicated forms and numerical integration generally need to be used. For objects moving through liquids, Eq. 5–6 works well for everyday objects at normal speeds (like a boat in water).

Let us consider an object that falls from rest, through air or other fluid, under the action of gravity and a resistive force proportional to v. The forces acting on the object are the force of gravity, mg, acting downward, and the drag force, $-bv$, acting upward (Fig. 5–27a). Since the velocity $\vec{\mathbf{v}}$ points downward, let us take the positive direction as downward. Then the net force on the object can be written

$$\Sigma F = mg - bv.$$

From Newton's second law $\Sigma F = ma$, we have

$$mg - bv = m\frac{dv}{dt}, \qquad \textbf{(5–7)}$$

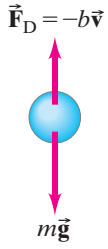
where we have written the acceleration according to its definition as rate of change of velocity, $a = dv/dt$. At $t = 0$, we set $v = 0$ and the acceleration $dv/dt = g$. As the object falls and increases in speed, the resistive force increases, and this reduces the acceleration, dv/dt (see Fig. 5–27b). The velocity continues to increase, but at a slower rate. Eventually, the velocity becomes so large that the magnitude of the resistive force, bv, approaches that of the gravitational force, mg; when the two are equal, we have

$$mg - bv = 0. \qquad \textbf{(5–8)}$$

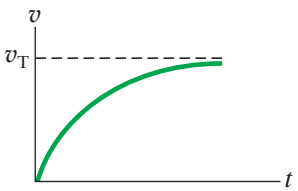
At this point $dv/dt = 0$ and the object no longer increases in speed. It has reached its **terminal velocity** and continues to fall at this constant velocity until it hits the ground. This sequence of events is shown in the graph of Fig. 5–27b.

PHYSICS APPLIED
Air resistance

FIGURE 5–27 (a) Forces acting on an object falling downward. (b) Graph of the velocity of an object falling due to gravity when the air resistance drag force is $F_{\mathrm{D}} = -bv$. Initially, $v = 0$ and $dv/dt = g$, but as time goes on dv/dt (= slope of curve) decreases because of F_{D}. Eventually, v approaches a maximum value, v_{T}, the terminal velocity, which occurs when F_{D} has magnitude equal to mg.



(a)



(b)

[†]Any buoyant force (Chapter 13) is ignored in this Section.

The value of the terminal velocity v_T can be obtained from Eq. 5–8,

$$v_T = \frac{mg}{b}. \tag{5–9}$$

If the resistive force is assumed proportional to v^2, or an even higher power of v, the sequence of events is similar and a terminal velocity reached, although it will not be given by Eq. 5–9.

EXAMPLE 5–17 **Force proportional to velocity.** Determine the velocity as a function of time for an object falling vertically from rest when there is a resistive force linearly proportional to v. Take downward as positive.

APPROACH This is a derivation and we start with Eq. 5–7, which we rewrite as

$$\frac{dv}{dt} = g - \frac{b}{m}v.$$

SOLUTION In this equation there are two variables, v and t. We collect variables of the same type on one or the other side of the equation:[†]

$$\frac{1}{1 - \frac{bv}{mg}}dv = g\,dt.$$

Now we integrate this equation from $t = 0$ when $v = 0$ to a later time t when the speed is v:

$$\int_0^v \frac{1}{1 - \frac{bv}{mg}}dv = g\int_0^t dt$$

which gives

$$-\frac{mg}{b}\left[\ln\left(1 - \frac{bv}{mg}\right) - \ln 1\right] = gt$$

where we have used (Appendix B) $\int dx/(c + ax) = (1/a)\ln(c + ax)$, where $c = 1$ and $a = -b/mg$. (Recall ln is the natural logarithm, Appendix A–7.) We rewrite the last displayed equation, setting $\ln 1 = 0$, as

$$\ln\left(1 - \frac{bv}{mg}\right) = -\frac{b}{m}t.$$

Next we raise each side of this equation to the exponential [the natural log and the exponential are inverse operations of each other: $e^{\ln x} = x$, or $\ln\left(e^x\right) = x$] and obtain

$$1 - \frac{bv}{mg} = e^{-\frac{b}{m}t}$$

so

$$v = \frac{mg}{b}\left(1 - e^{-\frac{b}{m}t}\right).$$

This relation gives the velocity v as a function of time and corresponds to the graph of Fig. 5–27b. As a check, note that at $t = 0$, we get $v = (mg/b)(1 - 1) = 0$, and

$$a(t = 0) = \frac{dv}{dt} = \frac{mg}{b}\frac{d}{dt}\left(1 - e^{-\frac{b}{m}t}\right) = \frac{mg}{b}\left(\frac{b}{m}\right)e^{-0} = g,$$

as expected (see also Eq. 5–7). At large t, $e^{-\frac{b}{m}t}$ approaches zero, so v approaches mg/b, which is the terminal velocity, v_T, as we saw earlier. If we set $\tau = m/b$, then $v = v_T\left(1 - e^{-t/\tau}\right)$. So $\tau = m/b$ is the time required for the velocity to reach 63% of the terminal velocity (since $e^{-1} = 0.37$). Figure 5–27b shows a plot of speed v vs. time t, where the terminal velocity $v_T = mg/b$.

[†]We rearranged it like this also so that the logarithm that we obtain below will be dimensionless, just as θ is for $\sin\theta$ or $\cos\theta$. See Appendix A–7.

Summary

When two objects slide over one another, the force of **friction** that each object exerts on the other can be written approximately as $F_{fr} = \mu_k F_N$, where F_N is the **normal force** (the force each object exerts on the other perpendicular to their contact surfaces), and μ_k is the coefficient of **kinetic friction**. If the objects are at rest relative to each other, even though one is pushed or pulled relative to the other, then F_{fr} is just large enough to hold them at rest and satisfies the inequality $F_{fr} \leq \mu_s F_N$, where μ_s is the coefficient of **static friction**.

An object moving in a circle of radius r with constant speed v is said to be in **uniform circular motion**. It has a **radial acceleration** a_R that is directed radially toward the center of the circle (also called **centripetal acceleration**), and has magnitude

$$a_R = \frac{v^2}{r}. \qquad \textbf{(5–1)}$$

The velocity vector and the acceleration vector $\vec{\mathbf{a}}_R$ are continually changing in direction, but are perpendicular to each other at each moment.

A force is needed to keep an object revolving uniformly in a circle, and the direction of this force is toward the center of the circle. This force may be due to gravity (as for the Moon), to tension in a cord, to a component of the normal force, or to another type of force or a combination of forces.

But there is no outward "centrifugal force" acting on an object moving in circular motion.

When the speed of circular motion is not constant, the acceleration has two components, tangential as well as radial. The force too has tangential and radial components.

[*A **drag force** acts on an object moving through a fluid, such as air or water. The drag force F_D can often be approximated by $F_D = -bv$ or $F_D \propto v^2$, where v is the speed of the object relative to the fluid.]

Questions

1. A heavy crate rests on the bed of a flatbed truck. When the truck accelerates, the crate remains where it is on the truck, so it, too, accelerates. What force causes the crate to accelerate?
2. A block is given a push so that it slides up a ramp. After the block reaches its highest point, it slides back down, but the magnitude of its acceleration is less on the descent than on the ascent. Why?
3. *Cross-country skiers* prefer their skis to have a large coefficient of static friction but a small coefficient of kinetic friction. Explain why. [*Hint*: Think of uphill and downhill.]
4. In the equation for static friction in Section 5–1, what is the significance of the $<$ sign? When should you use the equals sign in the static friction equation?
5. It is not easy to walk on an icy sidewalk without slipping. Even your gait looks different than on dry pavement. Describe what you need to do differently on the icy surface and why.
6. Why is the stopping distance of a truck much shorter than for a train going the same speed?
7. A car rounds a curve at a steady 50 km/h. If it rounds the same curve at a steady 70 km/h, will its acceleration be any different? Explain.
8. Will the acceleration of a car be the same when a car travels around a sharp curve at a constant 60 km/h as when it travels around a gentle curve at the same speed? Explain.
9. Describe all the forces acting on a child riding a horse on a merry-go-round. Which of these forces provides the centripetal acceleration of the child?
10. A child on a sled comes flying over the crest of a small hill, as shown in Fig. 5–28. His sled does not leave the ground, but he feels the normal force between his chest and the sled decrease as he goes over the hill. Explain this decrease using Newton's second law.

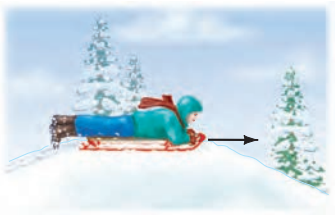


FIGURE 5–28
Question 10.

11. Sometimes it is said that water is removed from clothes in a spin dryer by centrifugal force throwing the water outward. Is this correct? Discuss.
12. Technical reports often specify only the rpm for centrifuge experiments. Why is this inadequate?
13. A girl is whirling a ball on a string around her head in a horizontal plane. She wants to let go at precisely the right time so that the ball will hit a target on the other side of the yard. When should she let go of the string?
14. The game of *tetherball* is played with a ball tied to a pole with a string. When the ball is struck, it whirls around the pole as shown in Fig. 5–29. In what direction is the acceleration of the ball, and what causes the acceleration?

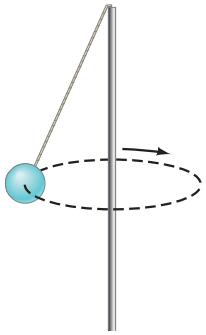


FIGURE 5–29
Question 14.

15. Astronauts who spend long periods in outer space could be adversely affected by weightlessness. One way to simulate gravity is to shape the spaceship like a cylindrical shell that rotates, with the astronauts walking on the inside surface (Fig. 5–30). Explain how this simulates gravity. Consider (*a*) how objects fall, (*b*) the force we feel on our feet, and (*c*) any other aspects of gravity you can think of.

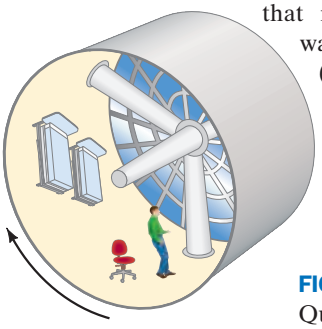


FIGURE 5–30
Question 15.

16. A bucket of water can be whirled in a vertical circle without the water spilling out, even at the top of the circle when the bucket is upside down. Explain.

17. A car maintains a constant speed v as it traverses the hill and valley shown in Fig. 5–31. Both the hill and valley have a radius of curvature R. At which point, A, B, or C, is the normal force acting on the car (*a*) the largest, (*b*) the smallest? Explain. (*c*) Where would the driver feel heaviest and (*d*) lightest? Explain. (*e*) How fast can the car go without losing contact with the road at A?

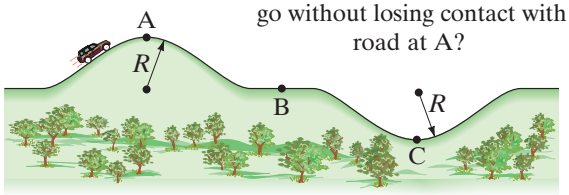


FIGURE 5–31 Question 17.

18. Why do bicycle riders lean in when rounding a curve at high speed?

19. Why do *airplanes bank* when they turn? How would you compute the banking angle given the airspeed and radius of the turn? [*Hint*: Assume an aerodynamic "lift" force acts perpendicular to the wings.]

***20.** For a drag force of the form $F = -bv$, what are the units of b?

***21.** Suppose two forces act on an object, one force proportional to v and the other proportional to v^2. Which force dominates at high speed?

MisConceptual Questions

1. A 50-N crate sits on a horizontal floor where the coefficient of static friction between the crate and the floor is 0.50. A 20-N force is applied to the crate acting to the right. What is the resulting static friction force acting on the crate?
(*a*) 20 N to the right.
(*b*) 20 N to the left.
(*c*) 25 N to the right.
(*d*) 25 N to the left.
(*e*) None of the above; the crate starts to move.

2. What is the maximum acceleration a car can undergo if the coefficient of static friction between the tires and the ground is 0.8? Assume (i) 2-wheel drive, (ii) 4-wheel drive. [*Hint*: Consider the normal force on each tire, assumed equal.]
(*a*) 0.4 *g*, 0.4 *g*.
(*b*) 0.4 *g*, 0.8 *g*.
(*c*) 0.8 *g*, 0.4 *g*.
(*d*) 0.8 *g*, 0.8 *g*.
(*e*) Cannot tell without more information.

3. Which of the following point towards the center of the circle in uniform circular motion?
(*a*) Acceleration.
(*b*) Velocity, acceleration, net force.
(*c*) Velocity, acceleration.
(*d*) Velocity, net force.
(*e*) Acceleration, net force.

4. In a rotating vertical cylinder (Rotor-ride) a rider finds herself pressed with her back to the rotating wall. Which is the correct free-body diagram for her (Fig. 5–32)?

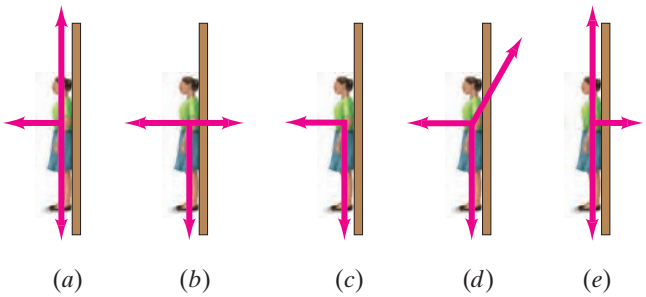


FIGURE 5–32 MisConceptual Question 4.

5. A Ping-Pong ball is shot into a circular tube that is lying flat (horizontal) on a tabletop. When the Ping-Pong ball exits the tube, which path will it follow in Fig. 5–33?

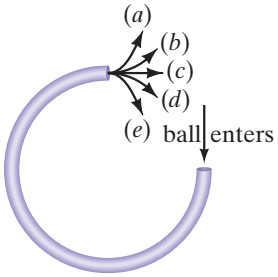


FIGURE 5–33
MisConceptual Question 5.

6. A child has a ball on the end of a cord, and whirls the ball in a vertical circle. Assuming the speed of the ball is constant (an approximation), when would the tension in the cord be greatest?
(*a*) At the top of the circle.
(*b*) At the bottom of the circle.
(*c*) A little after the bottom of the circle when the ball is climbing.
(*d*) A little before the bottom of the circle when the ball is descending quickly.
(*e*) Nowhere; the cord is pulled the same amount at all points.

7. While driving fast around a sharp right turn, you find yourself pressing against the left car door. What is happening?
(*a*) Centrifugal force is pushing you into the door.
(*b*) The door is exerting a rightward force on you.
(*c*) Both of the above.
(*d*) Neither of the above.

8. A car drives at steady speed around a perfectly circular track.
(*a*) The car's acceleration is zero.
(*b*) The net force on the car is zero.
(*c*) Both the acceleration and net force on the car point outward.
(*d*) Both the acceleration and net force on the car point inward.
(*e*) If there is no friction, the acceleration is outward.

9. What supplies the force that keeps a car on the road as it rounds a curve?
(*a*) The force of the steering mechanism on the tires.
(*b*) The forward component of the force of the tires on the road.
(*c*) A sideways component of the force of the tires on the road.
(*d*) A component of the force of the road on the tires that is directed along the road.
(*e*) A component of the force of the road on the tires that is directed sideways.

10. You are flung sideways when your car travels around a sharp curve because
(*a*) you tend to continue moving in a straight line.
(*b*) there is a centrifugal force acting on you.
(*c*) the car exerts an outward force on you.
(*d*) of gravity.

11. A roller coaster rounds the bottom of a circular loop at a nearly constant speed. At this point the net force on the coaster cart is
(*a*) zero.
(*b*) directed upward.
(*c*) directed downward.
(*d*) Cannot tell without knowing the exact speed.

12. A car is traveling around a banked curve at a speed greater than the design speed. What is needed to help the car make the turn?
(*a*) There must be a frictional force pointing up the bank.
(*b*) There must be a frictional force pointing down the bank.
(*c*) No additional force is needed; the banking of the road is sufficient.

13. A car travels around an unbanked curve of constant radius at a steadily decreasing speed. In what direction is the acceleration?
(*a*) In the direction of motion.
(*b*) Opposite the direction of motion.
(*c*) Perpendicular to the direction of motion.
(*d*) None of the above.

Problems

5–1 Friction and Newton's Laws

1. (I) If the coefficient of kinetic friction between a 26-kg crate and the floor is 0.30, what horizontal force is required to move the crate at a steady speed across the floor? What horizontal force is required if μ_k is zero?

2. (I) A force of 35.0 N is required to start a 4.0-kg box moving across a horizontal concrete floor. (*a*) What is the coefficient of static friction between the box and the floor? (*b*) If the 35.0-N force continues, the box accelerates at 0.60 m/s^2. What is the coefficient of kinetic friction?

3. (II) Suppose you are standing on a train accelerating at 0.20 *g*. What minimum coefficient of static friction must exist between your feet and the floor if you are not to slide?

4. (II) The coefficient of static friction between hard rubber and normal street pavement is about 0.90. On how steep a hill (maximum angle) can you leave a car parked?

5. (II) (*a*) A box sits at rest on a rough 28° inclined plane. Draw the free-body diagram, showing all the forces acting on the box. (*b*) How would the diagram change if the box were sliding down the plane? (*c*) How would it change if the box were sliding up the plane after an initial shove?

6. (II) A 25.0-kg box is released on a 27° incline and accelerates down the incline at 0.36 m/s^2. Find the friction force impeding its motion. What is the coefficient of kinetic friction?

7. (II) A car can decelerate at $a_x = -3.80$ m/s^2 without skidding when coming to rest on a level road. What would its deceleration be if the road is inclined at 9.3° and the car moves uphill? Assume the same static friction coefficient.

8. (II) A wet bar of soap slides freely down a ramp 7.5 m long inclined at 8.0°. Starting from rest, how long does it take to reach the bottom? Assume $\mu_k = 0.060$.

9. (II) Show that if a skier moves at constant speed straight down a slope of angle θ (Example 5–6), then the coefficient of kinetic friction between skis and snow is $\mu_k = \tan\theta$. Ignore air resistance.

10. (II) A box is given a push so that it slides across the floor. How far will it go, given that the coefficient of kinetic friction is 0.15 and the push imparts an initial speed of 3.5 m/s?

11. (II) (*a*) Show that the minimum stopping distance for an automobile traveling on a level road at speed v is equal to $v^2/(2\,\mu_s g)$, where μ_s is the coefficient of static friction between the tires and the road, and g is the acceleration of gravity. (*b*) What is this distance for a 1200-kg car traveling 95 km/h if $\mu_s = 0.65$?

12. (II) A 1360-kg car pulls a 350-kg trailer. The car exerts a horizontal force of 3.6×10^3 N against the ground in order to accelerate. What force does the car exert on the trailer? Assume an effective friction coefficient of 0.15 for the trailer.

13. (II) A car starts down a long slippery hill inclined at 15°. The brakes are applied at the top of the hill where the car's speed is 22.0 m/s. If the coefficient of static friction between the road and the tires is 0.35 and the car slows at its maximum possible rate, how far will the car travel down the hill before stopping?

14. (II) Drag-race tires in contact with an asphalt surface have a very high coefficient of static friction. Assuming a constant acceleration and no slipping of tires, estimate the coefficient of static friction needed for a drag racer to cover 1.0 km in 12 s, starting from rest.

15. (II) The girl and sled in Conceptual Example 5–4 have a combined mass of 31 kg. The coefficient of kinetic friction between the sled and the snow is 0.12. (*a*) If you push the sled, as in Fig. 5–6a at an angle of 25°, what magnitude of force is needed to maintain the sled's motion at a constant speed? (*b*) If you pull the sled, as in Fig. 5–6b at an angle of 25°, what magnitude of force is needed to maintain the sled's motion at a constant speed?

16. (II) Police investigators, examining the scene of an accident involving a car and an old truck, measure 72-m-long skid marks for the truck, which nearly came to a stop before colliding with the car at rest. The coefficient of kinetic friction between rubber and the pavement is about 0.80. Estimate the initial speed of the truck assuming a level road.

17. (II) Piles of snow on slippery roofs can become dangerous projectiles as they melt. Consider a chunk of snow at the ridge of a roof with a slope of 24°. (*a*) What is the minimum value of the coefficient of static friction that will keep the snow from sliding down? (*b*) As the snow begins to melt, the coefficient of static friction decreases and the snow finally slips. Assuming that the distance from a chunk of snow to the edge of the roof is 6.0 m and the coefficient of kinetic friction is 0.20, calculate the speed of the snow chunk when it slides off the roof. (*c*) If the edge of the roof is 10.0 m above ground, how far out does the snow hit the ground? (*d*) Estimate the speed of the snow when it hits the ground.

18. (II) A small box is held in place against a rough vertical wall by someone pushing on it with a force directed upward at 22° above the horizontal. The coefficients of static and kinetic friction between the box and wall are 0.40 and 0.30, respectively. The box slides down unless the applied force has magnitude 23 N. What is the mass of the box?

19. (II) Two crates, of mass 65 kg and 125 kg, are in contact and at rest on a horizontal surface (Fig. 5–34). A 650-N force is exerted on the 65-kg crate. If the coefficient of kinetic friction is 0.18, calculate (*a*) the acceleration of the system, and (*b*) the force that each crate exerts on the other. (*c*) Repeat with the crates reversed.



FIGURE 5–34 Problem 19.

20. (II) The crate shown in Fig. 5–35 lies on a plane tilted at an angle $\theta = 25.0°$ to the horizontal, with $\mu_k = 0.19$. (*a*) Determine the acceleration of the crate as it slides down the plane. (*b*) If the crate starts from rest 8.75 m up the plane from its base, what will be the crate's speed when it reaches the bottom of the incline?

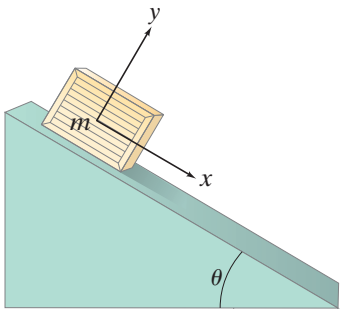


FIGURE 5–35 Crate on inclined plane. Problems 20 and 21.

21. (II) A crate is given an initial speed of 3.0 m/s up the 25.0° plane shown in Fig. 5–35. (*a*) How far up the plane will it go? (*b*) How much time elapses before it returns to its starting point? Assume $\mu_k = 0.17$.

22. (II) A flatbed truck is carrying a heavy crate. The coefficient of static friction between the crate and the bed of the truck is 0.75. What is the maximum rate at which the driver can decelerate and still avoid having the crate slide against the cab of the truck?

23. (II) In Fig. 5–36 the coefficient of static friction between mass m_A and the table is 0.40, whereas the coefficient of kinetic friction is 0.30. (*a*) What minimum value of m_A will keep the system from starting to move? (*b*) What value(s) of m_A will keep the system moving at constant speed? [Ignore masses of the cord and the (frictionless) pulley.]

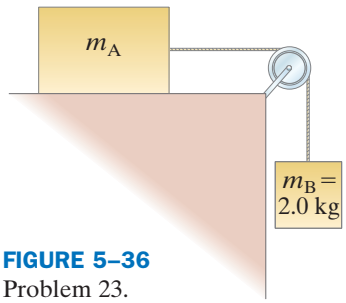


FIGURE 5–36 Problem 23.

24. (II) A 68-kg snowboarder has an initial velocity of 5.0 m/s at the top of a 28° incline (Fig. 5–37). After sliding down the 110-m-long incline (assume a coefficient of kinetic friction $\mu_k = 0.18$), the snowboarder has attained a velocity v. The snowboarder then slides along a flat surface (on which $\mu_k = 0.15$) and comes to rest after a distance x. Use Newton's second law to find the snowboarder's acceleration while on the incline and while on the flat surface. Then use these accelerations to determine x. Ignore air resistance.

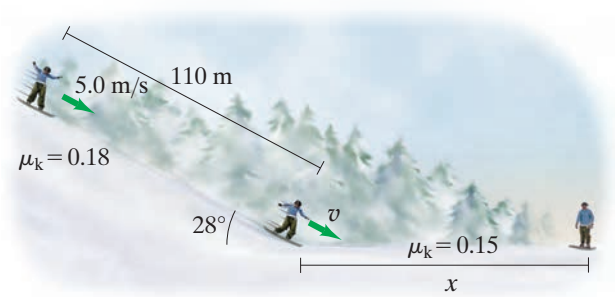


FIGURE 5–37 Problem 24.

25. (II) A package of mass m is dropped vertically onto a horizontal conveyor belt whose speed is $v = 1.5$ m/s, and the coefficient of kinetic friction between the package and the belt is $\mu_k = 0.70$. (*a*) For how much time does the package slide on the belt (until it is at rest relative to the belt)? (*b*) How far does the package move during this time?

26. (II) Two masses $m_A = 2.0$ kg and $m_B = 5.0$ kg are on inclines and are connected together by a string as shown in Fig. 5–38. The coefficient of kinetic friction between each mass and its incline is $\mu_k = 0.30$. If m_A moves up, and m_B moves down, determine their acceleration. Ignore masses of the cord and the pulley (assume frictionless).

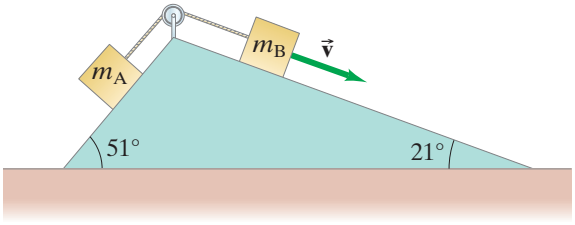


FIGURE 5–38 Problem 26.

27. (II) A child slides down a slide with a 28° incline, and at the bottom her speed is precisely half what it would have been if the slide had been frictionless. Calculate the coefficient of kinetic friction between the slide and the child.

28. (II) (*a*) Suppose the coefficient of kinetic friction between m_A and the ramp in Fig. 5–39 is $\mu_k = 0.15$, and that $m_A = m_B = 2.7$ kg. As m_B moves down, determine the magnitude of the acceleration of m_A and m_B, given $\theta = 34°$. (*b*) What smallest value of μ_k will keep the system from accelerating? [Ignore masses of cord and (frictionless) pulley.]

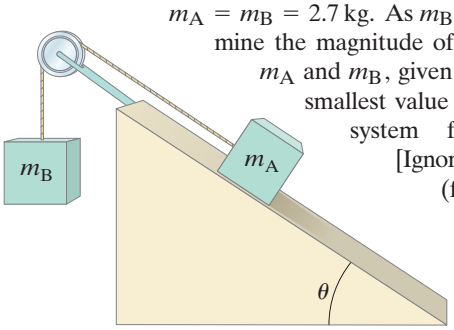


FIGURE 5–39 Problem 28.

29. (II) Two blocks made of different materials connected together by a thin cord slide down a ramp inclined at an angle θ to the horizontal, Fig. 5–40 (block B is above block A). The masses of the blocks are m_A and m_B, and the coefficients of friction are μ_A and μ_B. If $m_A = m_B = 4.0\,\text{kg}$, and $\mu_A = 0.20$ and $\mu_B = 0.30$, determine (*a*) the acceleration of the blocks and (*b*) the tension in the cord, for an angle $\theta = 32°$.

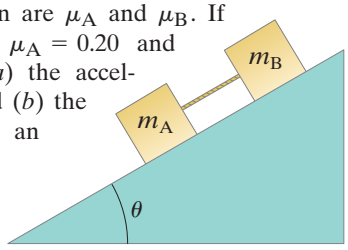


FIGURE 5–40
Problems 29 and 30.

30. (III) For two blocks, connected by a cord and sliding down the incline shown in Fig. 5–40 (see Problem 29), describe the motion (*a*) if $\mu_A < \mu_B$, and (*b*) if $\mu_A > \mu_B$. (*c*) Determine a formula for the acceleration of each block and the tension F_T in the cord in terms of m_A, m_B, and θ; interpret your results in light of your answers to (*a*) and (*b*).

31. (III) A 2.5-kg block is placed on a table as shown in Fig. 5–41. The coefficient of kinetic friction between the block and the table is 0.35. The block is connected by massless ropes over pulleys (whose mass and friction can be ignored) to a 5.0-kg block on the right, and a 3.0-kg block on the left, as shown in Fig. 5–41. Find the acceleration of the block on the table if it is moving to the right.

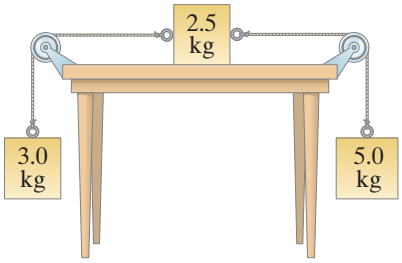


FIGURE 5–41
Problem 31.

32. (III) A 3.0-kg block sits on top of a 5.0-kg block which is on a horizontal surface. The blocks are connected (Fig. 5–42) by a cord over a pulley (ignore mass and friction). The 5.0-kg block is pulled to the right with a force $\vec{\mathbf{F}}$. The coefficient of static friction between all surfaces is 0.60 and the kinetic coefficient is 0.40. (*a*) What is the minimum value of F needed to move the two blocks? (*b*) If the force is 10% greater than your answer for (*a*), what is the acceleration of each block?

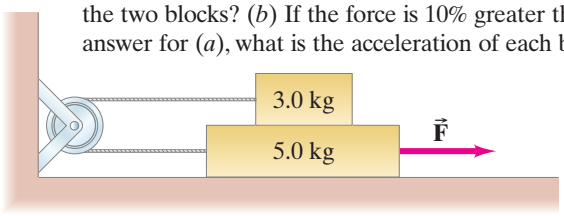


FIGURE 5–42 Problem 32.

33. (III) A 4.0-kg block is stacked on top of a 12.0-kg block, which is accelerating along a horizontal table at $a = 5.2\,\text{m/s}^2$ (Fig. 5–43). Let $\mu_k = \mu_s = \mu$. (*a*) What minimum coefficient of friction μ between the two blocks will prevent the 4.0-kg block from sliding off? (*b*) If μ is only half this minimum value, what is the acceleration of the 4.0-kg block with respect to the table, and (*c*) with respect to the 12.0-kg block? (*d*) What is the force that must be applied to the 12.0-kg block in (*a*) and in (*b*), assuming that the table is frictionless?

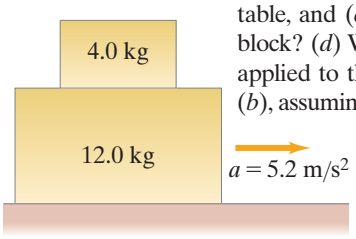


FIGURE 5–43
Problem 33.

34. (III) A small block of mass m rests on the rough, sloping side of a triangular block of mass M which itself rests on a horizontal frictionless table as shown in Fig. 5–44. If the coefficient of static friction is μ, determine the minimum horizontal force F applied to M that will cause the small block m to start moving up the incline.

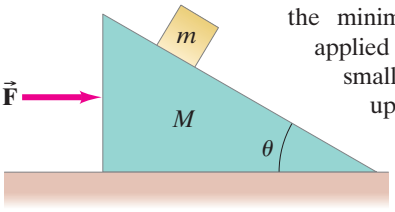


FIGURE 5–44
Problem 34.

5–2 to 5–4 Uniform Circular Motion

35. (I) What is the maximum speed with which a 1200-kg car can round a turn of radius 90.0 m on a flat road if the coefficient of static friction between tires and road is 0.65? Is this result independent of the mass of the car?

36. (I) A child sitting 1.40 m from the center of a merry-go-round moves with a speed of 1.30 m/s. Calculate (*a*) the centripetal acceleration of the child and (*b*) the net horizontal force exerted on the child (mass = 22.5 kg).

37. (I) A horizontal force of 310 N is exerted on a 2.0-kg ball as it rotates (at arm's length) uniformly in a horizontal circle of radius 0.90 m. Calculate the speed of the ball.

38. (I) A jet plane traveling 1890 km/h (525 m/s) pulls out of a dive by moving in an arc of radius 4.80 km. What is the plane's acceleration in g's?

39. (II) How large must the coefficient of static friction be between the tires and the road if a car is to round a level curve of radius 85 m at a speed of 95 km/h?

40. (II) A 0.45-kg ball, attached to the end of a horizontal cord, is revolved in a circle of radius 1.3 m on a frictionless horizontal surface. If the cord will break when the tension in it exceeds 75 N, what is the maximum speed the ball can have?

41. (II) On an ice rink two skaters of equal mass grab hands and spin in a mutual circle once every 2.5 s. If we assume their arms are each 0.80 m long and their individual masses are 55.0 kg, how hard are they pulling on one another?

***42.** (II) How fast (in rpm) must a centrifuge rotate if a particle 8.00 cm from the axis of rotation is to experience an acceleration of 125,000 g's?

43. (II) A car drives in a spiral path of gradually decreasing radius in a large empty parking lot. The speed is held constant at 10.0 m/s, and the car starts at a distance of 50.0 m from the center of the spiral. If the coefficient of static friction is 0.80, how far is the car from the center when it first starts to skid?

44. (II) Redo Example 5–11, precisely this time, by not ignoring the weight of the ball which revolves on a string 0.600 m long. In particular, find the magnitude of $\vec{\mathbf{F}}_T$, and the angle it makes with the horizontal. [*Hint*: Set the horizontal component of $\vec{\mathbf{F}}_T$ equal to ma_R; also, since there is no vertical motion, what can you say about the vertical component of $\vec{\mathbf{F}}_T$?]

45. (II) Highway curves are marked with a suggested speed. If this speed is based on what would be safe in wet weather, estimate the radius of curvature for an unbanked curve marked 50 km/h. Use Table 5–1 (coefficient of friction).

46. (II) Is it possible to whirl a bucket of water fast enough in a vertical circle so that the water won't fall out? If so, what is the minimum speed? Define all quantities needed.

47. (II) At what minimum speed must a roller coaster be traveling so that passengers upside down at the top of a circle (Fig. 5–45) do not fall out? Assume a radius of curvature of 7.6 m.



FIGURE 5–45
Problem 47.

48. (II) A bucket of mass 2.00 kg is whirled in a vertical circle of radius 1.10 m. At the lowest point of its motion the tension in the rope supporting the bucket is 25.0 N. (*a*) Find the speed of the bucket. (*b*) How fast must the bucket move at the top of the circle so that the rope does not go slack?

49. (II) A car drives straight down toward the bottom of a valley and up the other side on a road whose bottom has a radius of curvature of 125 m. At the very bottom, the normal force on the driver is twice his weight. At what speed was the car traveling?

50. (II) How many revolutions per minute would a 22-m-diameter Ferris wheel need to make for the passengers to feel "weightless" at the topmost point?

51. (II) A 1250-kg car rounds a curve of radius 68 m banked at an angle of 14°. If the car is traveling at 85 km/h, will a friction force be required? If so, how much and in what direction?

52. (II) A jet pilot takes his aircraft in a vertical loop (Fig. 5–46). (*a*) If the jet is moving at a speed of 1200 km/h at the lowest point of the loop, determine the minimum radius of the circle so that the centripetal acceleration at the lowest point does not exceed 5.0 *g*'s. (*b*) Calculate the 78-kg pilot's effective weight (the force with which the seat pushes up on him) at the bottom of the circle, and (*c*) at the top of the circle (assume the same speed).

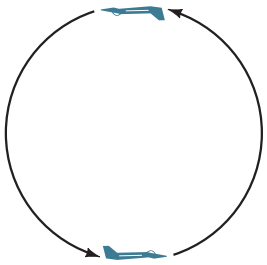


FIGURE 5–46
Problem 52.

53. (II) While fishing, you get bored and start to swing a sinker weight around in a circle below you on a 0.45-m piece of fishing line. The weight makes a complete circle every 0.65 s. What is the angle that the fishing line makes with the vertical? [*Hint*: See Fig. 5–20.]

54. (II) A proposed *space station* consists of a circular tube that will rotate about its center (like a tubular bicycle tire), Fig. 5–47. The circle formed by the tube has a diameter of about 1.1 km. What must be the rotation speed (revolutions per day) if an effect nearly equal to gravity at the surface of the Earth, 0.90 *g*, is to be felt by astronauts walking inside? Which part of the tube do they walk on?

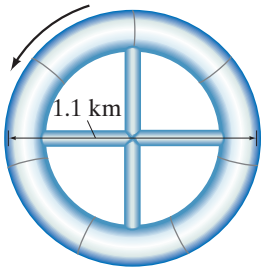


FIGURE 5–47
Problem 54.

55. (II) A bug walks outward from the center of a turntable that is rotating at 65 rpm. If the coefficient of static friction between the bug and the turntable is 0.42, how far out from the center does the bug get before it starts to slip on the turntable? What happens to the bug after it moves farther out past that point?

56. (II) A coin is placed 10.5 cm from the axis of a rotating turntable of variable speed. When the speed of the turntable is slowly increased, the coin remains fixed on the turntable until a rate of 38.0 rpm (revolutions per minute) is reached, at which point the coin slides off. What is the coefficient of static friction between the coin and the turntable?

57. (II) A 975-kg car (including driver) crosses the rounded top of a hill (radius = 88.0 m) at 14.0 m/s. Determine (*a*) the normal force exerted by the road on the car, (*b*) the normal force exerted by the car on the 72.0-kg driver, and (*c*) the car speed at which the car would leave the road (Fig. 5–48).



FIGURE 5–48 Problem 57.

58. (II) Two blocks, with masses m_A and m_B, are connected to each other and to a central post by thin rods as shown in Fig. 5–49. The blocks revolve about the post at the same frequency f (revolutions per second) on a frictionless horizontal surface at distances r_A and r_B from the post. Derive an algebraic expression for the tension in each rod.

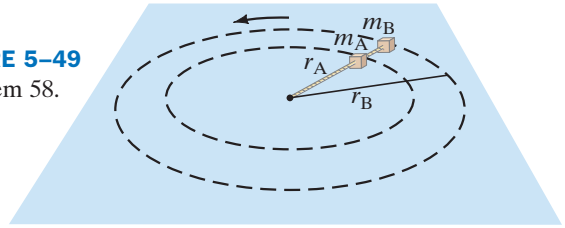


FIGURE 5–49
Problem 58.

59. (II) A device for training astronauts and jet fighter pilots is designed to rotate a trainee in a horizontal circle of radius 11.0 m. If the force felt by a trainee is 7.45 times her own weight, how fast is she revolving? Express your answer in both m/s and rev/s.

60. (II) Tarzan plans to cross a gorge by swinging in an arc from a hanging vine (Fig. 5–50). If his arms are capable of exerting a force of 1350 N on the vine, what is the maximum speed he can tolerate at the lowest point of his swing? His mass is 78 kg and the vine is 4.8 m long.



FIGURE 5–50
Problem 60.

61. (II) A pilot performs an evasive maneuver by diving vertically at a constant 310 m/s. If he can withstand an acceleration of 9.0 *g*'s without blacking out, at what altitude must he begin to pull his plane out of the dive (moving in a vertical circular path) to avoid crashing into the sea?

62. (III) The position of a particle moving in the *xy* plane is given by $\vec{\mathbf{r}} = (2.0\,\text{m})\cos[(3.0\,\text{rad/s})t]\,\hat{\mathbf{i}} + (2.0\,\text{m})\sin[(3.0\,\text{rad/s})t]\,\hat{\mathbf{j}}$, where r is in meters and t is in seconds. (*a*) Show that this represents circular motion of radius 2.0 m centered at the origin. (*b*) Calculate the velocity and acceleration vectors as functions of time. (*c*) Calculate the speed and magnitude of the acceleration. (*d*) Show that $a = v^2/r$. (*e*) Show that the acceleration vector always points toward the center of the circle.

63. (III) If a curve with a radius of 85 m is properly banked for a car traveling 65 km/h, what must be the coefficient of static friction for a car not to skid when traveling at 95 km/h?

64. (III) A curve of radius 75 m is banked for a design speed of 85 km/h. If the coefficient of static friction is 0.30 (wet pavement), at what range of speeds can a car safely make the curve? [*Hint*: Consider the direction of the friction force when the car goes too slow or too fast.]

5–5 Nonuniform Circular Motion

65. (I) Determine the tangential and centripetal components of the net force exerted on the car (by the ground) in Example 5–16 when its speed is 12 m/s. The car's mass is 950 kg.

66. (II) A car at the Indianapolis 500 accelerates uniformly, going from rest to 270 km/h in a semicircular arc with a radius of 220 m. Determine the tangential and radial acceleration of the car when it is halfway through the arc, assuming constant tangential acceleration. If the curve were flat, what coefficient of static friction would be necessary between the tires and the road to provide this acceleration with no slipping or skidding?

67. (II) For each of the cases described below, sketch and label the total acceleration vector, the radial acceleration vector, and the tangential acceleration vector. (*a*) A car speeds up as it rounds a curve of constant radius. (*b*) A car is going a constant 65 km/h as it rounds a curve of constant radius. (*c*) A car slows down while rounding a curve of constant radius.

68. (II) A particle starting from rest revolves with uniformly increasing speed in a clockwise circle in the *xy* plane. The center of the circle is at the origin of an *xy* coordinate system. At $t = 0$, the particle is at $x = 0.0, y = 2.0$ m. At $t = 2.0$ s, it has made one-quarter of a revolution and is at $x = 2.0$ m, $y = 0.0$. Determine (*a*) its speed at $t = 2.0$ s, (*b*) the average velocity vector, and (*c*) the average acceleration vector during this interval.

69. (II) An object moves in a circle of radius 22 m with its speed given by $v = 3.2 + 1.8t^2$, with v in meters per second and t in seconds. At $t = 3.0$ s, find (*a*) the tangential acceleration and (*b*) the radial acceleration.

70. (III) A particle rotates in a circle of radius 3.80 m. At a particular instant its acceleration is 1.15 m/s^2 in a direction that makes an angle of 38.0° to its direction of motion. Determine its speed (*a*) at this moment and (*b*) 3.00 s later, assuming constant tangential acceleration.

71. (III) An object of mass m is constrained to move in a circle of radius r. Its tangential acceleration as a function of time is given by $a_{\tan} = b + ct^2$, where b and c are constants. If $v = v_0$ at $t = 0$, determine the tangential and radial components of the force, $F_{\tan}$ and F_R, acting on the object at any time $t > 0$.

*5–6 Velocity-Dependent Forces

***72.** (I) Use *dimensional analysis* (Section 1–7) in Example 5–17 to determine if the time constant τ is $\tau = m/b$ or $\tau = b/m$.

***73.** (II) The terminal velocity of a 3×10^{-5} kg raindrop is about 9 m/s. Assuming a drag force $F_D = -bv$, determine (*a*) the value of the constant b and (*b*) the time required for such a drop, starting from rest, to reach 63% of terminal velocity.

***74.** (II) Determine a formula for the position and acceleration of a falling object as a function of time if the object starts from rest at $t = 0$ and undergoes a resistive force $F = -bv$, as in Example 5–17.

***75.** (II) The drag force on large objects such as cars, planes, and sky divers moving through air is more nearly $F_D = -cv^2$. (*a*) For this quadratic dependence on v, determine a formula for the terminal velocity v_T, of a vertically falling object. (*b*) Determine the value of the constant c for a 75-kg sky diver who has a terminal velocity of about 60 m/s. (*c*) Sketch a curve like that of Fig. 5–27b for this case of $F_D \propto v^2$. For the same terminal velocity, would this curve lie above or below that in Fig. 5–27? Explain why.

***76.** (III) An object moving vertically has $\vec{\mathbf{v}} = \vec{\mathbf{v}}_0$ at $t = 0$. Determine a formula for its velocity as a function of time assuming a resistive force $F = -bv$ as well as gravity for two cases: (*a*) $\vec{\mathbf{v}}_0$ is downward and (*b*) $\vec{\mathbf{v}}_0$ is upward.

***77.** (III) A bicyclist can coast down a 7.8° hill at a steady 9.5 km/h. If the drag force is proportional to the square of the speed v, so that $F_D = -cv^2$, calculate (*a*) the value of the constant c and (*b*) the average force that must be applied in order to descend the hill at 25 km/h. The mass of the cyclist plus bicycle is 80.0 kg. Ignore other types of friction.

***78.** (III) A block of mass m slides along a horizontal surface lubricated with a thick oil which provides a drag force proportional to the square root of velocity:

$$F_D = -bv^{\frac{1}{2}}.$$

If $v = v_0$ at $t = 0$, determine v and x as functions of time.

***79.** (III) A motorboat traveling at a speed of 2.4 m/s shuts off its engines at $t = 0$. How far does it travel before coming to rest if it is noted that after 3.0 s its speed has dropped to half its original value? Assume that the drag force of the water is proportional to v.

General Problems

80. A coffee cup on the horizontal dashboard of a car slides forward when the driver decelerates from 45 km/h to rest in 3.5 s or less, but not if she decelerates in a longer time. What is the coefficient of static friction between the cup and the dash? Assume the road and the dashboard are level (horizontal).

81. A 1.6-kg silverware drawer does not slide readily. The owner gradually pulls with more and more force, and when the applied force reaches 9.0 N, the drawer suddenly opens, throwing all the utensils to the floor. What is the coefficient of static friction between the drawer and the cabinet?

82. A roller coaster reaches the top of the steepest hill with a speed of 6.0 km/h. It then descends the hill, which is at an average angle of 45° and is 45.0 m long. What will its speed be when it reaches the bottom? Assume $\mu_k = 0.14$.

83. An 18.0-kg box is released on a 32.0° incline and accelerates down the incline at 0.220 m/s^2. Find the friction force impeding its motion. How large is the coefficient of friction?

84. A flat puck (mass M) is revolved in a circle on a frictionless air hockey table top, and is held in this orbit by a light cord which is connected to a dangling mass (mass m) through a central hole as shown in Fig. 5–51. Show that the speed of the puck is given by $v = \sqrt{mgR/M}$.

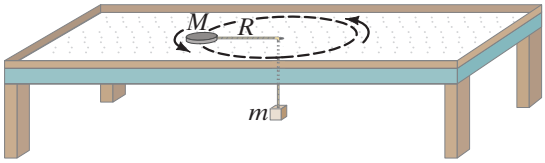


FIGURE 5–51 Problem 84.

85. In a "Rotor-ride" at a carnival, people rotate in a vertical cylindrically walled "room." See Fig. 5–52. If the room radius is 5.5 m, and the rotation frequency 0.50 revolutions per second when the floor drops out, what minimum coefficient of static friction keeps the people from slipping down? People on this ride said they were "pressed against the wall." Is there really an outward force pressing them against the wall? If so, what is its source? If not, what is the proper description of their situation (besides nausea)? [*Hint*: Draw a free-body diagram for a person.]



FIGURE 5–52 Problem 85.

86. At time $t = 0$, a car at the starting line of a circular race track of radius 410 m is moving at 11 m/s. The car's speed increases at a constant rate of 0.22 m/s^2. (*a*) At what time does the centripetal acceleration of the car equal g? (*b*) How many laps has the car made on the track at this time?

87. Determine the tangential and centripetal components of the net force exerted on a car (by the ground) when its speed is 27 m/s, and it has accelerated to this speed from rest in 11 s on a curve of radius 450 m. The car's mass is 1150 kg.

88. The 70.0-kg climber in Fig. 5–53 is supported in the "chimney" by the friction forces exerted on his shoes and back. The static coefficients of friction between his shoes and the wall, and between his back and the wall, are 0.80 and 0.60, respectively. What is the minimum normal force he must exert? Assume the walls are vertical and that the static friction forces are both at their maximum. Ignore his grip on the rope.



FIGURE 5–53 Problem 88.

89. A small mass m is set on the surface of a sphere, Fig. 5–54. If the coefficient of static friction is $\mu_s = 0.70$, at what angle ϕ would the mass start sliding?

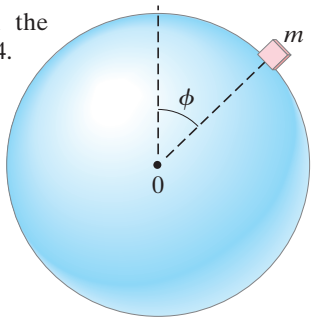


FIGURE 5–54 Problem 89.

90. A train traveling at a constant speed rounds a curve of radius 215 m. A lamp suspended from the ceiling swings out to an angle of 18.5° throughout the curve. What is the speed of the train? [*Hint*: See Example 4–15.]

91. Calculate (*a*) the centripetal acceleration of the Earth in its orbit around the Sun, and (*b*) the net force exerted on the Earth. What exerts this force on the Earth? Assume that the Earth's orbit is a circle of radius 1.50×10^{11} m, and it takes 365 days to make one complete orbit.

92. Consider a train that rounds a curve with a radius of 530 m at a speed of 160 km/h (approximately 100 mi/h). (*a*) Calculate the friction force needed on a train passenger of mass 55 kg if the track is not banked and the train does not tilt. (*b*) Calculate the friction force on the passenger if the train tilts at an angle of 8.0° toward the center of the curve.

93. A ball of mass 0.75 kg is attached to a vertical post by strings of length 1.1 m and 1.7 m, as shown in Fig. 5–55. If the ball is set moving in a horizontal circle, find (*a*) the maximum speed for the lower string to be slack (tension in cord = 0), and (*b*) the tension in each string if the ball is moving at twice the speed found in part (*a*).

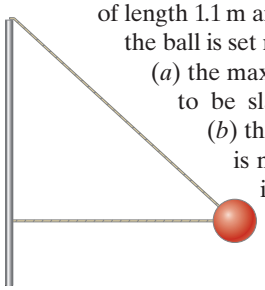


FIGURE 5–55 Problem 93.

94. A small block of mass m is given an initial speed v_0 up a ramp inclined at angle θ to the horizontal. It travels a distance d up the ramp and comes to rest. (*a*) Determine a formula for the coefficient of kinetic friction between block and ramp. (*b*) What can you say about the value of the coefficient of static friction?

95. A 28.0-kg block is connected to an empty 2.00-kg bucket by a cord running over a frictionless pulley (Fig. 5–56). The coefficient of static friction between the table and the block is 0.42 and the coefficient of kinetic friction between the table and the block is 0.34. Sand is gradually added to the bucket until the system just begins to move. (*a*) Calculate the mass of sand added to the bucket. (*b*) Calculate the acceleration of the system.

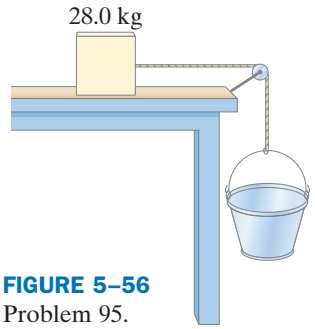


FIGURE 5–56 Problem 95.

96. A car is heading along a flat slippery road at a speed of 95 km/h. The minimum distance within which it can stop (because of ABS) is 66 m. Estimate the radius of the sharpest curve the car can negotiate on an icy flat surface at the same speed without skidding out.

97. An airplane traveling at 510 km/h needs to reverse its course. The pilot decides to accomplish this by banking the wings at an angle of 38° while moving in a level circular path. (*a*) Find the time needed to reverse course. (*b*) Describe any additional force the passengers experience during the turn. [*Hint*: Assume an aerodynamic "lift" force that acts perpendicularly to the flat wings; see Fig. 5–57.]

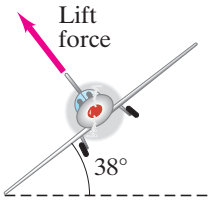


FIGURE 5–57 Problem 97.

98. A banked curve of radius R in a new highway is designed so that a car traveling at speed v_0 can negotiate the turn safely on glare ice (zero friction). If a car travels too slowly, then it will slip toward the center of the circle. If it travels too fast, it will slip away from the center of the circle. If the coefficient of static friction increases, it becomes possible for a car to stay on the road while traveling at a speed within a range from v_{min} to v_{max}. Derive formulas for v_{min} and v_{max} as functions of μ_s, v_0, and R.

99. *Earth is not quite an inertial frame.* We often make measurements in a reference frame fixed on the Earth, assuming Earth is an inertial reference frame [Section 4–2]. But the Earth rotates, so this assumption is not quite valid. Show that this assumption is off by 3 parts in 1000 by calculating the acceleration of an object at Earth's equator due to Earth's daily rotation, and compare to $g = 9.80\ \text{m/s}^2$, the acceleration due to gravity.

100. A small bead of mass m is constrained to slide without friction inside a circular vertical hoop of radius r which rotates about a vertical axis (Fig. 5–58) at a frequency f. (*a*) Determine the angle θ where the bead will be in equilibrium within the hoop—that is, where it will have no tendency to move up or down along the hoop. (*b*) If $f = 2.00$ rev/s and $r = 25.0$ cm, what is θ? (*c*) Can the bead ride as high as the center of the circle ($\theta = 90°$)? Explain.

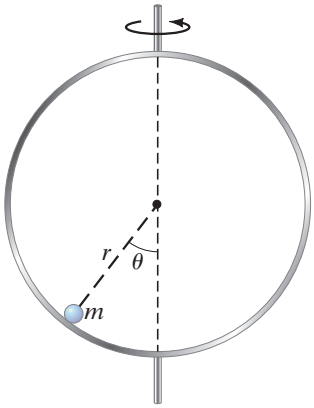


FIGURE 5–58 Problem 100.

***101.** Use *dimensional analysis* (Section 1–7) to obtain the form for the centripetal acceleration, $a_R = v^2/r$.

102. A car starts rolling down a 1-in-4 hill (1-in-4 means that for each 4 m traveled along the road, the elevation change is 1 m). How fast is it going when it reaches the bottom after traveling 55 m? (*a*) Ignore friction. (*b*) Assume an effective coefficient of *rolling friction* equal to 0.10.

103. The sides of a cone make an angle ϕ with the vertical, with its point down. A small cube of mass m is placed on the inside of the cone, and the cone is revolved at a frequency f (revolutions per second) about its symmetry axis. If the coefficient of static friction is μ_s, at what positions on the cone can the cube be placed without sliding on the cone? (Give the maximum and minimum distances, r, from the axis).

104. A 68-kg water skier is being accelerated by a ski boat on a flat ("glassy") lake. The coefficient of kinetic friction between the skier's skis and the water surface is $\mu_k = 0.25$ (Fig. 5–59). (*a*) What is the skier's acceleration if the rope pulling the skier behind the boat applies a horizontal tension force of magnitude $F_T = 240$ N to the skier ($\theta = 0°$)? (*b*) What is the skier's horizontal acceleration if the rope pulling the skier exerts a force of $F_T = 240$ N on the skier at an upward angle $\theta = 12°$? (*c*) Explain why the skier's acceleration in part (*b*) is greater than that in part (*a*).

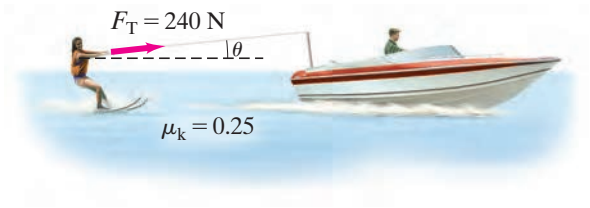


FIGURE 5–59 Problem 104.

105. A ball of mass $m = 1.0$ kg at the end of a thin cord of length $r = 0.80$ m revolves in a vertical circle about point O, as shown in Fig. 5–60. During the time we observe it, the only forces acting on the ball are gravity and the tension in the cord. The motion is circular but not uniform because of the force of gravity. The ball increases in speed as it descends and slows down as it rises on the other side of the circle. At the moment the cord makes an angle $\theta = 30°$ below the horizontal, the ball's speed is 6.0 m/s. At this point, determine the tangential acceleration, the radial acceleration, and the tension in the cord, F_T. Take θ increasing downward as shown.

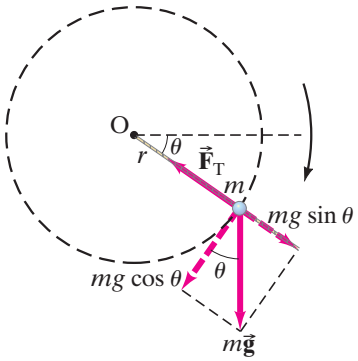


FIGURE 5–60 Problem 105.

106. A car drives at a constant speed around a banked circular track with a diameter of 145 m. The motion of the car can be described in a coordinate system with its origin at the center of the circle. At a particular instant the car's acceleration in the horizontal plane is given by

$$\vec{\mathbf{a}} = (-15.7\hat{\mathbf{i}} - 23.2\hat{\mathbf{j}})\ \text{m/s}^2.$$

(*a*) What is the car's speed? (*b*) Where (x and y) is the car at this instant?

ANSWERS TO EXERCISES

A: (*c*).
B: The box will slow down and stop.
C: No—the acceleration is not constant (in direction).
D: (*a*), it doubles.
E: (*d*).
F: (*a*).
G: No. Because $a = F_{\text{fr}}/m = \mu mg/m = \mu g$.
H: Yes. θ does not depend on m (Example 5–15).
I: (*a*) No change; (*b*) 4 times larger.



The astronauts in the upper left of this photo are working on a space vehicle. As they orbit the Earth—at a rather high speed—they experience apparent weightlessness. The Moon, in the background, also is orbiting the Earth at high speed. What keeps the Moon and the space vehicle (and its astronauts) from moving off in a straight line away from Earth? It is the force of gravity. Newton's law of universal gravitation states that all objects attract all other objects with a force proportional to their masses and inversely proportional to the square of the distance between them.

CHAPTER 6

Gravitation and Newton's Synthesis

CHAPTER-OPENING QUESTION—Guess now!

A space station revolves around the Earth as a satellite, 100 km above Earth's surface. What is the net force on an astronaut at rest inside the space station?

(a) Equal to her weight on Earth.
(b) A little less than her weight on Earth.
(c) Less than half her weight on Earth.
(d) Zero (she is weightless).
(e) Somewhat larger than her weight on Earth.

Sir Isaac Newton not only put forth the three great laws of motion that serve as the foundation for the study of dynamics. He also conceived of another great law to describe one of the basic forces in nature, gravitation; and he applied it to understand the motion of the planets. This new law, published in 1687 in his book *Philosophiae Naturalis Principia Mathematica* (the *Principia*, for short), is called Newton's law of universal gravitation. It was the capstone of Newton's analysis of the physical world. Indeed, Newtonian mechanics, with its three laws of motion and the law of universal gravitation, was accepted for centuries as a mechanical basis for the way the universe works.

CONTENTS

6–1 Newton's Law of Universal Gravitation
6–2 Vector Form of Newton's Law of Universal Gravitation
6–3 Gravity Near the Earth's Surface
6–4 Satellites and "Weightlessness"
6–5 Planets, Kepler's Laws, and Newton's Synthesis
6–6 Moon Rises an Hour Later Each Day
6–7 Types of Forces in Nature
*6–8 Gravitational Field
*6–9 Principle of Equivalence; Curvature of Space; Black Holes

6–1 Newton's Law of Universal Gravitation

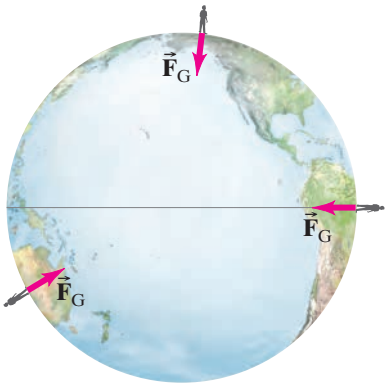


FIGURE 6–1 Anywhere on Earth, whether in Alaska, Australia, or Peru, the force of gravity acts downward toward the center of the Earth.

Among his many great accomplishments, Sir Isaac Newton examined the motion of the heavenly bodies—the planets and the Moon. In particular, he wondered about the nature of the force that must act to keep the Moon in its nearly circular orbit around the Earth.

Newton was also thinking about the problem of gravity. Since falling objects accelerate, Newton had concluded that they must have a force exerted on them, a force we call the force of gravity. Whenever an object has a force exerted *on* it, that force is exerted *by* some other object. But what *exerts* the force of gravity? Every object on the surface of the Earth feels the force of gravity F_G, and no matter where the object is, the force is directed toward the center of the Earth (Fig. 6–1). Newton concluded that it must be the Earth itself that exerts the gravitational force on objects at its surface.

According to legend, Newton noticed an apple drop from a tree. He is said to have been struck with a sudden inspiration: If gravity acts at the tops of trees, and even at the tops of mountains, then perhaps it acts all the way to the Moon! With this idea that it is Earth's gravity that holds the Moon in its orbit, Newton developed his great theory of gravitation. But there was controversy at the time. Many thinkers had trouble accepting the idea of a force "acting at a distance." Typical forces act through contact—your hand pushes a cart and pulls a wagon, a bat hits a ball, and so on. But gravity acts without contact, said Newton: the Earth exerts a force on a falling apple and on the Moon, even though there is no contact, and the two objects may even be very far apart.

Newton set about determining the magnitude of the gravitational force that the Earth exerts on the Moon as compared to the gravitational force on objects at the Earth's surface. The centripetal acceleration of the Moon is calculated from $a_R = v^2/r$ (see Example 5–9) and gives $a_R = 0.00272\,\text{m/s}^2$. In terms of the acceleration due to gravity at the Earth's surface, $g = 9.80\,\text{m/s}^2$,

$$a_R = \frac{0.00272\,\text{m/s}^2}{9.80\,\text{m/s}^2}\,g \approx \frac{1}{3600}\,g.$$

That is, the acceleration of the Moon toward the Earth is about $\frac{1}{3600}$ as great as the acceleration of objects at the Earth's surface. The Moon is 384,000 km from the Earth, which is about 60 times the Earth's radius of 6380 km. That is, the Moon is 60 times farther from the Earth's center than are objects at the Earth's surface. But $60 \times 60 = 60^2 = 3600$. Again that number 3600! Newton concluded that the gravitational force F_G exerted by the Earth on any object decreases with the *square* of its distance, r, from the Earth's center:

$$F_G \propto \frac{1}{r^2}.$$

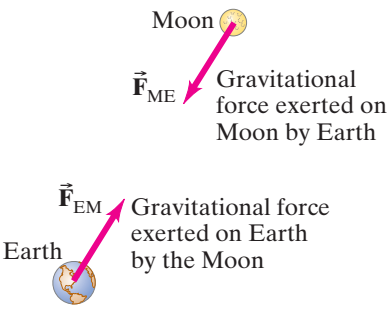
The Moon is 60 Earth radii away, so it feels a gravitational force only $\frac{1}{60^2} = \frac{1}{3600}$ times as strong as it would if it were at a point on the Earth's surface.

Newton realized that the force of gravity on an object depends not only on distance but also on the object's mass. In fact, it is directly proportional to its mass, as we have seen (Eq. 4–3). According to Newton's third law, when the Earth exerts its gravitational force on any object, such as the Moon, that object exerts an equal and opposite force on the Earth (Fig. 6–2). Because of this *symmetry*, Newton reasoned, the magnitude of the force of gravity must be proportional to *both* the masses. Thus

$$F_G \propto \frac{m_E m_{obj}}{r^2},$$

where m_E and m_{obj} are the masses of the Earth and the other object, respectively, and r is the distance from the Earth's center to the center of the other object.

FIGURE 6–2 The gravitational force one object exerts on a second object is directed toward the first object; and, by Newton's third law, is equal and opposite to the force exerted by the second object on the first. In the case shown, the gravitational force on the Moon due to Earth, $\vec{F}_{ME}$, is equal and opposite to the gravitational force on Earth due to the Moon, $\vec{F}_{EM}$. That is, $\vec{F}_{ME} = -\vec{F}_{EM}$.



Newton went further in his analysis of gravity. He examined the orbits of the planets, and using Kepler's laws (Section 6–5) he concluded that the force needed to hold different planets in their orbits around the Sun diminishes as the inverse square of their distance from the Sun. This led him to believe that the gravitational force acts between the Sun and each of the planets to keep them in their orbits. And if gravity acts between these objects, why not between all objects? Thus he proposed his **law of universal gravitation**, which we can state as follows:

NEWTON'S LAW OF UNIVERSAL GRAVITATION

Every particle in the universe attracts every other particle with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between them. This force acts along the line joining the two particles.

The magnitude of the gravitational force can be written as

$$F_G = G\frac{m_1 m_2}{r^2}, \qquad \textbf{(6–1)}$$

NEWTON'S LAW OF UNIVERSAL GRAVITATION

where m_1 and m_2 are the masses of the two particles, r is the distance between them, and G is a universal constant. G is called the **gravitational constant**, and it must be measured experimentally.

Equation 6–1 is called an **inverse square law** because the force is inversely proportional to r^2.

The value of G must be very small, since we are not aware of any force of attraction between ordinary-sized objects, such as between two baseballs. The gravitational force between two ordinary objects was first measured by Henry Cavendish in 1798, over 100 years after Newton published his law. To detect and measure the incredibly small force between ordinary objects, he used an apparatus like that shown in Fig. 6–3. Cavendish confirmed Newton's hypothesis that two objects attract one another and that Eq. 6–1 accurately describes this force. In addition, because Cavendish could measure F_G, m_1, m_2, and r accurately, he was able to determine the value of the constant G as well. The accepted value today is

$$G = 6.67 \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2.$$

(See Table inside front cover for values of all constants to highest known precision.)

Strictly speaking, Eq. 6–1 gives the magnitude of the gravitational force that one particle exerts on a second particle that is a distance r away. For an extended object (that is, not a point), we must consider how to measure the distance r. A correct calculation treats each extended body as a collection of particles, and the total force is the sum of the forces due to all the particles. The sum over all these particles is often done using integral calculus, which Newton himself invented. When extended bodies are small compared to the distance between them (as for the Earth–Sun system), little inaccuracy results from considering them as point particles.

Newton was able to show (see derivation in Appendix E) that the *gravitational force exerted on a particle outside a uniform sphere is the same as if the entire mass of the sphere was concentrated at its center*. Thus Eq. 6–1 gives the correct force between two uniform spheres where r is the distance between their centers.

EXAMPLE 6–1 ESTIMATE Can you attract another person gravitationally?

A 50-kg person and a 70-kg person are sitting on a bench close to each other. Estimate the magnitude of the gravitational force each exerts on the other.

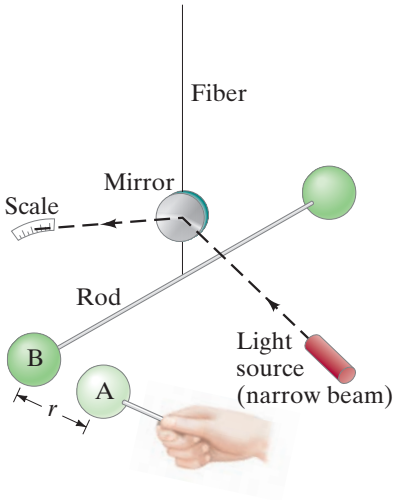
APPROACH This is an estimate: we let the distance between the centers of the two people be about $\frac{1}{2}$ m (about as close as you can get).

SOLUTION We use Eq. 6–1, which gives

$$F_G = G\frac{m_1 m_2}{r^2} \approx \frac{(6.67 \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(50\,\text{kg})(70\,\text{kg})}{(0.5\,\text{m})^2} \approx 10^{-6}\,\text{N},$$

rounded off to an order of magnitude. Such a force is unnoticeably small unless extremely sensitive instruments are used $\left(< \frac{1}{100{,}000}\text{ of a pound}\right)$.

FIGURE 6–3 Schematic diagram of Cavendish's apparatus. Two spheres are attached to a light horizontal rod, which is suspended at its center by a thin fiber. When a third sphere (labeled A) is brought close to one of the suspended spheres (B), the gravitational force causes sphere B to move, and this twists the fiber slightly. The tiny movement is magnified by the use of a narrow light beam directed at a mirror mounted on the fiber. The beam reflects onto a scale. Previous determination of how large a force will twist the fiber a given amount then allows the experimenter to determine the magnitude of the gravitational force between two objects, A and B.



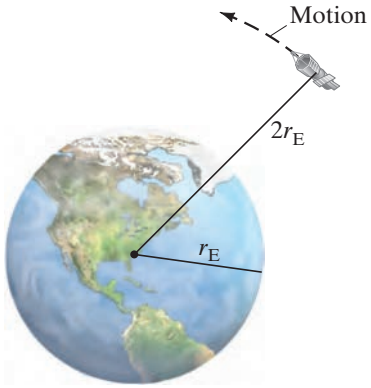
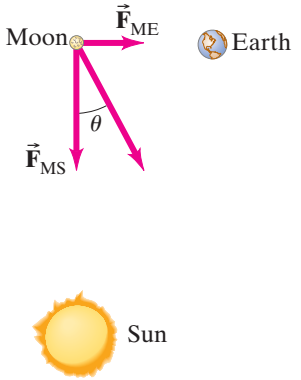


FIGURE 6–4 Example 6–2; a spacecraft in a circular orbit at 6380 km ($= r_E$) above the Earth, so $r = 2r_E$.

FIGURE 6–5 Example 6–3. Orientation of Sun (S), Earth (E), and Moon (M) at right angles to each other (not to scale).



EXAMPLE 6–2 **Spacecraft at $2r_E$.** What is the force of gravity acting on a 2000-kg spacecraft when it orbits two Earth radii from the Earth's center (that is, it stays a distance $r_E = 6380\,\text{km}$ above the Earth's surface, Fig. 6–4)? The mass of the Earth is $m_E = 5.98 \times 10^{24}\,\text{kg}$. Assume a circular orbit.

APPROACH We could plug all the numbers into Eq. 6–1, but there is a simpler approach. The spacecraft is twice as far from the Earth's center as when it is at the surface of the Earth. Therefore, since the force of gravity F_G decreases as the square of the distance (and $\frac{1}{2^2} = \frac{1}{4}$), the force of gravity on the spacecraft will be only one-fourth its weight at the Earth's surface.

SOLUTION At the surface of the Earth, $F_G = mg$. At a distance from the Earth's center of $2r_E$, F_G is $\frac{1}{4}$ as great:

$$F_G = \tfrac{1}{4}mg = \tfrac{1}{4}(2000\,\text{kg})(9.80\,\text{m/s}^2) = 4900\,\text{N}.$$

EXAMPLE 6–3 **Force on the Moon.** Find the net force on the Moon ($m_M = 7.35 \times 10^{22}\,\text{kg}$) due to the gravitational attraction of both the Earth ($m_E = 5.98 \times 10^{24}\,\text{kg}$) and the Sun ($m_S = 1.99 \times 10^{30}\,\text{kg}$), assuming they are at right angles to each other as in Fig. 6–5.

APPROACH The forces on our object, the Moon, are the gravitational force exerted on the Moon by the Earth F_{ME} and the force exerted by the Sun F_{MS}, as shown in the free-body diagram of Fig. 6–5. We use the law of universal gravitation to find the magnitude of each force, and then add the two forces as vectors.

SOLUTION The Earth is $3.84 \times 10^5\,\text{km} = 3.84 \times 10^8\,\text{m}$ from the Moon, so F_{ME} (the gravitational force on the Moon due to the Earth) is

$$F_{ME} = \frac{(6.67\times10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(7.35\times10^{22}\,\text{kg})(5.98\times10^{24}\,\text{kg})}{(3.84\times10^{8}\,\text{m})^2} = 1.99\times10^{20}\,\text{N}.$$

The Earth is $1.50 \times 10^8\,\text{km}$ from the Sun and so is the Moon (the Earth-Moon distance is small compared to this), so F_{MS} (the gravitational force on the Moon due to the Sun) is

$$F_{MS} = \frac{(6.67\times10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(7.35\times10^{22}\,\text{kg})(1.99\times10^{30}\,\text{kg})}{(1.50\times10^{11}\,\text{m})^2} = 4.34\times10^{20}\,\text{N}.$$

The two forces act at right angles in the case we are considering (Fig. 6–5), so we can apply the Pythagorean theorem to find the magnitude of the total force:

$$F = \sqrt{(1.99 \times 10^{20}\,\text{N})^2 + (4.34 \times 10^{20}\,\text{N})^2} = 4.77 \times 10^{20}\,\text{N}.$$

The force acts at an angle θ (Fig. 6–5) given by $\theta = \tan^{-1}(1.99/4.34) = 24.6°$.

NOTE The two forces, F_{ME} and F_{MS}, have the same order of magnitude ($10^{20}\,\text{N}$). This may be surprising. Is it reasonable? The Moon is much farther from the Sun than from Earth (a factor of $10^{11}\,\text{m}/10^8\,\text{m} \approx 10^3$), but the Sun is much more massive (a factor of $10^{30}\,\text{kg}/10^{25}\,\text{kg} \approx 10^5$). Mass divided by distance squared ($10^5/10^6$) comes out within an order of magnitude, and we have ignored factors of 3 or more. Yes, it is reasonable.

⚠ CAUTION

Distinguish between Newton's second law and the law of universal gravitation

Note carefully that the law of universal gravitation describes a *particular* force (gravity), whereas Newton's second law of motion ($F = ma$) tells how an object accelerates due to *any* type of force.

Spherical Shells

Newton was able to show, using the calculus he invented for the purpose, that a thin uniform spherical shell exerts a force on a particle *outside* it as if all the shell's mass were at its center; and that such a thin uniform shell exerts *zero* force on a particle *inside* the shell. (The derivation is given in Appendix E.)

The Earth can be modeled as a series of concentric shells starting at its center, each shell uniform but perhaps having a different density to take into account Earth's varying density in various layers. As a simple example, suppose the Earth were uniform throughout: What is the gravitational force on a particle placed exactly halfway from Earth's center to its surface? Only the mass *inside* this radius $r = \frac{1}{2} r_E$ would exert a net force on this particle. The mass of a sphere is proportional to its volume $V = \frac{4}{3}\pi r^3$, so the mass m inside $r = \frac{1}{2} r_E$ is $\left(\frac{1}{2}\right)^3 = \frac{1}{8}$ the mass of the entire Earth. The gravitational force on the particle at $r = \frac{1}{2} r_E$, which is proportional to m/r^2 (Eq. 6–1), is reduced to $\left(\frac{1}{8}\right)/\left(\frac{1}{2}\right)^2 = \frac{1}{2}$ the gravitational force it would experience at Earth's surface. An elegant way to deal with these results is the use of Gauss's law, discussed in Chapter 22 (Example 22–4).

6–2 Vector Form of Newton's Law of Universal Gravitation

We can write Newton's law of universal gravitation in vector form as

$$\vec{\mathbf{F}}_{12} = -G\frac{m_1 m_2}{r_{21}^2}\hat{\mathbf{r}}_{21}, \tag{6–2}$$

where $\vec{\mathbf{F}}_{12}$ is the vector force *on* particle 1 (of mass m_1) exerted *by* particle 2 (of mass m_2), which is a distance r_{21} away. The unit vector $\hat{\mathbf{r}}_{21}$ points from particle 2 toward particle 1 along the line joining them so that $\hat{\mathbf{r}}_{21} = \vec{\mathbf{r}}_{21}/r_{21}$, where $\vec{\mathbf{r}}_{21}$ is the displacement vector as shown in Fig. 6–6. The minus sign in Eq. 6–2 is necessary because the force on particle 1 due to particle 2 points toward m_2, in the direction opposite to $\hat{\mathbf{r}}_{21}$. The displacement vector $\vec{\mathbf{r}}_{12}$ is a vector of the same magnitude as $\vec{\mathbf{r}}_{21}$, but it points in the opposite direction so that

$$\vec{\mathbf{r}}_{12} = -\vec{\mathbf{r}}_{21}.$$

By Newton's third law, the force $\vec{\mathbf{F}}_{21}$ acting on m_2 exerted by m_1 must have the same magnitude as $\vec{\mathbf{F}}_{12}$ but acts in the opposite direction (Fig. 6–7), so that

$$\vec{\mathbf{F}}_{21} = -\vec{\mathbf{F}}_{12} = G\frac{m_1 m_2}{r_{21}^2}\hat{\mathbf{r}}_{21} = -G\frac{m_2 m_1}{r_{12}^2}\hat{\mathbf{r}}_{12}.$$

The force of gravity exerted on one particle by a second particle is always directed toward the second particle, as in Fig. 6–6. When many particles interact, the total gravitational force on a given particle is the vector sum of the forces exerted by each of the others. For example, the total force on particle number 1 is

$$\vec{\mathbf{F}}_1 = \vec{\mathbf{F}}_{12} + \vec{\mathbf{F}}_{13} + \vec{\mathbf{F}}_{14} + \cdots + \vec{\mathbf{F}}_{1n} = \sum_{i=2}^{n} \vec{\mathbf{F}}_{1i} \tag{6–3}$$

where $\vec{\mathbf{F}}_{1i}$ means the force on particle 1 exerted by particle i, and n is the total number of particles.

This vector notation can be very helpful, especially when sums over many particles are needed. However, in many cases we do not need to be so formal and we can deal with directions by making careful diagrams.

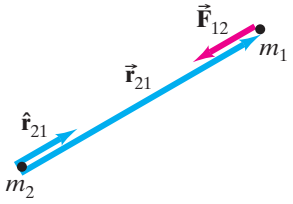


FIGURE 6–6 The displacement vector $\vec{\mathbf{r}}_{21}$ points from particle of mass m_2 to particle of mass m_1. The unit vector shown, $\hat{\mathbf{r}}_{21}$ is in the same direction as $\vec{\mathbf{r}}_{21}$, but is defined as having length one.

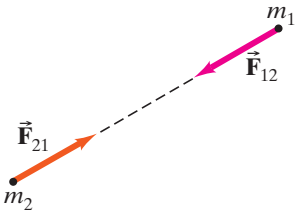


FIGURE 6–7 By Newton's third law, the gravitational force on particle 1 exerted by particle 2, $\vec{\mathbf{F}}_{12}$, is equal and opposite to that on particle 2 exerted by particle 1, $\vec{\mathbf{F}}_{21}$; that is $\vec{\mathbf{F}}_{21} = -\vec{\mathbf{F}}_{12}$.

6–3 Gravity Near the Earth's Surface

When Eq. 6–1 is applied to the gravitational force between the Earth and an object at its surface, m_1 becomes the mass of the Earth m_E, m_2 becomes the mass of the object m, and r becomes the distance of the object from the Earth's center, which is the radius of the Earth r_E. This force of gravity due to the Earth is the weight of the object on Earth, which we have been writing as mg.

Thus,

$$mg = G\frac{mm_E}{r_E^2}.$$

We can solve this for g, the acceleration of gravity at the Earth's surface:

$$g = G\frac{m_E}{r_E^2}. \tag{6–4}$$

CAUTION
Distinguish G from g

Thus, the acceleration of gravity at the surface of the Earth, g, is determined by m_E and r_E. (Don't confuse G with g; they are very different quantities, but are related by Eq. 6–4.)

Until G was measured, the mass of the Earth was not known. But once G was measured, Eq. 6–4 could be used to calculate the Earth's mass, and Cavendish was the first to do so. Since $g = 9.80\,\text{m/s}^2$ and the radius of the Earth is $r_E = 6.38 \times 10^6\,\text{m}$, then, from Eq. 6–4, we obtain the mass of the Earth to be

$$m_E = \frac{gr_E^2}{G} = \frac{(9.80\,\text{m/s}^2)(6.38 \times 10^6\,\text{m})^2}{6.67 \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2} = 5.98 \times 10^{24}\,\text{kg}.$$

Equation 6–4 can be applied to other planets, where g, m, and r would refer to that planet.



FIGURE 6–8 Example 6–4. Mount Everest, 8850 m (29,035 ft) above sea level; in the foreground, the author with sherpas at 5500 m (18,000 ft).

EXAMPLE 6–4 **ESTIMATE** **Gravity on Everest.** Estimate the effective value of g on the top of Mt. Everest, 8850 m (29,035 ft) above sea level (Fig. 6–8). That is, what is the acceleration due to gravity of objects allowed to fall freely at this altitude? Ignore the mass of the mountain itself.

APPROACH The force of gravity (and the acceleration due to gravity g) depends on the distance from the center of the Earth, so there will be an effective value g' on top of Mt. Everest which will be smaller than g at sea level. We assume the Earth is a uniform sphere (a reasonable "estimate").

SOLUTION We use Eq. 6–4, with r_E replaced by $r = 6380\,\text{km} + 8.9\,\text{km} = 6389\,\text{km} = 6.389 \times 10^6\,\text{m}$:

$$g' = G\frac{m_E}{r^2} = \frac{(6.67 \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(5.98 \times 10^{24}\,\text{kg})}{(6.389 \times 10^6\,\text{m})^2} = 9.77\,\text{m/s}^2,$$

which is a reduction of about 3 parts in a thousand (0.3%).

TABLE 6–1
Acceleration Due to Gravity at Various Locations on Earth

Location	Elevation (m)	g (m/s^2)
New York	0	9.803
San Francisco	0	9.800
Denver	1650	9.796
Pikes Peak	4300	9.789
Sydney, Australia	0	9.798
Equator	0	9.780
North Pole (calculated)	0	9.832

Note that Eq. 6–4 does not give precise values for g at different locations because the Earth is not a perfect sphere. The Earth not only has mountains and valleys, and it bulges at the equator, but also its mass is not distributed precisely uniformly (see Table 6–1). The Earth's rotation also affects the value of g (see Example 6–5). However, for most practical purposes when an object is near the Earth's surface, we will simply use $g = 9.80\,\text{m/s}^2$ and write the weight of an object as mg.

EXERCISE A Suppose you could double the mass of a planet but kept its volume the same. How would the acceleration of gravity, g, at the surface change?

PHYSICS APPLIED
Geology—mineral and oil exploration

The value of g can vary locally on the Earth's surface because of the presence of irregularities and rocks of different densities. Such variations in g, known as "gravity anomalies," are very small—on the order of 1 part per 10^6 or 10^7 in the value of g. But they can be measured ("gravimeters" today can detect variations in g to 1 part in 10^9). Geophysicists use such measurements as part of their investigations into the structure of the Earth's crust, and in mineral and oil exploration. Mineral deposits, for example, often have a greater density than surrounding material; because of the greater mass in a given volume, g can have a slightly greater value on top of such a deposit than at its flanks. "Salt domes," under which petroleum is often found, have a lower than average density and searches for a slight reduction in the value of g in certain locales have led to the discovery of oil.

EXAMPLE 6–5 **Effect of Earth's rotation on g.** Assuming the Earth is a perfect sphere, determine how the Earth's rotation affects the value of g at the equator compared to its value at the poles.

APPROACH Figure 6–9 shows a person of mass m standing on a doctor's scale at two places on the Earth. At the North Pole there are two forces acting on the mass m (the person): the force of gravity, $\vec{\mathbf{F}}_G = m\vec{\mathbf{g}}$, and the force with which the scale pushes up on the person, $\vec{\mathbf{w}}$. We call this latter force w because it is what the scale reads as the weight of the person, and by Newton's third law it equals the force down on the scale due to the person. Since the person is not accelerating, Newton's second law tells us

$$mg - w = 0,$$

so $w = mg$. Thus the weight w that the scale registers equals mg, which is no surprise. Next, at the equator, there *is* an acceleration because points on the equator are rotating. The same magnitude of the force of gravity $F_G = mg$ acts downward (we are letting g represent the acceleration of gravity in the absence of rotation and we ignore the slight bulging of the equator). The scale pushes upward with a force w'; w' is also the force with which the person pushes down on the scale (Newton's third law) and hence is the weight registered on the scale. From Newton's second law we now have (see Fig. 6–9)

$$mg - w' = m\frac{v^2}{r_E},$$

because the person of mass m now has a centripetal acceleration due to Earth's rotation; $r_E = 6.38 \times 10^6\,\text{m}$ is the Earth's radius and v is the speed of m due to the Earth's daily rotation.

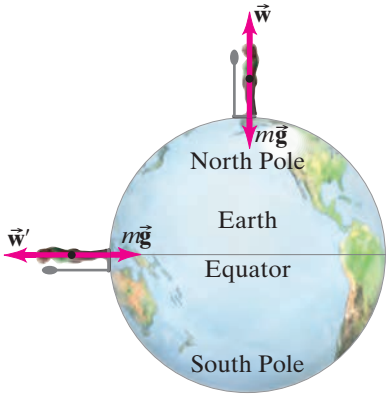


FIGURE 6–9 Example 6–5: the pull of gravity at the North Pole and at the equator, and what you weigh at those places (w, w').

SOLUTION First we determine the speed v of an object at rest on the Earth's equator, rembering that Earth makes one rotation (distance = circumference of Earth $= 2\pi r_E$) in 1 day $= (24\,\text{h})(60\,\text{min/h})(60\,\text{s/min}) = 8.64 \times 10^4\,\text{s}$:

$$v = \frac{2\pi r_E}{1\,\text{day}} = \frac{(6.283)(6.38 \times 10^6\,\text{m})}{(8.64 \times 10^4\,\text{s})}$$
$$= 4.64 \times 10^2\,\text{m/s}.$$

The effective weight is $w' = mg'$ where g' is the effective value of g, and so $g' = w'/m$. Solving the equation above for w', we have

$$w' = m\left(g - \frac{v^2}{r_E}\right),$$

so

$$g' = \frac{w'}{m} = g - \frac{v^2}{r_E}.$$

Hence

$$\Delta g = g - g' = \frac{v^2}{r_E} = \frac{(4.64 \times 10^2\,\text{m/s})^2}{(6.38 \times 10^6\,\text{m})}$$
$$= 0.0337\,\text{m/s}^2,$$

which is about $\Delta g \approx 0.003\,g$, a difference of 0.3%.

NOTE In Table 6–1 we see that the difference in g at the pole and equator is actually greater than this: $(9.832 - 9.780)\,\text{m/s}^2 = 0.052\,\text{m/s}^2$. This discrepancy is due mainly to the Earth being slightly fatter at the equator (by 21 km) than at the poles. Do you weigh less at the equator?

NOTE The calculation of the effective value of g at latitudes other than at the poles or equator is a two-dimensional problem because $\vec{\mathbf{F}}_G$ acts radially toward the Earth's center whereas the centripetal acceleration is directed perpendicular to the axis of rotation, parallel to the equator and that means that a plumb line (the effective direction of g) is not precisely vertical except at the equator and the poles.

Earth as Inertial Reference Frame

We often make the assumption that reference frames fixed on the Earth are inertial reference frames. Our calculation in Example 6–5 above shows that this assumption can result in errors no larger than about 0.3% in the use of Newton's second law, for example. We discuss the effects of Earth's rotation and reference frames in more detail in Chapter 11, including the Coriolis effect.

6–4 Satellites and "Weightlessness"

Satellite Motion

PHYSICS APPLIED
Artificial Earth satellites

Artificial satellites circling the Earth are now commonplace (Fig. 6–10). A satellite is put into orbit by accelerating it to a sufficiently high tangential speed with the use of rockets, as shown in Fig. 6–11. If the speed is too high, the spacecraft will not be confined by the Earth's gravity and will escape, never to return. If the speed is too low, it will return to Earth. Satellites are usually put into circular (or nearly circular) orbits, because such orbits require the least takeoff speed.



FIGURE 6–10 A satellite, the International Space Station, circling the Earth.

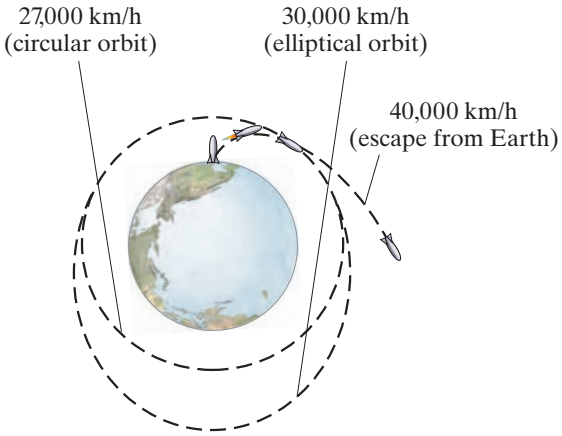


FIGURE 6–11 Artificial satellites launched at different speeds.

FIGURE 6–12 A moving satellite "falls" out of a straight-line path toward the Earth.



It is sometimes asked: "What keeps a satellite up?" The answer is: its high speed. If a satellite in orbit stopped moving, it would fall directly to Earth. But at the very high speed a satellite has, it would quickly fly out into space (Fig. 6–12) if it weren't for the gravitational force of the Earth pulling it into orbit. In fact, a satellite in orbit *is* falling (accelerating) toward Earth, but its high tangential speed keeps it from hitting Earth.

For satellites that move in a circle (at least approximately), the needed acceleration is centripetal and equals v^2/r. The force that gives a satellite this acceleration is the force of gravity exerted by the Earth, and since a satellite may be at a considerable distance from the Earth, we must use Newton's law of universal gravitation (Eq. 6–1) for the force acting on it. When we apply Newton's second law, $\Sigma F_R = ma_R$ in the radial direction, we find

$$G\frac{mm_E}{r^2} = m\frac{v^2}{r}, \qquad \textbf{(6–5)}$$

where m is the mass of the satellite. This equation relates the distance of the satellite from the Earth's center, r, to its speed, v, in a circular orbit. Note that only one force—gravity—is acting on the satellite, and that r is the sum of the Earth's radius r_E plus the satellite's height h above the Earth: $r = r_E + h$. If we solve Eq. 6–5 for v, we find $v = \sqrt{Gm_E/r}$ and we see that a satellite's speed does not depend on its own mass, m. Satellites of different mass orbiting at the same distance above Earth have the same speed; this also means they have the same *orbital period* (time to travel one revolution around the Earth).

EXAMPLE 6–6 **Geostationary satellite.** A *geostationary* satellite is one that stays above the same point on the Earth, which is possible only if it is above a point on the equator. Why? Because the center of a satellite orbit is always at the center of the Earth; so it is not possible to have a satellite orbiting above a fixed point on the Earth at any latitude other than 0°. Geostationary satellites are commonly used for TV and radio transmission, for weather forecasting, and as communication relays.† Determine (*a*) the height above the Earth's surface such a satellite must orbit, and (*b*) such a satellite's speed. (*c*) Compare to the speed of a satellite orbiting 200 km above Earth's surface.

PHYSICS APPLIED
Geostationary satellites

APPROACH To remain above the same point on Earth as the Earth rotates, the satellite must have a period of 24 hours. We can apply Newton's second law, $F = ma$, where $a = v^2/r$ if we assume the orbit is circular.

SOLUTION (*a*) The only force on the satellite is the gravitational force due to the Earth. (We can ignore the gravitational force exerted by the Sun. Why?) We apply Eq. 6–5 assuming the satellite moves in a circle:

$$G\frac{m_{\text{Sat}}\,m_{\text{E}}}{r^2} = m_{\text{Sat}}\frac{v^2}{r}.$$

This equation has two unknowns, r and v. So we need a second equation. Our satellite revolves around the Earth with the same period that the Earth rotates on its axis, namely once in 24 hours. Thus the speed of the satellite must be

$$v = \frac{2\pi r}{T},$$

where $T = 1\text{ day} = (24\text{ h})(3600\text{ s/h}) = 86{,}400\text{ s}$. We substitute this into the "satellite equation" above and obtain (after canceling m_{Sat} on both sides):

$$G\frac{m_{\text{E}}}{r^2} = \frac{(2\pi r)^2}{rT^2}.$$

After canceling an r, we can solve for r^3:

$$r^3 = \frac{Gm_{\text{E}}T^2}{4\pi^2} = \frac{(6.67\times10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(5.98\times10^{24}\,\text{kg})(86{,}400\,\text{s})^2}{4\pi^2}$$
$$= 7.54\times10^{22}\,\text{m}^3.$$

We take the cube root and find

$$r = 4.22\times10^7\,\text{m},$$

or 42,200 km from the Earth's center. We subtract the Earth's radius of 6380 km to find that a geostationary satellite must orbit about 36,000 km (about $6r_{\text{E}}$) above the Earth's surface.

(*b*) We solve for v in the satellite equation, Eq. 6–5:

$$v = \sqrt{\frac{Gm_{\text{E}}}{r}} = \sqrt{\frac{(6.67\times10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(5.98\times10^{24}\,\text{kg})}{(4.22\times10^7\,\text{m})}} = 3070\,\text{m/s},$$

or about 11,000 km/h ($\approx$ 7000 mi/h). We get the same result if we use $v = 2\pi r/T$.

(*c*) The equation in part (*b*) for v shows $v \propto \sqrt{1/r}$. So for $r = r_{\text{E}} + h = 6380\text{ km} + 200\text{ km} = 6580\text{ km}$, we get

$$v' = v\sqrt{\frac{r}{r'}} = (3070\,\text{m/s})\sqrt{\frac{(42{,}200\,\text{km})}{(6580\,\text{km})}} = 7770\,\text{m/s},$$

or about 28,000 km/h ($\approx$ 17,000 mi/h).

NOTE *Polar satellites*, which pass over the north and south poles, cover all parts of the Earth and are also used in applications like weather forecasting.

†Geostationary satellites are useful because receiving and transmitting antennas at a given place on Earth can stay fixed on such a satellite (no tracking and no switching satellites is needed).

EXERCISE B Two satellites orbit the Earth in circular orbits of the same radius. One satellite is twice as massive as the other. Which of the following statements is true about the speeds of these satellites? (*a*) The heavier satellite moves twice as fast as the lighter one. (*b*) The two satellites have the same speed. (*c*) The lighter satellite moves twice as fast as the heavier one. (*d*) The heavier satellite moves four times as fast as the lighter one.

Weightlessness

People and other objects in a satellite circling the Earth are said to experience apparent weightlessness. Let us first look at a simpler case: a falling elevator. In Fig. 6–13a, an elevator is at rest with a bag hanging from a spring scale. The scale reading indicates the downward force exerted on it by the bag. This force, exerted *on* the scale, is equal and opposite to the force exerted *by* the scale upward on the bag, and we call its magnitude w. Two forces act on the bag: the downward gravitational force and the upward force exerted by the scale equal to w. Because the bag is not accelerating ($a = 0$) when we apply $\Sigma F = ma$ to the bag in Fig. 6–13a we obtain

$$w - mg = 0,$$

where mg is the weight of the bag. Thus, $w = mg$, and since the scale indicates the force w exerted on it by the bag, it registers a force equal to the weight of the bag, as we expect.

Now let the elevator have an acceleration, a. Applying Newton's second law, $\Sigma F = ma$, to the bag as seen from an inertial reference frame (the elevator itself is not now an inertial frame) we have

$$w - mg = ma.$$

Solving for w, we have

$$w = mg + ma. \qquad [a \text{ is } + \text{ upward}]$$

We have chosen the positive direction up. Thus, if the acceleration a is up, a is positive; and the scale, which measures w, will read more than mg. We call w' the *apparent weight* of the bag, which in this case would be greater than its actual weight (mg). If the elevator accelerates downward, a will be negative and w', the apparent weight, will be less than mg. The direction of the velocity $\vec{\mathbf{v}}$ doesn't matter. Only the direction of the acceleration $\vec{\mathbf{a}}$ (and its magnitude) influences the scale reading.

Suppose the elevator's acceleration is $\frac{1}{2}g$ upward; then we find

$$w' = mg + m(\tfrac{1}{2}g) = \tfrac{3}{2}mg.$$

That is, the scale reads $1\frac{1}{2}$ times the actual weight of the bag (Fig. 6–13b). The apparent weight of the bag is $1\frac{1}{2}$ times its real weight. The same is true of the person: her apparent weight (equal to the normal force exerted on her by the elevator floor) is $1\frac{1}{2}$ times her real weight. We can say that she is experiencing $1\frac{1}{2}g$'s, just as astronauts experience so many g's at a rocket's launch.

If, instead, the elevator's acceleration is $a = -\frac{1}{2}g$ (downward), then $w' = mg - \frac{1}{2}mg = \frac{1}{2}mg$. That is, the scale reads half the actual weight. If the elevator is in *free fall* (for example, if the cables break), then $a = -g$ and $w' = mg - mg = 0$. The scale reads zero. See Fig. 6–13c. The bag appears weightless. If the person in the elevator accelerating at $-g$ let go of a box, it would not fall to the floor. True, the box would be falling with acceleration g. But so would the floor of the elevator and the person. The box would hover right in front of the person. This phenomenon is called **apparent weightlessness** because in the reference frame of the person, objects don't fall or seem to have weight—yet gravity does not disappear. Gravity is still acting on each object, whose weight is still mg. The person and other objects seem weightless only because the elevator is accelerating in free fall, and there is no contact force on the person to make her feel the weight.

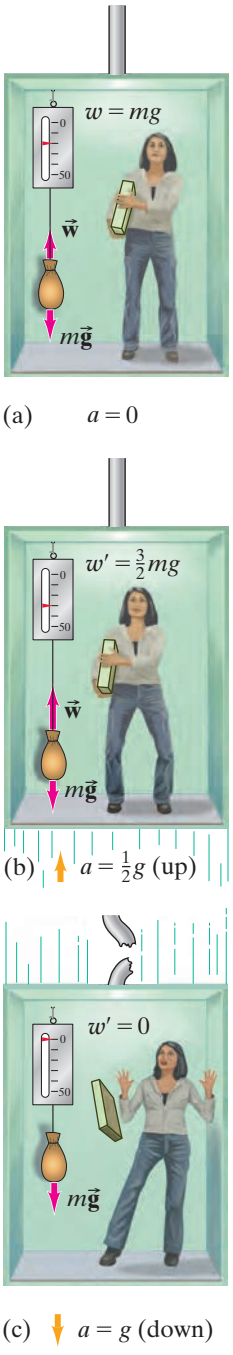


FIGURE 6–13 (a) An object in an elevator at rest exerts a force on a spring scale equal to its weight. (b) In an elevator accelerating upward at $\frac{1}{2}g$, the object's apparent weight is $1\frac{1}{2}$ times larger than its true weight. (c) In a freely falling elevator, the object experiences "weightlessness": the scale reads zero.

The "weightlessness" experienced by people in a satellite orbit close to the Earth (Fig. 6–14) is the same apparent weightlessness experienced in a freely falling elevator. It may seem strange, at first, to think of a satellite as freely falling. But a satellite is indeed falling toward the Earth, as was shown in Fig. 6–12. The force of gravity causes it to "fall" out of its natural straight-line path. The acceleration of the satellite must be the acceleration due to gravity at that point, because the only force acting on it is gravity. Thus, although the force of gravity acts on objects within the satellite, the objects experience an apparent weightlessness because they, and the satellite, are accelerating together as in free fall.

EXERCISE C Return to the Chapter-Opening Question, page 145, and answer it again now. Try to explain why you may have answered differently the first time.

Figure 6–15 shows some examples of "free fall," or apparent weightlessness, experienced by people on Earth for brief moments.

A completely different situation occurs if a spacecraft is out in space far from the Earth, the Moon, and other attracting bodies. The force of gravity due to the Earth and other celestial bodies will then be quite small because of the distances involved, and persons in such a spacecraft would experience real weightlessness.

EXERCISE D Could astronauts in a spacecraft far out in space easily play catch with a bowling ball ($m = 7\,\text{kg}$)?



FIGURE 6–14 This astronaut is outside the International Space Station. He must feel very free because he is experiencing apparent weightlessness.



FIGURE 6–15 Experiencing "weightlessness" on Earth.

6–5 Planets, Kepler's Laws, and Newton's Synthesis

Where did we first get the idea of planets? Have you ever escaped the lights of the city to gaze late at night at the multitude of stars in the night sky? It is a moving experience. Thousands of years ago, the ancients saw this sight every cloudless night, and were fascinated. They noted that the vast majority of stars, bright or dim, seemed to maintain fixed positions relative to each other. The ancients imagined these **fixed stars** as being attached to a huge inverted bowl, or sphere. This **celestial sphere** revolved around the Earth almost exactly once a day (Fig. 6–16), from east to west. Among all the stars that were visible to the naked eye (there were no telescopes until much later, about 1600), the ancients saw five stars that changed position relative to the fixed stars over weeks and months. These five wandering stars were called **planets** (Greek for wandering). Planets were thus visible at night as tiny points of light like other stars.

The ancient idea that the Sun, Moon, and planets revolve around the Earth is called the **geocentric** view (geo = Earth in Greek). It was developed into a fine theoretical system by Ptolemy in the second century B.C. Today we view the solar system as a **heliocentric** system (helios = Sun in Greek), where the Earth is just another planet, between Venus and Mars, orbiting around the Sun.

FIGURE 6–16 Time exposure showing movement of stars over a period of several hours.



Although a heliocentric view was proposed in ancient times, it was largely ignored until Renaissance Italy of the fifteenth century. The real theory change began with the heliocentric theory of Nicolaus Copernicus (1473–1543) and then was greatly advanced by the experimental observations of Galileo around 1610 using his newly developed 30× telescope. Galileo observed that the planet Jupiter has moons (like a miniature solar system) and that Venus has phases like our Moon, not explainable by Ptolemy's geocentric system. [Galileo's famous encounter with the Church had little to do with religious faith, but rather with politics, a bit of rudeness, and that the Pope at that time was also a secular ruler. Today it is generally agreed that science and faith are different approaches and are not in conflict.]

Kepler's Laws

Also about 1600, more than a half century before Newton proposed his three laws of motion and his law of universal gravitation, the German astronomer Johannes Kepler (1571–1630) had worked out a detailed description of the motion of the planets around the Sun. Kepler's work resulted in part from the many years he spent examining data collected (without a telescope) by Tycho Brahe (1546–1601) on the positions of the planets in their motion through the night sky.

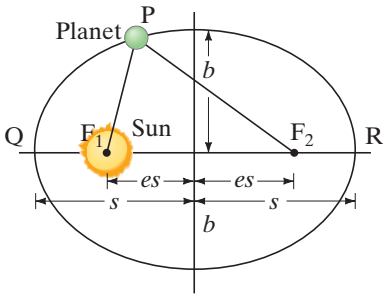


FIGURE 6–17 *Kepler's first law*. An ellipse is a closed curve such that the sum of the distances from any point P on the curve to two fixed points (called the foci, F_1 and F_2) remains constant. That is, the sum of the distances, $F_1P + F_2P$, is the same for all points on the curve. A circle is a special case of an ellipse in which the two foci coincide, at the center of the circle. The semimajor axis is s (that is, the long axis is $2s$) and the semiminor axis is b, as shown. The *eccentricity*, e, is defined as the ratio of the distance from either focus to the center divided by the semimajor axis s. Thus es is the distance from the center to either focus, as shown. For a circle, $e = 0$. The Earth and most of the other planets have nearly circular orbits. For Earth $e = 0.017$.

Among Kepler's writings were three empirical findings that we now refer to as **Kepler's laws of planetary motion**. These are summarized as follows, with additional explanation in Figs. 6–17 and 6–18.

Kepler's first law: The path of each planet about the Sun is an ellipse with the Sun at one focus (Fig. 6–17).

Kepler's second law: Each planet moves so that an imaginary line drawn from the Sun to the planet sweeps out equal areas in equal periods of time (Fig. 6–18).

Kepler's third law: The ratio of the squares of the periods of any two planets revolving about the Sun is equal to the ratio of the cubes of their semimajor axes. [The semimajor axis is half the long (major) axis of the orbit, as shown in Fig. 6–17, and represents the planet's mean distance from the Sun.[†]] That is, if T_1 and T_2 represent the periods (the time needed for one revolution about the Sun) for any two planets, and s_1 and s_2 represent their semimajor axes, then

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{s_1}{s_2}\right)^3.$$

We can rewrite Kepler's third law as

$$\frac{s_1^3}{T_1^2} = \frac{s_2^3}{T_2^2},$$

meaning that s^3/T^2 should be the same for each planet. Present-day data are given in Table 6–2; see the last column.

In Examples and Problems we usually will assume the orbits are circles, although it is not quite true in general.

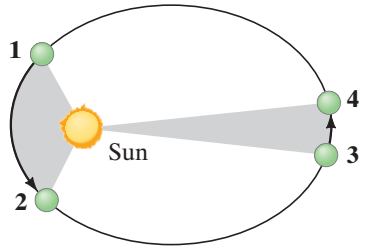


FIGURE 6–18 *Kepler's second law*. The two shaded regions have equal areas. The planet moves from point 1 to point 2 in the same time as it takes to move from point 3 to point 4. Planets move fastest when closest to the Sun. Exaggerated scale.

[†]The semimajor axis is equal to the planet's mean distance from the Sun in the sense that it equals half the sum of the planet's nearest and farthest distances from the Sun (points Q and R in Fig. 6–17). Most planetary orbits are close to circles, and for a circle the semimajor axis is the radius of the circle.

EXAMPLE 6–7 **Where is Mars?** Mars' period (its "year") was noted by Kepler to be about 687 days (Earth days), which is $(687\text{ d}/365\text{ d/yr}) = 1.88\text{ yr}$ (Earth years). Determine the mean distance of Mars from the Sun using the Earth as a reference.

APPROACH We are given the ratio of the periods of Mars and Earth. We can find the mean distance from Mars to the Sun using Kepler's third law, given the Earth–Sun distance as 1.50×10^{11} m (Table 6–2; also Table inside front cover).

SOLUTION Let the distance of Mars from the Sun be s_{MS}, and the Earth–Sun distance be $s_{ES} = 1.50 \times 10^{11}$ m. From Kepler's third law:

$$\frac{s_{MS}}{s_{ES}} = \left(\frac{T_M}{T_E}\right)^{\frac{2}{3}} = \left(\frac{1.88\text{ yr}}{1\text{ yr}}\right)^{\frac{2}{3}} = 1.52.$$

So Mars is 1.52 times the Earth's distance from the Sun, or 2.28×10^{11} m.

TABLE 6–2 Planetary Data Applied to Kepler's Third Law

Planet	Mean Distance to Sun, s (10^6 km)	Period, T (Earth yr)	s^3/T^2 $\left(10^{24}\,\frac{\text{km}^3}{\text{yr}^2}\right)$
Mercury	57.9	0.241	3.34
Venus	108.2	0.615	3.35
Earth	149.6	1.000	3.35
Mars	227.9	1.88	3.35
Jupiter	778.3	11.86	3.35
Saturn	1427	29.5	3.34
Uranus	2870	84.0	3.35
Neptune	4497	165	3.34
[Pluto]†	5900	248	3.34

†Pluto, since its discovery in 1930, was considered a ninth planet. But its small mass and the recent discovery of other objects beyond Neptune with similar masses has led to calling these smaller objects, including Pluto, "dwarf planets." We keep it in the Table to indicate its great distance, and its consistency with Kepler's third law.

Kepler's Third Law Derived, Sun's Mass, Perturbations

We can derive Kepler's second law, as we shall see in Chapter 11. Here we derive Kepler's third law using Newton's law of universal gravitation (Kepler lived before Newton, and based his laws on experiment); we do it for the special case of a circular orbit, in which case the semimajor axis is the radius r of the circle. (Most planetary orbits are close to a circle.) First, we write Newton's second law of motion, $\Sigma F = ma$. For F we use the law of universal gravitation (Eq. 6–1) for the force between the Sun and a planet of mass m_1, and for a the centripetal acceleration, v^2/r. We assume the mass of the Sun M_S is much greater than the mass of its planets, so we ignore the effects of the planets on each other. Then

$$\Sigma F = ma$$

$$G\frac{m_1 M_S}{r_1^2} = m_1\frac{v_1^2}{r_1}.$$

Here m_1 is the mass of a particular planet, r_1 its distance from the Sun, and v_1 its speed in orbit; M_S is the mass of the Sun, since it is the gravitational attraction of the Sun that keeps each planet in its orbit. The period T_1 of the planet is the time required for one complete orbit, which is a distance equal to $2\pi r_1$, the circumference of a circle. Thus $v_1 = 2\pi r_1/T_1$. We substitute this into the previous equation:

$$G\frac{m_1 M_S}{r_1^2} = m_1\frac{4\pi^2 r_1}{T_1^2}.$$

We rearrange this to get

$$\frac{T_1^2}{r_1^3} = \frac{4\pi^2}{GM_S}. \qquad \textbf{(6–6)}$$

We derived this for planet 1 (say, Mars). The same derivation would apply for a second planet (say, Saturn) orbiting the Sun,

$$\frac{T_2^2}{r_2^3} = \frac{4\pi^2}{GM_S},$$

where T_2 and r_2 are the period and orbit radius, respectively, for the second planet. Since the right sides of the two previous equations are equal, we have $T_1^2/r_1^3 = T_2^2/r_2^3$ or, rearranging,

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{r_1}{r_2}\right)^3, \qquad \textbf{(6–7)}$$

Kepler's third law

which is Kepler's third law. Equations 6–6 and 6–7 are valid also for elliptical orbits if we replace r with the semimajor axis s.

PHYSICS APPLIED
Determining the Sun's mass

EXAMPLE 6–8 **The Sun's mass determined.** Determine the mass of the Sun given the Earth's distance from the Sun as $r_{ES} = 1.5 \times 10^{11}$ m.

APPROACH Equation 6–6 relates the mass of the Sun M_S to the period and distance of any planet. We use the Earth.

SOLUTION The Earth's period is $T_E = 1\text{ yr} = (365\frac{1}{4}\text{ d})(24\text{ h/d})(3600\text{ s/h}) = 3.16 \times 10^7$ s. We solve Eq. 6–6 for M_S:

$$M_S = \frac{4\pi^2 r_{ES}^3}{G T_E^2} = \frac{4\pi^2 (1.5 \times 10^{11}\text{ m})^3}{(6.67 \times 10^{-11}\text{ N}\cdot\text{m}^2/\text{kg}^2)(3.16 \times 10^7\text{ s})^2} = 2.0 \times 10^{30}\text{ kg}.$$

EXERCISE E Suppose there were a planet in circular orbit exactly halfway between the orbits of Mars and Jupiter. What would its period be in Earth-years? Use Table 6–2.

PHYSICS APPLIED
Perturbations and discovery of planets

Accurate measurements on the orbits of the planets indicated that they did not precisely follow Kepler's laws. For example, slight deviations from perfectly elliptical orbits were observed. Newton was aware that this was to be expected because any planet would be attracted gravitationally not only by the Sun but also (to a much lesser extent) by the other planets. Such deviations, or **perturbations**, in the orbit of Saturn were a hint that helped Newton formulate the law of universal gravitation, that *all* objects attract each other gravitationally. Observation of other perturbations later led to the discovery of Neptune. Deviations in the orbit of Uranus could not all be accounted for by perturbations due to the other known planets. Careful calculation in the nineteenth century indicated that these deviations could be accounted for if another planet existed farther out in the solar system. The position of this planet was predicted from the deviations in the orbit of Uranus, and telescopes focused on that region of the sky quickly found it; the new planet was called Neptune. Similar but much smaller perturbations of Neptune's orbit led to the discovery of Pluto in 1930.

Other Centers for Kepler's Laws

CAUTION
Compare orbits of objects only around the same center

The derivation of Eq. 6–7, Kepler's third law, compared two planets revolving around the Sun. But the derivation is general enough to be applied to other systems. For example, we could apply Eq. 6–7 to compare an artificial satellite and our Moon, both revolving around Earth (then M_S would be replaced by M_E, the mass of the Earth). Or we could apply Eq. 6–7 to compare two moons revolving around Jupiter. But be careful: Kepler's third law, Eq. 6–7, applies only to objects orbiting the *same* attracting center. Do not use Eq. 6–7 to compare, say, the Moon's orbit around the Earth to the orbit of Mars around the Sun: they depend on different attracting centers.

Distant Planetary Systems

PHYSICS APPLIED
Planets around other stars

Starting in the mid-1990s, planets revolving around distant stars (Fig. 6–19) were inferred from the regular "wobble" in position of certain stars. The wobble was nicely explained as being due to the gravitational attraction of planet(s) revolving around the star. Many such "extrasolar" planets are now known.

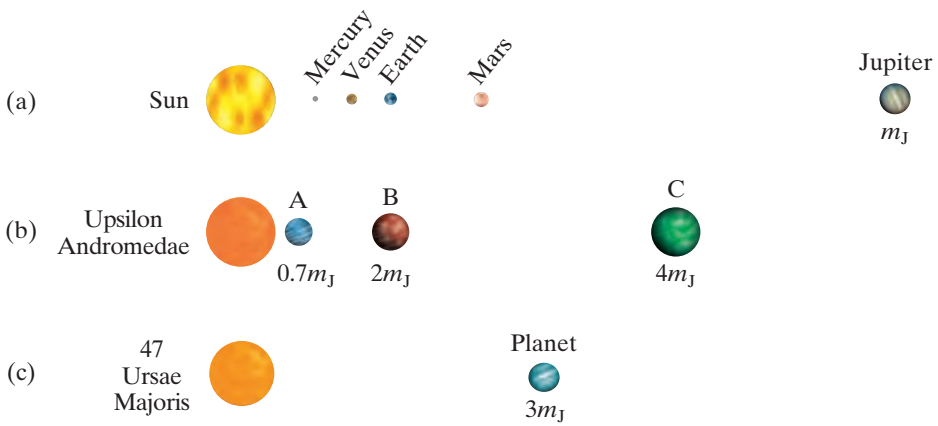


FIGURE 6–19 Our solar system (a) is compared to recently discovered planets orbiting (b) the star Upsilon Andromedae with at least three planets, and (c) the star 47 Ursae Majoris; m_J is the mass of Jupiter. (Sizes are not to scale.)

Newton's Synthesis

Kepler arrived at his laws through careful analysis of experimental data. Fifty years later, Newton showed that Kepler's laws could be derived mathematically from the law of universal gravitation and the laws of motion. Newton also showed that for any reasonable form for the gravitational force law, only one that depends on the inverse square of the distance is consistent with Kepler's laws. He thus used Kepler's laws as evidence in favor of his law of universal gravitation, Eq. 6–1.

The development by Newton of the law of universal gravitation and the three laws of motion was a major intellectual achievement. With these laws, he was able to describe the motion of objects on Earth and of the far-away planets seen in the night sky. The motions of the planets through the heavens and of objects on Earth were seen to follow the same laws (not recognized previously). For this reason, and also because Newton integrated the results of earlier scientists into his system, we sometimes speak of **Newton's synthesis**.

The laws formulated by Newton are referred to as **causal laws**. By **causality** we mean that one occurrence can cause another. When a rock strikes a window, we infer the rock *causes* the window to break. This idea of "cause and effect" relates to Newton's laws: the acceleration of an object was seen to be *caused* by the net force acting on it.

As a result of Newton's theories, the universe came to be viewed by many as a machine whose parts move in a **deterministic** way. This deterministic view of the universe had to be modified in the twentieth century (Chapter 38).

Ocean Tides

The reason for the Earth's tides, and why there are two high and two low tides (Fig. 6–20) about every 25 hours, was a subject of interest to thinkers for thousands of years. Newton gave the first clear explanation by arguing that the gravitational attraction of the Moon causes the tides. Because the gravitational force decreases with the square of the distance, the Moon pulls harder on that part of the Earth closest to it. The water in the oceans is pulled toward the Moon, resulting in a high tide on the side of the Earth facing the Moon. Simultaneously, a high tide occurs on the opposite side of the Earth (Fig. 6–21). Why this second high tide? The far side of the Earth is farther from the Moon and so experiences less gravitational attraction by the Moon than does the center of the Earth. (We might visualize this as the Earth being pulled away from the water on the far side.)[†]

High tides do not correspond precisely to the time when the Moon is directly overhead, because other forces are acting: the gravitational pull of the Sun, the rotation of the Earth, friction between the solid earth and the moving water.

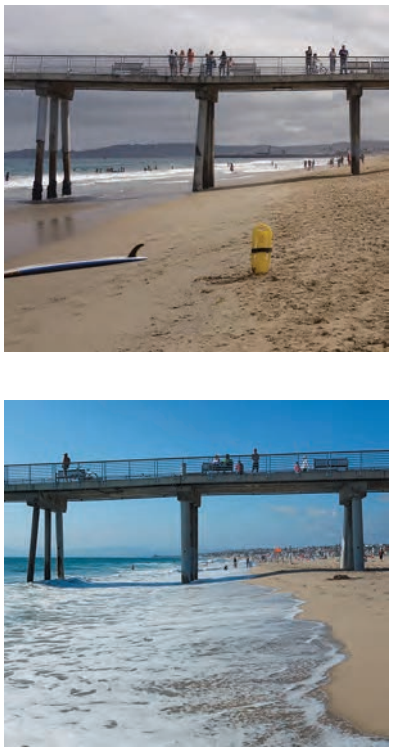


FIGURE 6–20 Low tide and high tide at the same place.

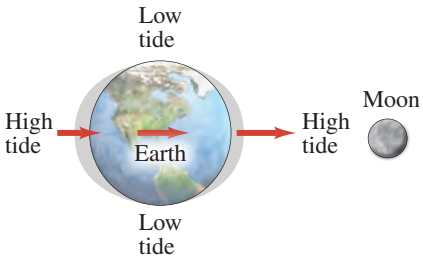


FIGURE 6–21 Tides on the Earth due to the movement of water caused by the Moon's gravitational force (red arrows). [Not to scale.]

*Does the Sun Orbit the Earth?

You may have heard someone say "People used to think the Sun goes around the Earth; but today we know the Earth goes around the Sun." The expression "today we know" is too strong. Yes, we usually like to say the Earth orbits the Sun. But we can also say the Sun goes around the Earth. Indeed we observe the Sun rising in the morning, moving until it is overhead at noon, and setting in the evening. It sure looks like the Sun is going around the Earth. And it is. You can believe your eyes. (You should!)

Frame of reference is the point. From the reference frame of the Earth, the Sun goes around the Earth. Its period is 1 day = 24 h. In the reference frame of the Sun, the Earth orbits the Sun (period = 365.256 days) as do the other planets. It is a matter of frame of reference.

But not only. We emphasize the description from the Sun's reference frame because the Sun is so massive that it exerts a strong gravitational force on all the planets, much greater than each planet exerts on other planets. So the description of the solar system with the orbits of the planets as ellipses relative to the Sun is the clearest description.

The Sun itself (and the Earth with it) revolves around the center of our Galaxy (see Fig. 44–2 or 6–39), which itself moves relative to other galaxies. Indeed, there is no one reference frame that we can consider as preferred or central.

[†]According to a historical anecdote (probably apocryphal), Napoleon was supposedly told by his admirals that he had nothing to fear from the British navy because it was run by superstitious people who believed the tides had something to do with the Moon!

*Lagrange Point

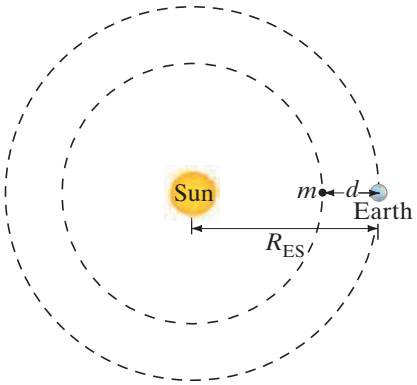


FIGURE 6–22 Finding the position of the Lagrange Point L1 for a satellite that can remain along the revolving line between the Sun and Earth, at distance d from the Earth. Thus a mass m at L1 has the same period around the Sun as the Earth has. (Not to scale.)

EXAMPLE 6–9 Lagrange Point. The mathematician Joseph-Louis Lagrange discovered five special points in the vicinity of the Earth's orbit about the Sun where a small satellite (mass m) can orbit the Sun with the same period T as Earth's ($= 1$ year). One of these "Lagrange Points," called L1, lies between the Earth (mass M_E) and Sun (mass M_S), on the line connecting them (Fig. 6–22). That is, the Earth and the satellite are always separated by a distance d. If the Earth's orbital radius is R_{ES}, then the satellite's orbital radius is $(R_{ES} - d)$. Determine d.

APPROACH We use Newton's law of universal gravitation and set it equal to the mass times the centripetal acceleration. But how could an object with a smaller orbit than Earth's have the same period as Earth? Kepler's third law clearly tells us a smaller orbit around the Sun results in a smaller period. But that law depends on only the Sun's gravitational attraction. Our satellite of mass m is pulled by both the Sun and the Earth. We assume circular orbits ($v = 2\pi r/T$).

SOLUTION Because the satellite is assumed to have negligible mass in comparison to the masses of the Earth and Sun, to an excellent approximation the Earth's orbit will be determined solely by the Sun. Applying Newton's second law to the Earth ($F = Ma$ becomes $F_G = M_E v^2/r = M_E(2\pi r/T)^2/r$) gives

$$\frac{GM_E M_S}{R_{ES}^2} = M_E\frac{v^2}{R_{ES}} = \frac{M_E}{R_{ES}}\frac{(2\pi R_{ES})^2}{T^2}$$

or

$$\frac{GM_S}{R_{ES}^2} = \frac{4\pi^2 R_{ES}}{T^2}. \qquad \textbf{(i)}$$

Next we apply Newton's second law similarly to the satellite m (which has the same period T as Earth), including the pull of *both* Sun and Earth:

$$\frac{GM_S m}{(R_{ES}-d)^2} - \frac{GM_E m}{d^2} = m\frac{v^2}{R_{ES}-d} = m\frac{(2\pi)^2(R_{ES}-d)^2}{(R_{ES}-d)T^2},$$

which, rewriting the left and right expressions, becomes

$$\frac{GM_S}{R_{ES}^2}\left(1-\frac{d}{R_{ES}}\right)^{-2} - \frac{GM_E}{d^2} = \frac{4\pi^2 R_{ES}}{T^2}\left(1-\frac{d}{R_{ES}}\right).$$

We assume $d \ll R_{ES}$ (and hope it works out) so that we can use the binomial expansion $(1+x)^n \approx 1 + nx$, if $x \ll 1$. Setting $x = d/R_{ES} \ll 1$, we have

$$\frac{GM_S}{R_{ES}^2}\left(1+2\frac{d}{R_{ES}}\right) - \frac{GM_E}{d^2} = \frac{4\pi^2 R_{ES}}{T^2}\left(1-\frac{d}{R_{ES}}\right). \qquad \textbf{(ii)}$$

Substituting GM_S/R_{ES}^2 in place of $4\pi^2 R_{ES}/T^2$ [from Eq. (i)] into Eq. (ii) we find

$$\frac{GM_S}{R_{ES}^2}\left(1+2\frac{d}{R_{ES}}\right) - \frac{GM_E}{d^2} = \frac{GM_S}{R_{ES}^2}\left(1-\frac{d}{R_{ES}}\right).$$

Simplifying, we have

$$\frac{GM_S}{R_{ES}^2}\left(3\frac{d}{R_{ES}}\right) = \frac{GM_E}{d^2}.$$

We solve for d to find

$$d = \left(\frac{M_E}{3M_S}\right)^{\frac{1}{3}} R_{ES}.$$

Substituting in values we find

$$d = 1.0\times10^{-2}R_{ES} = 1.5\times10^6\,\text{km}.$$

NOTE Since $d/R_{ES} = 10^{-2}$, we were justified in using the binomial expansion.

NOTE Placing a satellite at L1 has two advantages: the satellite's view of the Sun is never eclipsed by the Earth, and it is always close enough to Earth to transmit data easily. The L1 point of the Earth–Sun system is currently home to the Solar and Heliospheric Observatory (SOHO) satellite, Fig. 6–23.

FIGURE 6–23 Artist's rendition of the Solar and Heliospheric Observatory (SOHO) satellite in orbit.



6–6 Moon Rises an Hour Later Each Day

From the Earth's reference frame, our Moon revolves on average in 24 h, 50 min, which means the Moon rises nearly an hour later each day; and it is at its highest point in the sky about an hour later each day. When the Moon is on the direct opposite side of Earth from the Sun, the Sun's light fully illuminates the Moon and we call it a **Full moon** (Fig. 6–24a). When the Moon is on the same side of the Earth as the Sun, and nearly aligned with both, we see the Moon as a thin sliver—most or all of it is in shadow (= a **New moon**). The phases of the Moon (new, first quarter, full, third quarter) take it from one Full moon to the next Full moon in 29.53 days (= **synodic period**) on average, as seen from the Earth as reference frame (Fig. 6–24e). In the Sun's frame of reference, the Moon revolves around the Earth in 27.32 days (**sidereal period**, Fig. 6–24d). This small difference arises because, when the Moon has made one complete revolution around the Earth, the Earth itself has moved in its orbit relative to the Sun. So the Moon needs more time ($\approx$ 2 days) to be fully aligned with the Sun and Earth and be a Full moon, Fig. 6–24e. The red dot in Figs. 6–24a, b, and e represents an observer at the same location on Earth, which in (a) is when the Full moon is rising and the Sun is just setting.

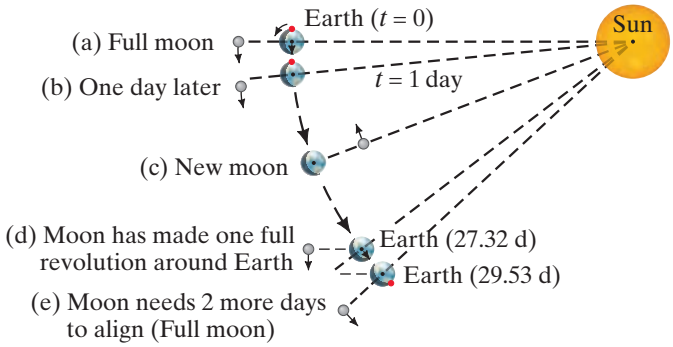


FIGURE 6–24 Looking down, in the Sun's reference frame, on the plane of Earth's orbit around the Sun (not to scale), above Earth's north pole, showing our Moon making one revolution about Earth: (a) at a Full moon (the red dot is an observer at about 6 PM who can just see the Full moon rise); (b) exactly one day later (for the red dot to see the Moon rise, the Earth must rotate another 50 min); (c) after making a "half revolution" relative to Earth, the Moon is in line with the Sun, on the Sun's side, and is a New moon; (d) after the Moon makes one complete revolution around Earth relative to the Sun (sidereal period); (e) at the next Full moon (synodic period). At (a) and (e) there could be a **lunar eclipse** (Earth's shadow falling on the Moon) but this rarely happens because the plane of Moon's orbit is inclined to the plane of Earth's orbit, so the Moon is usually above or below the Earth's orbital plane. At (c) there could be a **solar eclipse**, also rare.

6–7 Types of Forces in Nature

We have already discussed that Newton's law of universal gravitation, Eq. 6–1, describes how a particular type of force—gravity—depends on the masses of the objects involved and the distance between them. Newton's second law, $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$, on the other hand, tells how an object will accelerate due to *any* type of force. But what are the types of forces that occur in nature besides gravity?

In the twentieth century, physicists came to recognize four fundamental forces in nature: (1) the gravitational force; (2) the electromagnetic force (we shall see later that electric and magnetic forces are intimately related); (3) the strong nuclear force (which holds protons and neutrons together to form atomic nuclei); and (4) the weak nuclear force (involved in radioactivity). In this Chapter, we discussed the gravitational force in detail. The electromagnetic force will be discussed in Chapters 21 to 31. The strong and weak nuclear forces, which are discussed in Chapters 41 to 43, operate at the level of the atomic nucleus and are much less obvious in our daily lives.

Physicists have been working on theories that would unify these four forces—that is, to consider some or all of these forces as different manifestations of the same basic force. The electromagnetic and weak nuclear forces have been theoretically united to form *electroweak* theory, in which the electromagnetic and weak forces are seen as two aspects of a single *electroweak force*. It is part of the "Standard Model" of fundamental particle interactions which includes the prediction of the Higgs boson, for which strong experimental evidence was announced in 2012 (see Chapter 43). Further attempts to unify the forces, such as in *grand unified theories* (GUT), are hot research topics today.

But where do everyday forces fit? Ordinary forces, other than gravity, such as pushes, pulls, and other *contact forces* like the normal force and friction, are today considered to be due to the electromagnetic force acting at the atomic level. For example, the force your fingers exert on a pencil is the result of electrical repulsion between the outer electrons of the atoms of your finger and those of the pencil.

*6–8 Gravitational Field

Most of the forces we meet in everyday life are contact forces: you push or pull on a lawn mower, a tennis racket exerts a force on a tennis ball when they make contact, or a ball exerts a force on a window when they make contact. But the gravitational force acts over a distance: there is a force even when the two objects are not in contact. The Earth, for example, exerts a force on a falling apple. It also exerts a force on the Moon, 384,000 km away. And the Sun exerts a gravitational force on the Earth. The idea of a force *acting at a distance* was a difficult one for early thinkers. Newton himself felt uneasy with this concept when he published his law of universal gravitation.

CAUTION
Forces acting over a distance

Another point of view that helps with these conceptual difficulties is the concept of the **field**, developed in the nineteenth century by Michael Faraday (1791–1867) to aid understanding of electric and magnetic forces which also act over a distance. Only later was it applied to gravity. According to the field concept, a **gravitational field** surrounds every object that has mass, and this field permeates all of space. A second object at a particular location near the first object experiences a force because of the gravitational field that exists there. The *field* exerts the pull.

To be quantitative, we can define the **gravitational field** as the gravitational force per unit mass at any point in space. If we want to measure the gravitational field at any point, we place a small "test" mass m at that point and measure the force F exerted on it (making sure only gravitational forces are acting). Then the gravitational field at that point is defined as

$$g = \frac{F}{m}. \qquad \text{[gravitational field]} \quad \textbf{(6–8)}$$

The units of g are N/kg.

From Eq. 6–8 we see that the gravitational field an object experiences has magnitude equal to the acceleration due to gravity at that point. (When we speak of acceleration, however, we use units m/s^2, which is equivalent to N/kg, since $1\,N = 1\,kg \cdot m/s^2$.)

If the gravitational field is due to a single spherically symmetric (or small) object of mass M, such as when m is near the Earth's surface, then the gravitational field at a distance r from M has magnitude

$$g = \frac{1}{m} G\frac{mM}{r^2} = G\frac{M}{r^2}.$$

We see that the gravitational field g does not depend on the test mass m.

In interplanetary space, the gravitational field at any point is the (vector) sum of terms due to the Earth, Sun, Moon, and other bodies that contribute. The gravitational field does not depend on the value of our test mass, m, but only on the masses (and locations) of the bodies that create the field there.

The field concept will be discussed in more detail relative to other forces in later Chapters.

*6–9 Principle of Equivalence; Curvature of Space; Black Holes

We have dealt with two aspects of mass. In Chapter 4, we defined mass as a measure of the inertia of a body. Newton's second law relates the force acting on a body to its acceleration and its **inertial mass**, as we call it. We might say that inertial mass represents a resistance to any force.

In this Chapter we have dealt with mass as a property related to the gravitational force—that is, mass as a quantity that determines the strength of the gravitational force between two bodies. This we call the **gravitational mass**.

It is not obvious that the inertial mass of a body should be equal to its gravitational mass. The force of gravity might have depended on a different property of a body, just as the electrical force depends on a property called electric charge. Newton's and Cavendish's experiments indicated that the two types of mass are equal for any body, and modern experiments confirm it to a precision of about 1 part in 10^{12}.

Albert Einstein (1879–1955) called this equivalence between gravitational and inertial masses the **principle of equivalence**, and he used it as a foundation for his *general theory of relativity* (*c*. 1916). The principle of equivalence can be stated in another way:

> *there is no experiment observers can perform to distinguish if an acceleration arises because of a gravitational force or because their reference frame is accelerating*.

If you were far out in space and an apple fell to the floor of your spacecraft, you might assume a gravitational force was acting on the apple. But it would also be possible that the apple fell because your spacecraft accelerated upward (relative to an inertial reference frame). The effects would be indistinguishable, according to the principle of equivalence, because the apple's inertial and gravitational masses—that determine how a body "reacts" to outside influences—are indistinguishable.

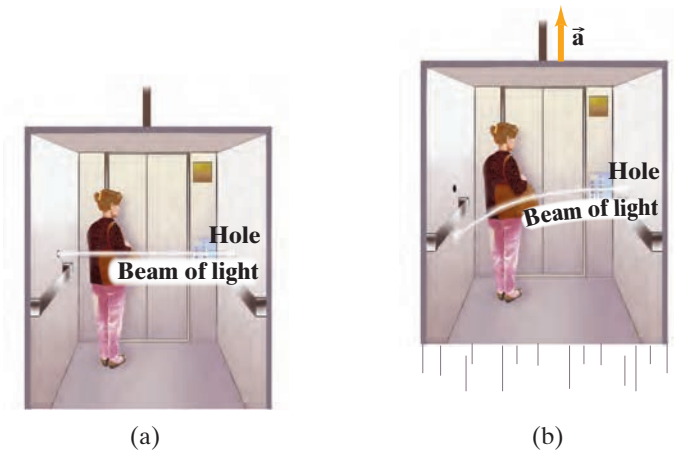


FIGURE 6–25 (a) Light beam goes straight across an elevator that is not accelerating. (b) The light beam bends (exaggerated) in an elevator accelerating in an upward direction.

The principle of equivalence can be used to show that light ought to be deflected due to the gravitational force of a massive object. Let us consider a thought experiment in an elevator in free space where virtually no gravity acts. If a light beam enters a hole in the side of the elevator, the beam travels straight across the elevator and makes a spot on the opposite side if the elevator is at rest (Fig. 6–25a). If the elevator is accelerating upward as in Fig. 6–25b, the light beam is observed to curve downward. Why? Because during the time the light travels from one side of the elevator to the other, the elevator has moved upward at ever-increasing speed.

According to the equivalence principle, an upwardly accelerating reference frame is equivalent to a downward gravitational field. Hence, we can picture the curved light path in Fig. 6–25b as being the effect of a gravitational field.

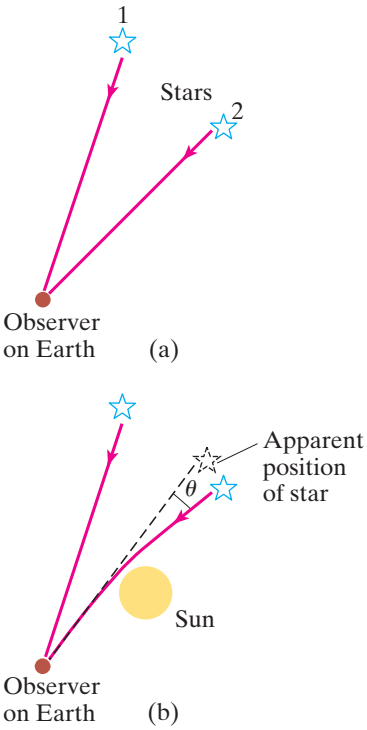


FIGURE 6–26 (a) Two stars in the sky, observed from Earth. (b) If the light from one of these stars passes very near the Sun, whose gravity bends the rays, the star will appear higher than it actually is (follow the ray backwards). [Not to scale.]

Thus we expect gravity to exert a force on a beam of light and to bend it out of a straight-line path!

Einstein's general theory of relativity predicts that light should be affected by gravity. It was calculated that light from a distant star would be deflected by 1.75″ of arc (tiny but detectable) as it passed near the Sun, as shown in Fig. 6–26. Such a deflection was measured and confirmed in 1919 during an eclipse of the Sun. (The eclipse reduced the brightness of the Sun so that the stars in line with its edge at that moment would be visible.)

That a light beam can follow a curved path suggests that **space itself is curved** and that it is gravitational mass that causes the curvature. The curvature is greatest near very massive objects. To visualize this curvature of space, we might think of space as being like a thin rubber sheet; if a heavy weight is hung from it, it curves as shown in Fig. 6–27. The weight corresponds to a huge mass that causes space (space itself!) to curve.

The extreme curvature of space-time shown in Fig. 6–27 could be produced by a **black hole**, a star that becomes so dense and massive that gravity would be so strong that even light could not escape it. Light would be pulled back in by the force of gravity. Since no light could escape from such a massive star, we could not see it—it would be black. An object might pass by it and be deflected by its gravitational field, but if the object came too close it would be swallowed up, never to escape. Hence the name black holes. Experimentally there is good evidence for their existence. One likely possibility is a giant black hole at the center of our Galaxy and probably at the center of other galaxies.

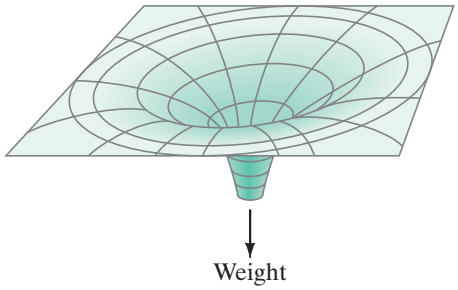


FIGURE 6–27 Rubber-sheet analogy for space (technically space-time) curved by matter.

Summary

Newton's **law of universal gravitation** states that every particle in the universe attracts every other particle with a force proportional to the product of their masses and inversely proportional to the square of the distance between them:

$$F_G = G\frac{m_1 m_2}{r^2}. \qquad \textbf{(6–1)}$$

The direction of this force is along the line joining the two particles, and is always attractive. It is this gravitational force that keeps the Moon revolving around the Earth and the planets revolving around the Sun.

The total gravitational force on any object is the vector sum of the forces exerted by all other objects; frequently the effects of all but one or two objects can be ignored.

Satellites revolving around the Earth are acted on by gravity, but "stay up" because of their high tangential speed.

Newton's three laws of motion, plus his law of universal gravitation, constituted a wide-ranging theory of the universe. With them, motion of objects on Earth and in the heavens could be accurately described. And they provided a theoretical base for **Kepler's laws** of planetary motion.

High and low **tides** are due mainly to the gravitational force exerted by the Moon.

The Moon rises about an hour later each day.

The four fundamental forces in nature are (1) the gravitational force, (2) electromagnetic force, (3) strong nuclear force, and (4) weak nuclear force. The first two fundamental forces are responsible for nearly all "everyday" forces.

[*According to the **field** concept, a **gravitational field** surrounds every object that has mass, and it permeates all of space. The gravitational field at any point in space is the vector sum of the fields due to all massive objects and can be defined as

$$g = \frac{F}{m} \qquad \textbf{(6–8)}$$

where F is the force acting on a small "test" mass m placed at that point.]

[*Inertial mass and gravitational mass are the same for any body, an experimental fact known as the **equivalence principle**. Even a light beam is affected by gravity, which suggests that space itself is curved. **Black holes** are such massive "stars" that not even light can escape.]

Questions

1. Does an apple exert a gravitational force on the Earth? If so, how large a force? Consider an apple (*a*) attached to a tree and (*b*) falling.
2. The Sun's gravitational pull on the Earth is much larger than the Moon's. Yet the Moon's is mainly responsible for the tides. Explain. [*Hint*: Consider the difference in gravitational pull from one side of the Earth to the other.]
3. Will an object weigh more at the equator or at the poles? What two effects are at work? Do they oppose each other?
4. Why is more fuel required for a spacecraft to travel from the Earth to the Moon than for it to return from the Moon to the Earth?
5. The gravitational force on the Moon due to the Earth is only about half the force on the Moon due to the Sun (see Example 6–3). Why isn't the Moon pulled away from the Earth?
6. How did the scientists of Newton's era determine the distance from the Earth to the Moon, despite not knowing about spaceflight or the speed of light? [*Hint*: Think about why two eyes are useful for depth perception.]
7. If it were possible to drill a hole all the way through the Earth along a diameter, then it would be possible to drop a ball through the hole. When the ball was right at the center of the Earth, what would be the total gravitational force exerted on it by the Earth?
8. Why is it not possible to put a satellite in geostationary orbit above the North Pole?
9. Which pulls harder gravitationally, the Earth on the Moon, or the Moon on the Earth? Which accelerates more?
10. Would it require less speed to launch a satellite (*a*) toward the east or (*b*) toward the west? Consider the Earth's rotation direction, and explain your choice.
11. An antenna loosens and becomes detached from a satellite in a circular orbit around the Earth. Describe the antenna's subsequent motion. If it will land on the Earth, describe where; if not, describe how it could be made to land on the Earth.
12. Describe how careful measurements of the variation in g in the vicinity of an ore deposit might be used to estimate the amount of ore present.
13. The Sun is below us at midnight, nearly in line with the Earth's center. Are we then heavier at midnight, due to the Sun's gravitational force on us, than we are at noon? Explain.
14. When will your apparent weight be the greatest, as measured by a scale in a moving elevator: when the elevator (*a*) accelerates downward, (*b*) accelerates upward, (*c*) is in free fall, or (*d*) moves upward at constant speed? In which case would your apparent weight be the least? When would it be the same as when you are on the ground? Explain.
15. If the Earth's mass were double what it actually is, in what ways would the Moon's orbit be different?
16. The source of the Mississippi River is closer to the center of the Earth than is its outlet in Louisiana (because the Earth is fatter at the equator than at the poles). Explain how the Mississippi can flow "uphill."
17. People sometimes ask, "What keeps a satellite up in its orbit around the Earth?" How would you respond?
18. Explain how a runner experiences "free fall" or "apparent weightlessness" between steps.
19. If you were in a satellite orbiting the Earth, how might you cope with walking, drinking, or putting a pair of scissors on a table?
20. Is the centripetal acceleration of Mars in its orbit around the Sun larger or smaller than the centripetal acceleration of the Earth? Explain.
21. The mass of the "planet" Pluto was not known until it was discovered to have a moon. Explain how this enabled an estimate of Pluto's mass.
22. The Earth moves faster in its orbit around the Sun in January than in July. Is the Earth closer to the Sun in January, or in July? Explain. [*Note*: This is not much of a factor in producing the *seasons*—the main factor is the tilt of the Earth's axis relative to the plane of its orbit.]
23. Kepler's laws tell us that a planet moves faster when it is closer to the Sun than when it is farther from the Sun. What causes this change in speed of the planet?

*24. Does your body directly sense a gravitational field? (Compare to what you would feel in free fall.)

*25. Discuss the conceptual differences between g as acceleration due to gravity and g as gravitational field.

MisConceptual Questions

1. The Moon does not crash into the Earth because
   (*a*) the net force on it is zero.
   (*b*) it is beyond the main pull of the Earth's gravity.
   (*c*) it is being pulled by the Sun as well as by the Earth.
   (*d*) it is freely falling but it has a high tangential velocity.
2. In the International Space Station which orbits Earth, astronauts experience apparent weightlessness because
   (*a*) the station is so far away from the center of the Earth.
   (*b*) the station is kept in orbit by a centrifugal force that counteracts the Earth's gravity.
   (*c*) the astronauts and the station are in free fall towards the center of the Earth.
   (*d*) there is no gravity in space.
   (*e*) the station's high speed nullifies the effects of gravity.
3. Which pulls harder gravitationally, the Earth on the Moon, or the Moon on the Earth? Which accelerates more?
   (*a*) The Earth on the Moon; the Earth.
   (*b*) The Earth on the Moon; the Moon.
   (*c*) The Moon on the Earth; the Earth.
   (*d*) The Moon on the Earth; the Moon.
   (*e*) Both the same; the Earth.
   (*f*) Both the same; the Moon.
4. Two satellites orbit the Earth in circular orbits of the same radius. One satellite is twice as massive as the other. Which statement is true about the speeds of these satellites?
   (*a*) The heavier satellite moves twice as fast as the lighter one.
   (*b*) The two satellites have the same speed.
   (*c*) The lighter satellite moves twice as fast as the heavier one.
   (*d*) The ratio of their speeds depends on the details of how they were launched.

5. A space vehicle in orbit around the Earth carries its payload, a satellite, with its mechanical arm. Suddenly, the arm malfunctions and releases the payload. What will happen to the payload?
(*a*) It will fall straight down and hit the Earth.
(*b*) It will follow a curved path and eventually hit the Earth.
(*c*) It will remain in the same orbit with the space vehicle.
(*d*) It will drift out into deep space.

6. As you travel away from Earth's surface,
(*a*) your weight decreases and your mass remains the same.
(*b*) both your weight and mass decrease.
(*c*) both your weight and mass remain the same.
(*d*) your weight is always greater than your mass.

7. A person who weighs 450 N stands on a scale in an elevator. The scale will read zero if
(*a*) the elevator is accelerating up at 9.8 m/s^2.
(*b*) the elevator is moving at a constant speed of 9.8 m/s.
(*c*) the elevator is accelerating down at 9.8 m/s^2.
(*d*) The scale will never read zero.

8. What force keeps the Moon up in orbit around the Earth?
(*a*) No force. The Moon is constantly falling as it orbits the Earth at high speed.
(*b*) Centrifugal force.
(*c*) Centripetal force.
(*d*) The combined effects of centrifugal force and gravity.
(*e*) The Earth's gravity combined with its magnetic field.

9. The Full moon appears full because
(*a*) the Moon is at the best angle to reflect the city lights from Earth.
(*b*) the round part of the Moon faces Earth at that time of month.
(*c*) the Earth is between the Moon and the Sun and we see reflected sunlight.
(*d*) the Moon is closest to the Earth at that time of month.

10. According to Kepler's third law, if we know the ratio s^3/T^2 for one of the moons that orbit Jupiter, this ratio should be the same for
(*a*) the Moon orbiting the Earth.
(*b*) any moon orbiting Jupiter.
(*c*) Jupiter's orbit about the Sun.
(*d*) all of the above.
(*e*) none of the above.

*11. Your weight as you stand on the surface of the Earth is due to
(*a*) the Earth's rotation (if it stopped, you would fly off into space).
(*b*) the Earth's magnetic field.
(*c*) the Earth's gravitational field.
(*d*) the contact force between you and the Earth.

*12. Earth's gravitational field extends
(*a*) a little past the Moon.
(*b*) only as far as the Sun.
(*c*) to the edge of the solar system.
(*d*) beyond the edge of the solar system.

Problems

6–1 to 6–3 Law of Universal Gravitation

1. (I) Calculate the force of Earth's gravity on a spacecraft 2.00 Earth radii above the Earth's surface if its mass is 1650 kg.
2. (I) At the surface of a certain planet, the gravitational acceleration g has a magnitude of 12.0 m/s^2. A 28.0-kg brass ball is transported to this planet. What is (*a*) the mass of the brass ball on the Earth and on the planet, and (*b*) the weight of the brass ball on the Earth and on the planet?
3. (I) Calculate the acceleration due to gravity on the Moon, which has radius 1.74×10^6 m and mass 7.35×10^{22} kg.
4. (I) A hypothetical planet has a radius 2.5 times that of Earth, but has the same mass. What is the acceleration due to gravity near its surface?
5. (I) A hypothetical planet has a mass 2.80 times that of Earth, but has the same radius. What is g near its surface?
6. (II) If you doubled the mass and tripled the radius of a planet, by what factor would g at its surface change?
7. (II) At what distance from the Earth's center will a spacecraft traveling directly from the Earth to the Moon experience zero net force because the Earth and Moon pull in opposite directions with equal force?
8. (II) Two objects attract each other gravitationally with a force of 6.5×10^{-9} N when they are 0.25 m apart. Their total mass is 6.50 kg. Find their individual masses.
9. (II) An object 1600 m above the Earth's surface has lost what percent of its weight (compared to on Earth)?
10. (II) Four masses are arranged as shown in Fig. 6–28. Determine the x and y components of the gravitational force on the mass at the origin (m). Write the force in vector notation $(\hat{\mathbf{i}}, \hat{\mathbf{j}})$.

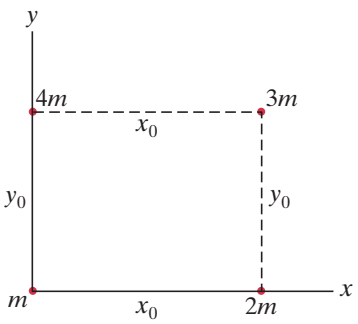


FIGURE 6–28
Problem 10.

11. (II) (*a*) Do you weigh more at the equator than in New York, or less? (*b*) By what percentage? See Table 6–1.
12. (II) Suppose the maximum distance you can throw a particular ball on Earth is 45.0 m. When you travel to another planet, the maximum distance for the same ball is 36.5 m. If the planet is the same size as Earth, what is its mass?
13. (II) Suppose the mass of the Earth were doubled, but it kept the same density (mass/volume) and spherical shape. How would the weight of objects at the Earth's surface change?
14. (II) Determine the mass of the Sun using the known value for the orbital period of the Earth and its distance from the Sun. [*Hint*: The force on the Earth due to the Sun is related to the centripetal acceleration of the Earth.] Compare your answer to that obtained using Kepler's laws, Example 6–8.

15. (II) Estimate the acceleration due to gravity at the surface of Europa (one of the moons of Jupiter) given that its mass is 4.9×10^{22} kg and making the assumption that its mass per unit volume is the same as Earth's.

16. (II) Given that the acceleration of gravity at the surface of Mars is 0.38 of what it is on Earth, and that Mars' radius is 3400 km, determine the mass of Mars.

17. (II) You are explaining to friends why an astronaut feels weightless orbiting the Earth in a space station, and they respond that they thought gravity was just a lot weaker up there. Convince them that it isn't so by calculating how much weaker (in %) gravity is 380 km above the Earth's surface.

18. (II) Every few hundred years most of the planets line up on the same side of the Sun. Calculate the total force on the Earth due to Venus, Jupiter, and Saturn, assuming all four planets are in a line, Fig. 6–29. The masses are $m_V = 0.815\, m_E$, $m_J = 318\, m_E$, $m_{Sat} = 95.1\, m_E$, and the mean distances of the four planets from the Sun are 108, 150, 778, and 1430 million km. What fraction of the Sun's force on the Earth is this?

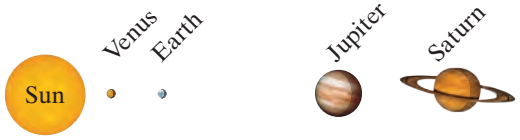


FIGURE 6–29 Problem 18 (not to scale).

19. (II) Three meteorites in outer space lie along a straight line. At a certain instant in time, the meteorite on the right is 53 m from the center meteorite and the lefthand meteorite is 21 m from the center meteorite. If the mass of the righthand meteorite is 1300 kg and the net gravitational force on the center meteorite is zero, what is the mass of the lefthand meteorite?

20. (II) Determine the distance from the Earth's center to a point outside the Earth where the gravitational acceleration due to the Earth is $\frac{1}{10}$ of its value at the Earth's surface.

21. (II) A certain neutron star has four times the mass of our Sun packed into a sphere about 10 km in radius. Estimate the acceleration due to gravity at the surface of this monster.

22. (II) A certain *white dwarf* star was once an average star like our Sun. But now it is in the last stage of its evolution and is the size of our Moon but has the mass of our Sun. (*a*) Estimate the acceleration due to gravity on the surface of this star. (*b*) How much would a 65-kg person weigh on this star? Give as a percentage of the person's weight on Earth. (*c*) What would be the speed of a baseball dropped from a height of 1.0 m when it hit the surface?

23. (II) Two identical point masses, each of mass M, always remain separated by a distance of $2R$. A third mass m is then placed a distance x along the perpendicular bisector of the original two masses, as shown in Fig. 6–30. Show that the gravitational force on the third mass is directed inward along the perpendicular bisector and has a magnitude of

$$F = \frac{2GMmx}{(x^2 + R^2)^{\frac{3}{2}}}.$$

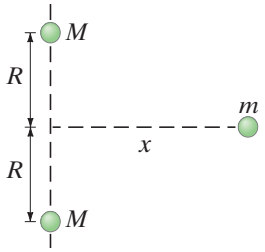


FIGURE 6–30
Problem 23.

24. (III) A mass M is ring shaped with radius r. A small mass m is placed at a distance x along the ring's axis as shown in Fig. 6–31. Show that the gravitational force on the mass m due to the ring is directed inward along the axis and has magnitude

$$F = \frac{GMmx}{(x^2 + r^2)^{\frac{3}{2}}}.$$

[*Hint*: Think of the ring as made up of many small point masses dM; sum over the forces due to each dM, and use *symmetry*.]

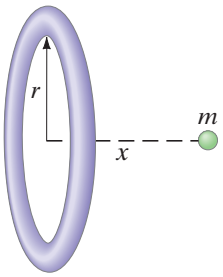


FIGURE 6–31
Problem 24.

25. (III) (*a*) Use the binomial expansion

$$(1 \pm x)^n = 1 \pm nx + \frac{n(n-1)}{2}x^2 \pm \cdots$$

to show that the value of g is altered by approximately

$$\Delta g \approx -2g\frac{\Delta r}{r_E}$$

at a height Δr above the Earth's surface, where r_E is the radius of the Earth, as long as $\Delta r \ll r_E$. (*b*) What is the meaning of the minus sign in this relation? (*c*) Use this result to compute the effective value of g at 125 km above the Earth's surface. Compare to a direct use of Eq. 6–1.

26. (III) The center of a 1.00-km-diameter spherical pocket of oil is 1.00 km beneath the Earth's surface. Estimate by what percentage the value of g would differ, directly above the pocket of oil, from the expected value of g for a uniform Earth. Assume the density of oil is 8.0×10^2 kg/m^3.

27. (III) Determine the magnitude and direction of the effective value of $\vec{\mathbf{g}}$ at a latitude of 45° on the Earth. Assume the Earth is a rotating sphere.

28. (III) It can be shown (Appendix E) that for a uniform sphere the force of gravity at a point inside the sphere depends only on the mass closer to the center than that point. The net force of gravity due to points outside the radius of the point cancels. How far would you have to drill into the Earth to reach a point where your weight is reduced by 5.0%? Approximate the Earth as a uniform sphere.

6–4 Satellites and Weightlessness

29. (I) A rocket releases a satellite into a circular orbit 780 km above the Earth. How fast must the rocket be moving (relative to Earth's center) when the release occurs?

30. (I) Calculate the speed of a satellite moving in a stable circular orbit around the Earth at a height of 5400 km.

31. (II) What will a spring scale read for the weight of a 58.0-kg woman in an elevator that moves (*a*) upward with constant speed 4.4 m/s, (*b*) downward with constant speed 4.4 m/s, (*c*) with an upward acceleration 0.18 g, (*d*) with a downward acceleration 0.18 g, and (*e*) in free fall?

32. (II) You know your mass is 62 kg, but when you stand on a bathroom scale in an elevator, it says your mass is 72 kg. What is the acceleration of the elevator, and in which direction?

33. (II) Astronomers using the Hubble Space Telescope deduced the presence of an extremely massive core in the distant *galaxy* M87, so dense that it could be a black hole (from which no light escapes). They did this by measuring the speed of gas clouds orbiting the core to be 780 km/s at a distance of 60 light-years $(= 5.7 \times 10^{17}\ \text{m})$ from the core. Deduce the mass of the core, and compare it to the mass of our Sun.

34. (II) An 18.0-kg monkey hangs from a cord suspended from the ceiling of an elevator. The cord can withstand a tension of 215 N and breaks as the elevator accelerates. What was the elevator's minimum acceleration (magnitude and direction)?

35. (II) Two satellites orbit Earth at altitudes of 8500 km and 17,000 km above the Earth's surface. (*a*) Which satellite is faster, and by what factor? (*b*) Which has the shorter period, and by what factor?

36. (II) Calculate the period of a satellite orbiting the Moon, 95 km above the Moon's surface. The Moon's radius is 1740 km.

37. (II) In a remote region of space, far from any other objects, a small ball is in a 30.0-cm-radius circular orbit around a 7.0-kg bowling ball. How long does it take the small ball to complete one orbit?

38. (II) Determine the time it takes for a satellite to orbit the Earth in a circular **near-Earth orbit**. A "near-Earth" orbit is at a height above the surface of the Earth that is very small compared to the radius of the Earth. [*Hint*: You may take the acceleration due to gravity as essentially the same as that on the surface.] Does your result depend on the mass of the satellite?

39. (II) What is the apparent weight of a 75-kg astronaut 2800 km from the center of the Moon in a space vehicle (*a*) moving at constant speed in a circular orbit, and (*b*) accelerating toward the Moon at 1.8 m/s^2? State "direction" in each case.

40. (II) Suppose a satellite is in a circular orbit not far above Earth's surface, and circles the Earth about once every 92 minutes. Find the centripetal acceleration of the satellite in its orbit. Express your answer in terms of g, the gravitational acceleration at the Earth's surface.

41. (II) At what rate must a cylindrical spaceship rotate (Fig. 6–32) if occupants are to experience simulated gravity of 0.70 g? Assume the spaceship's diameter is 28 m, and give your answer as the time needed for one revolution.

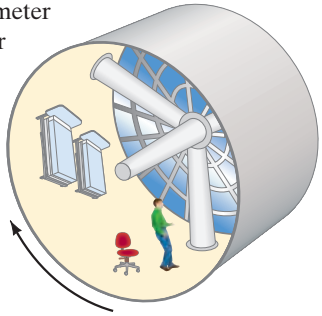


FIGURE 6–32
Problem 41.

42. (III) (*a*) Show that if a satellite orbits very near the surface of a planet with period T, the density (= mass per unit volume) of the planet is $\rho = m/V = 3\pi/GT^2$. (*b*) Estimate the density of the Earth, given that a satellite near the surface orbits with a period of 85 min. Approximate the Earth as a uniform sphere.

43. (III) Three small objects of identical mass M form the vertices of an equilateral triangle of side ℓ and rotate in circular orbits about the center of the triangle. They are held in place by their mutual gravitation. What is the speed of each?

44. (III) An inclined plane, fixed to the inside of an elevator, makes a 38° angle with the floor. A mass m slides on the plane without friction. What is its acceleration relative to the plane if the elevator (*a*) accelerates upward at 0.50 g, (*b*) accelerates downward at 0.50 g, (*c*) falls freely, and (*d*) moves upward at constant speed?

6–5 Kepler's Laws

45. (I) Neptune is an average distance of 4.5×10^9 km from the Sun. Estimate the length of the Neptunian year using the fact that the Earth is 1.50×10^8 km from the Sun on average.

46. (I) The *asteroid* Icarus, though only a few hundred meters across, orbits the Sun like the planets. Its period is 410 d. What is its mean distance from the Sun?

47. (I) Use Kepler's laws and the period of the Moon (27.4 d) to determine the period of an artificial satellite orbiting very near the Earth's surface.

48. (I) Determine the mass of the Earth from the known period and distance of the Moon.

49. (II) Table 6–3 gives the mean distance, period, and mass for the four largest moons of Jupiter (those discovered by Galileo in 1609). Determine the mass of Jupiter: (*a*) using the data for Io; (*b*) using data for each of the other three moons. Are the results consistent?

TABLE 6–3 Principal Moons of Jupiter
(Problems 49, 50, and 56)

Moon	Mass (kg)	Period (Earth days)	Mean distance from Jupiter (km)
Io	8.9×10^{22}	1.77	422×10^3
Europa	4.9×10^{22}	3.55	671×10^3
Ganymede	15×10^{22}	7.16	1070×10^3
Callisto	11×10^{22}	16.7	1883×10^3

50. (II) Determine the mean distance from Jupiter for each of Jupiter's principal moons, using Kepler's third law. Use the mean distance of Io and the periods given in Table 6–3. Compare your results to the values in Table 6–3.

51. (II) A science fiction writer imagines a planet that has three times the mass of the Earth orbiting a star that has twice the mass of our Sun. The writer wants the planet's "year" to be 10% longer than Earth's orbital period. At what mean distance should she place the planet from its star?

52. (II) Planet A and planet B are in circular orbits around a distant star. Planet A is 5.0 times farther from the star than is planet B. What is the ratio of their speeds v_A/v_B?

53. (II) *Halley's comet* orbits the Sun roughly once every 76 years. It comes very close to the surface of the Sun on its closest approach (Fig. 6–33). Estimate the farthest distance of the comet from the Sun. Is it still "in" the solar system? What planet's orbit is nearest when it is at that farthest distance?

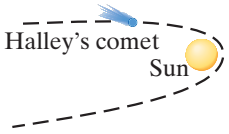


FIGURE 6–33
Problem 53.

54. (II) The **asteroid belt** between Mars and Jupiter consists of many fragments (which some space scientists think came from a planet that once orbited the Sun but was destroyed). (*a*) If the mean orbital radius of the asteroid belt (where the planet would have been) is about three times farther from the Sun than the Earth is, how long would it have taken this hypothetical planet to orbit the Sun? (*b*) Can we use these data to deduce the mass of this planet?

55. (II) (*a*) Use Kepler's second law to show that the ratio of the speeds of a planet at its nearest and farthest points from the Sun is equal to the inverse ratio of the near and far distances: $v_N/v_F = d_F/d_N$. (*b*) Given that the Earth's distance from the Sun varies from 1.47 to 1.52×10^{11} m, determine the minimum and maximum velocities of the Earth in its orbit around the Sun.

56. (III) The orbital periods T and mean orbital distances r for Jupiter's four largest moons are given in Table 6–3. (*a*) Starting with Kepler's third law in the form (see Eq. 6–6)

$$T^2 = \left(\frac{4\pi^2}{Gm_J}\right)r^3,$$

where m_J is the mass of Jupiter, show that this relation implies that a plot of $\log(T)$ vs. $\log(r)$ will yield a straight line. Explain what Kepler's third law predicts about the slope and y-intercept of this straight-line plot. (*b*) Using the data for Jupiter's four moons, plot $\log(T)$ vs. $\log(r)$ and show that you get a straight line. Determine the slope of this plot and compare it to the value you expect if the data are consistent with Kepler's third law. Determine the y-intercept of the plot and use it to compute the mass of Jupiter.

57. (III) The *comet Hale–Bopp* has an orbital period of 2400 years. (*a*) What is its mean distance from the Sun? (*b*) At its closest approach, the comet is about 1.0 astronomical unit from the Sun (1 AU = distance from Earth to the Sun). What is the farthest distance? (*c*) What is the ratio of the speed at the closest point to the speed at the farthest point?

6–6 Moon Rise

58. (III) (*a*) Explain why a Full moon always rises at sunset. (*b*) Explain how the position of the Moon in Fig. 6–24b cannot be seen yet by the person at the red dot (shown at 6 PM). (*c*) Explain why the red dot is where it is in parts (b) and (e), and show where it should be in part (d). (*d*) [PRETTY HARD:] Determine the average period of the Moon around the Earth (sidereal period) starting with the synodic period of 29.53 days as observed from Earth. [*Hint*: First determine the angle of the Moon in Fig. 6–24e relative to "horizontal," as in part (a).]

*6–8 Gravitational Field

***59.** (II) Determine the magnitude and direction of the gravitational field midway between the Earth and Moon. Ignore effects of the Sun.

***60.** (II) (*a*) What is the gravitational field at the surface of the Earth due to the Sun? (*b*) Will this affect your weight significantly?

***61.** (III) Two identical particles, each of mass m, are located on the x axis at $x = +x_0$ and $x = -x_0$. (*a*) Determine a formula for the gravitational field due to these two particles for points on the y axis; that is, write g as a function of y, m, x_0, and so on. (*b*) At what point (or points) on the y axis is the magnitude of g a maximum value, and what is its value there? [*Hint*: Take the derivative dg/dy.]

General Problems

62. Calculate the centripetal acceleration of the Earth in its orbit around the Sun, and the net force exerted on the Earth. What exerts this force on the Earth? Assume that the Earth's orbit is a circle of radius 1.50×10^{11} m.

63. **Binary star:** Two equal-mass stars maintain a constant distance D apart (Fig. 6–34) of 8.0×10^{11} m and revolve about a point midway between them at a rate of one revolution every 14.2 yr. (*a*) Why don't the two stars crash into one another due to the gravitational force between them? (*b*) What must be the mass m of each star?

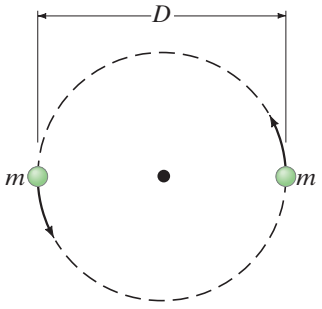


FIGURE 6–34
Problem 63.

64. How far above the Earth's surface will the acceleration due to gravity be half what it is at the surface?

65. How long would a day be if the Earth were rotating so fast that objects at the equator were apparently weightless?

66. Show that the rate of change of your weight is

$$-2G\frac{m_E m}{r^3}v$$

if you are traveling directly away from Earth at constant speed v. Your mass is m, and r is your distance from the center of the Earth at any moment.

67. The *rings of Saturn* are composed of chunks of ice that orbit the planet. The inner radius of the rings is 73,000 km, and the outer radius is 170,000 km. Find the period of an orbiting chunk of ice at the inner radius and the period of a chunk at the outer radius. Compare your numbers with Saturn's own rotation period of 10 hours and 39 minutes. The mass of Saturn is 5.7×10^{26} kg.

68. The **Navstar Global Positioning System** (GPS) utilizes a group of 24 satellites orbiting the Earth. Using "triangulation" and signals transmitted by these satellites, the position of a receiver on the Earth can be determined to within an accuracy of a few centimeters. The satellite orbits are distributed around the Earth, allowing continuous navigational "fixes." The satellites orbit at an altitude of approximately 11,000 nautical miles [1 nautical mile = 1.852 km = 6076 ft]. (*a*) Determine the speed of each satellite. (*b*) Determine the period of each satellite. [Originally, 1 *nautical mile* was defined as one minute ($\frac{1}{60}$ of a degree) of latitude on Earth's surface. 1 *knot* is a speed of 1 nautical mile/h.]

69. The *Near Earth Asteroid Rendezvous* (*NEAR*) spacecraft, after traveling 2.1 billion km, orbited the *asteroid* Eros with an orbital radius of about 20 km. Eros is roughly 40 km × 6 km × 6 km. Assume Eros has a density (mass/volume) of about 2.3×10^3 kg/m^3. (*a*) If Eros were a sphere with the same mass and density, what would its radius be? (*b*) What would g be at the surface of a spherical Eros? (*c*) Estimate the orbital period of *NEAR* around Eros, as if Eros were a sphere.

70. The *Lunar Reconnaissance Orbiter* (LRO) is a NASA Moon-orbiting spacecraft in an orbit with an altitude of 50 km above the Moon's surface. What is the period of the LRO as it orbits the Moon?

71. Astronomers have observed an otherwise normal star, called S2, closely orbiting an extremely massive but small object at the center of our *Milky Way Galaxy* called SgrA. S2 moves in an elliptical orbit around SgrA with a period of 15.2 yr and an eccentricity $e = 0.87$ (Fig. 6–17). In 2002, S2 reached its closest approach to SgrA, a distance of only 123 AU ($1\text{ AU} = 1.50 \times 10^{11}\text{ m}$ is the mean Earth–Sun distance). Determine the mass m of SgrA, the massive compact object (believed to be a supermassive black hole) at the center of our Galaxy. State m in kg and in terms of the mass of our Sun.

72. A satellite circles a spherical planet of unknown mass in a circular orbit of radius 1.6×10^7 m. The magnitude of the gravitational force exerted on the satellite by the planet is 120 N. (*a*) What would be the magnitude of the gravitational force exerted on the satellite by the planet if the radius of the orbit were increased to 3.0×10^7 m? (*b*) If the satellite circles the planet once every 2.0 h in the larger orbit, what is the mass of the planet?

73. A Ferris wheel (Fig. 6–35), 22.0 m in diameter, rotates once every 12.5 s. What is the ratio of a person's apparent weight to her real weight at (*a*) the top, and (*b*) the bottom?

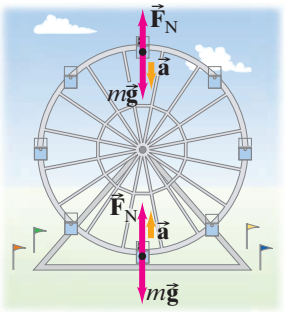


FIGURE 6–35 Problem 73.

74. Estimate the value of the gravitational constant G in Newton's law of universal gravitation using the following data: the acceleration due to gravity at the Earth's surface is about 10 m/s^2; the Earth has a circumference of about 40×10^6 m; rocks found on the Earth's surface typically have densities of about 3000 kg/m^3 and assume this density is constant throughout (even though you suspect it is not true).

75. Newton had the data listed in Table 6–4, plus the relative sizes of these objects: in terms of the Sun's radius r_S, the radii of Jupiter and Earth were $0.0997r_S$ and $0.0109r_S$. Newton used this information to determine that the average density ρ (= mass/volume) of Jupiter is slightly less than that of the Sun, while the average density of the Earth is four times that of the Sun. Thus, without leaving his home planet, Newton was able to predict that the composition of the Sun and Jupiter is markedly different than that of Earth. Reproduce Newton's calculation and find his values for the ratios ρ_J/ρ_{Sun} and ρ_E/ρ_{Sun} (the modern values for these ratios are 0.93 and 3.91, respectively).

TABLE 6–4 Problem 75

	Orbital Radius, R (in AU = 1.50×10^{11} m)	Orbital Period, T (Earth days)
Venus about Sun	0.724	224.70
Callisto about Jupiter	0.01253	16.69
Moon about Earth	0.003069	27.32

76. A uniform sphere has mass M and radius r. A spherical cavity (no mass) of radius $r/2$ is then carved within this sphere as shown in Fig. 6–36 (the cavity's surface passes through the sphere's center and just touches the sphere's outer surface). The centers of the original sphere and the cavity lie on a straight line, which defines the x axis. With what gravitational force will the hollowed-out sphere attract a point mass m which lies on the x axis a distance d from the sphere's center? [*Hint*: Subtract the effect of the "small" sphere (the cavity) from that of the larger entire sphere.]

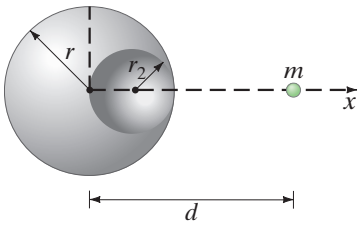


FIGURE 6–36 Problem 76.

77. Jupiter is about 320 times as massive as the Earth. Thus, it has been claimed that a person would be crushed by the force of gravity on a planet the size of Jupiter because people cannot survive more than a few g's. Calculate the number of g's a person would experience at Jupiter's equator, using the following data for Jupiter: mass $= 1.9 \times 10^{27}$ kg, equatorial radius $= 7.1 \times 10^4$ km, rotation period = 9 hr 55 min. Take the centripetal acceleration into account.

78. The gravitational force at different places on Earth due to the Sun and the Moon depends on each point's distance from the Sun or Moon, and this variation is what causes the **tides**. Use the values inside the front cover of this book for the Earth–Moon distance R_{EM}, the Earth–Sun distance R_{ES}, the Moon's mass m_M, the Sun's mass m_S, and the Earth's radius r_E. (*a*) First consider two small pieces of the Earth, each of mass m, one on the side of the Earth nearest the Moon, the other on the side farthest from the Moon. Show that the ratio of the Moon's gravitational forces on these two masses is

$$\left(\frac{F_{near}}{F_{far}}\right)_M = 1.0687.$$

(*b*) Next consider two small pieces of the Earth, each of mass m, one on the nearest point of Earth to the Sun, the other at the farthest point from the Sun. Show that the ratio of the Sun's gravitational forces on these two masses is

$$\left(\frac{F_{near}}{F_{far}}\right)_S = 1.000171.$$

(*c*) Show that the ratio of the Sun's average gravitational force on the Earth (F_S) compared to the Moon's average gravitational force on the Earth (F_M) is

$$\left(\frac{F_S}{F_M}\right)_{avg} = 178.$$

Note that the Moon's smaller force varies much more across the Earth's diameter than the Sun's larger force. (*d*) Estimate the resulting "force difference" (the cause of the tides)

$$\Delta F = F_{near} - F_{far} = F_{far}\left(\frac{F_{near}}{F_{far}} - 1\right) \approx F_{avg}\left(\frac{F_{near}}{F_{far}} - 1\right)$$

for the Moon and for the Sun. Show that the ratio of the tide-causing force differences due to the Moon compared to the Sun is

$$\frac{\Delta F_M}{\Delta F_S} \approx 2.3.$$

Thus the Moon's influence on tide production is over two times as great as the Sun's.

79. A particle is released at a height r_E (radius of Earth) above the Earth's surface. Determine its velocity when it hits the Earth. Ignore air resistance. [*Hint*: Use Newton's second law, the law of universal gravitation, the chain rule, and integrate.]

80. A plumb bob (a mass m hanging on a string) is deflected from the vertical by an angle θ due to a massive mountain nearby (Fig. 6–37). (*a*) Find an approximate formula for θ in terms of the mass of the mountain, m_M, the distance to its center, D_M, and the radius and mass of the Earth. (*b*) Make a rough estimate of the mass of Mt. Everest, assuming it has the shape of a cone 4000 m high and base of diameter 4000 m. Assume its mass per unit volume is 3000 kg per m^3. (*c*) Estimate the angle θ of the plumb bob if it is 5 km from the center of Mt. Everest.

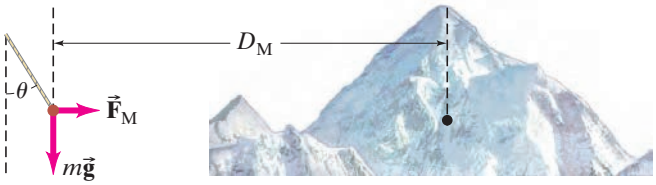


FIGURE 6–37 Problem 80.

81. A science-fiction tale describes an artificial "planet" in the form of a band completely encircling a sun (Fig. 6–38). The inhabitants live on the inside surface (where it is always noon). Imagine that this sun is exactly like our own, that the distance to the band is the same as the Earth–Sun distance (to make the climate livable), and that the ring rotates quickly enough to produce an apparent gravity of g as on Earth. What will be the period of revolution (this planet's year) in Earth days?

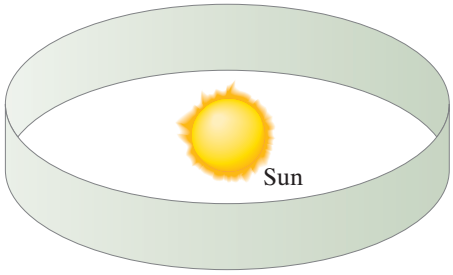


FIGURE 6–38 Problem 81.

82. A satellite of mass 5200 kg orbits the Earth and has a period of 6800 s. Determine (*a*) the radius of its circular orbit, (*b*) the magnitude of the Earth's gravitational force on the satellite, and (*c*) the altitude of the satellite.

83. Suppose all the mass of the Earth were compacted into a small spherical ball. What radius must the sphere have so that the acceleration due to gravity at the Earth's new surface would equal the acceleration due to gravity at the surface of the Sun?

84. An asteroid of mass m is in a circular orbit of radius r around the Sun with a speed v. It has an impact with another asteroid of mass M and is kicked into a new circular orbit with a speed of $1.5v$. What is the radius of the new orbit in terms of r?

85. Between the orbits of Mars and Jupiter, several thousand small objects called **asteroids** move in nearly circular orbits around the Sun. Consider an asteroid that is spherically shaped with radius r and density 2700 kg/m^3. (*a*) You find yourself on the surface of this asteroid and throw a baseball at a speed of 22 m/s (about 50 mi/h). If the baseball is to travel around the asteroid in a circular orbit, what is the largest radius asteroid on which you are capable of accomplishing this feat? (*b*) After you throw the baseball, you turn around and face the opposite direction and catch the baseball. How much time T elapses between your throw and your catch?

86. The Sun revolves around the center of our *Milky Way Galaxy* (Fig. 6–39) at a distance of about 30,000 light-years from the center $(1\,\text{ly} = 9.5 \times 10^{15}\,\text{m})$. If it takes about 200 million years to make one revolution, estimate the mass of our Galaxy. Assume that the mass distribution of our Galaxy is concentrated mostly in a central uniform sphere. If all the stars had about the mass of our Sun $(2 \times 10^{30}\,\text{kg})$, how many stars would there be in our Galaxy?

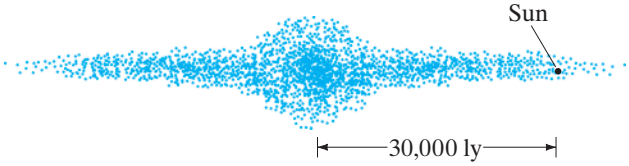


FIGURE 6–39 Edge-on view of our Galaxy. Problem 86.

ANSWERS TO EXERCISES

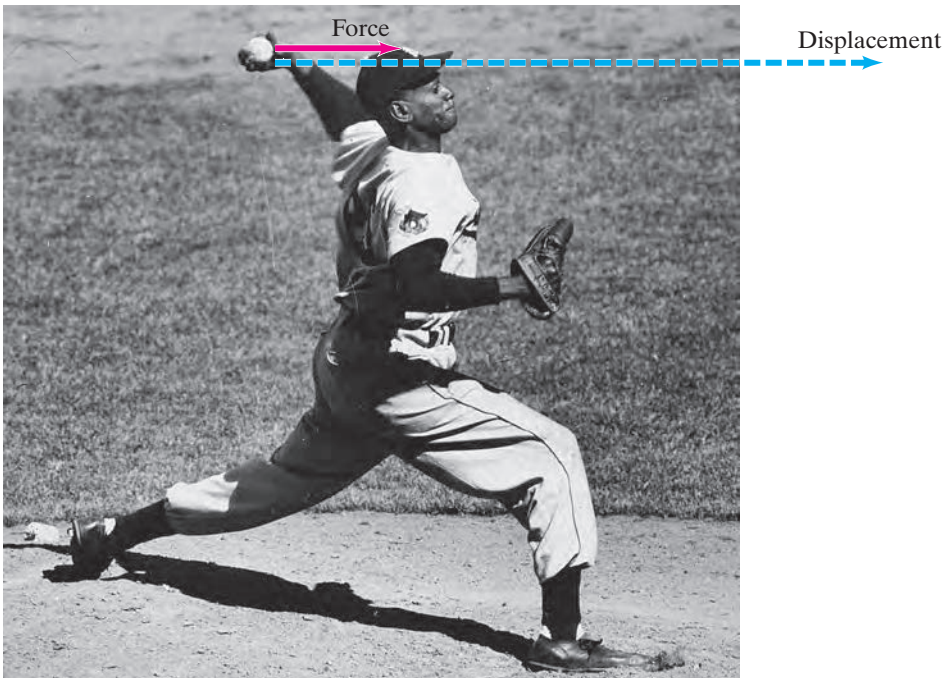
A: g would double.

B: (*b*).

C: (*b*).

D: No; even though they are experiencing weightlessness, the massive ball would require a large force to accelerate when thrown and also to decelerate when caught (inertial mass, Newton's second law of motion).

E: 6.17 yr.



This baseball pitcher (Satchel Paige) is about to accelerate the baseball from rest to a high velocity by exerting a force on it. He will be doing work on the ball as he exerts the force over a displacement of perhaps several meters, from behind his head until he releases the ball with arm outstretched in front of him. The total work done on the ball will be equal to the kinetic energy ($\frac{1}{2}mv^2$) acquired by the ball, a result known as the work-energy principle.

CHAPTER 7

Work and Energy

CONTENTS

7–1 Work Done by a Constant Force
7–2 Scalar Product of Two Vectors
7–3 Work Done by a Varying Force
7–4 Kinetic Energy and the Work-Energy Principle

CHAPTER-OPENING QUESTION—Guess now!
You push very hard on a heavy desk, trying to move it. You do work on the desk:

(a) Whether or not it moves, as long as you are exerting a force.
(b) Only if it starts moving.
(c) Only if it doesn't move.
(d) Never—it does work on you.
(e) None of the above.

Until now we have been studying the translational motion of an object in terms of Newton's three laws of motion. In that analysis, *force* has played a central role as the quantity determining the motion. In this Chapter and the next two (8 and 9), we discuss an alternative analysis of the translational motion of objects in terms of the quantities *energy* and *momentum*. The significance of energy and momentum is that they are *conserved*. That is, in quite general circumstances they remain constant. That conserved quantities exist gives us not only a deeper insight into the nature of the world, but also gives us another way to approach solving practical problems.

The conservation laws of energy and momentum are especially valuable in dealing with systems of many objects, in which a detailed consideration of the forces involved would be difficult or impossible. These laws apply to a wide range of phenomena. They even apply in the atomic and subatomic worlds, where Newton's laws are not sufficient.

This Chapter is devoted to the very important concept of *energy* and the closely related concept of *work*. These two quantities are scalars and so have no direction associated with them, which often makes them easier to work with than vector quantities such as acceleration and force.

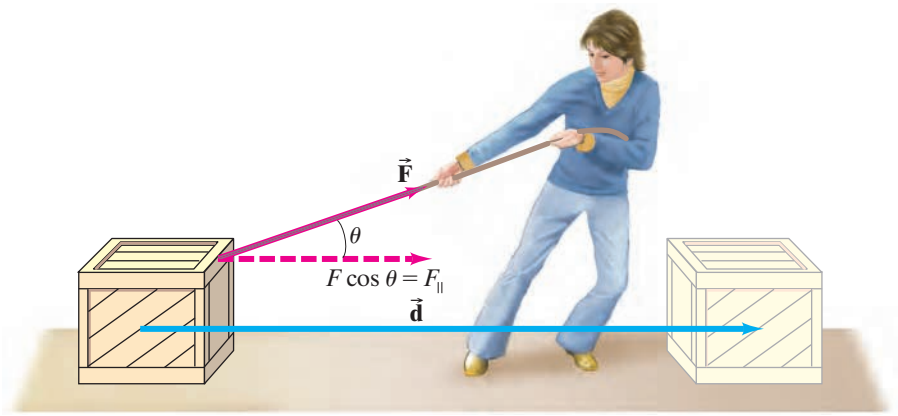


FIGURE 7–1 A person pulling a crate along the floor. The work done on the crate by the force $\vec{\mathbf{F}}$ is $W = Fd\cos\theta$, where $\vec{\mathbf{d}}$ is the displacement and θ is the angle between the directions of the force vector and the displacement vector.

7–1 Work Done by a Constant Force

The word *work* has a variety of meanings in everyday language. But in physics, work is given a very specific meaning to describe what is accomplished when a force acts on an object, and the object moves through a distance. We consider only translational motion for now and, unless otherwise explained, objects are assumed to be rigid with no complicating internal motion, and can be treated like particles. Then the **work** done on an object by a constant force (constant in both magnitude and direction) is defined to be *the product of the magnitude of the displacement times the component of the force parallel to the displacement*. In equation form, we can write

$$W = F_{\parallel} d,$$

where $F_{\parallel}$ is the component of the constant force $\vec{\mathbf{F}}$ parallel to the displacement $\vec{\mathbf{d}}$. We can also write

$$W = Fd\cos\theta, \qquad \textbf{(7–1)}$$

where F is the magnitude of the constant force, d is the magnitude of the displacement of the object, and θ is the angle between the directions of the force and the displacement (Fig. 7–1). The $\cos\theta$ factor appears in Eq. 7–1 because $F\cos\theta\ (= F_{\parallel})$ is the component of $\vec{\mathbf{F}}$ that is parallel to $\vec{\mathbf{d}}$. Work is a scalar quantity—it has only magnitude, which can be positive or negative.

Let us consider the case in which the motion and the force are in the same direction, so $\theta = 0$ and $\cos\theta = 1$; in this case, $W = Fd$. For example, if you push a loaded grocery cart a distance of 50 m by exerting a horizontal force of 30 N on the cart, you do $30\,\text{N} \times 50\,\text{m} = 1500\,\text{N}\cdot\text{m}$ of work on the cart.

As this example shows, in SI units work is measured in newton-meters (N · m). A special name is given to this unit, the **joule** (J): $1\,\text{J} = 1\,\text{N}\cdot\text{m}$.

[In the cgs system, the unit of work is called the **erg** and is defined as $1\,\text{erg} = 1\,\text{dyne}\cdot\text{cm}$. In British units, work is measured in **foot-pounds**. Their equivalence is $1\,\text{J} = 10^7\,\text{erg} = 0.7376\,\text{ft}\cdot\text{lb}$.]

A force can be exerted on an object and yet do no work. If you hold a heavy bag of groceries in your hands at rest, you do no work on it. You do exert a force on the bag, but the displacement of the bag is zero, so the work done by you on the bag is $W = 0$. You need both a force and a displacement to do work. You also do no work on the bag of groceries if you carry it as you walk horizontally across the floor at constant velocity, as shown in Fig. 7–2. No horizontal force is required to move the bag at a constant velocity. The person shown in Fig. 7–2 does exert an upward force $\vec{\mathbf{F}}_{\text{P}}$ on the bag equal to its weight, mg. But this upward force is perpendicular to the horizontal displacement of the bag and thus is doing no work. This conclusion comes from our definition of work, Eq. 7–1: $W = 0$, because $\theta = 90°$ and $\cos 90° = 0$. Thus, when a particular force is perpendicular to the displacement, no work is done by that force. When you start or stop walking, there is a horizontal acceleration and you do briefly exert a horizontal force, and thus do work on the bag.

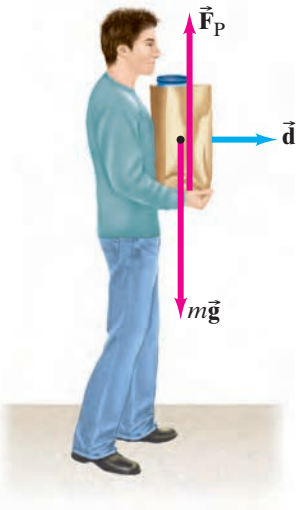


FIGURE 7–2 A person walking at constant velocity does no work on the bag of groceries since $\vec{\mathbf{F}}_{\text{P}}$ is perpendicular to the displacement $\vec{\mathbf{d}}$. ($F_{\text{P}} = mg$, the weight of the bag.)

⚠ CAUTION
Force without work

CAUTION
State that work is done on or by an object

When we deal with work, as with force, it is necessary to specify whether we are talking about work done *by* a specific object or done *on* a specific object. It is also important to specify whether the work done is due to one particular force (and which one), or the total (net) work done by the *net force* on the object.

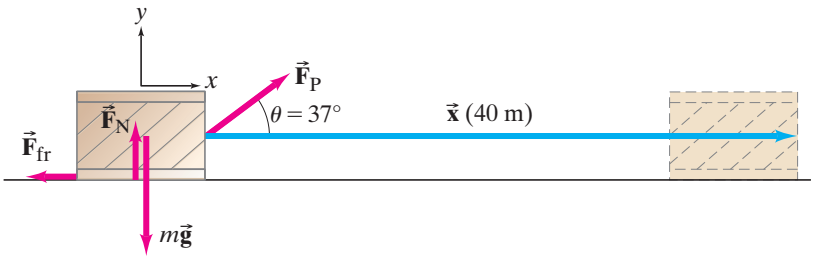


FIGURE 7–3 Example 7–1. A 50-kg crate is pulled along a floor.

EXAMPLE 7–1 **Work done on a crate.** A person pulls a 50-kg crate 40 m along a horizontal floor by a constant force $F_P = 100\,\text{N}$, which acts at a 37° angle as shown in Fig. 7–3. The floor is rough and exerts a friction force $F_{fr} = 50\,\text{N}$. Determine (*a*) the work done by each force acting on the crate, and (*b*) the net work done on the crate.

APPROACH We choose our coordinate system so that the vector that represents the 40-m displacement is $\vec{\mathbf{x}}$ (that is, along the x axis). Four forces act on the crate, as shown in the free-body diagram in Fig. 7–3: the force exerted by the person $\vec{\mathbf{F}}_P$; the friction force $\vec{\mathbf{F}}_{fr}$; the gravitational force exerted by the Earth, $\vec{\mathbf{F}}_G = m\vec{\mathbf{g}}$; and the normal force $\vec{\mathbf{F}}_N$ exerted upward by the floor. The net force on the crate is the vector sum of these four forces.

SOLUTION (*a*) The work done by gravity $(\vec{\mathbf{F}}_G)$ and by the normal force $(\vec{\mathbf{F}}_N)$ is zero, because they are perpendicular to the displacement $\vec{\mathbf{x}}$ ($\theta = 90°$ in Eq. 7–1):

$$W_G = mgx\cos 90° = 0$$
$$W_N = F_N x\cos 90° = 0.$$

The work done by $\vec{\mathbf{F}}_P$ is

$$W_P = F_P x\cos\theta = (100\,\text{N})(40\,\text{m})\cos 37° = 3200\,\text{J}.$$

The work done by the friction force is

$$W_{fr} = F_{fr} x\cos 180° = (50\,\text{N})(40\,\text{m})(-1) = -2000\,\text{J}.$$

The angle between the displacement $\vec{\mathbf{x}}$ and $\vec{\mathbf{F}}_{fr}$ is 180° because they point in opposite directions. Since the force of friction is opposing the motion (and $\cos 180° = -1$), the work done by friction on the crate is *negative*.

(*b*) The net work can be calculated in two equivalent ways.
(1) The net work done on an object is the algebraic sum of the work done by each force, since work is a scalar:

$$W_{net} = W_G + W_N + W_P + W_{fr}$$
$$= 0 + 0 + 3200\,\text{J} - 2000\,\text{J} = 1200\,\text{J}.$$

(2) The net work can also be calculated by first determining the net force on the object and then taking the component of this net force along the displacement: $(F_{net})_x = F_P\cos\theta - F_{fr}$. Then the net work is

$$W_{net} = (F_{net})_x x = (F_P\cos\theta - F_{fr})x$$
$$= (100\,\text{N}\cos 37° - 50\,\text{N})(40\,\text{m}) = 1200\,\text{J}.$$

In the vertical (y) direction, there is no displacement and no work done.

CAUTION
Negative work

In Example 7–1 we saw that friction did *negative work*. In general, the work done by a force is negative whenever the force (or the component of the force, $F_{\parallel}$) acts in the direction opposite to the direction of motion.

EXERCISE A A box is dragged a distance d across a floor by a force $\vec{\mathbf{F}}_P$ which makes an angle θ with the horizontal as in Fig. 7–1 or 7–3. If the magnitude of $\vec{\mathbf{F}}_P$ is held constant but the angle θ is increased, the work done by $\vec{\mathbf{F}}_P$ (*a*) remains the same; (*b*) increases; (*c*) decreases; (*d*) first increases, then decreases.

EXERCISE B Return to the Chapter-Opening Question, page 172, and answer it again now. Try to explain why you may have answered differently the first time.

PROBLEM SOLVING

Work

1. **Draw a free-body diagram** showing all the forces acting on the object you choose to study.
2. **Choose** an xy **coordinate system**. If the object is in motion, it may be convenient to choose one of the coordinate directions as the direction of one of the forces, or as the direction of motion. [Thus, for an object on an incline, you might choose one coordinate axis to be parallel to the incline.]
3. **Apply Newton's laws** to determine unknown forces.
4. Find the **work done *by* a specific force** *on* the object by using $W = Fd\cos\theta$ for a constant force. The work done is negative when a force opposes the displacement.
5. To find the **net work** done on the object, either (a) find the work done by each force and add the results algebraically; or (b) find the net force on the object, F_{net}, and then use it to find the net work done, which for constant net force is:

$$W_{\text{net}} = F_{\text{net}}\, d\cos\theta.$$

EXAMPLE 7–2 **Work on a backpack.** (a) Determine the work a hiker must do on a 15.0-kg backpack to carry it up a hill of height $h = 10.0\,\text{m}$, as shown in Fig. 7–4a. Determine also (b) the work done by gravity on the backpack, and (c) the net work done on the backpack. For simplicity, assume the motion is smooth and at constant velocity (i.e., acceleration is zero).

APPROACH We explicitly follow the steps of the Problem Solving Strategy above.

SOLUTION

1. **Draw a free-body diagram**. The forces on the backpack are shown in Fig. 7–4b: the force of gravity, $m\vec{\mathbf{g}}$, acting downward; and $\vec{\mathbf{F}}_{\text{H}}$, the force the hiker must exert upward to support the backpack. The acceleration is zero, so horizontal forces on the backpack are negligible.
2. **Choose a coordinate system**. We are interested in the vertical motion of the backpack, so we choose the y coordinate as positive vertically upward.
3. **Apply Newton's laws**. Newton's second law applied in the vertical direction to the backpack gives (with $a_y = 0$)

$$\Sigma F_y = ma_y$$
$$F_{\text{H}} - mg = 0.$$

So,

$$F_{\text{H}} = mg = (15.0\,\text{kg})(9.80\,\text{m/s}^2) = 147\,\text{N}.$$

4. **Work done by a specific force**. (a) To calculate the work done by the hiker on the backpack, we use Eq. 7–1, where θ is shown in Fig. 7–4c,

$$W_{\text{H}} = F_{\text{H}}(d\cos\theta),$$

and we note from Fig. 7–4a that $d\cos\theta = h$. So the work done by the hiker is

$$W_{\text{H}} = F_{\text{H}}(d\cos\theta) = F_{\text{H}}h = mgh = (147\,\text{N})(10.0\,\text{m}) = 1470\,\text{J}.$$

The work done depends only on the elevation change and not on the angle of the hill, θ. The hiker would do the same work to lift the pack vertically by height h.

(b) The work done by gravity on the backpack is (from Eq. 7–1 and Fig. 7–4c)

$$W_{\text{G}} = mg\,d\cos(180^\circ - \theta).$$

Since $\cos(180^\circ - \theta) = -\cos\theta$ (Appendix A–9), we have

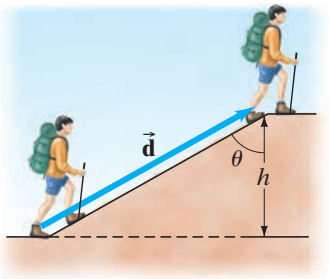
$$W_{\text{G}} = mg(-d\cos\theta)$$
$$= -mgh$$
$$= -(15.0\,\text{kg})(9.80\,\text{m/s}^2)(10.0\,\text{m}) = -1470\,\text{J}.$$

NOTE The work done by gravity (which is negative here) does not depend on the angle of the incline, only on the vertical height h of the hill.

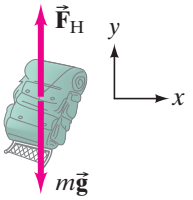
5. **Net work done**. (c) The *net* work done on the backpack is $W_{\text{net}} = 0$, because the net force on the backpack is zero (it is assumed not to accelerate significantly). We can also get the net work done by adding the work done by each force:

$$W_{\text{net}} = W_{\text{G}} + W_{\text{H}} = -1470\,\text{J} + 1470\,\text{J} = 0.$$

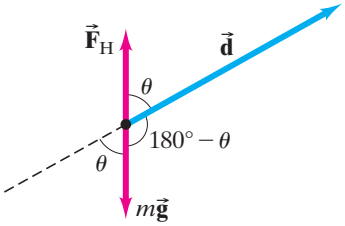
NOTE Even though the *net* work done by all the forces on the backpack is zero, the hiker *does* do work on the backpack equal to 1470 J.



(a)



(b)



(c)

FIGURE 7–4 Example 7–2.

PROBLEM SOLVING
Work done by gravity depends on height of hill (not on angle)

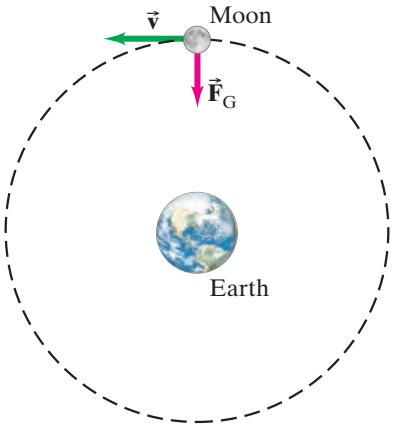


FIGURE 7–5 Example 7–3.

CONCEPTUAL EXAMPLE 7–3 **Does the Earth do work on the Moon?**
The Moon revolves around the Earth in a nearly circular orbit, with approximately constant tangential speed, kept there by the gravitational force exerted by the Earth. Does gravity do (*a*) positive work, (*b*) negative work, or (*c*) no work at all on the Moon?

RESPONSE The gravitational force $\vec{\mathbf{F}}_G$ exerted by the Earth on the Moon (Fig. 7–5) acts toward the Earth and provides its centripetal acceleration, inward along the radius of the Moon's orbit. The Moon's displacement at any moment is tangent to the circle, in the direction of its velocity, perpendicular to the radius and perpendicular to the force of gravity. Hence the angle θ between the force $\vec{\mathbf{F}}_G$ and the instantaneous displacement of the Moon is 90°, and the work done by gravity is therefore zero $(\cos 90° = 0)$. This is why the Moon, as well as artificial satellites, can stay in orbit without expenditure of fuel: no work needs to be done against the force of gravity.

7–2 Scalar Product of Two Vectors

Although work is a scalar, it involves the product of two quantities, force and displacement, both of which are vectors. Therefore, we now investigate the multiplication of vectors, which will be useful throughout the book, and apply it to work.

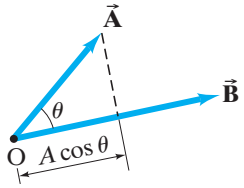
Because vectors have direction as well as magnitude, they cannot be multiplied in the same way that scalars are. Instead we must *define* what the operation of vector multiplication means. Among the possible ways to define how to multiply vectors, there are three ways that we find useful in physics: (1) multiplication of a vector by a scalar, which was discussed in Section 3–3; (2) multiplication of one vector by a second vector to produce a scalar; (3) multiplication of one vector by a second vector to produce another vector. The third type, called the *vector product*, will be discussed later, in Section 11–2.

We now discuss the second type, called the *scalar product*, or *dot product* (because a dot is used to indicate the multiplication). If we have two vectors, $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$, then their **scalar** (or **dot**) **product** is defined to be

Scalar product

$$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = AB\cos\theta, \qquad \textbf{(7–2)}$$

FIGURE 7–6 The scalar product, or dot product, of two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ is $\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = AB\cos\theta$. The scalar product can be interpreted as the magnitude of one vector (B in this case) times the projection of the other vector, $A\cos\theta$, onto $\vec{\mathbf{B}}$.



where A and B are the magnitudes of the vectors and θ is the angle ($< 180°$) between them when their tails touch, Fig. 7–6. Since A, B, and $\cos\theta$ are scalars, then so is the scalar product $\vec{\mathbf{A}} \cdot \vec{\mathbf{B}}$ (read "A dot B").

This definition, Eq. 7–2, fits perfectly with our definition of the work done by a constant force, Eq. 7–1. That is, we can write the work done by a constant force as the scalar product of force and displacement:

$$W = \vec{\mathbf{F}} \cdot \vec{\mathbf{d}} = Fd\cos\theta. \qquad \textbf{(7–3)}$$

Indeed, the definition of scalar product, Eq. 7–2, is so chosen because many physically important quantities, such as work (and others we will meet later), can be described as the scalar product of two vectors.

An equivalent definition of the scalar product is that it is the product of the magnitude of one vector (say B) and the component (or projection) of the other vector along the direction of the first $(A\cos\theta)$. See Fig. 7–6.

Since A, B, and $\cos\theta$ are scalars, it doesn't matter in what order they are multiplied. Hence the scalar product is **commutative**:

$$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = \vec{\mathbf{B}} \cdot \vec{\mathbf{A}}. \qquad \text{[commutative property]}$$

It is not difficult to show that the scalar product is also **distributive** (see Problem 33 for the proof):

$$\vec{\mathbf{A}} \cdot (\vec{\mathbf{B}} + \vec{\mathbf{C}}) = \vec{\mathbf{A}} \cdot \vec{\mathbf{B}} + \vec{\mathbf{A}} \cdot \vec{\mathbf{C}}. \qquad \text{[distributive property]}$$

Let us write our vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ in terms of their rectangular components using unit vectors (Section 3–5, Eq. 3–5) as

$$\vec{\mathbf{A}} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}$$
$$\vec{\mathbf{B}} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}}.$$

We will take the scalar product, $\vec{\mathbf{A}}\cdot\vec{\mathbf{B}}$, of these two vectors, remembering that the unit vectors, $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$, are perpendicular to each other

$$\hat{\mathbf{i}}\cdot\hat{\mathbf{i}} = \hat{\mathbf{j}}\cdot\hat{\mathbf{j}} = \hat{\mathbf{k}}\cdot\hat{\mathbf{k}} = 1$$
$$\hat{\mathbf{i}}\cdot\hat{\mathbf{j}} = \hat{\mathbf{i}}\cdot\hat{\mathbf{k}} = \hat{\mathbf{j}}\cdot\hat{\mathbf{k}} = 0.$$

Thus the scalar product can be written (using the distributive property) as

$$\vec{\mathbf{A}}\cdot\vec{\mathbf{B}} = (A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}})\cdot(B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}})$$
$$= A_xB_x + A_yB_y + A_zB_z. \qquad \textbf{(7–4)}$$

Scalar (or dot) product

Equation 7–4 is very useful.

If $\vec{\mathbf{A}}$ is perpendicular to $\vec{\mathbf{B}}$, then Eq. 7–2 tells us $\vec{\mathbf{A}}\cdot\vec{\mathbf{B}} = AB\cos 90° = 0$. But the converse, given that $\vec{\mathbf{A}}\cdot\vec{\mathbf{B}} = 0$, can come about in three different ways: $\vec{\mathbf{A}} = 0$, $\vec{\mathbf{B}} = 0$, or $\vec{\mathbf{A}} \perp \vec{\mathbf{B}}$.

EXAMPLE 7–4 **Using the dot product.** The force shown in Fig. 7–7 has magnitude $F_P = 20\,\text{N}$ and makes an angle $\theta = 30°$ to the ground. Calculate the work done by this force using Eq. 7–4 when the wagon is dragged 100 m along the ground.



FIGURE 7–7 Example 7–4. Work done by a force $\vec{\mathbf{F}}_P$ acting at an angle θ to the ground is $W = \vec{\mathbf{F}}_P\cdot\vec{\mathbf{d}}$.

APPROACH We choose the x axis horizontal to the right and the y axis vertically upward, and write $\vec{\mathbf{F}}_P$ and $\vec{\mathbf{d}}$ in terms of unit vectors.

SOLUTION

$$\vec{\mathbf{F}}_P = F_x\hat{\mathbf{i}} + F_y\hat{\mathbf{j}} = (F_P\cos 30°)\hat{\mathbf{i}} + (F_P\sin 30°)\hat{\mathbf{j}} = (17\,\text{N})\hat{\mathbf{i}} + (10\,\text{N})\hat{\mathbf{j}},$$

whereas $\vec{\mathbf{d}} = (100\,\text{m})\hat{\mathbf{i}}$. Then, using Eq. 7–4,

$$W = \vec{\mathbf{F}}_P\cdot\vec{\mathbf{d}} = (17\,\text{N})(100\,\text{m}) + (10\,\text{N})(0) + (0)(0) = 1700\,\text{J}.$$

Note that by choosing the x axis along $\vec{\mathbf{d}}$ we simplified the calculation because $\vec{\mathbf{d}}$ then has only one component.

7–3 Work Done by a Varying Force

If the force acting on an object is constant, the work done by that force can be calculated using Eq. 7–1. In many cases, however, the force varies in magnitude or direction during a process. For example, as a rocket moves away from Earth, work is done to overcome the force of gravity, which varies as the inverse square of the distance from the Earth's center. Other examples are the force exerted by a spring, which increases with the amount of stretch, or the work done by a varying force exerted to pull a box or cart up an uneven hill.

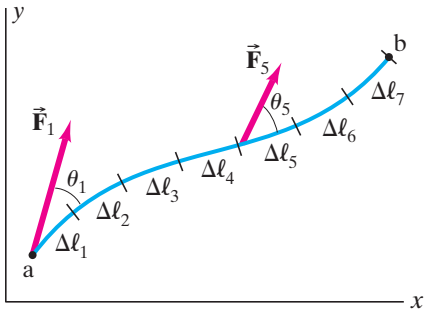


FIGURE 7–8 A particle acted on by a variable force, $\vec{\mathbf{F}}$, moves along a path in the xy plane as shown from point a to point b.

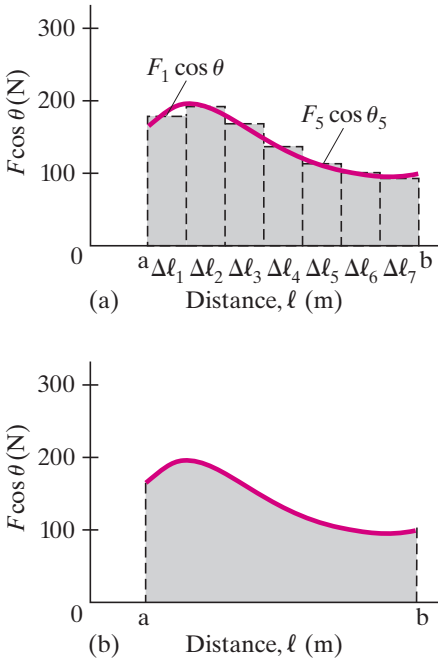
Figure 7–8 shows the path of an object in the xy plane as it moves from point a to point b. The path has been divided into short intervals each of length $\Delta\ell_1$, $\Delta\ell_2, \ldots, \Delta\ell_7$. A force $\vec{\mathbf{F}}$ acts at each point on the path, and is indicated at two points as $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_5$. During each small interval $\Delta\ell$, the force is approximately constant, assuming the intervals are taken small enough. For the first interval, the force does work ΔW of approximately (see Eq. 7–1)

$$\Delta W \approx F_1 \cos\theta_1 \Delta\ell_1.$$

In the second interval the work done is approximately $F_2 \cos\theta_2 \Delta\ell_2$, and so on. The total work done in moving the particle the total distance $\ell = \Delta\ell_1 + \Delta\ell_2 + \ldots + \Delta\ell_7$ is the sum of all these terms:

$$W \approx \sum_{i=1}^{7} F_i \cos\theta_i \Delta\ell_i. \tag{7–5}$$

FIGURE 7–9 Work done by a force F is (a) approximately equal to the sum of the areas of the rectangles (grey shading), (b) exactly equal to the area under the curve of $F\cos\theta$ vs. ℓ.



We can examine this graphically by plotting $F\cos\theta$ versus the distance ℓ along the path as shown in Fig. 7–9a. The distance ℓ has been subdivided into the same seven intervals (see the vertical dashed lines). The value of $F\cos\theta$ at the center of each interval is indicated by the *horizontal* dashed lines. Each of the shaded rectangles has an area $(F_i\cos\theta)(\Delta\ell_i)$, which is a good estimate of the work done during the interval. The estimate of the work done along the entire path, given by Eq. 7–5, equals the sum of the areas of all the rectangles. If we subdivide the distance into a greater number of intervals, so that each $\Delta\ell_i$ is smaller, the estimate of the work done becomes more accurate (the assumption that F is constant over each interval is more accurate).[†] Letting each $\Delta\ell_i$ approach zero (so we approach an infinite number of intervals), we obtain an exact result for the work done:

$$W = \lim_{\Delta\ell_i \to 0} \Sigma F_i \cos\theta_i \Delta\ell_i = \int_a^b F\cos\theta \, d\ell. \tag{7–6}$$

This limit as $\Delta\ell_i \to 0$ is the *integral* of $(F\cos\theta\, d\ell)$ from point a to point b. The symbol for the integral, $\int$, is an elongated S to indicate an infinite sum; and $\Delta\ell$ has been replaced by $d\ell$, meaning an infinitesimal distance.

In this limit as $\Delta\ell$ approaches zero, the total area of the rectangles (Fig. 7–9a) approaches the area between the $(F\cos\theta)$ curve and the ℓ axis from a to b as shown shaded in Fig. 7–9b. That is, *the work done by a variable force in moving an object between two points is equal to the area under the* $(F\cos\theta)$ *versus* (ℓ) *curve between those two points.*

In the limit as $\Delta\ell$ approaches zero, the infinitesimal distance $d\ell$ equals the magnitude of the infinitesimal displacement vector $d\vec{\boldsymbol{\ell}}$. The direction of the vector $d\vec{\boldsymbol{\ell}}$ is along the tangent to the path at that point, so θ is the angle between $\vec{\mathbf{F}}$ and $d\vec{\boldsymbol{\ell}}$ at any point. Thus we can rewrite Eq. 7–6, using dot-product notation:

$$W = \int_a^b \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}. \tag{7–7}$$

This is a *general definition of work* done by a force $\vec{\mathbf{F}}$. In this equation, a and b represent two points in space, (x_a, y_a, z_a) and (x_b, y_b, z_b). The integral in Eq. 7–7 is called a *line integral* since it is the integral of $F\cos\theta$ along the line that represents the path of the object. (Equation 7–1 for a constant force is a special case of Eq. 7–7.)

[†]We also discuss this "numerical integration" in Appendix C.

In rectangular coordinates, any force can be written

$$\vec{\mathbf{F}} = F_x\hat{\mathbf{i}} + F_y\hat{\mathbf{j}} + F_z\hat{\mathbf{k}}$$

and the displacement $d\vec{\boldsymbol{\ell}}$ is

$$d\vec{\boldsymbol{\ell}} = dx\,\hat{\mathbf{i}} + dy\,\hat{\mathbf{j}} + dz\,\hat{\mathbf{k}}.$$

Then the work done can be written (see Eq. 7–4)

$$W = \int_{x_a}^{x_b} F_x\,dx + \int_{y_a}^{y_b} F_y\,dy + \int_{z_a}^{z_b} F_z\,dz.$$

To actually use Eq. 7–6 or 7–7 to calculate the work, there are several options: (1) If $F\cos\theta$ is known as a function of position, a graph like that of Fig. 7–9b can be made and the area determined graphically. (2) Another possibility is to use numerical integration (numerical summing—see Appendix C), perhaps with the aid of a computer or calculator. (3) A third and important possibility is to use the analytical methods of integral calculus, when it is doable; to do so, we must be able to write $\vec{\mathbf{F}}$ as a function of position, $F(x, y, z)$, and we must know the path. Let's look at some specific examples.

Work Done by a Spring Force

Let us determine the work needed to stretch or compress a coiled spring, such as that shown in Fig. 7–10. For a person to hold a spring either stretched or compressed by a displacement x from its normal (relaxed) length requires a force F_P that is directly proportional to x. That is,

$$F_P = kx,$$

where k is a constant, called the **spring constant** (or **spring stiffness constant**), and is a measure of the stiffness of the particular spring. The spring itself exerts a force in the opposite direction (Fig. 7–10b or c):

$$F_S = -kx. \qquad \textbf{(7–8)}$$

This force is sometimes called a "restoring force" because the spring exerts its force in the direction opposite the displacement (hence the minus sign), and thus acts to restore the spring to its normal length. Equation 7–8 is known as the **spring equation** or **Hooke's law**, and is accurate for springs as long as x is not too great (see Section 12–5) and no permanent deformation occurs.

Let us calculate the work a person does to stretch (or compress) a spring from its normal (unstretched) length, $x_a = 0$, to an extra length, $x_b = x$. We assume the stretching is done slowly, so that the acceleration is essentially zero. The force $\vec{\mathbf{F}}_P$ is exerted parallel to the axis of the spring, along the x axis, so $\vec{\mathbf{F}}_P$ and $d\vec{\boldsymbol{\ell}}$ are parallel. Hence, since $d\vec{\boldsymbol{\ell}} = dx\,\hat{\mathbf{i}}$ in this case, the work done by the person is[†]

$$W_P = \int_{x_a=0}^{x_b=x} \left[F_P(x)\,\hat{\mathbf{i}}\right]\cdot\left[dx\hat{\mathbf{i}}\right] = \int_0^x F_P(x)\,dx = \int_0^x kx\,dx = \tfrac{1}{2}kx^2\Big|_0^x = \tfrac{1}{2}kx^2.$$

(As is frequently done, we have used x to represent both the variable of integration, and the particular value of x at the end of the interval $x_a = 0$ to $x_b = x$.) Thus we see that the work needed is proportional to the *square* of the distance stretched (or compressed), x.

This same result can be obtained by computing the area under the graph of F vs. x (with $\cos\theta = 1$ in this case) as shown in Fig. 7–11, and as we discussed on page 178 and Fig. 7–9. Since the area of a triangle of altitude $h = kx$ and base $b = x$ is $\frac{1}{2}hb = \frac{1}{2}kx^2$, then the work a person does to stretch or compress a spring an amount x is

$$W = \tfrac{1}{2}(kx)(x) = \tfrac{1}{2}kx^2,$$

which is the same result as before. Because $W \propto x^2$, it takes the same amount of work to stretch a spring or compress it the same displacement x.

[†]See the Table of Integrals, Appendix B.

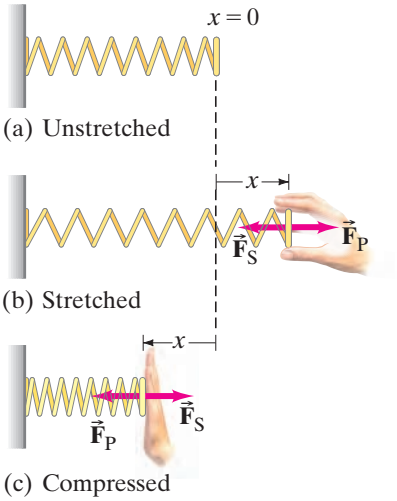


FIGURE 7–10 (a) Spring in normal (unstretched) position. (b) Spring is stretched by a person exerting a force $\vec{\mathbf{F}}_P$ to the right (positive direction). The spring pulls back with a force $\vec{\mathbf{F}}_S$ where $F_S = -kx$. (c) Person compresses the spring ($x < 0$), and the spring pushes back with a force $F_S = -kx$ where $F_S > 0$ because $x < 0$.

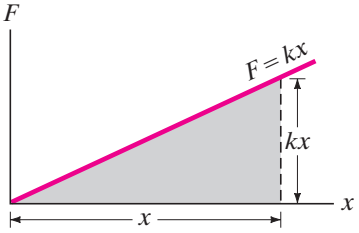


FIGURE 7–11 Work done to stretch a spring a distance x equals the triangular area under the curve $F = kx$. The area of a triangle is $\frac{1}{2}\times$ base $\times$ altitude, so $W = \frac{1}{2}(x)(kx) = \frac{1}{2}kx^2$.

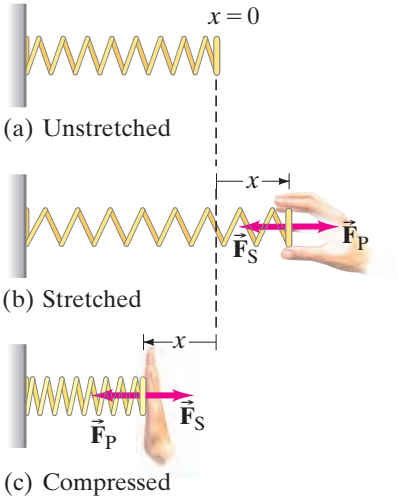


FIGURE 7–10 Repeated for Example 7-5.

EXAMPLE 7–5 **Work done on a spring.** (*a*) A person pulls on the spring in Fig. 7–10, stretching it 3.0 cm ($= x_{max}$), which requires a maximum force of 75 N. How much work does the person do? (*b*) If, instead, the person compresses the spring 3.0 cm, how much work does the person do?

APPROACH The force $F_P = kx$ holds at each point, including x_{max}. Hence F_{max} occurs at $x = x_{max}$.

SOLUTION (*a*) First we need to calculate the spring constant k:

$$k = \frac{F_{max}}{x_{max}} = \frac{75\text{ N}}{0.030\text{ m}} = 2.5 \times 10^3\text{ N/m}.$$

Then the work done by the person on the spring is

$$W = \tfrac{1}{2}kx_{max}^2 = \tfrac{1}{2}(2.5 \times 10^3\text{ N/m})(0.030\text{ m})^2 = 1.1\text{ J}.$$

(*b*) The force that the person exerts is still $F_P = kx$, though now both x and F_P are negative (x is positive to the right). The work done is

$$W_P = \int_{x=0}^{x=-0.030\text{ m}} F_P(x)\,dx = \int_0^{x=-0.030\text{ m}} kx\,dx = \tfrac{1}{2}kx^2\Big|_0^{-0.030\text{ m}}$$
$$= \tfrac{1}{2}(2.5 \times 10^3\text{ N/m})(-0.030\text{ m})^2 = 1.1\text{ J},$$

which is the same as for stretching it.

NOTE We cannot use $W = Fd$ (Eq. 7–1) for a spring because the force is not constant.

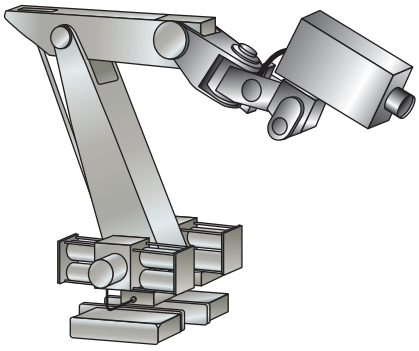


FIGURE 7–12 Robot arm positions a video camera.

A More Complex Force Law—Robot Arm

EXAMPLE 7–6 **Force as function of *x*.** A robot arm that controls the position of a video camera (Fig. 7–12) in an automated surveillance system is manipulated by a motor that exerts a force on the arm. Assume that the force is given by

$$F(x) = F_0\left(1 + \frac{1}{6}\frac{x^2}{x_0^2}\right),$$

where $F_0 = 2.0\text{ N}$, $x_0 = 0.0070\text{ m}$, and x is the position of the end of the arm. If the arm moves from $x_1 = 0.010\text{ m}$ to $x_2 = 0.050\text{ m}$, how much work did the motor do?

APPROACH The force applied by the motor is not a linear function of x. We can determine the integral $\int F(x)\,dx$, or the area under the $F(x)$ curve (shown as grey in Fig. 7–13).

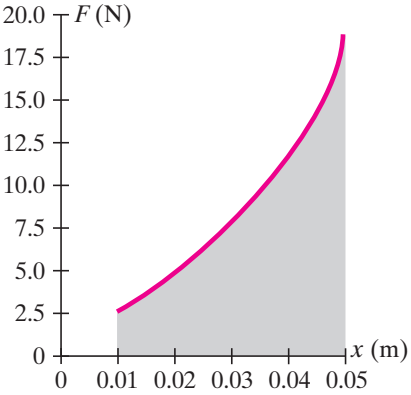
SOLUTION We integrate to find the work done by the motor:

$$W_M = F_0\int_{x_1}^{x_2}\left(1 + \frac{x^2}{6x_0^2}\right)dx = F_0\int_{x_1}^{x_2}dx + \frac{F_0}{6x_0^2}\int_{x_1}^{x_2}x^2\,dx$$
$$= F_0\left(x + \frac{1}{6x_0^2}\frac{x^3}{3}\right)\Bigg|_{x_1}^{x_2}.$$

We put in the values given and obtain

$$W_M = 2.0\text{ N}\left[(0.050\text{ m} - 0.010\text{ m}) + \frac{(0.050\text{ m})^3 - (0.010\text{ m})^3}{(3)(6)(0.0070\text{ m})^2}\right] = 0.36\text{ J}.$$

FIGURE 7–13 Example 7–6. The red curve represents the force on the robot arm as a function of position. The area under the curve (grey) equals the work done on the arm (and camera).



7–4 Kinetic Energy and the Work-Energy Principle

Energy is one of the most important concepts in science. Yet we cannot give a simple general definition of energy in only a few words. Nonetheless, each specific type of energy can be defined fairly simply. In this Chapter we define translational kinetic energy; in the next Chapter, we take up potential energy. In later Chapters we will examine other types of energy, such as that related to heat and electricity. The crucial aspect of energy is that the sum of all types, the *total energy*, is the same after any process as it was before: that is, energy is a conserved quantity.

For the purposes of this Chapter, we can define energy in the traditional way as "the ability to do work." This simple definition is not always applicable,[†] but it is valid for mechanical energy which we discuss in this Chapter and the next. We now define and discuss one of the basic types of energy, kinetic energy.

A moving object can do work on another object it strikes. A flying cannonball does work on a brick wall it knocks down; a moving hammer does work on a nail it drives into wood. In either case, a moving object exerts a force on a second object which undergoes a displacement. An object in motion has the ability to do work and thus can be said to have energy. The energy of motion is called **kinetic energy**, from the Greek word *kinetikos*, meaning "motion."

To obtain a quantitative definition for kinetic energy, let us consider a simple rigid object of mass m (treated as a particle) that is moving in a straight line with an initial speed v_1. To accelerate it uniformly to a speed v_2, a constant net force F_{net} is exerted on it parallel to its motion over a displacement d, Fig. 7–14.

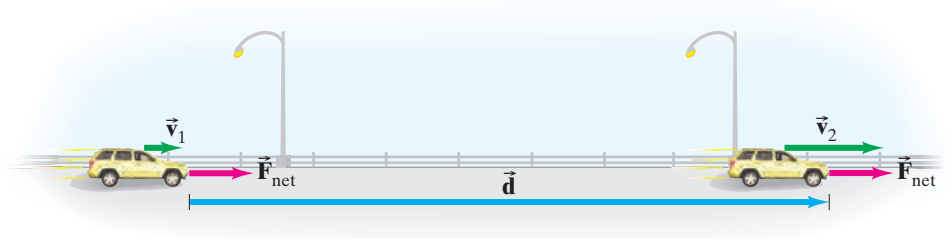


FIGURE 7–14 A constant net force F_{net} accelerates a car from speed v_1 to speed v_2 over a displacement d. The net work done is $W_{\text{net}} = F_{\text{net}}\, d$.

Then the net work done on the object is $W_{\text{net}} = F_{\text{net}}\, d$. We apply Newton's second law, $F_{\text{net}} = ma$, with the acceleration a assumed constant, and use Eq. 2–12c $\left(v_2^2 = v_1^2 + 2ad\right)$ rewritten as

$$a = \frac{v_2^2 - v_1^2}{2d},$$

where v_1 is the initial speed and v_2 the final speed. Substituting this into $F_{\text{net}} = ma$, we determine the work done:

$$W_{\text{net}} = F_{\text{net}}\, d = mad = m\left(\frac{v_2^2 - v_1^2}{2d}\right)d = m\left(\frac{v_2^2 - v_1^2}{2}\right)$$

or

$$W_{\text{net}} = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2. \qquad \textbf{(7–9)}$$

We *define* the quantity $\frac{1}{2}mv^2$ to be the **translational kinetic energy**, K, of the object:

$$K = \tfrac{1}{2}mv^2. \qquad \textbf{(7–10)}$$

Kinetic energy (defined)

We often just call $\frac{1}{2}mv^2$ the "kinetic energy" except when we want to distinguish it from rotational kinetic energy, which we discuss in Chapter 10.

Equation 7–9, derived here for one-dimensional motion with a constant force, is valid in general for translational motion of an object in three dimensions and even if the force varies, as we will show at the end of this Section.

[†]Energy associated with heat is often not available to do work, as we will discuss in Chapter 20.

We can rewrite Eq. 7–9 as:

$$W_{\text{net}} = K_2 - K_1$$

or

WORK-ENERGY PRINCIPLE

$$W_{\text{net}} = \Delta K = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2. \qquad \textbf{(7–11)}$$

Equation 7–11 (or Eq. 7–9) is a useful result known as the **work-energy principle**. It can be stated in words:

WORK-ENERGY PRINCIPLE

> **The net work done on an object is equal to the change in the object's kinetic energy.**

CAUTION

Work-energy valid only for net *work*

Notice that we made use of Newton's second law, $F_{\text{net}} = ma$, where F_{net} is the *net* force—the sum of all forces acting on the object. Thus, the work-energy principle is valid only if W is the *net work* done on the object—that is, the work done by all forces acting on the object.

The work-energy principle is a very useful reformulation of Newton's laws. It tells us that if (positive) net work W is done on an object, the object's kinetic energy increases by an amount W. The principle also holds true for the reverse situation: if the net work W done on an object is negative, the object's kinetic energy decreases by an amount W. That is, a net force exerted on an object opposite to the object's direction of motion decreases its speed and its kinetic energy. An example is a moving hammer (Fig. 7–15) striking a nail. The net force on the hammer ($-\vec{\mathbf{F}}$ in Fig. 7–15, where $\vec{\mathbf{F}}$ is assumed constant for simplicity) acts toward the left, whereas the displacement $\vec{\mathbf{d}}$ of the hammer is toward the right. So the net work done on the hammer, $W_{\text{h}} = (F)(d)(\cos 180°) = -Fd$, is negative and the hammer's kinetic energy decreases (usually to zero).

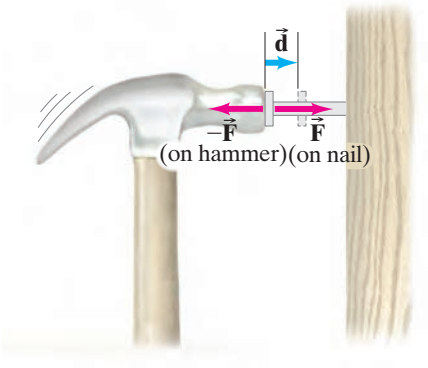


FIGURE 7–15 A moving hammer strikes a nail and comes to rest. The hammer exerts a force F on the nail; the nail exerts a force $-F$ on the hammer (Newton's third law). The work done on the nail by the hammer is positive $(W_{\text{n}} = Fd > 0)$. The work done on the hammer by the nail is negative $(W_{\text{h}} = -Fd)$.

Figure 7–15 also illustrates how energy can be considered the ability to do work. The hammer, as it slows down, does positive work on the nail: $W_{\text{n}} = (+F)(+d)(\cos 0°) = Fd$ and is positive. The decrease in kinetic energy of the hammer $(= Fd$ by Eq. 7–11) is equal to the work the hammer can do on another object, the nail in this case.

The translational kinetic energy $\left(= \tfrac{1}{2}mv^2\right)$ is directly proportional to the mass of the object, and it is also proportional to the *square* of the speed. Thus, if the mass is doubled, the kinetic energy is doubled. But if the speed is doubled, the object has four times as much kinetic energy and is therefore capable of doing four times as much work.

Because of the direct connection between work and kinetic energy, energy is measured in the same units as work: joules in SI units. [The energy unit is ergs in the cgs, and foot-pounds in the British system.] Like work, kinetic energy is a scalar quantity. The kinetic energy of a group of objects is the sum of the kinetic energies of the individual objects.

The work-energy principle can be applied to a particle, and also to an object that can be approximated as a particle, such as an object that is rigid or whose internal motions are insignificant. It is very useful in simple situations, as we will see in the following Examples. For now we assume the object undergoes only translational motion—no rotation.

EXAMPLE 7–7 **Kinetic energy and work done on a baseball.** A 145-g baseball is thrown so that it acquires a speed of 25 m/s. (*a*) What is its kinetic energy? (*b*) What was the net work done on the ball to make it reach this speed, if it started from rest?

APPROACH We use $K = \tfrac{1}{2}mv^2$, and the work-energy principle, Eq. 7–11.

SOLUTION (*a*) The kinetic energy of the ball after the throw is

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}(0.145\text{ kg})(25\text{ m/s})^2 = 45\text{ J}.$$

(*b*) Since the initial kinetic energy was zero, the net work done is just equal to the final kinetic energy, 45 J. (We ignored any rotation.)

EXAMPLE 7–8 **ESTIMATE** **Work on a car, to increase its kinetic energy.** How much net work is required to accelerate a 1000-kg car from 20 m/s to 30 m/s (Fig. 7–16)?

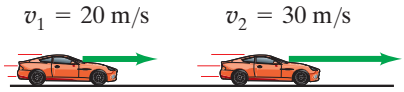


FIGURE 7–16 Example 7–8.

APPROACH A car is a complex system. The engine turns the wheels and tires which push against the ground, and the ground pushes back on the tires (see Example 4–4). We aren't interested right now in those complications. Instead, we can get a useful result using the work-energy principle, but only if we model the car as a particle or simple rigid object.

SOLUTION The net work needed is equal to the increase in kinetic energy:

$$W = K_2 - K_1 = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2$$
$$= \tfrac{1}{2}(1000\,\text{kg})(30\,\text{m/s})^2 - \tfrac{1}{2}(1000\,\text{kg})(20\,\text{m/s})^2 = 2.5 \times 10^5\,\text{J}.$$

EXERCISE C (*a*) Make a guess: will the work needed to accelerate the car in Example 7–8 from rest to 20 m/s be more than, less than, or equal to the work already calculated to accelerate it from 20 m/s to 30 m/s? (*b*) Make the calculation.

CONCEPTUAL EXAMPLE 7–9 **Work to stop a car.** A car traveling 60 km/h can brake to a stop within a distance d of 20 m (Fig. 7–17a). If the car is going twice as fast, 120 km/h, what is its stopping distance (Fig. 7–17b)? Assume the maximum braking force is approximately independent of speed.

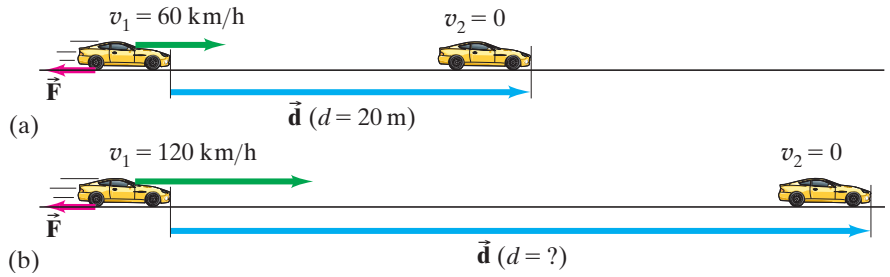


FIGURE 7–17 Example 7–9.

RESPONSE Again we model the car as if it were a particle. Because the net stopping force F is approximately constant, the work needed to stop the car, Fd, is proportional to the distance traveled. We apply the work-energy principle, noting that $\vec{\mathbf{F}}$ and $\vec{\mathbf{d}}$ are in opposite directions and that the final speed of the car is zero:

$$W_{\text{net}} = Fd\cos 180^\circ = -Fd.$$

Then

$$-Fd = \Delta K = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2$$
$$= 0 - \tfrac{1}{2}mv_1^2.$$

Thus, since the force and mass are constant, we see that the stopping distance, d, increases with the square of the speed:

$$d \propto v^2.$$

If the car's initial speed is doubled, the stopping distance is $(2)^2 = 4$ times as great, or 80 m.

PHYSICS APPLIED
Car's stopping distance $\propto$ to initial speed squared

EXERCISE D Can kinetic energy ever be negative?

EXERCISE E (*a*) If the kinetic energy of a baseball is doubled, by what factor has its speed increased? (*b*) If its speed is doubled, by what factor does its kinetic energy increase?

EXAMPLE 7–10 **A compressed spring.** A horizontal spring has spring constant $k = 360\ \text{N/m}$. (*a*) How much work is required to compress it from its uncompressed length ($x = 0$) to $x = 11.0\ \text{cm}$? (*b*) If a 1.85-kg block is placed against this compressed spring and the spring is released, what will be the speed of the block when it separates from the spring at $x = 0$? Ignore friction. (*c*) Repeat part (*b*) but assume that the block is moving on a table as in Fig. 7–18 and that some kind of constant drag force $F_D = 7.0\ \text{N}$ is acting to slow it down, such as friction (or perhaps your finger).

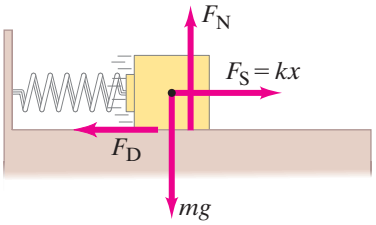


FIGURE 7–18 Example 7–10.

APPROACH We use our result from Section 7–3 that the net work, W, needed to stretch or compress a spring by a distance x is $W = \frac{1}{2}kx^2$. In (*b*) and (*c*) we use the work-energy principle.

SOLUTION (*a*) The work needed to compress the spring a distance $x = 0.110\ \text{m}$ is

$$W = \tfrac{1}{2}(360\ \text{N/m})(0.110\ \text{m})^2 = 2.18\ \text{J},$$

where we have converted all units to SI.

(*b*) In returning to its uncompressed length (from $x = 11.0\ \text{cm}$ to $x = 0$), the spring does 2.18 J of work on the block (same calculation as in part (*a*), only in reverse). According to the work-energy principle, the block acquires kinetic energy of 2.18 J. Since $K = \frac{1}{2}mv^2$, the block's speed must be

$$v = \sqrt{\frac{2K}{m}}$$
$$= \sqrt{\frac{2(2.18\ \text{J})}{1.85\ \text{kg}}} = 1.54\ \text{m/s}.$$

(*c*) There are two forces on the block: that exerted by the spring and that exerted by the drag force, $\vec{\mathbf{F}}_D$. Work done by a force such as friction is complicated. For one thing, heat (or, rather, "thermal energy") is produced—try rubbing your hands together. Nonetheless, the product $\vec{\mathbf{F}}_D \cdot \vec{\mathbf{d}}$ for the drag force, even when it is friction, can be used in the work-energy principle to give correct results for a particle-like object. The spring does 2.18 J of work on the block. The work done by the friction or drag force on the block, in the negative x direction, is

$$W_D = -F_D x = -(7.0\ \text{N})(0.110\ \text{m}) = -0.77\ \text{J}.$$

This work is negative because the drag force acts in the direction opposite to the displacement x. The net work done on the block is $W_{\text{net}} = 2.18\ \text{J} - 0.77\ \text{J} = 1.41\ \text{J}$. From the work-energy principle, Eq. 7–11 (with $v_2 = v$ and $v_1 = 0$), we have

$$v = \sqrt{\frac{2W_{\text{net}}}{m}}$$
$$= \sqrt{\frac{2(1.41\ \text{J})}{1.85\ \text{kg}}} = 1.23\ \text{m/s}$$

for the block's speed at the moment it separates from the spring ($x = 0$).

General Derivation of the Work-Energy Principle

We derived the work-energy principle, Eq. 7–11, for motion in one dimension with a constant force. It is valid even if the force is variable and the motion is in two or three dimensions, as we now show. Suppose the net force $\vec{\mathbf{F}}_{\text{net}}$ on a particle varies in both magnitude and direction, and the path of the particle is a curve as in Fig. 7–8. The net force may be considered to be a function of ℓ, the distance along the curve. The net work done is (Eq. 7–6):

$$W_{\text{net}} = \int \vec{\mathbf{F}}_{\text{net}} \cdot d\vec{\boldsymbol{\ell}} = \int F_{\text{net}} \cos\theta \, d\ell = \int F_{\parallel} \, d\ell,$$

where $F_{\parallel}$ represents the component of the net force parallel to the curve at any point. By Newton's second law,

$$F_{\parallel} = ma_{\parallel} = m\frac{dv}{dt},$$

where $a_{\parallel}$, the component of acceleration parallel to the curve at any point, is equal to the rate of change of speed, dv/dt. We can think of v as a function of ℓ, and using the chain rule for derivatives, we have

$$\frac{dv}{dt} = \frac{dv}{d\ell}\frac{d\ell}{dt} = \frac{dv}{d\ell}v,$$

since $d\ell/dt$ is the speed v. Thus (letting 1 and 2 refer to the initial and final quantities, respectively):

$$W_{\text{net}} = \int_1^2 F_{\parallel} d\ell = \int_1^2 m\frac{dv}{dt} d\ell = \int_1^2 mv\frac{dv}{d\ell} d\ell = \int_1^2 mv \, dv,$$

which integrates to

$$W_{\text{net}} = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2 = \Delta K.$$

This is again the work-energy principle, which we have now derived for motion in three dimensions with a variable net force, using the definitions of work and kinetic energy plus Newton's second law.

Notice in this derivation that only the component of $\vec{\mathbf{F}}_{\text{net}}$ parallel to the motion, $F_{\parallel}$, contributes to the work. Indeed, a force (or component of a force) acting perpendicular to the velocity vector does no work. Such a force changes only the direction of the velocity. It does not affect the magnitude of the velocity. One example of this is uniform circular motion in which an object moving with constant speed in a circle has a ("centripetal") force acting on it toward the center of the circle. This force does no work on the object, because (as we saw in Example 7–3) it is always perpendicular to the object's displacement $d\vec{\boldsymbol{\ell}}$.

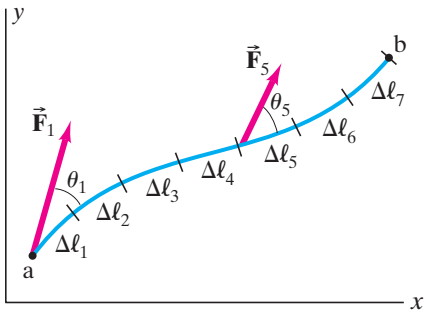


FIGURE 7–8 (Repeated.) A particle acted on by a variable force, $\vec{\mathbf{F}}$, moves along the path shown from point a to point b.

Summary

Work is done on an object by a force when the object moves through a distance, d. If the direction of a constant force $\vec{\mathbf{F}}$ makes an angle θ with the displacement $\vec{\mathbf{d}}$ of the object, the work done by this force is

$$W = Fd\cos\theta = \vec{\mathbf{F}} \cdot \vec{\mathbf{d}}, \qquad \textbf{(7–1, 7–3)}$$

where θ is the angle between $\vec{\mathbf{F}}$ and $\vec{\mathbf{d}}$.

The last expression is called the scalar product of $\vec{\mathbf{F}}$ and $\vec{\mathbf{d}}$. In general, the **scalar product** of any two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ is defined as

$$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = AB\cos\theta \qquad \textbf{(7–2)}$$

(also called the **dot product**) where θ is the angle between $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$. In rectangular coordinates we can also write

$$\vec{\mathbf{A}} \cdot \vec{\mathbf{B}} = A_x B_x + A_y B_y + A_z B_z. \qquad \textbf{(7–4)}$$

The work W done by a variable force $\vec{\mathbf{F}}$ on an object that moves from point a to point b is

$$W = \int_a^b \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}} = \int_a^b F\cos\theta \, d\ell, \qquad \textbf{(7–7)}$$

where $d\vec{\boldsymbol{\ell}}$ represents an infinitesimal displacement along the path of the object and θ is the angle between $d\vec{\boldsymbol{\ell}}$ and $\vec{\mathbf{F}}$ at each point of the object's path.

The restoring force exerted by a spring stretched or compressed a distance x is $F_S = -kx$, where k is the **spring constant**. The work done to compress or stretch a spring a distance x is $W = \frac{1}{2}kx^2$.

The translational **kinetic energy**, K, of an object of mass m moving with speed v is defined to be

$$K = \tfrac{1}{2}mv^2. \qquad \textbf{(7–10)}$$

The **work-energy principle** states that the net work done on an object (by the net force) equals the change in kinetic energy of the object:

$$W_{\text{net}} = \Delta K = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2. \qquad \textbf{(7–11)}$$

Questions

1. In what ways is the word "work" as used in everyday language the same as defined in physics? In what ways is it different? Give examples of both.
2. A woman swimming upstream is not moving with respect to the shore. Is she doing any work? If she stops swimming and merely floats, is work done on her?
3. Can a centripetal force ever do work on an object? Explain.
4. Why is it tiring to push hard against a solid wall even though you are doing no work on the wall?
5. Does the scalar product of two vectors depend on the choice of coordinate system?
6. Can a dot product ever be negative? If yes, under what conditions?
7. Give an example of friction doing positive work.
8. Does the dot product of two vectors have direction as well as magnitude?
9. Can the normal force on an object ever do work? Explain.
10. You have two springs that are identical except that spring 1 is stiffer than spring 2 $(k_1 > k_2)$. On which spring is more work done: (*a*) if the two springs are compressed using the same force; (*b*) if they are compressed the same distance?
11. If the speed of a particle doubles, by what factor does its kinetic energy increase?
12. In Example 7–10, it was stated that the block separates from the compressed spring when the spring reaches its equilibrium length $(x = 0)$. Explain why separation doesn't take place before (or after) this point.
13. Does the net work done on a particle depend on the choice of reference frame? How does this affect the work-energy principle?

MisConceptual Questions

1. You are carrying a 10-kg bag and moving at constant speed. In which case will you do the most work on the bag?
   (*a*) Walk horizontally for 5 m.
   (*b*) Run horizontally for 5 m.
   (*c*) Climb up a 5-m-long slope.
   (*d*) Climb up a 5-m-tall slope.
2. You push very hard on a car with a dead battery, trying to move it. You do work on the car:
   (*a*) only if it starts moving.
   (*b*) only if it doesn't move.
   (*c*) whether or not it moves, as long as you are exerting a force.
   (*d*) never—it does work on you.
   (*e*) None of the above.
3. A man holds a 1.00-kg box above his head for 2 minutes. The work done by the man during these 2 minutes is
   (*a*) 60 J.
   (*b*) 120 J.
   (*c*) 19.6 J.
   (*d*) 9.8 J.
   (*e*) 0 J.
4. You push a heavy crate *down* a ramp at a constant velocity. Only four forces act on the crate. Which force does the greatest magnitude of work on the crate?
   (*a*) The force of friction.
   (*b*) The force of gravity.
   (*c*) The normal force.
   (*d*) The force of you pushing.
   (*e*) The net force.
5. If you push twice as hard against a stationary brick wall, the amount of work you do
   (*a*) quadruples.
   (*b*) doubles.
   (*c*) is cut in half.
   (*d*) is zero.
   (*e*) remains constant but non-zero.
6. A *satellite* in circular orbit around the Earth moves at constant speed. This orbit is maintained by the force of gravity between the Earth and the satellite, yet no work is done on the satellite. How is this possible?
   (*a*) No work is done if there is no contact between objects.
   (*b*) No work is done because there is no gravity in space.
   (*c*) No work is done if the direction of motion is perpendicular to the force.
   (*d*) No work is done whenever objects move in a circle.
7. A delivery man carrying a package (Fig. 7–19) walks up the stairs to the second floor at constant speed (A), and along the hall at a constant speed (B). He accelerates to a run (C) and then (D) moves at a greater constant speed along the hall. During what portions of his motion is the delivery man doing work on the package? (Ignore friction.)
   (*a*) A only.
   (*b*) C only.
   (*c*) A and C only.
   (*d*) B and C only.
   (*e*) None of these portions.
   (*f*) Work is done throughout.

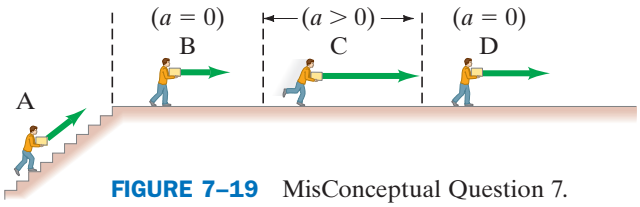


FIGURE 7–19 MisConceptual Question 7.

8. You and a friend decide to go sledding and climb to the top of a hill. There are two paths: a short steep path to the bottom, and a longer, gentle path to the bottom. If you go down the steep path and your friend takes the gentler path, who has the greater speed at the bottom? Ignore friction.
   (*a*) You do.
   (*b*) Your friend does.
   (*c*) You both have the same speed.
   (*d*) The one who is heavier does.

9. Which of the following statements about work and energy are true?
   (*a*) Kinetic energy and work can only be positive.
   (*b*) Kinetic energy can only be positive; work can be positive or negative.
   (*c*) Both kinetic energy and work can be positive or negative.

10. An old truck moving at 50 km/h skids to a stop with locked brakes over a distance d. Over what distance will the truck skid to a stop with locked brakes if it is traveling at 150 km/h?
   (*a*) d.
   (*b*) $3d$.
   (*c*) Larger than $3d$.

11. A car accelerates from rest to 30 km/h. Later, on a highway it accelerates from 30 km/h to 60 km/h. Which takes more energy, going from 0 to 30, or from 30 to 60?
   (*a*) 0 to 30 km/h.
   (*b*) 30 to 60 km/h.
   (*c*) Both are the same.

12. Two blocks of mass m_1 and m_2 ($m_1 > m_2$) slide on a frictionless floor and have the same kinetic energy when they hit a long rough stretch ($\mu > 0$), which slows them down to a stop. Which one goes farther?
   (*a*) m_1.
   (*b*) m_2.
   (*c*) They will go the same distance.
   (*d*) Can't tell without knowing the initial speed of each.

13. Two stones, one twice the mass of the other, are dropped from a cliff. Just before hitting the ground, what is the kinetic energy of the heavy stone compared to the light one?
   (*a*) 1/4 as much.
   (*b*) 1/2 as much.
   (*c*) The same.
   (*d*) 2× as much.
   (*e*) 4× as much.

14. If an object has kinetic energy, then it must be
   (*a*) at an elevated position.
   (*b*) falling.
   (*c*) moving.
   (*d*) at rest.

Problems

7–1 Work, Constant Force

1. (I) How much work is done by the gravitational force when a 280-kg pile driver falls 3.80 m?
2. (I) How high will a 1.85-kg rock go if thrown straight up by someone who does 70.0 J of work on it? Neglect air resistance.
3. (I) A 55.0-kg firefighter climbs a flight of stairs 28.0 m high at constant speed. How much work does she do?
4. (I) The head of a hammer with a mass of 1.2 kg is allowed to fall onto a nail from a height of 0.65 m. What is the maximum amount of work it could do on the nail? Why do people not just "let it fall" but add their own force to the hammer as it falls?
5. (II) How much work would movers do pushing a 46.0-kg crate 10.3 m horizontally across a rough floor without acceleration, if the effective coefficient of friction was 0.40?
6. (II) A 1200-N crate rests on the floor. How much work is required to move it at constant speed (*a*) 6.0 m along the floor against a friction force of 230 N, and (*b*) 6.0 m vertically?
7. (II) (*a*) What magnitude force is required to give a helicopter of mass M an acceleration of $0.10g$ upward? (*b*) What work is done by this force as the helicopter moves a distance h upward?
8. (II) What is the minimum work needed to push a 950-kg car 510 m up along a 9.0° incline? Ignore friction.
9. (II) Estimate the work you do to mow a lawn 10 m by 20 m with a 50-cm-wide mower. Assume you push with a horizontal force of about 15 N.
10. (II) In a certain library the first shelf is 15.0 cm off the ground, and the remaining four shelves are each spaced 38.0 cm above the previous one. If the average book has a mass of 1.25 kg with a height of 22.0 cm, and an average shelf holds 28 books (standing vertically), how much work is required to fill all the shelves, assuming the books are all laying flat on the floor to start?
11. (II) A **lever** such as that shown in Fig. 7–20 can be used to lift objects we might not otherwise be able to lift. Show that the ratio of output force, F_O, to input force, F_I, is related to the lengths ℓ_I and ℓ_O from the pivot by $F_O/F_I = \ell_I/\ell_O$. Ignore friction and the mass of the lever, and assume the work output equals the work input.

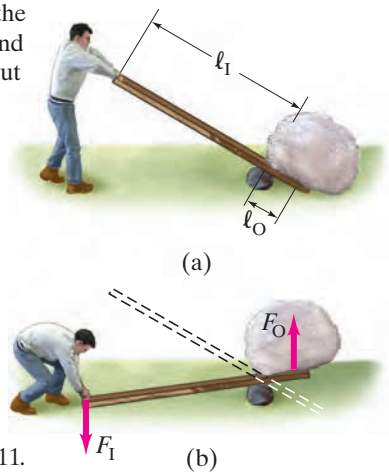


FIGURE 7–20
A lever. Problem 11.

12. (II) A box of mass 4.0 kg is accelerated from rest by a force across a floor at a rate of 2.0 m/s^2 for 8.0 s. Find the net work done on the box.
13. (II) Eight books, each 4.0 cm thick with mass 1.6 kg, lie flat on a table. How much work is required to stack them one on top of another?
14. (II) A *gondola* can carry 20 skiers, with a total mass of up to 2250 kg. The gondola ascends at a constant speed from the base of a mountain, at 2150 m, to the summit at 3625 m. (*a*) How much work does the motor do in moving a full gondola up the mountain? (*b*) How much work does gravity do on the gondola? (*c*) If the motor could generate 10% more work than found in (*a*), what is the acceleration of the gondola?
15. (II) A grocery cart with mass of 16 kg is pushed at constant speed up a 12° ramp by a force F_P which acts at an angle of 17° below the horizontal. Find the work done by each of the forces $(m\vec{\mathbf{g}}, \vec{\mathbf{F}}_N, \vec{\mathbf{F}}_P)$ on the cart if the ramp is 7.5 m long.

16. (II) A 380-kg piano slides 4.6 m down a 30° incline and is kept from accelerating by a man who is pushing back on it *parallel to the incline* (Fig. 7–21). Determine: (*a*) the force exerted by the man, (*b*) the work done on the piano by the man, (*c*) the work done on the piano by the force of gravity, and (*d*) the net work done on the piano. Ignore friction.



FIGURE 7–21
Problem 16.

17. (II) Recall from Chapter 4, Example 4–14, that you can use a *pulley* and ropes to decrease the force needed to raise a heavy load (see Fig. 7–22). But for every meter the load is raised, how much rope must be pulled up? Explain.

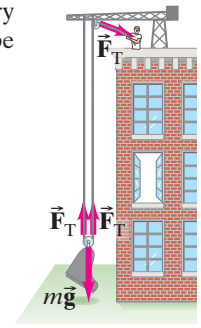


FIGURE 7–22
Problem 17.

7–2 Scalar Product

18. (I) What is the dot product of $\vec{\mathbf{A}} = 2.0x^2\hat{\mathbf{i}} - 4.0x\,\hat{\mathbf{j}} + 5.0\hat{\mathbf{k}}$ and $\vec{\mathbf{B}} = 11.0\,\hat{\mathbf{i}} + 2.5x\,\hat{\mathbf{j}}$?

19. (I) For any vector $\vec{\mathbf{V}} = V_x\hat{\mathbf{i}} + V_y\hat{\mathbf{j}} + V_z\hat{\mathbf{k}}$ show that

$$V_x = \hat{\mathbf{i}}\cdot\vec{\mathbf{V}}, \quad V_y = \hat{\mathbf{j}}\cdot\vec{\mathbf{V}}, \quad V_z = \hat{\mathbf{k}}\cdot\vec{\mathbf{V}}.$$

20. (I) Calculate the angle between the vectors:

$$\vec{\mathbf{A}} = 5.8\hat{\mathbf{i}} - 3.4\,\hat{\mathbf{j}} - 6.2\hat{\mathbf{k}} \text{ and } \vec{\mathbf{B}} = 8.2\hat{\mathbf{i}} + 4.3\,\hat{\mathbf{j}} - 7.0\hat{\mathbf{k}}.$$

21. (I) Show that $\vec{\mathbf{A}}\cdot(-\vec{\mathbf{B}}) = -\vec{\mathbf{A}}\cdot\vec{\mathbf{B}}$.

22. (I) Vector $\vec{\mathbf{V}}_1$ points along the z axis and has magnitude $V_1 = 75$. Vector $\vec{\mathbf{V}}_2$ lies in the xz plane, has magnitude $V_2 = 48$, and makes a $-48°$ angle with the x axis (points below the x axis). What is the scalar product $\vec{\mathbf{V}}_1\cdot\vec{\mathbf{V}}_2$?

23. (II) Given the vector $\vec{\mathbf{A}} = 3.0\hat{\mathbf{i}} + 1.8\,\hat{\mathbf{j}}$, find a vector $\vec{\mathbf{B}}$ in the xy plane that is perpendicular to $\vec{\mathbf{A}}$.

24. (II) A constant force $\vec{\mathbf{F}} = (2.0\hat{\mathbf{i}} + 4.0\,\hat{\mathbf{j}})$ N acts on an object as it moves along a straight-line path. If the object's displacement is $\vec{\mathbf{d}} = (1.0\hat{\mathbf{i}} + 5.0\,\hat{\mathbf{j}})$ m, calculate the work done by $\vec{\mathbf{F}}$ using these alternate ways of writing the dot product: (*a*) $W = Fd\cos\theta$; (*b*) $W = F_x d_x + F_y d_y$.

25. (II) If $\vec{\mathbf{A}} = 9.0\hat{\mathbf{i}} - 8.5\,\hat{\mathbf{j}}$, $\vec{\mathbf{B}} = -8.0\hat{\mathbf{i}} + 6.1\,\hat{\mathbf{j}} + 4.2\hat{\mathbf{k}}$, and $\vec{\mathbf{C}} = 5.8\hat{\mathbf{i}} - 9.2\,\hat{\mathbf{j}}$, determine (*a*) $\vec{\mathbf{A}}\cdot(\vec{\mathbf{B}}+\vec{\mathbf{C}})$; (*b*) $(\vec{\mathbf{A}}+\vec{\mathbf{C}})\cdot\vec{\mathbf{B}}$; (*c*) $(\vec{\mathbf{B}}+\vec{\mathbf{A}})\cdot\vec{\mathbf{C}}$.

26. (II) In Section 7–2, there are two different expressions given for the dot product—Equations 7–2 and 7–4. For vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{C}}$ in Problem 25, show that the two different expressions give the same result. You will have to find the magnitude and direction of both of the vectors.

27. (II) Show that if two nonparallel vectors have the same magnitude, their sum must be perpendicular to their difference.

28. (II) Given vectors $\vec{\mathbf{A}} = -4.8\hat{\mathbf{i}} + 6.8\,\hat{\mathbf{j}}$ and $\vec{\mathbf{B}} = 9.6\hat{\mathbf{i}} + 6.7\,\hat{\mathbf{j}}$, determine the vector $\vec{\mathbf{C}}$ that lies in the xy plane, is perpendicular to $\vec{\mathbf{B}}$, and whose dot product with $\vec{\mathbf{A}}$ is 18.0.

29. (II) Let $\vec{\mathbf{V}} = 20.0\hat{\mathbf{i}} + 26.0\,\hat{\mathbf{j}} - 14.0\hat{\mathbf{k}}$. What angles does this vector make with the x, y, and z axes?

30. (II) Use the scalar product to prove the *law of cosines* for a triangle:

$$c^2 = a^2 + b^2 - 2ab\cos\theta,$$

where a, b, and c are the lengths of the sides of a triangle and θ is the angle opposite side c.

31. (II) $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ are two vectors in the xy plane that make angles α and β with the x axis respectively. Evaluate the scalar product of $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ and deduce the following trigonometric identity: $\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$.

32. (II) Suppose $\vec{\mathbf{A}} = 1.0\hat{\mathbf{i}} + 1.0\,\hat{\mathbf{j}} - 2.0\hat{\mathbf{k}}$ and $\vec{\mathbf{B}} = -1.0\hat{\mathbf{i}} + 1.0\,\hat{\mathbf{j}} - 2.0\hat{\mathbf{k}}$. What is the angle between these two vectors? Explain the significance of the sign in your result.

33. (III) Show that the scalar product of two vectors is *distributive*: $\vec{\mathbf{A}}\cdot(\vec{\mathbf{B}}+\vec{\mathbf{C}}) = \vec{\mathbf{A}}\cdot\vec{\mathbf{B}} + \vec{\mathbf{A}}\cdot\vec{\mathbf{C}}$ [*Hint*: Use a diagram showing all three vectors in a plane and indicate dot products on the diagram.]

34. (III) Find a vector of unit length in the xy plane that is perpendicular to $4.0\hat{\mathbf{i}} + 3.0\,\hat{\mathbf{j}}$.

35. (III) The *arrangement of atoms* in zinc is an example of "hexagonal close-packed" structure. Three of the nearest neighbors are found at the following (x, y, z) coordinates, given in nanometers (10^{-9} m): atom 1 is at (0, 0, 0); atom 2 is at (0.230, 0.133, 0); atom 3 is at (0.077, 0.133, 0.247). Find the angle between two vectors: one that connects atom 1 with atom 2 and another that connects atom 1 with atom 3.

7–3 Work, Varying Force

36. (I) In pedaling a bicycle uphill, a cyclist exerts a downward force of 420 N during each stroke. If the diameter of the circle traced by each pedal is 36 cm, calculate how much work is done in each stroke.

37. (II) A spring has $k = 65$ N/m. Draw a graph like that in Fig. 7–11 and use it to determine the work needed to stretch the spring from $x = 3.0$ cm to $x = 7.5$ cm, where $x = 0$ refers to the spring's unstretched length.

38. (II) If the hill in Example 7–2 (Fig. 7–4) was not an even slope but rather an irregular curve as in Fig. 7–23, show that the same result would be obtained as in Example 7–2: namely, that the work done by gravity depends only on the height of the hill and not on its shape or the path taken.

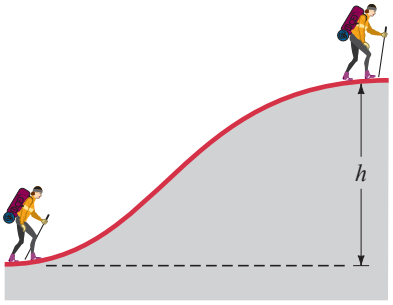


FIGURE 7–23
Problem 38.

39. (II) The net force exerted on a particle acts in the positive x direction. Its magnitude increases linearly from zero at $x = 0$, to 380 N at $x = 3.0$ m. It remains constant at 380 N from $x = 3.0$ m to $x = 7.0$ m, and then decreases linearly to zero at $x = 12.0$ m. Determine the work done to move the particle from $x = 0$ to $x = 12.0$ m graphically, by determining the area under the F_x versus x graph.

40. (II) In Fig. 7–9 assume the distance axis is the x axis and that a = 10.0 m and b = 30.0 m. Estimate the work done by this force in moving a 3.50-kg object from a to b.

41. (II) The force on a particle, acting along the x axis, varies as shown in Fig. 7–24. Determine the work done by this force to move the particle along the x axis: (a) from $x = 0.0$ to $x = 10.0$ m; (b) from $x = 0.0$ to $x = 15.0$ m.

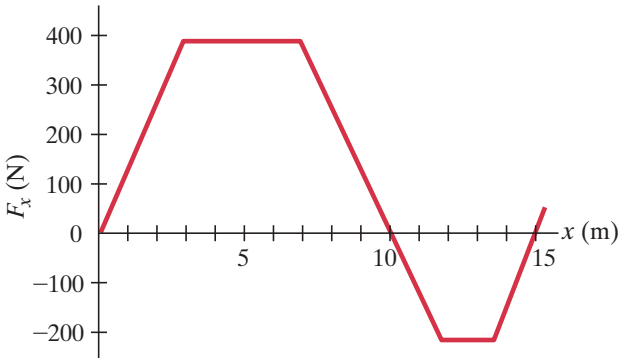


FIGURE 7–24 Problem 41.

42. (II) A child is pulling a wagon down the sidewalk. For 5.0 m the wagon stays on the sidewalk and the child pulls with a horizontal force of 22 N. Then one wheel of the wagon goes off onto the grass so the child has to pull with a horizontal force of 38 N at an angle of 12° to the side for the next 3.0 m. Finally the wagon gets back on the sidewalk so the child makes the rest of the trip, 8.5 m, with a force of 22 N. How much total work did the child do on the wagon?

43. (II) The resistance of a packing material to a sharp object penetrating it is a force proportional to the fourth power of the penetration depth x; that is, $\vec{\mathbf{F}} = -bx^4\,\hat{\mathbf{i}}$. Calculate the work done to force a sharp object a distance d into the material.

44. (II) At the top of a *pole vault*, an athlete actually can do work pushing on the pole before releasing it. Suppose the pushing force that the pole exerts back on the athlete is given by $F(y) = (1.5 \times 10^2\ \text{N/m})\,y - (1.9 \times 10^2\ \text{N/m}^2)\,y^2$ acting over a displacement from $y = 0$ (y is vertical) to $y = 0.20$ m. How much work is done on the athlete?

45. (II) Consider a force $F_1 = A/\sqrt{x}$ which acts on an object during its journey along the x axis from $x = 0.0$ to $x = 1.0$ m, where $A = 3.0\ \text{N} \cdot \text{m}^{1/2}$. Show that during this journey, even though F_1 is infinite at $x = 0.0$, the work W done on the object by this force is finite, and determine W.

46. (II) Assume that a force acting on an object is given by $\vec{\mathbf{F}} = ax\,\hat{\mathbf{i}} + by\,\hat{\mathbf{j}}$ where the constants $a = 3.0\ \text{N} \cdot \text{m}^{-1}$ and $b = 4.0\ \text{N} \cdot \text{m}^{-1}$. Determine the work done on the object by this force as it moves in a straight line from the origin to $\vec{\mathbf{r}} = (10.0\,\hat{\mathbf{i}} + 20.0\,\hat{\mathbf{j}})$ m.

47. (II) If it requires 5.0 J of work to stretch a particular spring by 2.0 cm from its equilibrium length, how much more work will be required to stretch it an additional 4.0 cm?

48. (II) An object, moving along the circumference of a circle with radius R, is acted upon by a force of constant magnitude F. The force is directed at all times at a 30° angle with respect to the tangent to the circle as shown in Fig. 7–25. Determine the work done by this force when the object moves along the half circle from A to B.

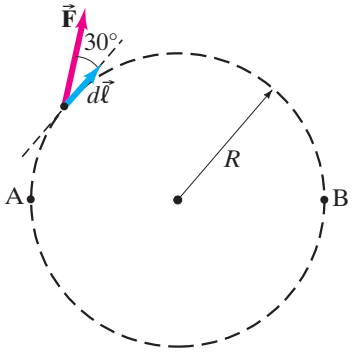


FIGURE 7–25 Problem 48.

49. (II) The net force along the linear path of a particle of mass 480 g has been measured at 10.0-cm intervals, starting at $x = 0.0$, to be 26.0, 28.5, 28.8, 29.6, 32.8, 40.1, 46.6, 42.2, 48.8, 52.6, 55.8, 60.2, 60.6, 58.2, 53.7, 50.3, 45.6, 45.2, 43.2, 38.9, 35.1, 30.8, 27.2, 21.0, 22.2, and 18.6, all in newtons. Determine the total work done on the particle over this entire range.

50. (II) A 17,000-kg jet takes off from an aircraft carrier with the help of a *catapult* (Fig. 7–26a). The gases thrust out from the jet's engines exert a constant force of 130 kN on the jet; the horizontal force exerted on the jet by the steam-powered catapult is plotted in Fig. 7–26b. Determine the work done on the jet: (a) by the gases expelled by its engines during launch of the jet; and (b) by the catapult during launch of the jet.

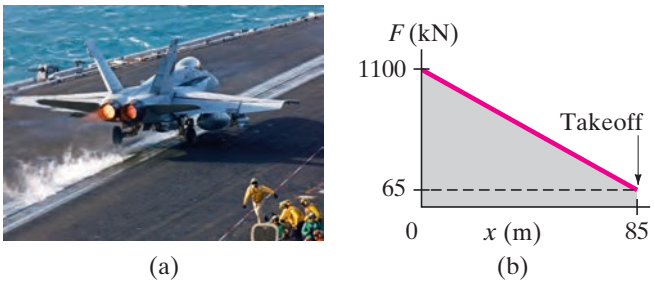


FIGURE 7–26 Problem 50.

51. (III) A spring of spring constant k and negligible mass is attached to a mass m and the other end of the spring is pulled vertically in order to lift the mass. Find an expression for the amount of work that must be done on the spring before the mass begins to leave the ground.

52. (III) A 2800-kg space vehicle, initially at rest, falls vertically from a height of 2900 km above the Earth's surface. Determine how much work is done by the force of gravity in bringing the vehicle to the Earth's surface.

53. (III) A 3.0-m-long steel *chain* is stretched out along the top level of a horizontal scaffold at a construction site, in such a way that 2.0 m of the chain remains on the top level and 1.0 m hangs vertically, Fig. 7–27. At this point, the force on the hanging segment is sufficient to pull the entire chain over the edge. Once the chain is moving, the kinetic friction is so small that it can be neglected. How much work is performed on the chain by the force of gravity as the chain falls from the point where 2.0 m remains on the scaffold to the point where the entire chain has left the scaffold? (Assume that the chain has a linear weight density of 24 N/m.)

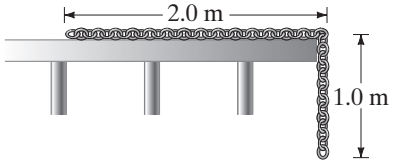


FIGURE 7–27
Problem 53.

7–4 Kinetic Energy; Work-Energy Principle

54. (I) At room temperature, an oxygen molecule, with mass of 5.31×10^{-26} kg, typically has a kinetic energy of about 6.21×10^{-21} J. How fast is it moving?

55. (I) (*a*) If the kinetic energy of a particle is tripled, by what factor has its speed increased? (*b*) If the speed of a particle is halved, by what factor does its kinetic energy change?

56. (I) How much work is required to stop an electron $(m = 9.11 \times 10^{-31}\,\text{kg})$ which is moving with a speed of 1.10×10^{6} m/s?

57. (II) A 1300-kg car moving 20.0 m/s slows to a stop at a steady rate in a distance of 30 m. It then speeds back up to 20.0 m/s at constant acceleration over a distance of 100 m. What is the net work done on the car (*a*) in slowing down? (*b*) in speeding up? (*c*) overall?

58. (II) Two bullets are fired at the same time with the same kinetic energy. If one bullet has twice the mass of the other, which has the greater speed and by what factor? Which can do more work, and by what factor?

59. (II) A baseball $(m = 145\,\text{g})$ traveling 32 m/s moves a fielder's glove backward 22 cm when the ball is caught. What was the average force exerted by the ball on the glove?

60. (II) An 85-g arrow is fired from a bow whose string exerts an average force of 105 N on the arrow over a distance of 75 cm. What is the speed of the arrow as it leaves the bow?

61. (II) If the speed of a car is increased by 50%, by what factor will its minimum braking distance on a flat road be increased, assuming all else is the same? Ignore the driver's reaction time.

62. (II) At an accident scene on a level road, investigators measure a *car's skid* mark to be 78 m long. It was a rainy day and the coefficient of friction was estimated to be 0.30. (*a*) Use these data to determine the speed of the car when the driver slammed on (and locked) the brakes. (*b*) Why does the car's mass not matter? (*c*) What is wrong with a car that skids (see page 131)?

63. (II) You back your 1200-kg car into a large fixed horizontal spring, compressing it by 1.8 m. You stop, and when you take your foot off the brake, your car accelerates and separates from the spring at a speed of 28 km/h. What is the spring constant of the spring?

64. (II) A 3.5-kg object moving in two dimensions initially has a velocity $\vec{\mathbf{v}}_1 = (10.0\,\hat{\mathbf{i}} + 20.0\,\hat{\mathbf{j}})$ m/s. A net force $\vec{\mathbf{F}}$ then acts on the object for 2.0 s, after which the object's velocity is $\vec{\mathbf{v}}_2 = (15.0\,\hat{\mathbf{i}} + 30.0\,\hat{\mathbf{j}})$ m/s. Determine the work done by $\vec{\mathbf{F}}$ on the object.

65. (II) A box is initially moving 5.0 m/s across the floor. It slides to rest due to friction in a distance of 8.0 m. What is the coefficient of friction between the box and the floor?

66. (II) (*a*) How much work is done by the horizontal force $F_P = 150$ N on the 18-kg block of Fig. 7–28 when the force pushes the block 5.0 m up along the 32° frictionless incline? (*b*) How much work is done by the gravitational force on the block during this displacement? (*c*) How much work is done by the normal force? (*d*) What is the speed of the block (assume that it is zero initially) after this displacement?

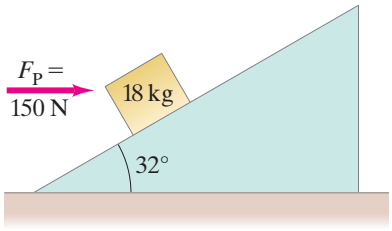


FIGURE 7–28
Problems 66 and 67.

67. (II) Repeat Problem 66 using a coefficient of friction $\mu_k = 0.13$.

68. (II) A car traveling at a velocity v can stop in a minimum distance d. What would be the car's minimum stopping distance if it were traveling at a velocity of $2v$?

69. (II) A train is moving along a track with constant speed v_1 relative to the ground. A person on the train holds a ball of mass m and throws it toward the front of the train with a speed v_2 relative to the train. Calculate the change in kinetic energy of the ball (*a*) in the Earth frame of reference, and (*b*) in the train frame of reference. (*c*) Relative to each frame of reference, how much work was done on the ball? (*d*) Explain why the results in part (*c*) are not the same for the two frames—after all, it's the same ball.

70. (II) A mass m is attached to a spring which is held stretched a distance x by a force F, Fig. 7–29, and then released. The spring pulls the mass to the left, towards its natural equilibrium length. Assuming there is no friction, determine the speed of the mass m when the spring returns: (*a*) to its normal length $(x = 0)$; (*b*) to half its original extension $(x/2)$.

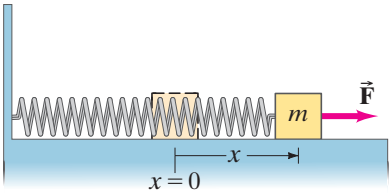


FIGURE 7–29
Problem 70.

71. (II) One car has twice the mass of a second car, but only half as much kinetic energy. When both cars increase their speed by 8.0 m/s, they then have the same kinetic energy. What were the original speeds of the two cars?

72. (II) A 265-kg load is lifted 18.0 m vertically with an acceleration $a = 0.140\,g$ by a single cable. Determine (*a*) the tension in the cable; (*b*) the net work done on the load; (*c*) the work done by the cable on the load; (*d*) the work done by gravity on the load; (*e*) the final speed of the load assuming it started from rest.

73. (III) An elevator cable breaks when a 975-kg elevator is 22.5 m above the top of a huge spring $(k = 8.00 \times 10^4\,\text{N/m})$ at the bottom of the shaft. Calculate (*a*) the work done by gravity on the elevator just before striking the spring; (*b*) the speed of the elevator just before striking the spring; (*c*) the amount the spring compresses (note that here work is done by both the spring and gravity).

General Problems

74. *Spiderman* uses his spider webs to save a runaway train moving about 60 km/h, Fig. 7–30. His web stretches a few city blocks (500 m) before the 10^4-kg train comes to a stop. Assuming the web acts like a spring, estimate the effective spring constant.



FIGURE 7–30 Problem 74.

75. A 36.0-kg crate, starting from rest, is pulled across a floor with a constant horizontal force of 225 N. For the first 11.0 m the floor is frictionless, and for the next 15.0 m the coefficient of friction is 0.20. What is the final speed of the crate after being pulled these 26.0 m?

76. A 75-kg meteorite buries itself 4.5 m into soft mud. The force between the meteorite and the mud is given by $F(x) = (640\ \text{N/m}^3)x^3$, where x is the depth in the mud. What was the speed of the meteorite when it initially impacted the mud?

77. The force required to compress an "*imperfect*" horizontal *spring* (doesn't follow Hooke's law) an amount x is given by $F = 150x + 12x^3$, where x is in meters and F in newtons. If the spring is compressed 2.0 m, what speed will it give to a 3.0-kg ball held against it and then released?

78. A 6.10-kg block is pushed 9.25 m up a smooth 37.0° inclined plane by a horizontal force of 85.0 N. If the initial speed of the block is 3.25 m/s up the plane, calculate (*a*) the initial kinetic energy of the block; (*b*) the work done by the 85.0-N force; (*c*) the work done by gravity; (*d*) the work done by the normal force; (*e*) the final kinetic energy of the block.

79. A varying force is given by $F = Ae^{-kx}$, where x is the position. A and k are constants that have units of N and m^{-1}, respectively. What is the work done when x goes from 0.10 m to infinity?

80. The force needed to hold a particular spring compressed by a fairly large amount x from its normal length is given by $F = kx + ax^3 + bx^4$. How much work must be done to compress this *nonlinear spring* by an amount X, starting from $x = 0$?

81. A force $\vec{\mathbf{F}} = (10.0\,\hat{\mathbf{i}} + 9.0\,\hat{\mathbf{j}} + 12.0\,\hat{\mathbf{k}})\,\text{kN}$ acts on a small object of mass 85 g. If the displacement of the object is $\vec{\mathbf{d}} = (5.0\,\hat{\mathbf{i}} + 4.0\,\hat{\mathbf{j}})\,\text{m}$, find the work done by the force. What is the angle between $\vec{\mathbf{F}}$ and $\vec{\mathbf{d}}$?

82. Two forces, $\vec{\mathbf{F}}_1 = (1.50\,\hat{\mathbf{i}} - 0.80\,\hat{\mathbf{j}} + 0.70\,\hat{\mathbf{k}})\,\text{N}$ and $\vec{\mathbf{F}}_2 = (-0.70\,\hat{\mathbf{i}} + 1.20\,\hat{\mathbf{j}})\,\text{N}$, are applied on a moving object of mass 0.20 kg. The displacement vector of the object while the two forces act is $\vec{\mathbf{d}} = (6.0\,\hat{\mathbf{i}} + 8.0\,\hat{\mathbf{j}} + 5.0\,\hat{\mathbf{k}})\,\text{m}$. What is the work done by the two forces?

83. In the game of *paintball*, players use guns powered by pressurized gas to propel 33-g gel capsules filled with paint at the opposing team. Game rules dictate that a paintball cannot leave the barrel of a gun with a speed greater than 85 m/s. Model the shot by assuming the pressurized gas applies a constant force F to a 33-g capsule over the length of the 32-cm barrel. Determine F by (*a*) using the work-energy principle, and (*b*) using the kinematic equations (Eqs. 2–12) and Newton's second law.

84. A softball having a mass of 0.25 kg is pitched horizontally at 120 km/h. By the time it reaches the plate, it may have slowed by 10%. Neglecting gravity, estimate the average force of air resistance during a pitch, if the distance between the plate and the pitcher is about 15 m.

85. We usually neglect the *mass of a spring* if it is small compared to the mass attached to it. But in some applications, the mass of the spring must be taken into account. Consider a spring of unstretched length ℓ and mass M_S uniformly distributed along the length of the spring. A mass m is attached to the end of the spring. One end of the spring is fixed and the mass m is allowed to vibrate horizontally without friction (Fig. 7–31). Each point on the spring moves with a velocity proportional to the distance from that point to the fixed end. For example, if the mass on the end moves with speed v_0, the midpoint of the spring moves with speed $v_0/2$. Show that the kinetic energy of the mass plus spring when the mass m is moving with velocity v is

$$K = \tfrac{1}{2}Mv^2$$

where $M = m + \frac{1}{3}M_S$ is the "effective mass" of the system. [*Hint*: Let D be the total length of the stretched spring. Then the velocity of an infinitesimal length dx of spring, of mass dM, located at x is $v(x) = v_0(x/D)$. Note also that $dM = dx\,(M_S/D)$.]

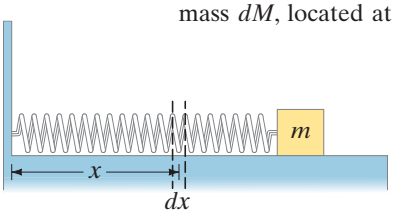


FIGURE 7–31 Problem 85.

86. A 25-g projectile is fired into a cube of ballistic gel at a velocity of 360 m/s. If the projectile penetrates 15 cm into the gel before stopping, find the average force exerted by the gel onto the projectile. Use (*a*) kinematics and dynamics (Newton's laws) and (*b*) the work-energy principle.

87. An airplane pilot fell 370 m after jumping from an aircraft without his parachute opening. He landed in a snowbank, creating a crater 1.1 m deep, but survived with only minor injuries. Assuming the pilot's mass was 82 kg and his terminal velocity was 45 m/s, estimate: (*a*) the work done by the snow in bringing him to rest; (*b*) the average force exerted on him by the snow to stop him; and (*c*) the work done on him by air resistance as he fell. Model him as a particle.

88. Many cars have "3 mi/h (5 km/h) *bumpers*" that are designed to compress and rebound elastically without any physical damage at speeds below 5 km/h. If the material of the bumpers permanently deforms after a compression of 1.5 cm, but remains like an elastic spring up to that point, what must be the effective spring constant of the bumper material, assuming the car has a mass of 1050 kg and is tested by driving it into a solid wall?

89. A package of mass m is placed onto a horizontal conveyor belt moving at speed v (Fig. 7–32). The coefficient of kinetic friction between package and belt is μ_k. (*a*) How long does it take the package to stop sliding on the belt? [*Hint*: Use $F = ma$.] (*b*) What is the package's displacement d during this time? (*c*) How much work is done by the conveyor belt's motor? (*d*) How much of this work is done against friction and how much to accelerate the package?

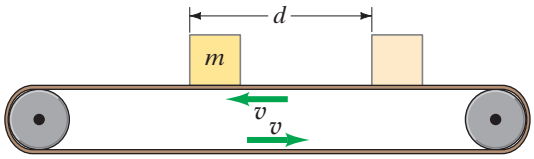


FIGURE 7–32 Problem 89.

90. A cyclist starts from rest and coasts down a 4.0° hill. The mass of the cyclist plus bicycle is 85 kg. After the cyclist has traveled 180 m, (*a*) what was the net work done by gravity on the cyclist? (*b*) How fast is the cyclist going? Ignore friction and air resistance (very unrealistic).

91. What should be the spring constant k of a spring designed to bring a 1500-kg car to rest from a speed of 90 km/h so that the occupants undergo a maximum acceleration of $5.0\,g$?

92. Assume a *cyclist* of weight mg can exert a force on the pedals equal to $0.90\,mg$ on the average. If the pedals rotate in a circle of radius 18 cm, the wheels have a radius of 34 cm, and the front and back *sprockets* on which the chain runs have 42 and 19 teeth respectively (Fig. 7–33), determine the maximum slope of a hill the cyclist can climb at constant speed. Assume the mass of the bike is 12 kg and that of the rider is 65 kg. Ignore friction. Assume the cyclist's average force is always: (*a*) downward; (*b*) tangential to pedal motion.



FIGURE 7–33 Problem 92.

93. A car passenger buckles himself in with a seat belt and holds his 18-kg toddler on his lap. Use the work-energy principle to answer the following questions. (*a*) While traveling 25 m/s, the driver has to make an emergency stop over a distance of 45 m. Assuming constant deceleration, how much force will the arms of the parent need to exert on the child during this deceleration period? Is this force achievable by an average parent? (*b*) Now assume that the car $(v = 25\text{ m/s})$ is in an accident and is brought to stop over a distance of 12 m. Assuming constant deceleration, how much force will the parent need to exert on the child? Is this force achievable by an average parent?

94. A simple pendulum consists of a small object of mass m (the "bob") suspended by a cord of length ℓ (Fig. 7–34) of negligible mass. A force $\vec{\mathbf{F}}$ is applied in the horizontal direction (so $\vec{\mathbf{F}} = F\hat{\mathbf{i}}$), moving the bob very slowly so the acceleration is essentially zero. (Note that the magnitude of $\vec{\mathbf{F}}$ will need to vary with the angle θ that the cord makes with the vertical at any moment.) (*a*) Determine the work done by this force, $\vec{\mathbf{F}}$, to move the pendulum from $\theta = 0$ to $\theta = \theta_0$. (*b*) Determine the work done by the gravitational force on the bob, $\vec{\mathbf{F}}_G = m\vec{\mathbf{g}}$, and the work done by the force $\vec{\mathbf{F}}_T$ that the cord exerts on the bob.

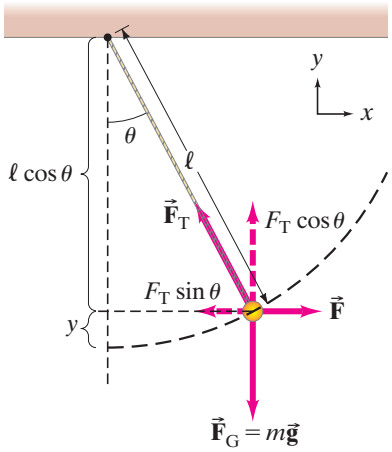


FIGURE 7–34 Problem 94.

95. As an object moves along the x axis from $x = 0.0\text{ m}$ to $x = 20.0\text{ m}$ it is acted upon by a force given by $F = (100 - (x - 10)^2)\text{ N}$. Determine the work done by the force on the object: (*a*) by first sketching the F vs. x graph and estimating the area under this curve; (*b*) by evaluating the integral $\int_{x=0.0\text{ m}}^{x=20\text{ m}} F\,dx$.

96. A small mass m hangs at rest from a vertical rope of length ℓ that is fixed to the ceiling. A force $\vec{\mathbf{F}}$ then pushes on the mass, perpendicular to the taut rope at all times, until the rope is oriented at an angle $\theta = \theta_0$ and the mass has been raised by a vertical distance h (Fig. 7–35). Assume the force's magnitude F is adjusted so that the mass moves at constant speed along its curved trajectory. Show that the work done by $\vec{\mathbf{F}}$ during this process equals mgh, which is equivalent to the amount of work it takes to slowly lift a mass m straight up by a height h. [*Hint*: When the angle is increased by $d\theta$ (in radians), the mass moves along an arc length $ds = \ell\,d\theta$.]

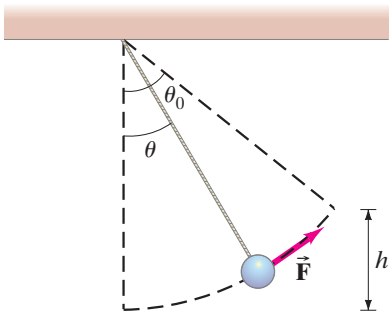


FIGURE 7–35 Problem 96.

97. A 1200-kg car rolling on a horizontal surface has speed $v = 66$ km/h when it strikes a horizontal coiled spring and is brought to rest in a distance of 2.2 m. What is the spring constant of the spring?

98. A hand exerts a constant horizontal force on a block that is free to slide on a frictionless surface (Fig. 7–36). The block starts from rest at point A, and by the time it has traveled a distance d to point B it is traveling with speed v_B. When the block has traveled another distance d to point C, what will its speed be? Explain your reasoning.

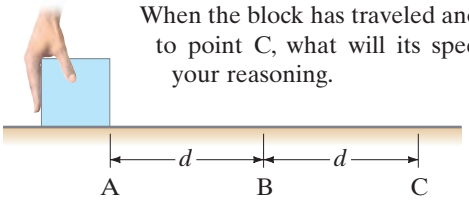


FIGURE 7–36
Question 98.

99. *Stretchable ropes* are used to safely arrest the fall of rock *climbers*. Suppose one end of a rope with unstretched length ℓ is anchored to a cliff and a climber of mass m is attached to the other end. When the climber is a height ℓ above the anchor point, he slips and falls under the force of gravity for a distance 2ℓ, after which the rope becomes taut and stretches a distance x as it stops the climber (see Fig. 7–37). Assume a stretchy rope behaves as a spring with spring constant k. (*a*) Applying the work-energy principle, show that

$$x = \frac{mg}{k}\left[1 + \sqrt{1 + \frac{4k\ell}{mg}}\right].$$

(*b*) Assuming $m = 85$ kg, $\ell = 8.0$ m and $k = 850$ N/m, determine x/ℓ (the fractional stretch of the rope) and kx/mg (the force that the rope exerts on the climber compared to his own weight) at the moment the climber's fall has been stopped.

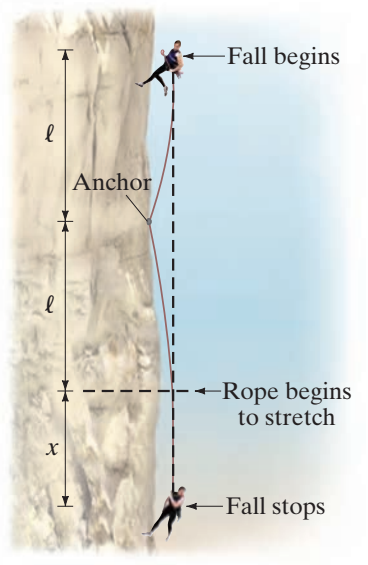


FIGURE 7–37
Problem 99.

100. A 2.0-kg block slides across a rough surface with a constant coefficient of kinetic friction of 0.50 (Fig. 7–38a). The block starts at $x = 0$ with an initial velocity of 4.9 m/s. Pushing the block is a force directed at 36.8° below the horizontal and whose magnitude increases with position as shown in Fig. 7–38b. (*a*) What is the initial kinetic energy of the block? (*b*) How much work is done by the applied force in the first 4.0 m? (*c*) What is the magnitude of the friction force at 0.0 m, and at 4.0 m? (*d*) Draw a line on the graph showing the magnitude of the friction force versus distance x. (*e*) What is the speed of the block at 4.0 m?

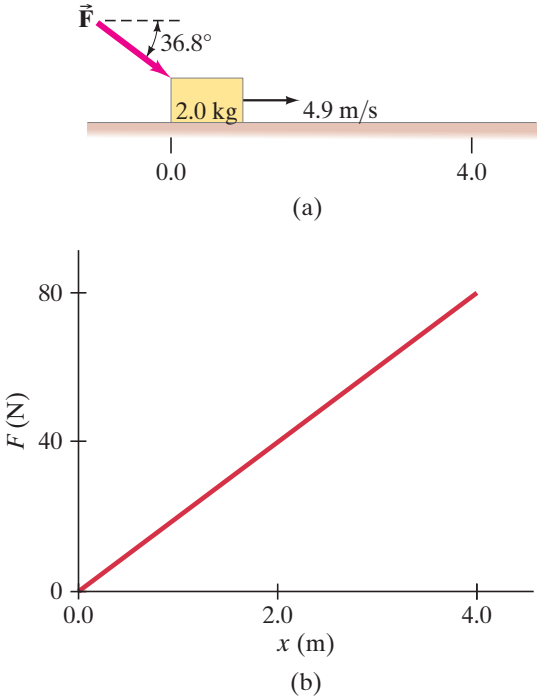


FIGURE 7–38 Problem 100.

ANSWERS TO EXERCISES

A: (*c*).
B: (*b*).
C: (*b*) 2.0×10^5 J. That is, less. (Because v is squared.)
D: No, because the speed v would be the square root of a negative number, which is not real.
E: (*a*) $\sqrt{2}$, (*b*) 4.

A polevaulter running toward the high bar has kinetic energy. When he plants the pole and puts his weight on it, his kinetic energy gets transformed: first into elastic potential energy of the bent pole and then into gravitational potential energy as his body rises. As he crosses the bar, the pole is straight and has given up all its elastic potential energy to the athlete's gravitational potential energy. Nearly all his initial kinetic energy has disappeared, becoming gravitational potential energy of his body at the great height of the bar (world record is over 6 m), which is exactly what he wants. In these, and all other energy transformations that continually take place in the world, the total energy is always conserved. Indeed, the conservation of energy is one of the most fundamental laws of physics, and finds applications in a wide range of fields.



CHAPTER 8

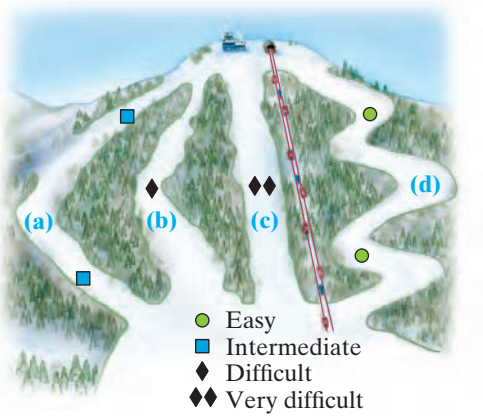
Conservation of Energy

CONTENTS

8 – 1 Conservative and Nonconservative Forces
8 – 2 Potential Energy
8 – 3 Mechanical Energy and Its Conservation
8 – 4 Problem Solving Using Conservation of Mechanical Energy
8 – 5 The Law of Conservation of Energy
8 – 6 Energy Conservation with Dissipative Forces: Solving Problems
8 – 7 Gravitational Potential Energy and Escape Velocity
8 – 8 Power
8 – 9 Potential Energy Diagrams; Stable and Unstable Equilibrium
*8–10 Gravitational Assist (Gravitational Slingshot)

CHAPTER-OPENING QUESTION—Guess now!

A skier starts at the top of a hill. On which run does her gravitational potential energy change the most: **(a)**, **(b)**, **(c)**, or **(d)**; or are they **(e)** all the same? On which run would her speed at the bottom be the fastest if the runs are icy and we assume no friction or air resistance? Recognizing that there is always some friction, answer the above two questions again. List your four answers now.



This chapter continues the discussion of the concepts of work and energy begun in Chapter 7 and introduces additional types of energy, in particular potential energy. Now we will see why the concept of energy is so important. The reason, ultimately, is that energy is conserved—the total energy *always* remains constant in any process. That a quantity can be defined which remains constant, as far as our best experiments can tell, is a remarkable statement about nature. The law of conservation of energy is, in fact, one of the great unifying principles of science.

The law of conservation of energy also gives us another tool, another approach, to solving problems. There are many situations for which an analysis based on Newton's laws would be difficult or impossible—the forces may not be known or accessible to measurement. But often these situations can be dealt with using the law of conservation of energy.

In this Chapter we will mainly treat objects as if they were particles or rigid objects that undergo only translational motion, with no internal or rotational motion.

8–1 Conservative and Nonconservative Forces

We will find it important to categorize forces into two types: conservative and nonconservative. By definition, we call any force a **conservative force** if

DEFINITION OF CONSERVATIVE FORCE

> **the work done by the force on an object moving from one point to another depends only on the initial and final positions of the object, and is independent of the particular path taken.**

A conservative force can be a function *only of position*, and cannot depend on other variables like time or velocity.

We can readily show that the force of gravity is a conservative force. The gravitational force (exerted by the Earth) on an object of mass m near the Earth's surface is $\vec{\mathbf{F}} = m\vec{\mathbf{g}}$, where $\vec{\mathbf{g}}$ is a constant. The work done by this gravitational force on an object that falls a vertical distance h is $W_G = Fd = mgh$ (see Fig. 8–1a). Now suppose that instead of moving vertically downward or upward, an object follows some arbitrary path in the xy plane, as shown in Fig. 8–1b. The object starts at a vertical height y_1 and reaches a height y_2, where $y_2 - y_1 = h$. To calculate the work done by gravity, W_G, we use Eq. 7–7:

$$W_G = \int_1^2 \vec{\mathbf{F}}_G \cdot d\vec{\boldsymbol{\ell}} = \int_1^2 mg\cos\theta\, d\ell.$$

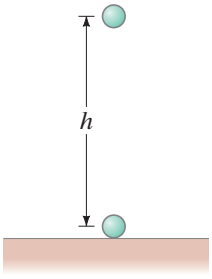
We now let $\phi = 180° - \theta$ be the angle between $d\vec{\boldsymbol{\ell}}$ and its vertical component dy, as shown in Fig. 8–1b. Then, since $\cos\theta = -\cos\phi$ (Appendix A–9) and $dy = d\ell\cos\phi$, we have

$$W_G = -\int_{y_1}^{y_2} mg\, dy = -mg(y_2 - y_1). \qquad \textbf{(8–1)}$$

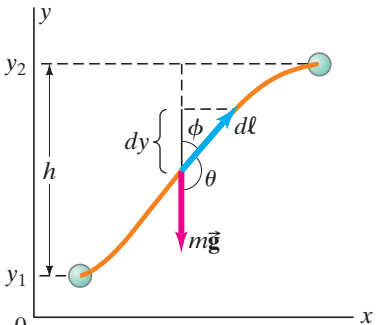
Since $(y_2 - y_1)$ is the vertical height h, we see that the work done depends only on the vertical height and does *not* depend on the particular path taken! Hence, by definition, gravity is a conservative force.

Note that in the case shown in Fig. 8–1b, $y_2 > y_1$ and therefore the work done by gravity is negative. If on the other hand $y_2 < y_1$, so that the object is falling, then W_G is positive.

FIGURE 8–1 Object of mass m: (a) falls a height h vertically; (b) is raised along an arbitrary two-dimensional path.



(a)



(b)

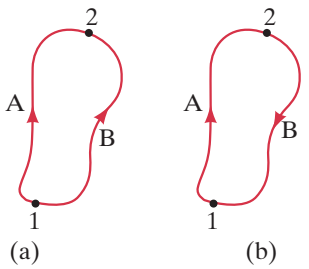


FIGURE 8–2 (a) A tiny object moves from point 1 to point 2 via two different paths, A and B. (b) The object makes a round trip, via path A from point 1 to point 2 and via path B back to point 1.

We can give the definition of a conservative force in another, completely equivalent way:

> **a force is conservative if the net work done by the force on an object moving around any closed path is zero**.

To see why this second definition is equivalent to our first definition on page 195, consider a small object that moves from point 1 to point 2 via either of two paths labeled A and B in Fig. 8–2a. If we assume a conservative force acts on the object, the work done by this force is the same whether the object takes path A or path B, by our first definition. This work to get from point 1 to point 2 we will call W. Now consider the round trip shown in Fig. 8–2b. The object moves from 1 to 2 via path A and our force does work W on the object. Our object then returns to point 1 via path B. How much work is done during the return? In going from 1 to 2 via path B the work done is W, which by definition equals $\int_1^2 \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}$. In doing the reverse, going from 2 to 1 via path B, the force $\vec{\mathbf{F}}$ at each point is the same, but $d\vec{\boldsymbol{\ell}}$ is directed in precisely the opposite direction. Consequently $\vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}$ has the opposite sign at each point so the total work done on our object in making the return trip from 2 to 1 must be $\int_2^1 \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}} = -W$. Hence the total work done in going from 1 to 2 and back to 1 is $W + (-W) = 0$, which proves the equivalence of the two above definitions for a conservative force.

The second definition of a conservative force illuminates an important aspect of such a force: the **work done by a conservative force is recoverable** in the sense that if positive work is done on an object on one part of a closed path, an equivalent amount of negative work will be done on the object on its return.

As we saw above, the force of gravity is conservative, and so is the elastic force $(F = -kx)$, as we see in the next Section.

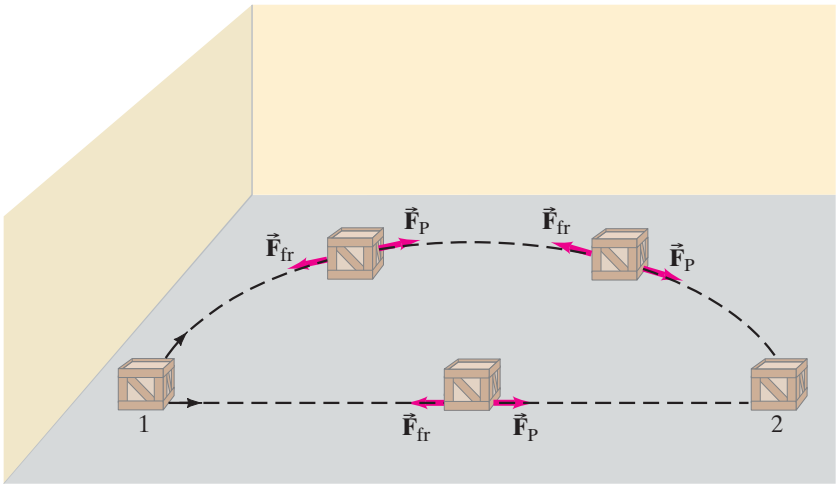


FIGURE 8–3 A crate is pushed slowly at constant speed across a rough floor from position 1 to position 2 via two paths: one straight and one curved. The pushing force $\vec{\mathbf{F}}_{\mathrm{P}}$ is in the direction of motion at each point. (And the friction force opposes the motion.) Hence for a constant magnitude pushing force, the work F_{P} does is $W = F_{\mathrm{P}}d$, so if the distance traveled d is greater (as for the curved path), then W is greater. The work done does not depend *only* on points 1 and 2; it also depends on the path taken.

TABLE 8–1 Conservative and Nonconservative Forces

Conservative Forces	Nonconservative Forces
Gravitational	Friction
Elastic	Air resistance
Electric	Tension in cord
	Motor or rocket propulsion
	Push or pull by a person

Many forces, such as friction and a push or pull exerted by a person, are not conservative: they are **nonconservative forces** because any work they do depends on the path. For example, if you push a crate across a floor from one point to another, the work you do depends on whether the path taken is straight, or is curved. As shown in Fig. 8–3, if a crate is pushed slowly from point 1 to point 2 along the longer semicircular path, you do more work against friction than if you push it along the straight path. You do more work on the curved path because the distance is greater and, unlike the gravitational force, the pushing force $\vec{\mathbf{F}}_{\mathrm{P}}$ is in the direction of motion at each point. Thus the work done by the person in Fig. 8–3 does not depend *only* on points 1 and 2; it depends also on the path taken. The force of kinetic friction (F_{fr}), also shown in Fig. 8–3, always opposes the motion; it too is a nonconservative force, and we discuss how to treat it later in this Chapter (Section 8–6). Table 8–1 lists a few conservative and nonconservative forces.

8–2 Potential Energy

In Chapter 7 we discussed the energy associated with a moving object, which is its kinetic energy $K = \frac{1}{2}mv^2$. Now we introduce **potential energy**, which is the energy associated with *forces* that depend on the position or configuration of an object (or objects) relative to the surroundings. Various types of potential energy can be defined, and each type is associated with a particular conservative force.

The wound-up spring of a toy is an example of potential energy. The spring acquired potential energy because work was done *on* it by the person winding the toy. As the spring unwinds, it exerts a force and does work to make the toy move.

Gravitational Potential Energy

A common example of potential energy is *gravitational potential energy*. A brick held high above the ground has potential energy because of its position relative to the Earth. The raised brick has the ability to do work: if it is released, it falls to the ground due to the gravitational force, and can do work on a stake, driving it into the ground.

Let us seek the form for the gravitational potential energy of an object near the surface of the Earth. For an object of mass m to be lifted vertically, an upward force at least equal to its weight, mg, must be exerted on it, say by a person's hand. To lift the object without acceleration, assuming gravity is the only other force acting on it, the person exerts an "external force," $F_{\text{ext}} = mg$. If it is raised a vertical height h, from position y_1 to y_2 in Fig. 8–4 (upward direction chosen positive), a person does work equal to the product of the "external" force she exerts, $F_{\text{ext}} = mg$ upward, multiplied by the vertical displacement h. That is,

$$W_{\text{ext}} = \vec{\mathbf{F}}_{\text{ext}} \cdot \vec{\mathbf{d}} = mgh\cos 0^\circ = mgh = mg(y_2 - y_1)$$

where both $\vec{\mathbf{F}}_{\text{ext}}$ and $\vec{\mathbf{d}}$ point upward. Gravity is also acting on the object as it moves from y_1 to y_2, and does work W_{G} on the object equal to

$$W_{\text{G}} = \vec{\mathbf{F}}_{\text{G}} \cdot \vec{\mathbf{d}} = mgh\cos 180^\circ = -mgh = -mg(y_2 - y_1),$$

where $\theta = 180^\circ$ because $\vec{\mathbf{F}}_{\text{G}}$ and $\vec{\mathbf{d}}$ point in opposite directions. Since $\vec{\mathbf{F}}_{\text{G}}$ is downward and $\vec{\mathbf{d}}$ is upward, W_{G} is negative. If the object follows an arbitrary path, as in Fig. 8–1b, the work done by gravity still depends only on the change in vertical height (Eq. 8–1): $W_{\text{G}} = -mg(y_2 - y_1) = -mgh$.

Next, if we allow the object at y_2 in Fig. 8–4 to fall freely from rest under the action of gravity, it acquires a velocity given by $v^2 = 2gh$ (Eq. 2–12c) after falling a height h. It then has kinetic energy $\frac{1}{2}mv^2 = \frac{1}{2}m(2gh) = mgh$, and if it strikes a stake it can do work on the stake equal to mgh.

Thus, to raise an object of mass m to a height h *requires* an amount of work equal to mgh. And once at height h, the object has the *ability* to do an amount of work equal to mgh. We can say that the work done in lifting the object has been stored as gravitational potential energy.

Indeed, we can define the *change* in gravitational potential energy U, when an object is moved from height y_1 to height y_2, as equal to the work done by the net external force that accomplishes this without acceleration:

$$\Delta U = U_2 - U_1 = W_{\text{ext}} = mg(y_2 - y_1). \qquad \textbf{(8–2a)}$$

A more direct way to define the *change in gravitational potential energy*, ΔU, is that it is equal to the *negative of the work done by gravity* itself in the process (note Eq. 8–1):

$$\Delta U = U_2 - U_1 = -W_{\text{G}} = mg(y_2 - y_1). \qquad \textbf{(8–2b)}$$

Equations 8–2 define the change in gravitational potential energy when an object of mass m moves between two points near the surface of the Earth. The **gravitational potential energy**, U, at any point a vertical height y above some reference point (the origin of the coordinate system) can be defined as

$$U_{\text{grav}} = mgy. \qquad \text{[gravity only]} \quad \textbf{(8–3)}$$

Note that the potential energy is associated with the force of gravity between the Earth and the mass m. Hence U_{grav} represents the gravitational potential energy, not simply of the mass m alone, but of the mass–Earth system.

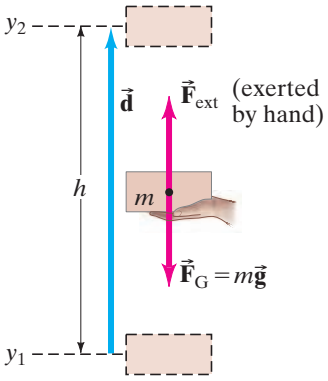


FIGURE 8–4 A person exerts an upward force $F_{\text{ext}} = mg$ to lift a brick from y_1 to y_2.

CAUTION
Change in potential energy is what is physically meaningful

CAUTION
Potential energy belongs to a system, not to a single object

Gravitational potential energy depends on the *vertical height* of the object *above some reference level*, $U = mgy$. Sometimes you may wonder from what point to measure y. The gravitational potential energy of a book held high above a table, for example, depends on whether we measure y from the top of the table, or from the floor, or from some other reference level. But what is physically important in any situation is the *change* in potential energy, ΔU, because that is what is related to the work done, and it is ΔU that can be measured. We can thus choose to measure y from any reference level that is convenient, but we must *choose the reference level at the start and be consistent throughout*. The *change* in potential energy between any two points does not depend on this choice.

Potential energy belongs to a *system*, and not to a single object alone. Potential energy is associated with a force, and a force on one object is always exerted by some other object. Thus potential energy is a property of the system as a whole. For an object raised to a height y above the Earth's surface, the change in gravitational potential energy is mgy. The system here is the object plus the Earth, and properties of both are involved: object (m) and Earth (g). In general, a **system** is one or more objects that we choose to study.

The choice of what makes up a system is always ours, and we often try to choose a simple system. Below, when we deal with the potential energy of an object in contact with a spring, our system will be the object and the spring.

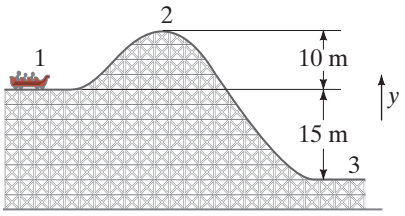


FIGURE 8–5 Example 8–1.

EXAMPLE 8–1 Potential energy changes for a roller coaster. A 1000-kg roller-coaster car moves from point 1, Fig. 8–5, to point 2 and then to point 3. (*a*) What is the gravitational potential energy at points 2 and 3 relative to point 1? That is, take $y = 0$ at point 1. (*b*) What is the change in potential energy when the car goes from point 2 to point 3? (*c*) Repeat parts (*a*) and (*b*), but take the reference point $(y = 0)$ to be at point 3.

APPROACH We are interested in the potential energy of the car–Earth system. We take upward as the positive y direction, and use the definition of gravitational potential energy to calculate the potential energy.

SOLUTION (*a*) We measure heights from point 1 $(y_1 = 0)$, which means initially that the gravitational potential energy is zero at y_1. At point 2, where $y_2 = 10\,\text{m}$,

$$U_2 = mgy_2 = (1000\,\text{kg})(9.8\,\text{m/s}^2)(10\,\text{m}) = 9.8 \times 10^4\,\text{J}.$$

At point 3, $y_3 = -15\,\text{m}$, since point 3 is below point 1. Therefore,

$$U_3 = mgy_3 = (1000\,\text{kg})(9.8\,\text{m/s}^2)(-15\,\text{m}) = -1.5 \times 10^5\,\text{J}.$$

(*b*) In going from point 2 to point 3, the potential energy change $(U_{\text{final}} - U_{\text{initial}})$ is

$$U_3 - U_2 = (-1.5 \times 10^5\,\text{J}) - (9.8 \times 10^4\,\text{J}) = -2.5 \times 10^5\,\text{J}.$$

The gravitational potential energy decreases by $2.5 \times 10^5\,\text{J}$.

(*c*) Now we set $y_3 = 0$. Then $y_1 = +15\,\text{m}$ at point 1, so the potential energy initially (at point 1) is

$$U_1 = (1000\,\text{kg})(9.8\,\text{m/s}^2)(15\,\text{m}) = 1.5 \times 10^5\,\text{J}.$$

At point 2, $y_2 = 25\,\text{m}$, so the potential energy is

$$U_2 = 2.5 \times 10^5\,\text{J}.$$

At point 3, $y_3 = 0$, so the potential energy is zero. The change in potential energy going from point 2 to point 3 is

$$U_3 - U_2 = 0 - 2.5 \times 10^5\,\text{J} = -2.5 \times 10^5\,\text{J},$$

which is the same as in part (*b*), as it should be.

NOTE Work done by gravity depends only on the vertical height, so changes in gravitational potential energy do not depend on the path taken.

EXERCISE A By how much does the gravitational potential energy change when a 1200-kg car climbs to the top of a 300-m-tall hill? (*a*) $3.6 \times 10^5\,\text{J}$; (*b*) $3.5 \times 10^6\,\text{J}$; (*c*) 4 J; (*d*) 40 J; (*e*) 39.2 J.

Potential Energy in General

We have defined the change in gravitational potential energy (Eq. 8–2) to be equal to the negative of the work done by gravity when an object moves from height y_1 to y_2, which we now write as (see page 195)

$$\Delta U = -W_G = -\int_1^2 \vec{\mathbf{F}}_G \cdot d\vec{\boldsymbol{\ell}}.$$

There are other types of potential energy besides gravitational. Each type of potential energy is associated with a particular conservative force, and can be defined just as for gravity. In general, we define the **change in potential energy associated with a particular conservative force $\vec{\mathbf{F}}$ as the negative of the work done by that force**,

$$\Delta U = U_2 - U_1 = -\int_1^2 \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}} = -W, \qquad \textbf{(8–4)}$$

when an object goes from position 1 to position 2. However, we cannot use this definition to define a potential energy for every force. It makes sense only for conservative forces such as gravity, for which the integral depends only on the end points and not on the path taken. It does not apply to nonconservative forces like friction, because the integral in Eq. 8–4 would *not* have a unique value depending only on the end points 1 and 2. Thus the concept of potential energy cannot be defined and is meaningless for a nonconservative force.

POTENTIAL ENERGY DEFINED

CAUTION

Potential energy can be defined only for conservative forces

We can also define the change of potential energy associated with a particular conservative force in an equivalent way as: *the work done by an external (different) force to move the object without acceleration from one position to another.* That is, the external force acts against our conservative force. An example is the work done by a hand lifting an object upward (without acceleration) equals the change in gravitational potential energy.

Elastic Potential Energy

Consider the simple coil spring shown in Fig. 8–6, whose mass is so small that we can ignore it. The spring has potential energy when it is compressed because, when it is released, it can do work on a ball. Like other elastic materials, a spring is described by Hooke's law (see Section 7–3) as long as the displacement x is not too great. Let us take our coordinate system so the end of the uncompressed spring is at $x = 0$ (Fig. 8–6a) and x is positive to the right. To hold the spring compressed (or stretched) a distance x from its natural (unstretched) length requires the person's hand to exert an external force $F_{ext} = kx$ on the spring (Fig. 8–6b), where k is the *spring constant* (Section 7–3). The spring itself exerts a force (Newton's third law) in the opposite direction on the hand (Fig. 8–6c):

$$F_S = -kx. \qquad \text{[spring force]}$$

The negative sign appears because the force $\vec{\mathbf{F}}_S$ is in the direction opposite to the displacement x. We can use the concept of potential energy for a simple (1–D) spring because F_S depends only on position x. From Eq. 8–4, the change in potential energy when the spring is compressed or stretched from $x_1 = 0$ (its uncompressed position) to $x_2 = x$ (where x can be + or −) is

$$\Delta U = U(x) - U(0) = -\int_1^2 \vec{\mathbf{F}}_S \cdot d\vec{\boldsymbol{\ell}} = -\int_0^x (-kx)\, dx = \tfrac{1}{2}kx^2.$$

Here, $U(x)$ means the potential energy at x, and $U(0)$ means U at $x = 0$. It is usually convenient to choose the potential energy at $x = 0$ to be zero: $U(0) = 0$, so the potential energy of a spring compressed or stretched an amount x from equilibrium is

$$U_{el}(x) = \tfrac{1}{2}kx^2. \qquad \text{[elastic spring]} \qquad \textbf{(8–5)}$$

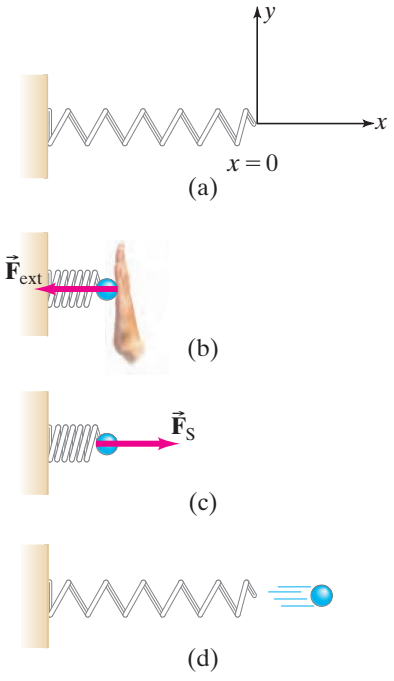


FIGURE 8–6 (a) Spring in natural (uncompressed) position ($x = 0$). (b) Spring is compressed ($x < 0$) by a person exerting an external force $\vec{\mathbf{F}}_{ext}$ to the left. (c) The spring pushes back with a force $F_S = -kx$, where $F_S > 0$ because $x < 0$. (d) The spring can do work on a ball when released.

Potential Energy Related to Force (1–D)

In the one-dimensional case, where a conservative force can be written as a function of x, say, the potential energy can be written as an indefinite integral

$$U(x) = -\int F(x)\, dx + C, \qquad \textbf{(8–6)}$$

where the constant C can be chosen for convenience since only a change in U is meaningful. We sometimes choose C so that at $x = 0$, $U(0) = 0$.

For other forces we may choose C so that when $x = \infty$ (infinitely far away), then $U(\infty) = 0$. Equation 8–6

$$U(x) = -\int F(x)\,dx + C, \tag{8–6}$$

tells us how to obtain $U(x)$ when given $F(x)$. If, instead, we are given $U(x)$, we can obtain $F(x)$ from Eq. 8–6 by taking the derivative of both sides, remembering that integration and differentiation are inverse operations:

$$\frac{d}{dx}\int F(x)\,dx = F(x).$$

Thus, from Eq. 8–6,

$$F(x) = -\frac{dU(x)}{dx}. \tag{8–7}$$

EXAMPLE 8–2 **Determine F from U.** Suppose $U(x) = -ax/(b^2 + x^2)$, where a and b are constants. What is F as a function of x?

APPROACH Since $U(x)$ depends only on x, this is a one-dimensional problem.

SOLUTION Equation 8–7 gives

$$F(x) = -\frac{dU}{dx} = -\frac{d}{dx}\left[-\frac{ax}{b^2 + x^2}\right] = \frac{a}{b^2 + x^2} - \frac{ax}{(b^2 + x^2)^2}2x = \frac{a(b^2 - x^2)}{(b^2 + x^2)^2}.$$

*Potential Energy in Three Dimensions

In three dimensions, we can write the relation between $\vec{\mathbf{F}}(x, y, z)$ and U as:

$$F_x = -\frac{\partial U}{\partial x}, \quad F_y = -\frac{\partial U}{\partial y}, \quad F_z = -\frac{\partial U}{\partial z},$$

or

$$\vec{\mathbf{F}}(x, y, z) = -\hat{\mathbf{i}}\frac{\partial U}{\partial x} - \hat{\mathbf{j}}\frac{\partial U}{\partial y} - \hat{\mathbf{k}}\frac{\partial U}{\partial z}.$$

Here, $\partial/\partial x$, $\partial/\partial y$ and $\partial/\partial z$ are called partial derivatives; $\partial/\partial x$, for example, means that although U may be a function of x, y, and z, written $U(x, y, z)$, we take the derivative only with respect to x with the other variables held constant.

8–3 Mechanical Energy and Its Conservation

Let us consider a conservative system (meaning only conservative forces do work) in which energy is transformed from kinetic to potential or vice versa. Again, we must consider a system because potential energy does not exist for an isolated object. Our system might be a mass m oscillating on the end of a spring or moving in the Earth's gravitational field.

According to the work-energy principle (Eq. 7–11), the net work W_{net} done *on* an object is equal to its change in kinetic energy K:

$$W_{\text{net}} = \Delta K.$$

(If more than one object of our system has work done on it, then W_{net} and ΔK can represent the sum for all of them.) If we assume a conservative system, we can write the work done on an object or objects by the conservative forces in terms of the change in total potential energy (see Eq. 8–4) between points 1 and 2:

$$\Delta U_{\text{total}} = -\int_1^2 \vec{\mathbf{F}}_{\text{net}} \cdot d\vec{\boldsymbol{\ell}} = -W_{\text{net}}. \tag{8–8}$$

We combine the previous two equations, letting U be the total potential energy:

$$\Delta K + \Delta U = 0 \quad \text{[conservative forces only]} \tag{8–9a}$$

or

$$(K_2 - K_1) + (U_2 - U_1) = 0. \quad \text{[conservative forces only]} \tag{8–9b}$$

We now define a quantity E, called the **total mechanical energy** of our system, as the sum of the kinetic energy plus the potential energy of the system at any moment

$$E = K + U.$$

We can rewrite Eq. 8–9b as

$$K_2 + U_2 = K_1 + U_1 \qquad \text{[conservative forces only]} \quad \textbf{(8–10a)}$$

or

$$E_2 = E_1 = \text{constant.} \qquad \text{[conservative forces only]} \quad \textbf{(8–10b)}$$

CONSERVATION OF MECHANICAL ENERGY

Equations 8–10 express a useful and profound principle regarding the total mechanical energy of a system: namely that it is a **conserved quantity**. The total mechanical energy E remains constant as long as no nonconservative forces do work. That is, the quantity $E = K + U$ at some initial time 1 is equal to $K + U$ at any later time 2.

To say it another way, consider Eq. 8–9a which tells us $\Delta U = -\Delta K$; that is, if the kinetic energy K of a system increases, then the potential energy U must decrease by an equivalent amount to compensate. Thus the total, $K + U$, remains constant:

If only conservative forces do work, the total mechanical energy of a system neither increases nor decreases in any process. It stays constant—it is conserved.

CONSERVATION OF MECHANICAL ENERGY

This is the **principle of conservation of mechanical energy** for conservative forces.

We now see the reason for the term "conservative force"—because for such forces, mechanical energy is conserved.

If only one object of a system† has significant kinetic energy, Eqs. 8–10 become

$$E = \tfrac{1}{2}mv^2 + U = \text{constant.} \qquad \text{[conservative forces only]} \quad \textbf{(8–11a)}$$

If we let v_1 and U_1 represent the velocity and potential energy at one instant, and v_2 and U_2 represent them at a second instant, then we can rewrite this as

$$\tfrac{1}{2}mv_1^2 + U_1 = \tfrac{1}{2}mv_2^2 + U_2. \qquad \text{[conservative system]} \quad \textbf{(8–11b)}$$

From this equation we can see again that it doesn't make any difference where we choose the potential energy to be zero: adding a constant to U merely adds the same constant to both sides of Eq. 8–11b, and these cancel. A constant also doesn't affect the force obtained using Eq. 8–7, $F = -dU/dx$, since the derivative of a constant is zero. Only *changes* in the potential energy matter.

CAUTION
It is ΔU that matters

8–4 Problem Solving Using Conservation of Mechanical Energy

A simple example of the conservation of mechanical energy (neglecting air resistance) is a rock allowed to fall due to Earth's gravity from a height h above the ground, as shown in Fig. 8–7. If the rock starts from rest, all of the initial energy is potential energy. As the rock falls, the potential energy mgy decreases (because y decreases), but the rock's kinetic energy increases to compensate, so that the sum of the two remains constant. At any point along the path, the total mechanical energy is given by

$$E = K + U = \tfrac{1}{2}mv^2 + mgy$$

where y is the rock's height above the ground at a given instant and v is its speed at that point. If we use subscript 1 for the rock at one point along its path (maybe the initial point), and subscript 2 when it is at some other point, then we can write

total mechanical energy at point 1 = total mechanical energy at point 2

or (see also Eq. 8–11b)

$$\tfrac{1}{2}mv_1^2 + mgy_1 = \tfrac{1}{2}mv_2^2 + mgy_2. \qquad \text{[gravity only]} \quad \textbf{(8–12)}$$

Just before the rock hits the ground, where we chose $y = 0$, all of the initial potential energy will have been transformed into kinetic energy.

†For an object falling under the influence of Earth's gravity, the kinetic energy of the Earth moving upwards can usually be ignored. For a mass oscillating at the end of a spring, the mass of the spring, and hence its kinetic energy, can often be ignored.

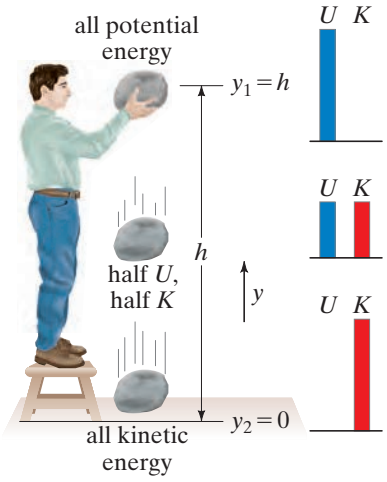


FIGURE 8–7 The rock's potential energy changes to kinetic energy as it falls. Note bar graphs representing the amount of potential energy U and kinetic energy K for three different positions.

EXAMPLE 8–3 **Falling rock.** If the original height of the rock in Fig. 8–7 is $y_1 = h = 3.0\,\text{m}$, calculate the rock's speed when it has fallen to 1.0 m above the ground.

APPROACH We apply the principle of conservation of mechanical energy, Eq. 8–12, with only gravity acting on the rock. We choose the ground as our reference level $(y = 0)$.

SOLUTION At the moment of release (point 1) the rock's position is $y_1 = 3.0\,\text{m}$ and it is at rest: $v_1 = 0$. We want to find v_2 when the rock is at position $y_2 = 1.0\,\text{m}$. Equation 8–12 gives

$$\tfrac{1}{2}mv_1^2 + mgy_1 = \tfrac{1}{2}mv_2^2 + mgy_2.$$

The m's cancel out and $v_1 = 0$ so

$$gy_1 = \tfrac{1}{2}v_2^2 + gy_2.$$

Solving for v_2 we find

$$v_2 = \sqrt{2g(y_1 - y_2)} = \sqrt{2(9.8\,\text{m/s}^2)[(3.0\,\text{m}) - (1.0\,\text{m})]} = 6.3\,\text{m/s}.$$

The rock's speed 1.0 m above the ground is 6.3 m/s downward.

NOTE The velocity of the rock is independent of the rock's mass.

EXERCISE B In Example 8–3, what is the rock's speed just before it hits the ground? (*a*) 6.5 m/s; (*b*) 7.0 m/s; (*c*) 7.7 m/s; (*d*) 8.3 m/s; (*e*) 9.8 m/s.

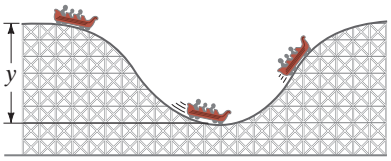


FIGURE 8–8 A roller-coaster car (shown here in 3 different positions on the track) moving without friction illustrates the conservation of mechanical energy.

Equation 8–12 can be applied to any object moving without friction under the action of gravity. For example, Fig. 8–8 shows a roller-coaster car starting from rest at the top of a hill, and coasting without friction to the bottom and up the hill on the other side. True, there is another force besides gravity acting on the car, the normal force exerted by the tracks. But the normal force acts perpendicular to the direction of motion at each point and so does zero work. We ignore rotational motion of the car's wheels and treat the car as a particle undergoing simple translation. Initially, the car has only potential energy. As it coasts down the hill, it loses potential energy and gains kinetic energy, but the sum of the two remains constant. At the bottom of the hill it has its maximum kinetic energy, and as it climbs up the other side the kinetic energy changes back to potential energy. When the car comes to rest again at the same height from which it started, all of its energy will be potential energy. Given that the gravitational potential energy is proportional to the vertical height, energy conservation tells us that (in the absence of friction) the car comes to rest at a height equal to its original height. If the two hills are the same height, the car will just barely reach the top of the second hill when it stops. If the second hill is lower than the first, not all of the car's kinetic energy will be transformed to potential energy so the car will have speed to continue over the top and down the other side. If the second hill is higher, the car will reach a maximum height on it equal to its original height on the first hill. This is true (in the absence of friction) no matter how steep the hill is, since potential energy depends only on the vertical height.

EXAMPLE 8–4 **Roller-coaster car speed using energy conservation.** Assuming the height of the hill in Fig. 8–8 is 40 m, and the roller-coaster car starts from rest at the top, calculate (*a*) the speed of the roller-coaster car at the bottom of the hill, and (*b*) at what height it will have half this speed. Take $y = 0$ at the bottom of the hill. Ignore frictional forces.

APPROACH We use conservation of mechanical energy. We choose point 1 to be where the car starts from rest $(v_1 = 0)$ at the top of the hill $(y_1 = 40\,\text{m})$. In part (*a*), point 2 is the bottom of the hill, which we choose as our reference level, so $y_2 = 0$. In part (*b*) we let y_2 be the unknown.

SOLUTION (*a*) We use Eq. 8–12 with $v_1 = 0$ and $y_2 = 0$, which gives

$$mgy_1 = \tfrac{1}{2}mv_2^2$$

or

$$v_2 = \sqrt{2gy_1} = \sqrt{2(9.8\,\text{m/s}^2)(40\,\text{m})} = 28\,\text{m/s}.$$

(*b*) Now y_2 is the unknown. We again use conservation of energy,

$$\tfrac{1}{2}mv_1^2 + mgy_1 = \tfrac{1}{2}mv_2^2 + mgy_2,$$

but now $v_2 = \frac{1}{2}(28\,\text{m/s}) = 14\,\text{m/s}$ and $v_1 = 0$. Solving for the unknown y_2 gives

$$y_2 = y_1 - \frac{v_2^2}{2g} = 40\,\text{m} - \frac{(14\,\text{m/s})^2}{2(9.8\,\text{m/s}^2)} = 30\,\text{m}.$$

That is, the car has a speed of 14 m/s when it is 30 *vertical* meters above the lowest point, both when descending the left-hand hill and when ascending the right-hand hill.

The mathematics of the roller-coaster Example 8–4 is almost the same as for the rock in Example 8–3. But there is an important difference between them. In Example 8–3 the motion is all vertical and could have been solved using force, acceleration, and the kinematic equations (Eqs. 2–12). For the "friction-free" roller coaster of Example 8–4, where the motion is not vertical, we could *not* have used Eqs. 2–12 because the acceleration a is not constant on the curved track. But energy conservation gave us the answer readily.

CONCEPTUAL EXAMPLE 8–5 **Speeds on two water slides.** Two water slides at a pool are shaped differently, but start at the same height h (Fig. 8–9). Two riders, Paul and Jeanne, start from rest at the same time on different slides. (*a*) Which rider, Paul or Jeanne, is traveling faster at the bottom? (*b*) Which rider makes it to the bottom first? Ignore friction and assume both slides have the same path length.

RESPONSE (*a*) Each rider's initial potential energy mgh gets transformed to kinetic energy, so the speed v at the bottom is obtained from $\frac{1}{2}mv^2 = mgh$. The mass cancels and so the speed will be the same, regardless of the mass of the rider. Since they descend the same vertical height, they will finish with the same speed. (*b*) Note that Jeanne is consistently at a lower elevation than Paul at any instant, until the end. This means she has converted her potential energy to kinetic energy earlier. Consequently, she is traveling faster than Paul at each moment during the trip, and so gets to the bottom first.

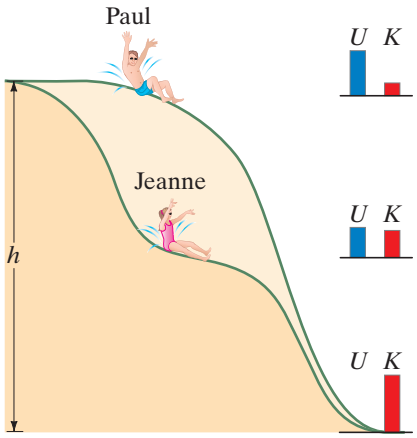


FIGURE 8–9 Example 8–5. On the right are bar graphs for potential and kinetic energy, as in Fig. 8–7.

EXERCISE C Two balls are released from the same height above the floor. Ball A falls freely through the air, whereas ball B slides on a curved frictionless track to the floor. How do the speeds of the balls compare when they reach the floor?

There are many interesting examples of the conservation of energy in sports, such as the pole vault illustrated in Fig. 8–10. We often have to make approximations, but the sequence of events in broad outline for the pole vault is as follows. The initial kinetic energy of the running athlete is transformed into elastic potential energy of the bending pole and, as the athlete leaves the ground, into gravitational potential energy. When the vaulter reaches the top and the pole has straightened out again, the energy has all been transformed into gravitational potential energy (we ignore the vaulter's low horizontal speed over the bar). The pole does not supply any energy, but it acts as a device to *store* energy and thus aid in the transformation of kinetic energy into gravitational potential energy, which is the net result. The energy required to pass over the bar depends on how high the center of mass (CM) of the vaulter must be raised. (The CM is where we can treat the object as a "point" object for translational motion, as discussed in Chapter 9.)

FIGURE 8–10 Transformation of energy during a pole vault: $K \rightarrow U_{\text{el}} \rightarrow U_{\text{G}}$.



PHYSICS APPLIED
Sports



FIGURE 8–11 By bending her body, a pole vaulter can keep her center of mass so low that it may even pass below the bar.

By bending their bodies, pole vaulters keep their CM so low that it can actually pass slightly beneath the bar (Fig. 8–11), thus enabling them to cross over a higher bar than would otherwise be possible.

EXAMPLE 8–6 ESTIMATE **Pole vault.** Estimate the kinetic energy and the speed required for a 70-kg pole vaulter to just pass over a bar 5.0 m high. Assume the vaulter's center of mass is initially 0.90 m off the ground and reaches its maximum height at the level of the bar itself.

APPROACH We equate the total energy just before the vaulter places the end of the pole onto the ground (and the pole begins to bend and store potential energy) with the vaulter's total energy when passing over the bar (we ignore the small amount of kinetic energy at this point). We choose the initial position of the vaulter's center of mass to be $y_1 = 0$. The vaulter's center of mass must then be raised to a height $y_2 = 5.0\,\text{m} - 0.9\,\text{m} = 4.1\,\text{m}$.

SOLUTION We use Eq. 8–12,

$$\tfrac{1}{2}mv_1^2 + 0 = 0 + mgy_2$$

so

$$K_1 = \tfrac{1}{2}mv_1^2 = mgy_2 = (70\,\text{kg})(9.8\,\text{m/s}^2)(4.1\,\text{m}) = 2.8\times10^3\,\text{J}.$$

The speed is

$$v_1 = \sqrt{\frac{2K_1}{m}} = \sqrt{\frac{2(2800\,\text{J})}{70\,\text{kg}}} = 8.9\,\text{m/s} \approx 9\,\text{m/s}.$$

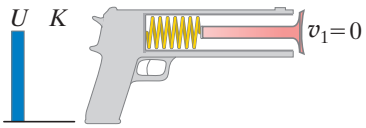
NOTE This is an approximation because we have ignored such things as the vaulter's speed while crossing over the bar, and mechanical energy transformed when the pole is planted in the ground. All would increase the needed initial kinetic energy.

As another example of the conservation of mechanical energy, let us consider an object of mass m connected to a compressed horizontal spring (Fig. 8–6) whose own mass can be neglected and whose spring constant is k. When the spring is released, the mass m has speed v at any moment. The potential energy (PE) of the system (object plus spring) is $\frac{1}{2}kx^2$, where x is the displacement of the spring from its unstretched length. If neither friction nor any other force is acting, conservation of mechanical energy tells us that

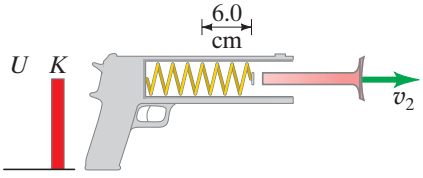
$$\tfrac{1}{2}mv_1^2 + \tfrac{1}{2}kx_1^2 = \tfrac{1}{2}mv_2^2 + \tfrac{1}{2}kx_2^2, \qquad \text{[elastic PE only]} \quad \textbf{(8–13)}$$

where the subscripts 1 and 2 refer to the velocity and displacement at two different moments.

FIGURE 8–12 Example 8–7. (a) A dart is pushed against a spring, compressing it 6.0 cm. The dart is then released, and in (b) it leaves the spring at velocity v_2.



(a) $E = \frac{1}{2}kx_1^2$



(b) $E = \frac{1}{2}mv_2^2$

EXAMPLE 8–7 **Toy dart gun.** A dart of mass 0.100 kg is pressed against the spring of a toy dart gun as shown in Fig. 8–12a. The spring (with spring constant $k = 250\,\text{N/m}$ and ignorable mass) is compressed 6.0 cm and released. If the dart detaches from the spring when the spring reaches its natural length $(x = 0)$, what speed does the dart acquire?

APPROACH The dart is initially at rest (point 1), so $K_1 = 0$. We ignore friction and use conservation of mechanical energy; the only potential energy is elastic.

SOLUTION We use Eq. 8–13 with point 1 being at the maximum compression of the spring, so $v_1 = 0$ (dart not yet released) and $x_1 = -0.060\,\text{m}$. Point 2 we choose to be the instant the dart flies off the end of the spring (Fig. 8–12b), so $x_2 = 0$ and we want to find v_2. Thus Eq. 8–13 can be written

$$0 + \tfrac{1}{2}kx_1^2 = \tfrac{1}{2}mv_2^2 + 0.$$

Then

$$v_2^2 = \frac{kx_1^2}{m} = \frac{(250\,\text{N/m})(-0.060\,\text{m})^2}{(0.100\,\text{kg})} = 9.0\,\text{m}^2/\text{s}^2,$$

and $v_2 = \sqrt{v_2^2} = 3.0\,\text{m/s}$.

EXAMPLE 8–8 **Two kinds of potential energy.** A ball of mass $m = 2.60\,\text{kg}$, starting from rest, falls a vertical distance $h = 55.0\,\text{cm}$ before striking a vertical coiled spring, which it compresses an amount $Y = 15.0\,\text{cm}$ (Fig. 8–13). Determine the spring constant k of the spring. Assume the spring has negligible mass, and ignore air resistance. Measure all distances from the point where the ball first touches the uncompressed spring (let $y = 0$ at this point).

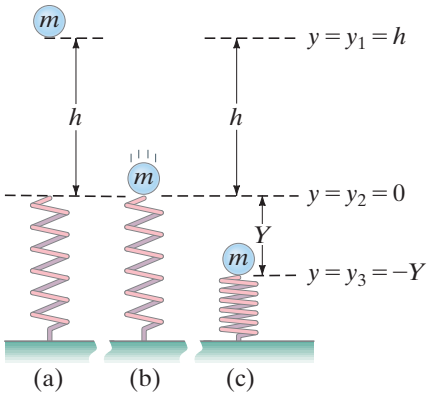


FIGURE 8–13 Example 8–8.

APPROACH The forces acting on the ball are the gravitational pull of the Earth and the elastic force exerted by the spring. Both forces are conservative, so we can use conservation of mechanical energy, including both types of potential energy. We must be careful, however: gravity acts throughout the fall (Fig. 8–13), whereas the elastic force does not act until the ball touches the spring (Fig. 8–13b). We choose y positive upward, and $y = 0$ at the end of the spring in its natural (uncompressed) state.

SOLUTION We divide this solution into two parts. (An alternate solution follows.) *Part 1*: Let us first consider the energy changes as the ball falls from a height $y_1 = h = 0.55\,\text{m}$ (Fig. 8–13a) to $y_2 = 0$, just as it touches the spring, Fig. 8–13b. Our system is the ball acted on by gravity plus the spring (which up to this point doesn't do anything). Thus

$$\begin{aligned} \tfrac{1}{2}mv_1^2 + mgy_1 &= \tfrac{1}{2}mv_2^2 + mgy_2 \\ 0 + mgh &= \tfrac{1}{2}mv_2^2 + 0. \end{aligned}$$

We solve for $v_2 = \sqrt{2gh} = \sqrt{2(9.80\,\text{m/s}^2)(0.550\,\text{m})} = 3.283\,\text{m/s} \approx 3.28\,\text{m/s}$. This is the speed of the ball just as it touches the top of the spring, Fig. 8–13b. *Part 2*: As the ball compresses the spring, Figs. 8–13b to c, there are two conservative forces on the ball—gravity and the spring force. So our conservation of energy equation is

$$\begin{aligned} E_2\,(\text{ball touches spring}) &= E_3\,(\text{spring compressed}) \\ \tfrac{1}{2}mv_2^2 + mgy_2 + \tfrac{1}{2}ky_2^2 &= \tfrac{1}{2}mv_3^2 + mgy_3 + \tfrac{1}{2}ky_3^2. \end{aligned}$$

Substituting $y_2 = 0$, $v_2 = 3.283\,\text{m/s}$, $v_3 = 0$ (the ball comes to rest for an instant), and $y_3 = -Y = -0.150\,\text{m}$, we have

$$\tfrac{1}{2}mv_2^2 + 0 + 0 = 0 - mgY + \tfrac{1}{2}k(-Y)^2.$$

We know m, v_2, and Y, so we can solve for k:

$$\begin{aligned} k &= \frac{2}{Y^2}\left[\tfrac{1}{2}mv_2^2 + mgY\right] = \frac{m}{Y^2}\left[v_2^2 + 2gY\right] \\ &= \frac{(2.60\,\text{kg})}{(0.150\,\text{m})^2}\left[(3.283\,\text{m/s})^2 + 2(9.80\,\text{m/s}^2)(0.150\,\text{m})\right] = 1590\,\text{N/m}. \end{aligned}$$

Alternate Solution Instead of dividing the solution into two parts, we can do it all at once. After all, we get to choose what two points are used on the left and right of the energy equation. Let us write the energy equation for points 1 and 3 in Fig. 8–13. Point 1 is the initial point just before the ball starts to fall (Fig. 8–13a), so $v_1 = 0$, and $y_1 = h = 0.550\,\text{m}$. Point 3 is when the spring is fully compressed (Fig. 8–13c), so $v_3 = 0$, $y_3 = -Y = -0.150\,\text{m}$. The forces on the ball in this process are gravity and (at least part of the time) the spring. So conservation of energy tells us

PROBLEM SOLVING
Alternate Solution

$$\begin{aligned} \tfrac{1}{2}mv_1^2 + mgy_1 + \tfrac{1}{2}k(0)^2 &= \tfrac{1}{2}mv_3^2 + mgy_3 + \tfrac{1}{2}ky_3^2 \\ 0 + mgh + 0 &= 0 - mgY + \tfrac{1}{2}kY^2 \end{aligned}$$

where we have set $y = 0$ for the spring at point 1 because it is not acting: the spring is neither compressed nor stretched. We solve for k:

$$k = \frac{2mg(h + Y)}{Y^2} = \frac{2(2.60\,\text{kg})(9.80\,\text{m/s}^2)(0.550\,\text{m} + 0.150\,\text{m})}{(0.150\,\text{m})^2} = 1590\,\text{N/m}$$

just as in our first method of solution.

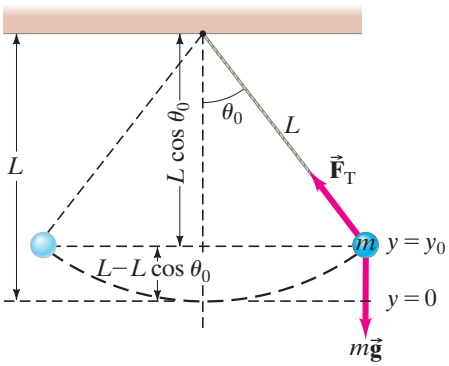


FIGURE 8–14 Example 8–9: a simple pendulum; y is measured positive upward, from the lowest point of the swing $(y = 0)$.

EXAMPLE 8–9 **A swinging pendulum.** The simple pendulum shown in Fig. 8–14 consists of a small bob of mass m suspended by a massless cord of length L. The bob is released (without a push) at $t = 0$, where the cord makes an angle $\theta = \theta_0$ to the vertical. (*a*) Describe the motion of the bob in terms of kinetic energy and potential energy. Then determine the speed of the bob (*b*) as a function of position θ as it swings back and forth, and (*c*) at the lowest point of the swing. (*d*) Find the tension in the cord, $\vec{\mathbf{F}}_T$. Ignore friction and air resistance.

APPROACH The bob has two forces acting on it at any moment: gravity, mg, and the force the cord (in tension) exerts on it, $\vec{\mathbf{F}}_T$. The tension force always acts perpendicular to the motion, so it does no work $(\vec{\mathbf{F}}_T \cdot d\vec{\ell} = 0)$. We need be concerned only with gravity, for which we can write the potential energy. We use the law of conservation of mechanical energy, except in (*d*) where we use Newton's second law.

SOLUTION (*a*) At the moment of release, the bob is at rest, so its kinetic energy $K = 0$. As the bob swings down, it loses potential energy and gains kinetic energy. At the lowest point its kinetic energy is a maximum and the potential energy is a minimum. The bob continues its swing until it reaches an equal height and angle (θ_0) on the opposite side, at which point the potential energy is a maximum and $K = 0$. It continues the swinging motion as $U \rightarrow K \rightarrow U$ and so on, but it can never go higher than $\theta = \pm\theta_0$ (conservation of mechanical energy).

(*b*) The cord is assumed to be massless, so we need to consider only the bob's kinetic energy, and the gravitational potential energy. The mechanical energy of the system is

$$E = \tfrac{1}{2}mv^2 + mgy,$$

where y is the vertical height of the bob at any moment. We take $y = 0$ at the lowest point of the bob's swing. Hence at $t = 0$,

$$y = y_0 = L - L\cos\theta_0 = L(1 - \cos\theta_0)$$

as can be seen in Fig. 8–14. At the moment the bob is released at $y = y_0$,

$$E = mgy_0,$$

since $v = v_0 = 0$. At any other point along the swing

$$E = \tfrac{1}{2}mv^2 + mgy = mgy_0.$$

We solve this for v:

$$v = \sqrt{2g(y_0 - y)}.$$

In terms of the angle θ of the cord, we can write

$$(y_0 - y) = (L - L\cos\theta_0) - (L - L\cos\theta) = L(\cos\theta - \cos\theta_0)$$

so

$$v = \sqrt{2gL(\cos\theta - \cos\theta_0)}.$$

(*c*) At the lowest point, $y = 0$, so

$$v = \sqrt{2gy_0}$$

or

$$v = \sqrt{2gL(1 - \cos\theta_0)}.$$

(*d*) The tension in the cord is the force $\vec{\mathbf{F}}_T$ that the cord exerts on the bob. As we've seen, there is no work done by this force. But we can calculate the force F_T by using Newton's second law $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ and by noting that at any point the acceleration of the bob in the inward radial direction is v^2/L, since the bob is constrained to move in an arc of a circle of radius L (see Eq. 5–1). In the radial direction, $\vec{\mathbf{F}}_T$ acts inward, and a component of gravity equal to $mg\cos\theta$ acts outward along a radius. Hence

$$m\frac{v^2}{L} = F_T - mg\cos\theta.$$

We solve for F_T and use the result of part (*b*) for v^2:

$$F_T = m\left(\frac{v^2}{L} + g\cos\theta\right) = 2mg(\cos\theta - \cos\theta_0) + mg\cos\theta$$
$$= (3\cos\theta - 2\cos\theta_0)\,mg.$$

8–5 The Law of Conservation of Energy

We have seen that mechanical energy is conserved when conservative forces act. We have seen energy transformed from potential to kinetic (or vice versa), as when a rock (Fig. 8–7) or water (Fig. 8–15a) falls downward, and the total energy is conserved. We have also seen energy conserved when energy is transferred from one object to another, such as a compressed spring giving kinetic energy to a ball or dart (Fig. 8–6 or 8–12). Note that the spring does work on the ball or the dart. Indeed, we can state in general that

work is done when energy is transformed from one type to another or transferred from one object to another.

We now take into account nonconservative forces such as friction, since they are important in real situations. For example, consider again the roller-coaster car in Fig. 8–8, but this time let us include friction. The car will not in this case reach the same height on the second hill as it had on the first hill because of friction.

In this, and in other natural processes, the mechanical energy (sum of the kinetic and potential energies) does not remain constant but decreases. Because frictional forces reduce the mechanical energy (but *not* the total energy), they are called **dissipative forces**. Historically, the presence of dissipative forces hindered the formulation of a comprehensive conservation of energy law until well into the nineteenth century. It was only then that heat, which is always produced when there is friction (try rubbing your hands together), was interpreted in terms of energy. Quantitative studies in the nineteenth century (Chapter 19) demonstrated that if heat is considered as a transfer of energy (sometimes called **thermal energy**), then the total energy is conserved in any process. A block sliding freely across a table, for example, comes to rest because of friction. Its initial kinetic energy is all transformed into thermal energy. The block and table are a little warmer as a result of this process.

According to the atomic theory, thermal energy represents kinetic energy of rapidly moving molecules. We shall see in Chapter 18 that a rise in temperature corresponds to an increase in the average kinetic energy of the molecules. Because thermal energy represents the energy of atoms and molecules that make up an object, it is often called **internal energy**. Internal energy, from the atomic point of view, includes kinetic energy of molecules as well as potential energy (usually electrical in nature) due to the relative positions of atoms within molecules. For example the energy stored in food, or in a fuel such as gasoline, is regarded as electric potential energy (chemical bonds). The energy is released through chemical reactions (Fig. 8–15b). This is analogous to a compressed spring which, when released, can do work.

To establish the more general law of conservation of energy, nineteenth-century physicists had to recognize electrical, chemical, and other forms of energy in addition to heat, and to explore if in fact they could fit into a conservation law. For each type of force, conservative or nonconservative, it has always been possible to define a type of energy that corresponds to the work done by such a force, including nuclear energy. And it has been found experimentally that the total energy E always remains constant. That is, the change in total energy, kinetic plus potential plus all other forms of energy, equals zero:

$$\Delta K + \Delta U + [\text{change in all other forms of energy}] = 0. \qquad \textbf{(8–14)}$$

This is one of the most important principles in physics. It is called the **law of conservation of energy** and can be stated as:

The total energy is neither increased nor decreased in any process. Energy can be transformed from one form to another, and transferred from one object to another, but the total amount remains constant.

For conservative mechanical systems, this law can be derived from Newton's laws (Section 8–3) and thus is equivalent to them. But in its full generality, the validity of the law of conservation of energy, encompassing all forms of energy including those associated with nonconservative forces like friction, rests on experimental observation.

FIGURE 8–15 (a) Gravitational potential energy of water at the top of Yosemite Falls is transformed into kinetic energy as the water falls. (Some energy is transformed into heat by air resistance, and some into sound.) (b) The burning of fuel (a chemical reaction) releases energy to boil water in this old steam train in Scotland. The steam produced expands against a piston to do work in turning the wheels.



(a)



(b)

LAW OF CONSERVATION OF ENERGY

LAW OF CONSERVATION OF ENERGY

Even though Newton's laws have been found to fail in the submicroscopic world of the atom, the law of conservation of energy has been found to hold there and in every experimental situation so far tested.

8–6 Energy Conservation with Dissipative Forces: Solving Problems

In Section 8–4 we discussed several Examples of the law of conservation of energy for conservative systems. Now let us consider in detail some examples that involve nonconservative forces.

Suppose, for example, that the roller-coaster car rolling on the hills of Fig. 8–8 is subject to frictional forces. In going from some point 1 to a second point 2, the energy dissipated by the friction force $\vec{\mathbf{F}}_{\text{fr}}$ acting on the car (treating it as a particle) is $\int_1^2 \vec{\mathbf{F}}_{\text{fr}} \cdot d\vec{\boldsymbol{\ell}}$. If $\vec{\mathbf{F}}_{\text{fr}}$ is constant in magnitude, then $W_{\text{fr}} = -F_{\text{fr}}\ell$, where ℓ is the actual distance along the path traveled by the car from point 1 to point 2. (The minus sign appears because $\vec{\mathbf{F}}_{\text{fr}}$ and $d\vec{\boldsymbol{\ell}}$ are in opposite directions.) From the work-energy principle (Eq. 7–11), the net work W_{net} done on an object is equal to the change in its kinetic energy K:

$$\Delta K = W_{\text{net}}.$$

We can separate W_{net} into two parts:

$$W_{\text{net}} = W_{\text{C}} + W_{\text{NC}},$$

where W_{C} is the work done by conservative forces (gravity for our car) and W_{NC} is the work done by nonconservative forces (friction for our car). We saw in Eq. 8–4 that the work done by a conservative force can be written in terms of potential energy:

$$W_{\text{C}} = \int_1^2 \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}} = -\Delta U.$$

Thus $\Delta K = W_{\text{net}}$ means $\Delta K = W_{\text{C}} + W_{\text{NC}} = -\Delta U + W_{\text{NC}}$. Hence

GENERAL WORK-ENERGY PRINCIPLE

$$\Delta K + \Delta U = W_{\text{NC}}. \qquad \textbf{(8–15a)}$$

This equation represents the general form of the work-energy principle. It also represents the conservation of energy. For our car, W_{NC} is the work done by the friction force and represents production of thermal energy. Equation 8–15a is valid in general. Based on our derivation from the work-energy principle, W_{NC} on the right side of Eq. 8–15a must be the total work done by all forces that are not included in the potential energy term, ΔU, on the left side.

Let us rewrite Eq. 8–15a for our roller-coaster car example of Fig. 8–8, setting $W_{\text{NC}} = -F_{\text{fr}}\ell$ as we discussed above:

$$W_{\text{NC}} = -F_{\text{fr}}\ell = \Delta K + \Delta U = (\tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2) + (mgy_2 - mgy_1)$$

or

$$\tfrac{1}{2}mv_1^2 + mgy_1 = \tfrac{1}{2}mv_2^2 + mgy_2 + F_{\text{fr}}\ell. \qquad \left[\begin{array}{c}\text{gravity and}\\ \text{friction acting}\end{array}\right] \qquad \textbf{(8–15b)}$$

On the left we have the mechanical energy of the system initially. It equals the mechanical energy at any subsequent point along the path plus the amount of thermal (or internal) energy produced in the process. Total energy is conserved.

Other nonconservative forces can be treated similarly. If you are not sure about the sign of the extra term $\left(\int \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}\right)$, use your intuition: is the mechanical energy increased or decreased in the process?

EXERCISE D Return to the Chapter-Opening Question, page 194, and answer it again now. Try to explain why you may have answered differently the first time.

PROBLEM SOLVING
Use energy, or Newton's laws?

You may wonder sometimes whether to approach a problem using work and energy, or instead to use Newton's laws. As a rough guideline, if the force(s) involved are constant, either approach may succeed. If the forces are not constant, and/or the path is not simple, energy may be the better approach.

Problem solving is not a process that can be done by simply following a set of rules. The following Problem Solving Strategy involving conservation of energy is *not* a prescription, but is a summary to help you get started solving problems involving energy.

PROBLEM SOLVING

Conservation of Energy

1. **Draw a picture** of the physical situation.
2. Determine **the system** for which you will apply energy conservation: the object or objects and the forces acting.
3. Ask yourself what quantity you are looking for, and **choose initial** (point 1) **and final** (point 2) **positions**.
4. If the object under investigation changes its height during the problem, then **choose a reference frame** with a convenient $y = 0$ level for gravitational potential energy; the lowest point is often a good choice.
   If springs are involved, choose the uncompressed (and unstretched) spring position to be x (or y) $= 0$.
5. **Is mechanical energy conserved**? If no friction or other nonconservative forces act, then conservation of mechanical energy holds:
   $$K_1 + U_1 = K_2 + U_2.$$
6. **Apply conservation of energy**. If friction (or other nonconservative forces) are present, then an additional term of the form $\int \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}$ will be needed. For a constant friction force acting over a distance ℓ
   $$K_1 + U_1 = K_2 + U_2 + F_{\text{fr}}\ell.$$
   For other nonconservative forces use your intuition for the sign of $\int \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}$: is the total mechanical energy increased or decreased in the process?
7. Use the equation(s) you develop to **solve** for the unknown quantity.

EXAMPLE 8–10 **ESTIMATE** **Friction on the roller-coaster car.** The roller-coaster car in Example 8–4 reaches a vertical height of only 25 m on the second hill, where it slows to a momentary stop, Fig. 8–16. It traveled a total distance of 400 m along the track. Determine the thermal energy produced and estimate the average friction force (assume it is roughly constant) on the car, whose mass is 1000 kg.

APPROACH We explicitly follow the Problem Solving Strategy above.

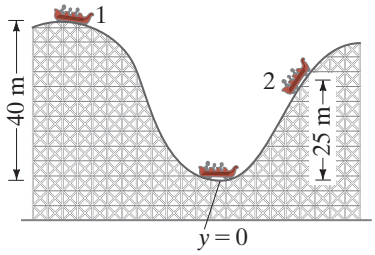
SOLUTION

1. **Draw a picture**. See Fig. 8–16.
2. **The system**. The system is the roller-coaster car and the Earth (which exerts the gravitational force), plus the track which exerts the friction and normal forces on the car.† The normal force does no work because it is always perpendicular to the motion, so it does not affect the energy. Gravity is accounted for as potential energy, and friction as a term $F_{\text{fr}}\ell$.
3. **Choose initial and final positions**. We take point 1 to be the instant when the car started coasting (at the top of the first hill), and point 2 to be the instant it stopped at a height of 25 m on the second hill.
4. **Choose a reference frame**. We choose the lowest point in the motion to be $y = 0$ for the gravitational potential energy.
5. **Is mechanical energy conserved**? No. Friction is present.
6. **Apply conservation of energy**. There is friction acting on the car, so we use conservation of energy in the form of Eq. 8–15b, with $v_1 = 0$, $y_1 = 40\,\text{m}$, $v_2 = 0$, $y_2 = 25\,\text{m}$, and $\ell = 400\,\text{m}$. Thus
   $$0 + (1000\,\text{kg})(9.8\,\text{m/s}^2)(40\,\text{m}) = 0 + (1000\,\text{kg})(9.8\,\text{m/s}^2)(25\,\text{m}) + F_{\text{fr}}\ell.$$
7. **Solve**. We solve the above equation for $F_{\text{fr}}\ell$, the energy dissipated to thermal energy:
   $$F_{\text{fr}}\ell = mg\,\Delta h = (1000\,\text{kg})(9.8\,\text{m/s}^2)(40\,\text{m} - 25\,\text{m}) = 147{,}000\,\text{J}.$$
   The friction force, which acts over a distance of 400 m, averages out to be
   $$F_{\text{fr}} = (1.47 \times 10^5\,\text{J})/400\,\text{m} = 370\,\text{N}.$$

NOTE This result is only a rough average: the friction force at various points depends on the normal force, which varies with the slope angle.

†Alternatively, we could take only the car and Earth as our system, in which case friction would be an external force that reduces the mechanical energy, Eq. 8–15a.

FIGURE 8–16 Example 8–10. Because of friction, a roller-coaster car does not reach the original height on the second hill. (Not to scale.)



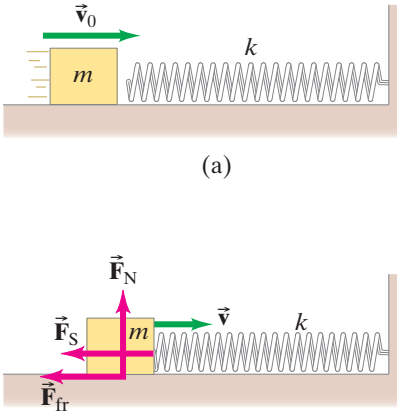


FIGURE 8–17 Example 8–11.

EXAMPLE 8–11 **Friction with a spring.** A block of mass m sliding along a rough horizontal surface is traveling at a speed v_0 when it strikes a massless spring head-on (see Fig. 8–17) and compresses the spring a maximum distance X. If the spring has stiffness constant k, determine the coefficient of kinetic friction between block and surface.

APPROACH At the moment of collision, the block has $K = \frac{1}{2}mv_0^2$ and the spring is presumably uncompressed, so $U = 0$. Initially the mechanical energy of the system is $\frac{1}{2}mv_0^2$. By the time the spring reaches maximum compression, $K = 0$ and $U = \frac{1}{2}kX^2$. In the meantime, the friction force $(= \mu_k F_N = \mu_k mg)$ has transformed an amount of energy $F_{fr}X = \mu_k mgX$ into thermal energy. We take as our system the block, the spring, and the surface (frictional work).

SOLUTION Conservation of energy allows us to write

$$\text{energy (initial)} = \text{energy (final)}$$

$$\tfrac{1}{2}mv_0^2 = \tfrac{1}{2}kX^2 + \mu_k mgX.$$

We solve for μ_k and find

$$\mu_k = \frac{v_0^2}{2gX} - \frac{kX}{2mg}.$$

NOTE Gravitational potential energy does not change because the motion is horizontal (y is constant in mgy).

*Work-Energy versus Energy Conservation

The law of conservation of energy is more general and more powerful than the work-energy principle. The work-energy principle should not be viewed as a statement of conservation of energy, but is nonetheless useful for mechanical problems; and whether you use it, or use the more powerful conservation of energy, can depend on your *choice of the system* under study. If you choose as your system a particle or rigid object on which external forces do work, then you can use the work-energy principle: the work done by the external forces on your object equals the change in its kinetic energy. But if you choose a system on which no external forces do work, then apply conservation of energy to that system directly.

Consider, for example, a spring connected to a block on a frictionless table (Fig. 8–18). If you choose the block as your system, then the work done on the block by the spring equals the change in kinetic energy of the block: the work-energy principle. (Energy conservation does not apply to this system—the block's energy changes.) If instead you choose the block plus the spring as your system, no external forces do work (since the spring is part of the chosen system). To this system you need to apply conservation of energy: if you compress the spring and then release it, the spring still exerts a force[†] on the block, but the subsequent motion can be discussed in terms of kinetic energy $(\frac{1}{2}mv^2)$ plus potential energy $(\frac{1}{2}kx^2)$, whose total remains constant.

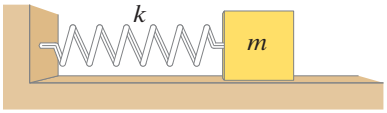


FIGURE 8–18 A spring connected to a block on a frictionless table. If you choose your system to be the block plus spring, then

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2$$

is conserved.

8–7 Gravitational Potential Energy and Escape Velocity

We have been dealing with gravitational potential energy so far in this Chapter assuming the force of gravity is constant, $\vec{\mathbf{F}} = m\vec{\mathbf{g}}$. This is an accurate assumption for ordinary objects near the surface of the Earth. But to deal with gravity more generally, for points not close to the Earth's surface, we must remember that the gravitational force exerted by the Earth on a particle of mass m decreases inversely

[†]The force the spring exerts on the block, and the force the block exerts back on the spring, are not "external" forces—they are within the system in this case.

as the square of the distance r from the Earth's center. The precise relationship is given by Newton's law of universal gravitation (Sections 6–1 and 6–2):

$$\vec{\mathbf{F}} = -G\frac{mM_E}{r^2}\hat{\mathbf{r}} \qquad [r > r_E]$$

where M_E is the mass of the Earth and $\hat{\mathbf{r}}$ is a unit vector (at the position of m) directed radially away from the Earth's center.[†] The minus sign indicates that the force on m is directed toward the Earth's center, in the direction opposite to $\hat{\mathbf{r}}$. This equation can also be used to describe the gravitational force on a mass m in the vicinity of other heavenly bodies, such as the Moon, a planet, or the Sun, in which case M_E must be replaced by that body's mass.

Suppose an object of mass m moves from one position to another along an arbitrary path (Fig. 8–19) so that its distance from the Earth's center changes from r_1 to r_2. The work done by the gravitational force is

$$W = \int_1^2 \vec{\mathbf{F}}\cdot d\vec{\boldsymbol{\ell}} = -GmM_E\int_1^2 \frac{\hat{r}\cdot d\vec{\boldsymbol{\ell}}}{r^2},$$

where $d\vec{\boldsymbol{\ell}}$ represents an infinitesimal displacement. But $\hat{\mathbf{r}}\cdot d\vec{\boldsymbol{\ell}} = dr$, which is the component of $d\vec{\boldsymbol{\ell}}$ along $\hat{\mathbf{r}}$ (see Fig. 8–19), then

$$W = -GmM_E\int_{r_1}^{r_2}\frac{dr}{r^2} = GmM_E\left(\frac{1}{r_2}-\frac{1}{r_1}\right)$$

or

$$W = \frac{GmM_E}{r_2} - \frac{GmM_E}{r_1}.$$

Because the value of this integral depends only on the position of the end points (r_1 and r_2) and not on the path taken, the gravitational force is a conservative force. We can therefore use the concept of potential energy for the gravitational force. Since the change in potential energy is always defined (Section 8–2) as the negative of the work done by the force, we have

$$\Delta U = U_2 - U_1 = -\frac{GmM_E}{r_2} + \frac{GmM_E}{r_1}. \qquad \textbf{(8–16)}$$

From Eq. 8–16 the potential energy at any distance r from the Earth's center can be written:

$$U(r) = -\frac{GmM_E}{r} + C,$$

where C is a constant. It is usual to choose $C = 0$, so that

$$U(r) = -\frac{GmM_E}{r}. \qquad \left[\begin{array}{c}\text{gravity}\\ (r > r_E)\end{array}\right] \qquad \textbf{(8–17)}$$

With this choice for C, $U = 0$ at $r = \infty$. As an object approaches the Earth, its potential energy decreases and is always negative (Fig. 8–20).

Equation 8–16 reduces to Eq. 8–2, $\Delta U = mg(y_2 - y_1)$, for objects near the surface of the Earth (see Problem 43).

For a particle of mass m, which experiences only the force of the Earth's gravity, the total energy is conserved because gravity is a conservative force. Therefore we can write

$$\tfrac{1}{2}mv_1^2 - G\frac{mM_E}{r_1} = \tfrac{1}{2}mv_2^2 - G\frac{mM_E}{r_2} = \text{constant}. \qquad \left[\begin{array}{c}\text{gravity}\\ \text{only}\end{array}\right] \qquad \textbf{(8–18)}$$

[†]We write $r > r_E$, where r_E is the radius of Earth, because this equation only applies beyond the Earth's surface. Inside the Earth, things are more complicated—see Appendix E.

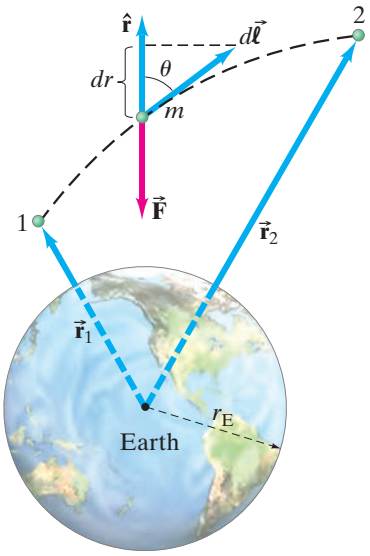
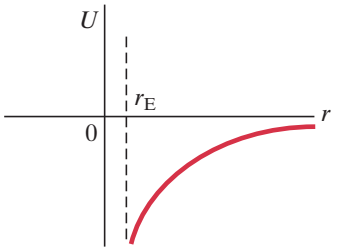


FIGURE 8–19 Arbitrary path of particle of mass m moving from point 1 to point 2.

FIGURE 8–20 Gravitational potential energy plotted as a function of r, the distance from Earth's center. Valid only for points $r > r_E$, the radius of the Earth.



EXAMPLE 8–12 **Package dropped from high-speed rocket.** A box of mass m is allowed to fall from a rocket traveling outward from Earth at a speed of 1800 m/s when 1600 km above the Earth's surface. The box eventually falls to the Earth. Estimate its speed just before impact. Ignore air resistance.

APPROACH We use conservation of energy. The box initially has a speed relative to Earth equal to the speed of the rocket from which it falls.

SOLUTION Conservation of energy in this case is expressed by Eq. 8–18:

$$\tfrac{1}{2}mv_1^2 - G\frac{mM_E}{r_1} = \tfrac{1}{2}mv_2^2 - G\frac{mM_E}{r_2}$$

where $v_1 = 1.80 \times 10^3\,\text{m/s}$, $r_1 = (1.60 \times 10^6\,\text{m}) + (6.38 \times 10^6\,\text{m}) = 7.98 \times 10^6\,\text{m}$, and $r_2 = 6.38 \times 10^6\,\text{m}$ (the radius of the Earth). We solve for v_2:

$$\begin{aligned} v_2 &= \sqrt{v_1^2 - 2GM_E\left(\frac{1}{r_1} - \frac{1}{r_2}\right)} \\ &= \sqrt{(1.80 \times 10^3\,\text{m/s})^2 - 2(6.67 \times 10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2)(5.98 \times 10^{24}\,\text{kg}) \times \left(\frac{1}{7.98 \times 10^6\,\text{m}} - \frac{1}{6.38 \times 10^6\,\text{m}}\right)} \\ &= 5320\,\text{m/s}. \end{aligned}$$

NOTE In reality, the falling box's speed will be considerably less than this because of air resistance. Note that the direction of the velocity never entered into this calculation, and this is one of the advantages of the energy method. The rocket could have been heading away from the Earth, or toward it, or at some other angle, and the result would be the same.

Escape Velocity

When an object is projected into the air from the Earth, it will return to Earth unless its speed is very high. But if the speed *is* high enough, it will continue out into space never to return to Earth (barring other forces or collisions). The minimum initial velocity needed to prevent an object from returning to the Earth is called the **escape velocity** from Earth, v_{esc}. To determine v_{esc} from the Earth's surface (ignoring air resistance), we use Eq. 8–18 with $v_1 = v_{esc}$ and $r_1 = r_E = 6.38 \times 10^6\,\text{m}$, the radius of the Earth. Since we want the minimum speed for escape, we need the object to reach $r_2 = \infty$ with merely zero speed, $v_2 = 0$. Applying Eq. 8–18 we have

$$\tfrac{1}{2}mv_{esc}^2 - G\frac{mM_E}{r_E} = 0 + 0$$

or

$$v_{esc} = \sqrt{2GM_E/r_E} = 1.12 \times 10^4\,\text{m/s} \qquad \textbf{(8–19)}$$

or 11.2 km/s. This escape velocity is the same for all objects no matter what their mass is. Note that although a mass can escape from the Earth (or solar system) never to return, nonetheless the force on it due to the Earth's gravitational field $(F \propto 1/r^2)$ is never actually zero for any real (finite) value of r.

EXAMPLE 8–13 **Escaping the Earth or the Moon.** (*a*) Compare the escape velocities of a rocket from the Earth and from the Moon. (*b*) Compare the energies required to launch the rockets. For the Moon, $M_M = 7.35 \times 10^{22}\,\text{kg}$ and $r_M = 1.74 \times 10^6\,\text{m}$, and for Earth, $M_E = 5.98 \times 10^{24}\,\text{kg}$ and $r_E = 6.38 \times 10^6\,\text{m}$.

APPROACH We use Eq. 8–19, replacing M_E and r_E with M_M and r_M for finding v_{esc} from the Moon.

SOLUTION (*a*) Using Eq. 8–19, the ratio of the escape velocities is

$$\frac{v_{esc}(\text{Earth})}{v_{esc}(\text{Moon})} = \sqrt{\frac{M_E}{M_M}\frac{r_M}{r_E}} = 4.7.$$

To escape Earth requires a speed 4.7 times that required to escape the Moon. (*b*) The fuel that must be burned provides energy proportional to v^2 $(K = \frac{1}{2}mv^2)$; so to launch a rocket to escape Earth requires $(4.7)^2 = 22$ times as much energy as to escape from the Moon.

8–8 Power

Power is defined as the *rate at which work is done*. *Average power*, $\overline{P}$, equals the work W done divided by the time t it takes to do it:

$$\overline{P} = \frac{W}{t}. \qquad \textbf{(8–20a)}$$

The *instantaneous power*, P, is

$$P = \frac{dW}{dt}. \qquad \textbf{(8–20b)}$$

The work done in a process is equal to the energy transformed from one type to another (a rock falling), or transferred from one object to another. For example, as the potential energy stored in the spring of Fig. 8–6c is transformed to kinetic energy of the ball, the spring is doing work on the ball. Similarly, when you throw a ball or push a grocery cart, *whenever work is done, energy is being transformed from one type to another, or is being transferred from one object to another*.

Hence we can also say that power is the *rate at which energy is transformed*. Thus, average power can also be defined as:

$$\overline{P} = \frac{W}{t} = \frac{\text{energy transformed}}{\text{time}},$$

and instantaneous power as

$$P = \frac{dE}{dt}. \qquad \textbf{(8–20c)}$$

The power rating of an engine refers to how much chemical or electrical energy can be transformed into mechanical energy per unit of time. In SI units, power is measured in joules per second, and this unit is given a special name, the **watt** (W): 1 W = 1 J/s. We are most familiar with the watt for electrical devices, such as the rate at which a lightbulb or heater changes electric energy into light or thermal energy. But the watt is used for other types of energy transformation as well.

In the British system, the unit of power is the foot-pound per second ($\text{ft} \cdot \text{lb/s}$). For practical purposes a larger unit is often used, the **horsepower**. One horsepower[†] (hp) is defined as 550 $\text{ft} \cdot \text{lb/s}$, which equals 746 W. An engine's power is usually specified in hp or in kilowatts: $1\,\text{kW} \approx 1\frac{1}{3}\,\text{hp}$ $\left(1\,\text{kW} = (1000\,\text{W})/(746\,\text{W/hp}) \approx 1\frac{1}{3}\,\text{hp}\right)$.

To see the distinction between energy and power, consider the following example. A person is limited in the work he or she can do, not only by the total energy required, but also by how fast this energy is transformed: that is, by power. For example, a person may be able to walk a long distance or climb many flights of stairs before having to stop because so much energy has been expended. On the other hand, a person who runs very quickly up stairs may feel exhausted after only a flight or two. He or she is limited in this case by power, the rate at which his or her body can transform chemical energy into mechanical energy.

EXAMPLE 8–14 **Stair-climbing power.** A 60-kg jogger runs up a long flight of stairs in 4.0 s (Fig. 8–21). The vertical height of the stairs is 4.5 m. (*a*) Estimate the jogger's power output in watts and horsepower. (*b*) How much energy did this require?

APPROACH The work done by the jogger is against gravity, and equals $W = mgy$. To find her average output, we divide W by the time it took.

SOLUTION (*a*) The average power output was

$$\overline{P} = \frac{W}{t} = \frac{mgy}{t} = \frac{(60\,\text{kg})\left(9.8\,\text{m/s}^2\right)(4.5\,\text{m})}{4.0\,\text{s}} = 660\,\text{W}.$$

Since there are 746 W in 1 hp, the jogger is doing work at a rate of just under 1 hp. A human cannot do work at this rate for very long.

(*b*) The energy required is $E = \overline{P}t = (660\,\text{J/s})(4.0\,\text{s}) = 2600\,\text{J}$. This equals $W = mgy$.

NOTE The person had to transform more energy than this 2600 J. The total energy transformed by a person or an engine always includes some thermal energy (recall how hot you get running up stairs).



FIGURE 8–21 Example 8–14.

[†]The unit was chosen by James Watt (1736–1819) to specify the power of his steam engines. He found that a good horse can work all day at an average rate of about 360 $\text{ft} \cdot \text{lb/s}$. To not be accused of exaggeration in the sale of his steam engines, he multiplied this by $1\frac{1}{2}$ when he defined the hp.

PHYSICS APPLIED
Power needs of a car

Automobiles do work to overcome the force of friction and air resistance, to climb hills, and to accelerate. A car is limited by the rate it can do work, which is why automobile engines are rated in horsepower or kilowatts. A car needs power most when it is climbing hills and when accelerating. In the next Example, we will calculate how much power is needed in these situations for a car of reasonable size. Even when a car travels on a level road at constant speed, it needs some power just to do work to overcome the retarding forces of internal friction and air resistance. These forces depend on the conditions and speed of the car, but are typically in the range 400 N to 1000 N.

It is often convenient to write the power in terms of the net force $\vec{\mathbf{F}}$ applied to an object and its velocity $\vec{\mathbf{v}}$. Since $P = dW/dt$ and $dW = \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}$ (Eq. 7–7), then

$$P = \frac{dW}{dt} = \vec{\mathbf{F}} \cdot \frac{d\vec{\boldsymbol{\ell}}}{dt} = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}, \qquad \textbf{(8–21)}$$

where $d\vec{\boldsymbol{\ell}}/dt$ is the velocity of the object.

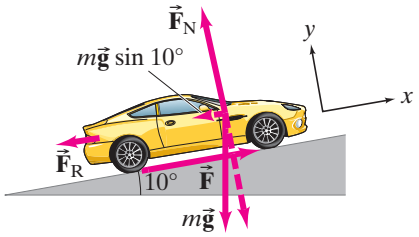


FIGURE 8–22 Example 8–15: Calculation of power needed for a car to climb a hill.

EXAMPLE 8–15 **Power needs of a car.** Calculate the power required of a 1400-kg car under the following circumstances: (*a*) the car climbs a 10° hill (a fairly steep hill) at a steady 80 km/h; and (*b*) the car accelerates along a level road from 90 to 110 km/h in 6.0 s to pass another car. Assume the average retarding force on the car is $F_R = 700\ \text{N}$ throughout. See Fig. 8–22.

APPROACH First we must be careful not to confuse $\vec{\mathbf{F}}_R$, which is due to air resistance and friction that retards the motion, with the force $\vec{\mathbf{F}}$ needed to accelerate the car, which is the frictional force exerted by the road on the tires—the reaction to the motor-driven tires pushing against the road. We must determine the magnitude of the force F before calculating the power.

SOLUTION (*a*) To move at a steady speed up the hill, the car must, by Newton's second law, exert a force F equal to the sum of the retarding force, 700 N, and the component of gravity parallel to the hill, mg sin 10°. Thus

$$\begin{aligned} F &= 700\ \text{N} + mg \sin 10° \\ &= 700\ \text{N} + (1400\ \text{kg})(9.80\ \text{m/s}^2)(0.174) = 3100\ \text{N}. \end{aligned}$$

Since $\bar{v} = 80\ \text{km/h} = 22\ \text{m/s}$[†] and is parallel to $\vec{\mathbf{F}}$, then (Eq. 8–21) the power is

$$\bar{P} = F\bar{v} = (3100\ \text{N})(22\ \text{m/s}) = 6.80 \times 10^4\ \text{W} = 68.0\ \text{kW} = 91\ \text{hp}.$$

(*b*) The car accelerates from 25.0 m/s to 30.6 m/s (90 to 110 km/h) on the flat. The car must exert a force that overcomes the 700-N retarding force plus that required to give it the acceleration

$$\bar{a}_x = \frac{(30.6\ \text{m/s} - 25.0\ \text{m/s})}{6.0\ \text{s}} = 0.93\ \text{m/s}^2.$$

We apply Newton's second law with x being the horizontal direction of motion (no component of gravity):

$$ma_x = \Sigma F_x = F - F_R.$$

We solve for the force required, F:

$$\begin{aligned} F &= ma_x + F_R \\ &= (1400\ \text{kg})(0.93\ \text{m/s}^2) + 700\ \text{N} = 1300\ \text{N} + 700\ \text{N} = 2000\ \text{N}. \end{aligned}$$

Since $P = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}$, the required power increases with speed and the motor in this case must be able to provide a maximum power output of

$$\bar{P} = (2000\ \text{N})(30.6\ \text{m/s}) = 6.12 \times 10^4\ \text{W} = 61.2\ \text{kW} = 82\ \text{hp}.$$

NOTE Even taking into account the fact that only 60 to 80% of the engine's power output reaches the wheels, it is clear from these calculations that an engine of 75 to 100 kW (100 to 130 hp) is adequate from a practical point of view.

We mentioned in Example 8–15 that only part of the energy output of a car engine reaches the wheels. Furthermore, in the engine itself, most of the energy from the burning of fuel does not do useful work. An important aspect

[†]Recall 1 km/h = 1000 m/3600 s = 0.278 m/s.

of all engines is their overall *efficiency* e, defined as the ratio of the useful power output of the engine, P_{out}, to the power input, P_{in} (provided by burning of fuel):

$$e = \frac{P_{\text{out}}}{P_{\text{in}}}.$$

The efficiency is always less than 1.0 because no engine can create energy, and no engine can even transform energy from one form to another without some energy going to friction and thermal energy. For example, a car engine converts chemical energy released in the burning of fuel into mechanical energy that turns the wheels. Nearly 85% of the input energy is "wasted" as thermal energy that goes into the environment via the exhaust pipe (and radiator). Thus car engines are roughly only about 15% efficient. We discuss efficiency in detail in Chapter 20.

8–9 Potential Energy Diagrams; Stable and Unstable Equilibrium

If only conservative forces do work on an object, we can learn a great deal about its motion simply by examining a potential energy diagram—the graph of $U(x)$ versus x. An example of a potential energy diagram is shown in Fig. 8–23. The rather complex curve represents some complicated potential energy $U(x)$. The total energy $E = K + U$ is constant and can be represented as a horizontal line on this graph. Four interesting values for E are shown, labeled E_0, E_1, E_2, and E_3. What the actual value of E will be for a given system depends on the initial conditions. (For example, the total energy E of a mass oscillating on the end of a spring depends on the amount the spring is initially compressed or stretched.) Kinetic energy $K = \frac{1}{2}mv^2$ cannot be less than zero (v would be imaginary), and because $E = U + K =$ constant, then $U(x)$ must always be less than or equal to E for all situations: $U(x) \leq E$. The minimum value that the total energy can have for the case shown in Fig. 8–23 is labeled E_0. For this value of E, the mass can only be at rest at $x = x_0$. The system has potential energy but no kinetic energy at this position.

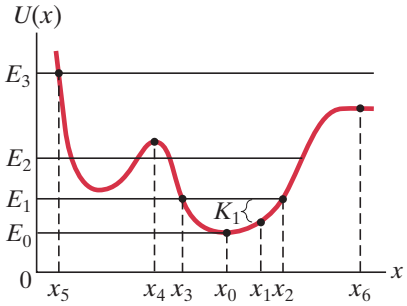


FIGURE 8–23 A potential energy diagram.

If the system's total energy E is greater than E_0, say it is E_1 on our plot, the system can have both kinetic and potential energy. Because energy is conserved,

$$K = E - U(x)$$

at any point x. Since the curve represents $U(x)$ at each x, the kinetic energy at any value of x is represented by the distance between the E line and the curve $U(x)$ at that value of x. In Fig. 8–23, the kinetic energy for an object at x_1, when its total energy is E_1, is indicated by the notation K_1.

An object with energy E_1 can oscillate only between the points x_2 and x_3. This is because if $x > x_2$ or $x < x_3$, the potential energy would be greater than E, meaning $K = \frac{1}{2}mv^2 < 0$ and v would be imaginary, and so impossible. At x_2 and x_3 the speed is zero, since $E = U$ at these points. Hence x_2 and x_3 are called the **turning points** of the motion. If the object is at x_0, say, moving to the right, its kinetic energy (and speed) decreases until it reaches zero at $x = x_2$. The object then reverses direction, proceeding to the left and increasing in speed until it passes x_0 again. It continues to move, decreasing in speed until it reaches $x = x_3$, where again $v = 0$, and the object again reverses direction.

If the object has energy $E = E_2$ in Fig. 8–23, there are four turning points. The object can move in only one of the two potential energy "valleys," depending on where it is initially. It cannot get from one valley to the other because of the barrier between them—for example at a point such as x_4, $U > E_2$, which means v would be imaginary.[†] For energy E_3, there is only one turning point since $U(x) < E_3$ for all $x > x_5$. Thus, if our object has $E = E_3$ and is moving initially to the left, it varies in speed as it passes the potential valleys (or **potential wells**) but eventually stops and turns around at $x = x_5$. The object then proceeds to the right indefinitely, never to return.

How do we know the object reverses direction at the turning points? Because of the force exerted on it. The force F is related to the potential energy U by Eq. 8–7, $F = -dU/dx$, so the force F equals the negative of the slope of the U-versus-x curve at any point x. At $x = x_2$ in Fig. 8–23, for example, the slope is positive so the force is negative, which means it acts to the left (toward decreasing values of x).

[†]Although this is true according to Newtonian physics, modern quantum mechanics predicts that objects can "tunnel" through such a barrier, and such processes have been observed at the atomic and subatomic level.

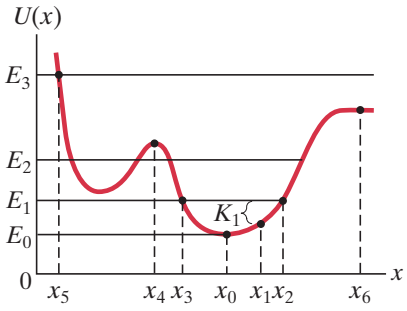


FIGURE 8–23 (Repeated.) A potential energy diagram.

At $x = x_0$ the slope is zero, so $F = 0$. At such a point the particle is said to be in **equilibrium**. This term means simply that the net force on the object is zero. Hence, its acceleration is zero, and so if it is initially at rest, it remains at rest. If the object at rest at $x = x_0$ were moved slightly to the left or right, a nonzero force would act on it in the direction to move it back toward x_0. An object that returns toward its equilibrium point when displaced slightly is said to be at a point of **stable equilibrium**. Any *minimum* in the potential energy curve represents a point of stable equilibrium.

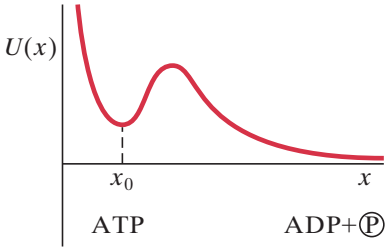
An object at $x = x_4$ would also be in equilibrium, since $F = -dU/dx = 0$. If the object were displaced a bit to either side of x_4, a force would act to pull the object *away* from the equilibrium point. Points like x_4, where the potential energy curve has a maximum, are points of **unstable equilibrium**. The object will *not* return to equilibrium if displaced slightly, but instead will move farther away.

When an object is in a region over which U is constant, such as near $x = x_6$ in Fig. 8–23, the force is zero over some distance. The object is in equilibrium and if displaced slightly to one side the force is still zero. The object is said to be in **neutral equilibrium** in this region.

*ATP – Storing Energy in Living Cells

Figure 8–24 is a simple potential energy diagram for a bond between a phosphate molecule (Ⓟ) and a second molecule (ADP) to form a molecule of ATP, **a**denosine **tri**phosphate. ATP is an important molecule in biological cells that stores energy for later use within cells. When the separation x between the 2 molecules is large (let $x \to \infty$ = infinity) the phosphate and ADP are separate and we choose $U(\infty) = 0$ in Fig. 8–24. With the help of other "catalyst" molecules, these two molecules can get close enough to be bonded together to form ATP, whose potential energy at the stable position $x = x_0$ is greater: $U(x_0) > U(\infty) = 0$. The bump of $U(x)$ on the right is a "barrier" that keeps ATP stable. ATP will not break apart unless an unlikely collision occurs that can provide massive kinetic energy so that the total energy becomes greater than the barrier height. Or, much more likely, certain other molecules (enzymes) come close which are able to lower the barrier, so the bond is broken and the phosphate can be freed, which releases the "stored" energy, $\Delta U = U(x_0) - U(\infty) > 0$. The energy released can be used to help crucial biochemical reactions take place in the cell.

FIGURE 8–24 Potential-energy diagram for the formation of ATP from ADP and phosphate (Ⓟ).



*8–10 Gravitational Assist (Gravitational Slingshot)

Spacecraft heading to other planets, or the outer reaches of the solar system and beyond, are slowed by the Sun's gravitational field but can have their speed increased by passing near a planet or a planet's moon. This is an important way to reduce the fuel a spacecraft carries and reduce the energy required of the rocket to accelerate all that fuel (in the spacecraft) at launch (photo, Fig. 4–10).

In Fig. 8–25a we are in the reference frame of the planet Jupiter (J), at rest relative to Jupiter's position. The spacecraft, which has kinetic energy K_{J0} when far from Jupiter (lower part Fig. 8–25a), is accelerated toward the planet by gravity and its trajectory bends toward the planet as its speed increases. As the spacecraft passes nearest the planet it reaches its maximum speed, v_J, and then as it moves away it slows down due to the planet's gravitational field pulling back on it. When far away again, and going in a different direction, its kinetic energy is again K_{J0}. The spacecraft's trajectory is *symmetric* here in *Jupiter's reference frame* about our chosen x axis. (We ignore the change in the planet's velocity because its mass M is vastly greater than the spacecraft's, m.) The subscript J means in Jupiter's reference frame; the subscript 0, such as $\vec{\mathbf{v}}_{J0}$, means the velocity of the spacecraft when far from Jupiter so that Jupiter's gravitational attraction is close to zero; the prime ($'$), such as $\vec{\mathbf{v}}'_{J0}$, means the velocity *after* the spacecraft has passed Jupiter, *after* the encounter.

Next we look at this encounter as *we* see it, in *our* reference frame of the Sun (and the Solar System as a whole). Figure 8–25b shows Jupiter moving to the right with velocity $\vec{\mathbf{u}}$ in its orbit around the Sun, at this moment taken to be along

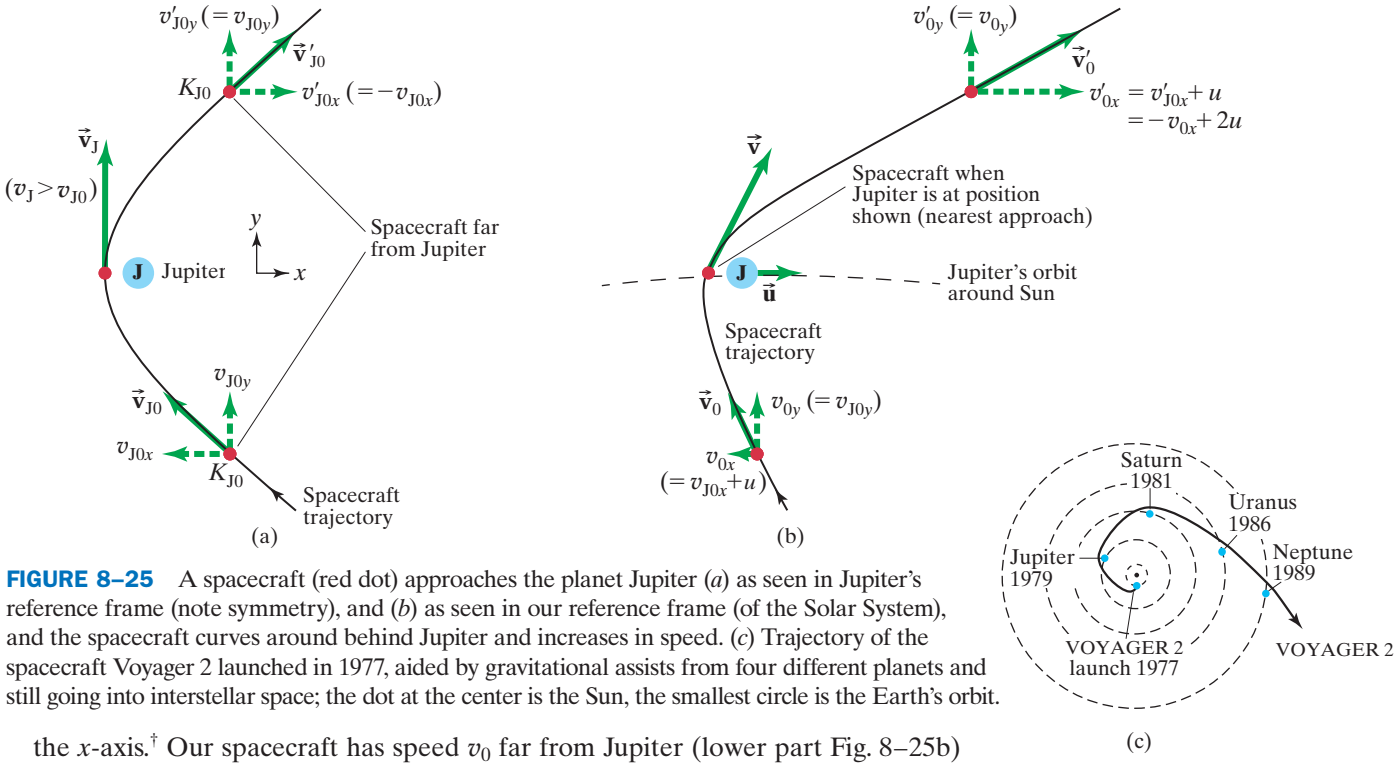


FIGURE 8–25 A spacecraft (red dot) approaches the planet Jupiter (*a*) as seen in Jupiter's reference frame (note symmetry), and (*b*) as seen in our reference frame (of the Solar System), and the spacecraft curves around behind Jupiter and increases in speed. (*c*) Trajectory of the spacecraft Voyager 2 launched in 1977, aided by gravitational assists from four different planets and still going into interstellar space; the dot at the center is the Sun, the smallest circle is the Earth's orbit.

the x-axis.† Our spacecraft has speed v_0 far from Jupiter (lower part Fig. 8–25b) and speeds up due to Jupiter's gravity, having velocity $\vec{\mathbf{v}}$ at closest approach to Jupiter. Later, far from Jupiter (upper right), the spacecraft has speed $v'_0 (> v_0)$, as shown. Carefully inspect velocity components explained in Fig. 8–25b, and note that v_{J0x} and v_{0x} are negative because they point along the negative x axis.

Let us calculate the change in kinetic energy of the spacecraft as a result of its encounter with Jupiter. We want to do this in our Solar System reference frame, Fig. 8–25b. First we need to look at the x-components of the spacecraft velocity (the y components don't change). The spacecraft's "final" x-component of velocity, v'_{0x}, in terms of its "initial" incoming value, v_{0x}, is equal to its value in the Jupiter frame (J) plus the relative velocity u of the two reference frames:

$$\begin{aligned} v'_{0x} &= v'_{J0x} + u \\ &= -v_{J0x} + u \qquad \text{[see upper part Fig. 8–25a]} \\ &= -(v_{0x} - u) + u = -v_{0x} + 2u \qquad \text{[see lower part Fig. 8–25b]} \end{aligned} \tag{8–22}$$

which is summarized in the upper part of Fig. 8–25b. The subscript J refers to Jupiter's reference frame; the subscripts $(_0)$ and $('_0)$ refer to the spacecraft far from Jupiter before and after the encounter. The change in the kinetic energy, final minus initial, is

$$\begin{aligned} \Delta K &= \tfrac{1}{2}m(v'^2_{0x} + v^2_{0y}) - \tfrac{1}{2}m(v^2_{0x} + v^2_{0y}) \\ &= \tfrac{1}{2}m(v^2_{0x} - 4uv_{0x} + 4u^2 - v^2_{0x}) = 2mu(u - v_{0x}). \end{aligned} \tag{8–23}$$

This result, Eq. 8–23, gives an increase in the spacecraft's kinetic energy because $v_{0x} < 0$ (Fig. 8–25b, bottom).

Energy is conserved because the planet (J) experiences an equal *loss* in kinetic energy, which affects the planet's motion unnoticeably because of its enormous mass.

If you want to *slow down* a spacecraft (perhaps to put it into orbit around a planet or moon), you will want to project it so it passes *in front* of the planet (J) and v_{0x} must be in the *same* direction as $\vec{\mathbf{u}}$ (> 0) and have magnitude greater than u. (See Problem 77.)

Figure 8–25c shows the trajectory of the Voyager 2 probe, launched in 1977 when the planets Jupiter, Saturn, Uranus and Neptune were in optimal positions to each provide an "assist", a rare alignment that will not happen again until the 22nd century. After all these decades, it is still able to send back data. It has now gone beyond the Solar System and entered interstellar space!

†In general, the velocity $\vec{\mathbf{u}}$ in Fig. 8–25b can have a y-component. We chose $\vec{\mathbf{u}}$ along the x axis to keep the math simpler so the effect of gravitational assist is readily seen.

Summary

A **conservative force** is one for which the work done by the force in moving an object from one position to another depends only on the two positions and not on the path taken. The work done by a conservative force is recoverable, which is not true for nonconservative forces, such as friction.

Potential energy, U, is energy associated with conservative forces that depend on the position or configuration of objects. Gravitational potential energy is

$$U_{\text{grav}} = mgy, \qquad \textbf{(8–3)}$$

where the mass m is near the Earth's surface, a height y above some reference point. Elastic potential energy is given by

$$U_{\text{el}} = \tfrac{1}{2}kx^2 \qquad \textbf{(8–5)}$$

for a spring with stiffness constant k stretched or compressed a displacement x from its equilibrium position. Other potential energies include chemical, electrical, and nuclear energy.

Potential energy is always associated with a conservative force, and the change in potential energy, ΔU, between two positions under the action of a conservative force $\vec{\mathbf{F}}$ is defined as the negative of the work done by the force:

$$\Delta U = U_2 - U_1 = -\int_1^2 \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}. \qquad \textbf{(8–4)}$$

Inversely, we can write, for the one-dimensional case,

$$F = -\frac{dU(x)}{dx}. \qquad \textbf{(8–7)}$$

Only *changes* in potential energy are physically meaningful, so the position where $U = 0$ can be chosen for convenience. Potential energy is not a property of an object but is associated with the interaction of two or more objects.

When only conservative forces act, the total **mechanical energy**, E, defined as the sum of kinetic and potential energies, is conserved:

$$E = K + U = \text{constant}. \qquad \textbf{(8–10)}$$

If nonconservative forces also act, additional types of energy are involved, such as thermal energy. It has been found experimentally that, when all forms of energy are included, the total energy is conserved. This is the **law of conservation of energy**:

$$\Delta K + \Delta U + \Delta(\text{other energy types}) = 0. \qquad \textbf{(8–14)}$$

The gravitational force as described by Newton's law of universal gravitation is a conservative force. The potential energy of an object of mass m due to the gravitational force exerted on it by the Earth is given by

$$U(r) = -\frac{GmM_{\text{E}}}{r}, \qquad \textbf{(8–17)}$$

where M_{E} is the mass of the Earth and r is the distance of the object from the Earth's center ($r \geq$ radius of earth).

Power is defined as the rate at which work is done or the rate at which energy is transformed from one form to another, or transferred from one object to another:

$$P = \frac{dW}{dt} = \frac{dE}{dt}, \qquad \textbf{(8–20)}$$

or

$$P = \vec{\mathbf{F}} \cdot \vec{\mathbf{v}}. \qquad \textbf{(8–21)}$$

The SI unit of power is the **watt** $(1\,\text{W} = 1\,\text{J/s})$.

Questions

1. List some everyday forces that are not conservative, and explain why they aren't.
2. You lift a heavy book from a table to a high shelf. List the forces on the book during this process, and state whether each is conservative or nonconservative.
3. Analyze the motion of a simple swinging pendulum in terms of energy, (*a*) ignoring friction, and (*b*) taking friction into account. (*c*) Explain why a grandfather clock has to be wound up.
4. Describe precisely what is "wrong" physically in the famous Escher drawing shown in Fig. 8–26.

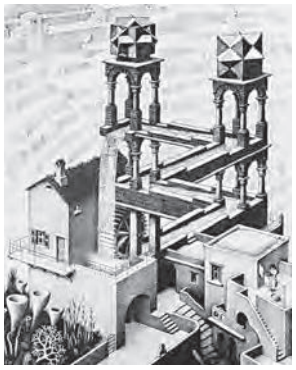


FIGURE 8–26 Question 4.

5. A coil spring of mass m rests upright on a table. If you compress the spring by pressing down with your hand and then release it, can the spring leave the table? Explain using the law of conservation of energy.
6. Experienced hikers prefer to step over a fallen log in their path rather than stepping on top and jumping down on the other side. Explain.
7. (*a*) Where does the kinetic energy come from when a car accelerates uniformly starting from rest? (*b*) How is the increase in kinetic energy related to the friction force the road exerts on the tires?
8. Can the total mechanical energy $E = K + U$ ever be negative? Explain.
9. Describe the energy transformations when a child hops around on a pogo stick.
10. (*a*) Only two conservative forces are discussed in this Chapter. What are they, and how are they accounted for when you are dealing with conservation of energy? (*b*) Not mentioned is the force of water on a swimmer. Is it conservative or nonconservative?
11. Recall from Chapter 4, Example 4–14, that you can use a *pulley* and ropes to decrease the force needed to raise a heavy load (see Fig. 8–27). But for every meter the piano is raised, how much rope must be pulled up at the top of the building? Account for this, using energy concepts.

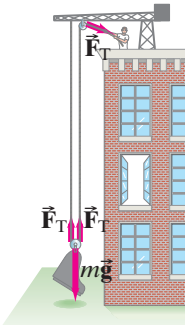


FIGURE 8–27 Question 11

12. Two identical arrows, one with twice the speed of the other, are fired into a bale of hay. Assuming the hay exerts a constant "frictional" force on the arrows, the faster arrow will penetrate how much farther than the slower arrow? Explain.

13. In Fig. 8–28, water balloons are tossed from the roof of a building, all with the same speed but with different launch angles. Which one has the highest speed when it hits the ground? Ignore air resistance, and explain your answer.

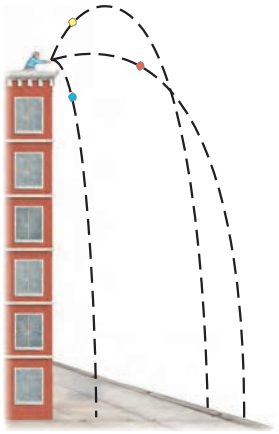


FIGURE 8–28
Question 13.

14. Suppose that you wish to launch a rocket from the surface of the Earth so that it escapes the Earth's gravitational field. You wish to use minimum fuel in doing this. From what point on the surface of the Earth should you make the launch and in what direction? Do the launch location and direction matter? Explain.

15. Suppose you lift a suitcase from the floor to a table. The work you do on the suitcase depends on which of the following: (*a*) whether you lift it straight up or along a more complicated path, (*b*) the time the lifting takes, (*c*) the height of the table, and/or (*d*) the weight of the suitcase? Explain.

16. Repeat Question 15 for the *power* needed instead of the work.

17. Why is it easier to climb a mountain via a zigzag trail rather than to climb straight up?

18. A pendulum is launched from a point that is a height h above its lowest point in two different ways (Fig. 8–29). During both launches, the pendulum is given an initial speed of 3.0 m/s. On the first launch, the initial velocity of the pendulum is directed upward along the "swing," and on the second launch it is directed downward along the swing. Which launch will cause the highest speed when the pendulum bob passes the lowest point of its swing, ignoring air resistance? Explain.

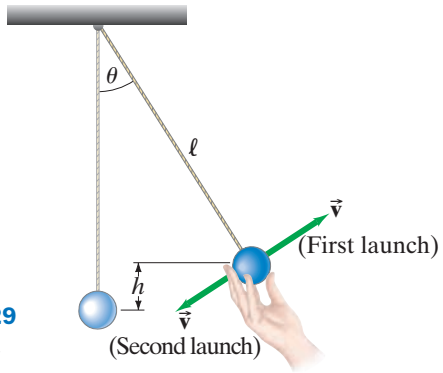


FIGURE 8–29
Question 18.

19. Figure 8–30 shows a potential energy curve, $U(x)$. (*a*) At which point does the force have greatest magnitude? (*b*) For each labeled point, state whether the force acts to the left or to the right, or is zero. (*c*) Where is there equilibrium and of what type is it?

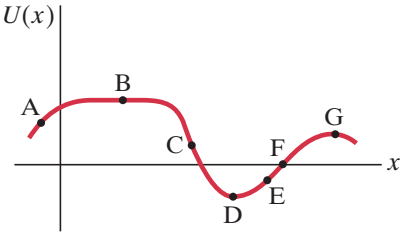


FIGURE 8–30
Question 19.

20. (*a*) Describe in detail the velocity changes of a particle that has energy E_3 in Fig. 8–23 as it moves from x_6 to x_5 and back to x_6. (*b*) Where is its kinetic energy the greatest and the least? Explain.

MisConceptual Questions

1. A bowling ball is dropped from a height h onto the center of a trampoline, which launches the ball back up into the air. How high will the ball rise?
   (*a*) Significantly less than h.
   (*b*) More than h. The exact amount depends on the mass of the ball and the springiness of the trampoline.
   (*c*) No more than h—probably a little less.
   (*d*) Cannot tell without knowing the characteristics of the trampoline.

2. A ball is thrown straight up. At what point does the ball have the most energy? Ignore air resistance.
   (*a*) At the highest point of its path.
   (*b*) When it is first thrown.
   (*c*) Just before it hits the ground.
   (*d*) When the ball is halfway to the highest point of its path.
   (*e*) Everywhere; the energy of the ball is the same at all of these points.

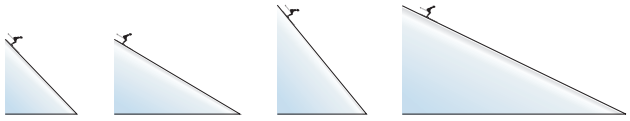
3. A man pushes a block up an incline at a constant speed. As the block moves up the incline,
   (*a*) its kinetic energy and potential energy both increase.
   (*b*) its kinetic energy increases and its potential energy remains the same.
   (*c*) its potential energy increases and its kinetic energy remains the same.
   (*d*) its potential energy increases and its kinetic energy decreases by the same amount.

4. As a skateboard rider rolls up a hill and slows down, she
   (*a*) gains kinetic energy (energy of motion).
   (*b*) gains gravitational potential energy (energy of position).
   (*c*) loses all of her original energy.
   (*d*) can't lose kinetic energy because energy remains constant.

5. A ball is thrown straight up. Neglecting air resistance, which statement is *not* true regarding the energy of the ball?
   (*a*) The potential energy decreases while the ball is going up.
   (*b*) The kinetic energy decreases while the ball is going up.
   (*c*) The sum of the kinetic energy and potential energy is constant.
   (*d*) The potential energy decreases when the ball is coming down.
   (*e*) The kinetic energy increases when the ball is coming down.

6. Two balls are thrown off a building with the same speed, one straight up and one at a 45° angle. Which statement is true if air resistance can be ignored?
   (*a*) Both hit the ground at the same time.
   (*b*) Both hit the ground with the same speed.
   (*c*) The one thrown at an angle hits the ground with a lower speed.
   (*d*) The one thrown at an angle hits the ground with a higher speed.
   (*e*) Both (*a*) and (*b*).

7. A skier starts from rest at the top of each of the hills shown in Fig. 8–31. On which hill will the skier have the highest speed at the bottom if we ignore friction: (*a*), (*b*), (*c*), (*d*), or (*e*) *c* and *d* equally?



(*a*) (*b*) (*c*) (*d*)

FIGURE 8–31 MisConceptual Questions 7 and 8.

8. Answer MisConceptual Question 7 assuming a small amount of friction.
9. A child's swing is pulled back by an angle θ and released from rest. If the child on the swing remains motionless (relative to the swing seat), what is the maximum angle the swing can reach on the other side? Neglect air resistance and friction.
   (*a*) The angle depends on the length of the swing, which is not specified.
   (*b*) The exact same angle, θ.
   (*c*) An angle less than θ.
   (*d*) An angle greater than θ.
10. Answer Question 9 taking into account air resistance and/or friction at the pivot point of the chain supporting the swing seat.
11. A person stands at the middle of a tall ladder, which is leaning against a wall. If the reference point $y = 0$ is taken to be at the top of the ladder, the person's gravitational potential energy is
   (*a*) positive.
   (*b*) negative.
   (*c*) zero.
12. What maximum work can a 250-W motor do?
   (*a*) Cannot tell—not enough information is given.
   (*b*) 250 J.
   (*c*) 0.335 hp.
   (*d*) 250 W.
   (*e*) Answers (*b*), (*c*), and (*d*) are all correct.
13. Is it possible for a system to have negative potential energy?
   (*a*) No, because the kinetic energy of a system must equal its potential energy.
   (*b*) No, because this would have no physical meaning.
   (*c*) Yes, since the choice of the zero of potential energy is arbitrary.
   (*d*) Yes, as long as the kinetic energy is positive.
   (*e*) Yes, as long as the total energy is positive.

Problems

8–1 and 8–2 Conservative Forces and Potential Energy

1. (I) By how much does the gravitational potential energy of a 58-kg pole vaulter change if her center of mass rises 4.0 m during the jump?
2. (I) You drop a basketball from a height of 5.0 m, which rebounds to a new height of 3.0 m. How much energy is lost to nonconservative forces? Give your answer as percentage of the initial energy.
3. (I) A spring has a spring constant k of 78.0 N/m. How much must this spring be compressed to store 45.0 J of potential energy?
4. (II) A 66.5-kg hiker starts at an elevation of 1150 m and climbs to the top of a peak 2660 m high. (*a*) What is the hiker's change in potential energy? (*b*) What is the minimum work the hiker must do? (*c*) Can the actual work done be greater than this? Explain.
5. (II) If it requries 7.5 J of work to stretch a particular spring by 2.0 cm from its equilibrium length, how much more work will be required to stretch it an additional 4.0 cm?
6. (II) A 1.60-m-tall person lifts a 1.65-kg book off the ground so it is 2.20 m above the ground. What is the potential energy of the book relative to (*a*) the ground, and (*b*) the top of the person's head? (*c*) How is the work done by the person related to the answers in parts (*a*) and (*b*)?
7. (II) A particle is constrained to move in one dimension along the x axis and is acted upon by a force given by

$$\vec{\mathbf{F}}(x) = -\frac{k}{x^3}\,\hat{\mathbf{i}}$$

where k is a constant with units appropriate to the SI system. Find the potential energy function $U(x)$, if U is arbitrarily defined to be zero at $x = 2.0\,\text{m}$, so that $U(2.0\,\text{m}) = 0$.

*8. (II) If $U = 3x^2 + 2xy + 4y^2z$, what is the force, $\vec{\mathbf{F}}$?
9. (II) A particular spring obeys the force law $\vec{\mathbf{F}} = (-kx + ax^3 + bx^4)\,\hat{\mathbf{i}}$, where k, a and b are constants. (*a*) Is this force conservative? Explain why or why not. (*b*) If it is conservative, determine the form of the potential energy function.
10. (III) The spring in Problem 9 is stretched out to a postition x_0 from the equilibrium position. A mass m is attached to the spring, and then released from rest. What is the speed of m at (*a*) $x = x_0$, (*b*) $x = x_0/2$, and (*c*) $x = 0$? (*d*) How do these speeds compare to the "ideal" spring, with $a = 0$ and $b = 0$? Explain your answer.

8–3 and 8–4 Conservation of Mechanical Energy

11. (I) A novice skier, starting from rest, slides down an icy frictionless 8.0° incline whose vertical height is 115 m. How fast is she going when she reaches the bottom?
12. (I) Jane, looking for Tarzan, is running at top speed (5.0 m/s) and grabs a vine hanging vertically from a tall tree in the jungle. How high can she swing upward? Does the length of the vine affect your answer?
13. (II) A sled is initially given a shove up a frictionless 18.0° incline. It reaches a maximum vertical height 1.22 m higher than where it started at the bottom. What was its initial speed?
14. (II) In the *high jump*, the kinetic energy of an athlete is transformed into gravitational potential energy without the aid of a pole. With what minimum speed must the athlete leave the ground in order to lift his center of mass 2.10 m and cross the bar with a speed of 0.50 m/s?

15. (II) A spring with $k = 88\ \text{N/m}$ hangs vertically next to a vertical ruler. The end of the spring is next to the 15-cm mark on the ruler. If a 2.5-kg mass is now attached to the end of the spring, and the mass is allowed to fall freely, where will the end of the spring line up with the ruler marks when the mass is at its lowest position?

16. (II) A 0.40-kg ball is thrown with a speed of 7.8 m/s at an upward angle of 36°. (*a*) What is its speed at its highest point, and (*b*) how high does it go? Use conservation of energy for part *b*.

17. (II) A 1400-kg car moving on a horizontal surface has speed $v = 85\ \text{km/h}$ when it strikes a horizontal coiled spring and is brought to rest in a distance of 2.2 m. What is the spring constant of the spring? Ignore any thermal energy produced in the collision.

18. (II) A 62-kg trampoline artist jumps upward from the top of a platform with a vertical speed of 4.5 m/s. (*a*) How fast is he going as he lands on the trampoline, 2.0 m below (Fig. 8–32)? (*b*) If the trampoline behaves like a spring of spring constant $5.8 \times 10^4\ \text{N/m}$, how far does he depress it?

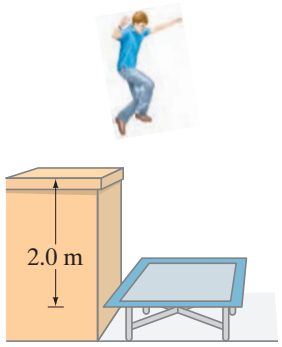


FIGURE 8–32
Problem 18.

19. (II) A vertical spring (ignore its mass), whose spring constant is 875 N/m, is attached to a table and is compressed down by 0.220 m. (*a*) What upward speed can it give to a 0.380-kg ball when released? (*b*) How high above its original position (spring compressed) will the ball fly?

20. (II) A roller-coaster car shown in Fig. 8–33 is pulled up to point 1 where it is released from rest. Assuming no friction, calculate the speed at points 2, 3, and 4.

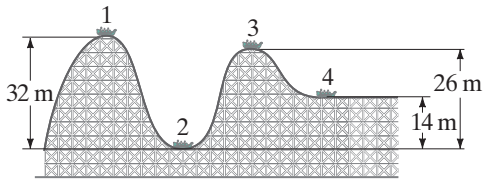


FIGURE 8–33 Problems 20 and 31.

21. (II) When a mass *m* sits at rest on a spring, the spring is compressed by a distance *d* from its undeformed length (Fig. 8–34a). Suppose instead that the mass is released from rest when it barely touches the undeformed spring (Fig. 8–34b). Find the distance *D* that the spring is compressed before it is able to stop the mass. Does $D = d$? If not, why not?

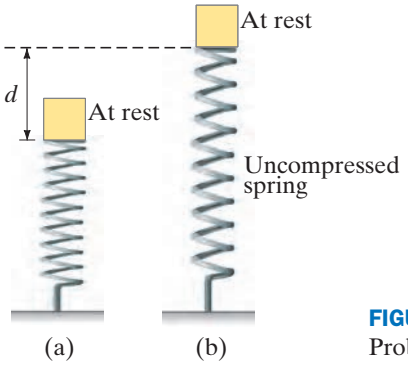


FIGURE 8–34
Problem 21.

22. (II) Two masses are connected by a string as shown in Fig. 8–35. Mass $m_A = 3.5\ \text{kg}$ rests on a frictionless inclined plane, while $m_B = 5.0\ \text{kg}$ is initially held at a height of $h = 0.75\ \text{m}$ above the floor. (*a*) If m_B is allowed to fall, what will be the resulting acceleration of the masses? (*b*) If the masses were initially at rest, use the kinematic equations (Eqs. 2–12) to find their velocity just before m_B hits the floor. (*c*) Use conservation of energy to find the velocity of the masses just before m_B hits the floor. You should get the same answer as in part (*b*).

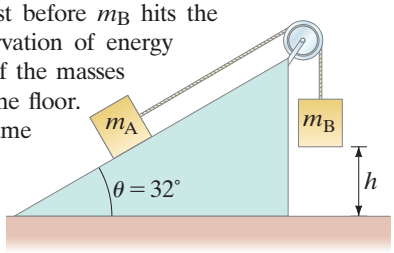


FIGURE 8–35
Problem 22.

23. (II) The 9.0-kg mass in Fig. 8–36 is held just barely in contact with a spring for which $k = 450\ \text{N/m}$. When that mass is released, it falls, compressing the spring and pulling the 3.0-kg mass up. How far does the 9.0-kg mass fall before momentarily coming to rest? Ignore friction in the pulley.

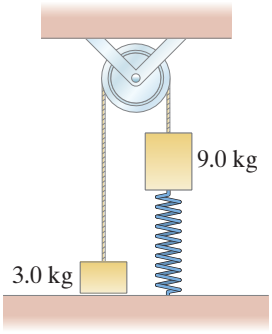


FIGURE 8–36
Problem 23.

24. (II) Chris jumps off a bridge with a 15-m-long *bungee cord* (a heavy stretchable cord) tied around his ankle, Fig. 8–37. He falls 15 m before the bungee cord begins to stretch. Chris's mass is 75 kg and we assume the cord obeys Hooke's law, $F = -kx$, with $k = 55\ \text{N/m}$. If we neglect air resistance, estimate what distance *d* below the bridge Chris's foot will be before coming to a stop. Ignore the mass of the cord (not realistic, however) and treat Chris as a particle.

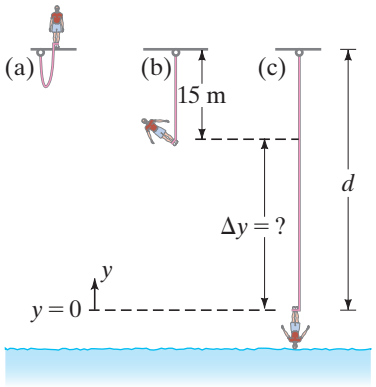


FIGURE 8–37
Problem 24. (a) Bungee jumper about to jump. (b) Bungee cord at its unstretched length. (c) Maximum stretch of cord.

25. (III) An engineer is designing a spring to be placed at the bottom of an elevator shaft. If the elevator cable breaks when the elevator is at a height *h* above the top of the spring, calculate the value that the spring constant *k* should have so that passengers undergo an acceleration of no more than 4.5 *g* when being brought to rest. Let *M* be the total mass of the elevator and passengers.

26. (III) A skier of mass m starts from rest at the top of a solid sphere of radius r and slides down its frictionless surface. (a) At what angle θ (Fig. 8–38) will the skier leave the sphere? (b) If friction is present, does the skier fly off at a greater or lesser angle?

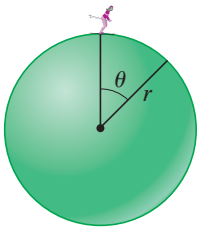


FIGURE 8–38
Problem 26.

8–5 and 8–6 Law of Conservation of Energy

27. (I) Two railroad cars, each of mass 72,000 kg, are traveling 85 km/h toward each other. They collide head-on and come to rest. How much thermal energy is produced in this collision?

28. (I) A 16.0-kg child descends a slide 2.20 m high and, starting from rest, reaches the bottom with a speed of 1.15 m/s. How much thermal energy due to friction was generated in this process?

29. (II) A ski starts from rest and slides down a 28° incline 85 m long. (a) If the coefficient of friction is 0.090, what is the ski's speed at the base of the incline? (b) If the snow is level at the foot of the incline and has the same coefficient of friction, how far will the ski travel along the level? Use energy methods.

30. (II) A 145-g baseball is dropped from a tree 14.0 m above the ground. (a) With what speed would it hit the ground if air resistance could be ignored? (b) If it actually hits the ground with a speed of 12.5 m/s, what is the average force of air resistance exerted on it?

31. (II) Suppose the roller-coaster car in Fig. 8–33 passes point 1 with a speed of 1.30 m/s. If the average force of friction is equal to 0.23 of its weight, with what speed will it reach point 2? The distance traveled is 45.0 m.

32. (II) An 18-kg sled starts up a 28° incline with a speed of 2.3 m/s. The coefficient of kinetic friction is $\mu_k = 0.25$. (a) How far up the incline does the sled travel? (b) What condition must you put on the coefficient of static friction if the sled is not to get stuck at the point determined in part (a)? (c) If the sled slides back down, what is its speed when it returns to its starting point?

33. (II) A projectile is fired at an upward angle of 34.0° from the top of a 135-m-high cliff with a speed of 165 m/s. What will be its speed when it strikes the ground below? (Use conservation of energy.)

34. (II) A *Lunar Module* could make a safe landing on the Moon if its vertical velocity at impact is 3.0 m/s or less. Suppose that you want to determine the greatest height h at which the pilot could shut off the engine if the velocity of the lander relative to the surface is (a) zero; (b) 2.0 m/s downward; (c) 2.0 m/s upward. Use conservation of energy to determine h in each case. The acceleration due to gravity at the surface of the Moon is 1.62 m/s^2.

35. (II) You slide down an 8.0-m-high icy hill ($\approx$ frictionless). At the bottom is a level stretch where the coefficient of kinetic friction is 0.30. How far would you travel across the level stretch?

36. (II) Consider the track shown in Fig. 8–39. The section AB is one quadrant of a circle of radius 2.0 m and is frictionless. B to C is a horizontal span 3.0 m long with a coefficient of kinetic friction $\mu_k = 0.25$. The section CD under the spring is frictionless. A block of mass 1.0 kg is released from rest at A. After sliding on the track, it compresses the spring by 0.20 m. Determine: (a) the velocity of the block at point B; (b) the thermal energy produced as the block slides from B to C; (c) the velocity of the block at point C; (d) the stiffness constant k for the spring.

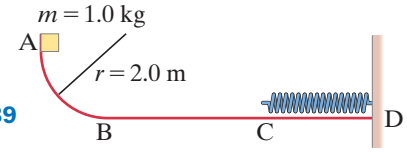


FIGURE 8–39
Problem 36.

37. (II) A 0.680-kg wood block is firmly attached to a very light horizontal spring ($k = 180$ N/m) as shown in Fig. 8–40. This block–spring system, when compressed 5.0 cm and released, stretches out 2.3 cm beyond the equilibrium position before stopping and turning back. What is the coefficient of kinetic friction between the block and the table?

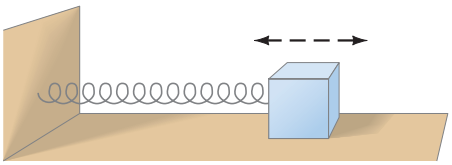


FIGURE 8–40
Problem 37.

38. (III) A spring ($k = 75$ N/m) has an equilibrium length of 1.00 m. The spring is compressed to a length of 0.50 m and a mass of 2.0 kg is placed at its free end on a frictionless slope which makes an angle of 41° with respect to the horizontal (Fig. 8–41). The spring is then released. (a) If the mass is *not* attached to the spring, how far up the slope will the mass move before coming to rest? (b) If the mass *is* attached to the spring, how far up the slope will the mass move before coming to rest? (c) Now the incline has a coefficient of kinetic friction μ_k. If the block, attached to the spring, is observed to stop just as it reaches the spring's equilibrium position, what is the coefficient of friction μ_k?

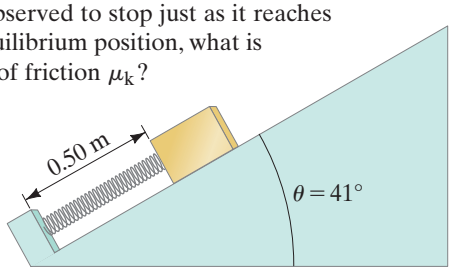


FIGURE 8–41
Problem 38.

39. (III) A 2.0-kg block slides along a horizontal surface with a coefficient of kinetic friction $\mu_k = 0.30$. The block has a speed $v = 1.6$ m/s when it strikes a massless spring head-on (as in Fig. 8–17). (a) If the spring has force constant $k = 120$ N/m, how far is the spring compressed? (b) What minimum value of the coefficient of static friction, μ_s, will assure that the spring remains compressed at the maximum compressed position? (c) If μ_s is less than this, what is the speed of the block when it detaches from the decompressing spring? [*Hint*: Detachment occurs when the spring reaches its natural length ($x = 0$); explain why.]

40. (III) Early test flights for the space shuttle used a "glider" (mass of 980 kg including pilot). After a horizontal launch at 480 km/h at a height of 3200 m, the glider eventually landed at sea level with a speed of 210 km/h. (*a*) What would its landing speed have been in the absence of air resistance? (*b*) What was the average force of air resistance exerted on it if it came in at a constant glide angle of 12° to the Earth's surface?

8–7 Gravitational Potential Energy

41. (I) The escape velocity from planet A is double that for planet B. The two planets have the same mass. What is the ratio of their radii, r_A/r_B?

42. (II) For a satellite of mass m_S in a circular orbit of radius r_S around the Earth, determine (*a*) its kinetic energy K, (*b*) its potential energy U ($U = 0$ at infinity), and (*c*) the ratio K/U.

43. (II) Show that Eq. 8–16 for gravitational potential energy reduces to Eq. 8–2, $\Delta U = mg(y_2 - y_1)$, for objects near the surface of the Earth. [*Hint*: set $r = r_E + y$, and assume $y \ll r_E$; see Eq. 6–4].

44. (II) Determine the *escape velocity from the Sun* for an object: (*a*) at the Sun's surface ($r = 7.0 \times 10^5$ km, $M = 2.0 \times 10^{30}$ kg); (*b*) at the average distance of the Earth (1.50×10^8 km). Compare (give factor for each) to the speed of the Earth in its orbit.

45. (II) Two Earth *satellites*, A and B, each of mass $m = 950$ kg, are launched into circular orbits around the Earth's center. Satellite A orbits at an altitude of 4800 km, and satellite B orbits at an altitude of 12,600 km. (*a*) What are the potential energies of the two satellites? (*b*) What are the kinetic energies of the two satellites? (*c*) How much work would it require to change the orbit of satellite A to match that of satellite B?

46. (II) A NASA *satellite* is said to have observed an *asteroid* that is on a collision course with the Earth. The asteroid has an estimated mass, based on its size, of 4×10^9 kg. It is approaching the Earth on a head-on course with a velocity of 660 m/s relative to the Earth and is now 5.0×10^6 km away. With what speed will it hit the Earth's surface, neglecting friction with the atmosphere?

47. (II) Take into account the Earth's rotational speed (1 rev/day) and determine the necessary "escape velocity," with respect to the Earth, for a rocket to escape if fired from the Earth at the equator in a direction (*a*) eastward; (*b*) westward; (*c*) vertically upward.

48. (II) (*a*) Determine a formula for the maximum height h that a rocket will reach if launched vertically from the Earth's surface with speed v_0 ($< v_{esc}$). Express in terms of v_0, r_E, M_E, and G. (*b*) How high does a rocket go if $v_0 = 8.35$ km/s? Ignore air resistance and the Earth's rotation.

49. (II) (*a*) Determine the rate at which the escape velocity from the Earth changes with distance r ($> r_E$) from the center of the Earth, dv_{esc}/dr. (*b*) Use the approximation $\Delta v \approx (dv/dr)\Delta r$ to determine the escape velocity for a spacecraft orbiting the Earth at a height of 320 km.

50. (II) A *meteorite* has a speed of 90.0 m/s when 750 km above the Earth. It is falling vertically (ignore air resistance) and strikes a bed of sand in which it is brought to rest in 3.25 m. (*a*) What is its speed just before striking the sand? (*b*) How much work does the sand do to stop the meteorite (mass = 575 kg)? (*c*) What is the average force exerted by the sand on the meteorite? (*d*) How much thermal energy is produced?

51. (II) How much work would be required to move a satellite of mass m from a circular orbit of radius $r_1 = 2r_E$ about the Earth to another circular orbit of radius $r_2 = 3r_E$? (r_E is the radius of the Earth.)

52. (II) A sphere of radius r_1 has a concentric spherical cavity of radius r_2 (Fig. 8–42). Assume this spherical shell of thickness $r_1 - r_2$ is uniform and has a total mass M. Show that the gravitational potential energy of a mass m at a distance r from the center of the shell ($r > r_1$) is given by

$$U = -\frac{GmM}{r}.$$

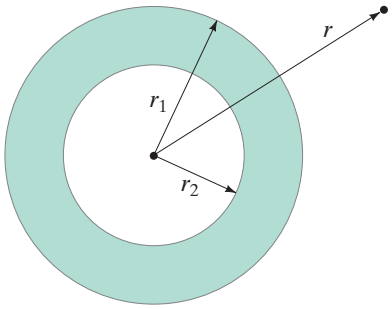


FIGURE 8–42
Problem 52.

53. (II) (*a*) Show that the total mechanical energy of a *satellite* (mass m) orbiting at a distance r from the center of the Earth (mass M_E) is

$$E = -\frac{1}{2}\frac{GmM_E}{r},$$

if $U = 0$ at $r = \infty$. (*b*) What is the ratio of kinetic energy to potential energy? (*c*) Although friction causes the value of E to decrease slowly, show that the orbit radius r decreases and kinetic energy increases if the orbit remains a circle.

54. (III) To *escape the solar system*, an interstellar spacecraft launched from Earth must overcome the gravitational attraction of both the Earth and Sun. Ignore the effects of other bodies in the solar system. (*a*) Show that the *escape velocity from the Solar System* is

$$v = \sqrt{v_E^2 + (v_S - v_0)^2} = 16.7 \text{ km/s},$$

where: v_E is the escape velocity from the Earth (Eq. 8–19); $v_S = \sqrt{2GM_S/r_{SE}}$ is the escape velocity from the gravitational field of the Sun (mass M_S) at the orbit of the Earth but far from the Earth's influence (r_{SE} is the Sun–Earth distance); and v_0 is the Earth's orbital velocity about the Sun. (*b*) Show that the energy required is 1.40×10^8 J per kilogram of spacecraft mass. [*Hint*: Write the energy equation for escape from Earth with v' as the spacecraft velocity, relative to Earth, but somewhat far from Earth (so its potential energy relative to Earth $\simeq 0$, but it is still $\simeq r_{SE}$ from the Sun). Then let $v' + v_0$ be the escape velocity from the Sun.]

8–8 Power

55. (I) What is the average power output of an elevator that lifts 845 kg a vertical height of 32.0 m in 11.0 s?

56. (I) How long will it take a 1750-W motor to lift a 335-kg piano to a sixth-story window 18.0 m above?

57. (I) An 85-kg football player traveling 5.0 m/s is stopped in 1.0 s by a tackler. (*a*) What is the original kinetic energy of the player? (*b*) What average power is required to stop him?

58. (II) If a car generates 18 hp when traveling at a steady 95 km/h, what must be the average force exerted on the car due to friction and air resistance?

59. (II) (*a*) Show that one British horsepower (550 ft · lb/s) is equal to 746 W. (*b*) What is the horsepower rating of a 75-W lightbulb? (*c*) How many kW is a 180-hp engine?

60. (II) An outboard motor for a boat is rated at 35 hp. If it can move a particular boat at a steady speed of 35 km/h, what is the total force resisting the motion of the boat?

61. (II) A shot-putter accelerates a 7.3-kg shot from rest to 14 m/s in 1.5 s. What average power was developed?

62. (II) A driver notices that her 950-kg car, when in neutral, slows down from 95 km/h to 65 km/h in about 7.0 s on a flat horizontal road. Approximately what power (watts and hp) is needed to keep the car traveling at a constant 80 km/h?

63. (II) A 975-kg sports car accelerates from rest to 95 km/h in 6.4 s. What is the average power delivered by the engine?

64. (II) During a workout, football players ran up the stadium stairs in 75 s. The distance along the stairs is 83 m and they are inclined at a 33° angle. If a player has a mass of 88 kg, estimate his average power output on the way up. Ignore friction and air resistance.

65. (II) A pump lifts 27.0 kg of water per minute through a height of 3.50 m. What minimum output rating (watts) must the pump motor have?

66. (II) A ski area claims that its lifts can move 45,000 people per hour. If the average lift carries people about 200 m (vertically) higher, estimate the maximum total power needed.

67. (II) A 65-kg skier grips a moving rope that is powered by an engine and is pulled at constant speed to the top of a 23° hill. The skier is pulled a distance $x = 250\,\text{m}$ along the incline and it takes 2.0 min to reach the top of the hill. If the coefficient of kinetic friction between the snow and skis is $\mu_k = 0.10$, what horsepower engine is required if 35 such skiers (max) are on the rope at one time?

68. (II) What minimum horsepower must a motor have to be able to drag a 370-kg box along a level floor at a speed of 1.20 m/s if the coefficient of friction is 0.45?

69. (III) The position of a 280-g object is given (in meters) by $x = 4.0t^3 - 8.0t^2 - 44t$, where t is in seconds. Determine the net rate of work done on this object (*a*) at $t = 2.0\,\text{s}$ and (*b*) at $t = 4.0\,\text{s}$. (*c*) What is the average net power input during the interval from $t = 0\,\text{s}$ to $t = 2.0\,\text{s}$, and in the interval from $t = 2.0\,\text{s}$ to 4.0 s?

70. (III) A bicyclist coasts down a 6.0° hill at a steady speed of 4.0 m/s. Assuming a total mass of 75 kg (bicycle plus rider), what must be the cyclist's power output to climb the same hill at the same speed?

8–9 Potential Energy Diagrams

71. (II) Draw a potential energy diagram, U vs. x, and analyze the motion of a mass m resting on a frictionless horizontal table and connected to a horizontal spring with stiffness constant k. The mass is pulled a distance to the right so that the spring is initially stretched a distance x_0 from equilibrium $(x = 0)$, and then the mass is released from rest.

72. (II) The spring of Problem 71 has a stiffness constant $k = 160\,\text{N/m}$. The mass $m = 6.0\,\text{kg}$ is released from rest when the spring is stretched $x_0 = 1.0\,\text{m}$ from equilibrium. Determine (*a*) the total energy of the system; (*b*) the kinetic energy when $x = \frac{1}{2}x_0$; (*c*) the maximum kinetic energy; (*d*) the maximum speed and at what positions it occurs; (*e*) the maximum acceleration and where it occurs.

73. (II) The graph of Fig. 8–43 shows the potential energy curve of a particle moving along the x axis under the influence of a conservative force. Note that the total energy $E > U(x)$, so that the particle's speed is never zero. (*a*) In which interval(s) of x is the force on the particle to the right? (*b*) At what value(s) of x is the magnitude of the force a minimum? (*c*) At what value of x is the magnitude of the force a maximum?

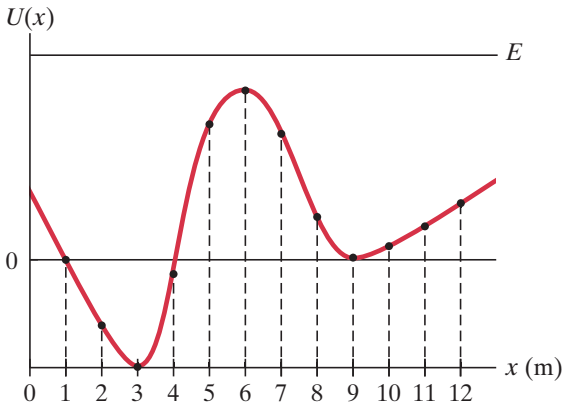


FIGURE 8–43 Problem 73.

74. (III) The potential energy of the two atoms in a diatomic (two-atom) molecule can be approximated as (*Lennard-Jones* potential)

$$U(r) = -\frac{a}{r^6} + \frac{b}{r^{12}},$$

where r is the distance between the two atoms and a and b are positive constants. (*a*) At what values of r is $U(r)$ a minimum? A maximum? (*b*) At what values of r is $U(r) = 0$? (*c*) Plot $U(r)$ as a function of r from $r = 0$ to r at a value large enough for all the features in (*a*) and (*b*) to show. (*d*) Describe the motion of one atom with respect to the second atom when $E < 0$, and when $E > 0$. (*e*) Let F be the force one atom exerts on the other. For what values of r is $F > 0$, $F < 0$, $F = 0$? (*f*) Determine F as a function of r.

75. (III) The *binding energy* of a two-particle system is defined as the energy required to separate the two particles from their state of lowest energy to $r = \infty$. Determine the binding energy for the molecule discussed in Problem 74.

76. (III) A particle moves where its potential energy is given by $U(r) = U_0[(2/r^2) - (1/r)]$. (*a*) Plot $U(r)$ versus r. Where does the curve cross the $U(r) = 0$ axis? At what value of r does the minimum value of $U(r)$ occur? (*b*) Suppose that the particle has an energy of $E = -0.050\,U_0$. Mark on your graph the approximate turning points of the motion of the particle. What is the maximum kinetic energy of the particle, and for what value of r does this occur?

*8–10 Gravitational Assist

***77.** (III) Show that to *slow down* a spacecraft (perhaps to put it into orbit around a planet or moon) v_{0x} must be in the *same* direction as $\vec{\mathbf{u}}$ and have magnitude greater than u. Show that this means the spacecraft will then pass in *front* of Jupiter. Draw a new diagram to replace Fig. 8–25b for this situation.

General Problems

78. (*a*) A 3.0-g locust reaches a speed of 3.0 m/s during its jump. What is its kinetic energy at this speed? (*b*) If the locust transforms energy with 35% efficiency, how much energy is required for the jump?

79. Water flows slowly over a dam at the rate of 320 kg/s and falls vertically 88 m before striking the turbine blades. Calculate (*a*) the speed of the water just before striking the turbine blades (neglect air resistance), and (*b*) the rate at which mechanical energy is transferred to the turbine blades, assuming 55% efficiency.

80. A bicyclist of mass 75 kg (including the bicycle) can coast down a 5.0° hill at a steady speed of 12 km/h. Pumping hard, the cyclist can descend the hill at a speed of 32 km/h. Using the same power, at what speed can the cyclist climb the same hill? Assume the force of friction is proportional to the square of the speed v; that is, $F_{fr} = bv^2$, where b is a constant.

81. A 55-kg skier starts from rest at the top of a ski jump, point A in Fig. 8–44, and travels down the ramp. If friction and air resistance can be neglected, (*a*) determine her speed v_B when she reaches the horizontal end of the ramp at B. (*b*) Determine the distance s to where she strikes the ground at C.

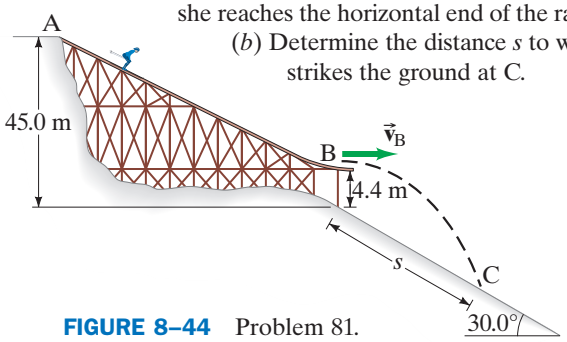


FIGURE 8–44 Problem 81.

82. A ball is attached to a horizontal cord of length ℓ whose other end is fixed, Fig. 8–45. (*a*) If the ball is released, what will be its speed at the lowest point of its path? (*b*) A peg is located a distance h directly below the point of attachment of the cord. If $h = 0.80\,\ell$, what will be the speed of the ball when it reaches the top of its circular path about the peg?

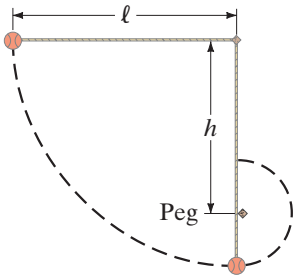


FIGURE 8–45
Problems 82 and 83.

83. Show that h must be greater than $0.60\,\ell$ if the ball in Fig. 8–45 is to make a complete circle about the peg.

84. A cyclist intends to cycle up a 7.50° hill whose vertical height is 125 m. The pedals turn in a circle of diameter 36.0 cm. Assuming the mass of bicycle plus rider is 75.0 kg, (*a*) calculate how much work she must do against gravity. (*b*) If each complete revolution of the pedals moves the bike 5.10 m along its path, calculate the average force that must be exerted on the pedals tangent to their circular path. Neglect work done by friction and other losses.

85. If you stand on a bathroom scale, the spring inside the scale compresses 0.60 mm, and it tells you your weight is 760 N. Now if you jump on the scale from a height of 1.0 m, what does the scale read at its peak? Assume Hooke's law holds.

86. Show that on a roller coaster with a circular vertical loop (Fig. 8–46), the difference in your "apparent weight" (see Section 6–4) at the top of the loop and the bottom of the loop is 6.0 times your weight. Ignore friction. Show also that as long as your speed is above the minimum needed (so the car holds the track), this answer doesn't depend on the size of the loop or how fast you go through it.

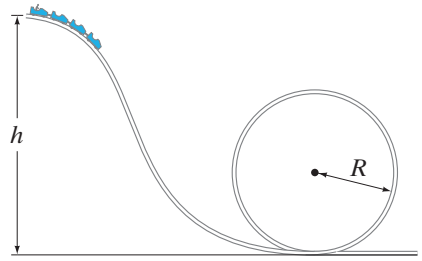


FIGURE 8–46
Problem 86.

87. A 65-kg hiker climbs to the top of a mountain 3900 m high. The climb is made in 4.6 h starting at an elevation of 2800 m. Calculate (*a*) the work done by the hiker against gravity, (*b*) the average power output in watts and in horsepower, and (*c*) assuming the body is 15% efficient, what rate of energy input was required.

88. Consider a force $F(x) = Ax^{3/2}$ acting on an object moving in a straight line. Assume that $A = 10.0\,\text{N/m}^z$. (*a*) What is the value of the exponent z? (*b*) Calculate the work done by this force as the object moves from $x = 0$ to $x = 3.0$ m.

89. The small mass m sliding without friction along the looped track shown in Fig. 8–47 is to remain on the track at all times, even at the very top of the loop of radius r. (*a*) In terms of the given quantities, determine the minimum release height h. (*b*) If the actual release height is $2h$, calculate the normal force exerted by the track at the bottom of the loop, then (*c*) by the track at the top of the loop, and (*d*) by the track after the block exits the loop onto the flat section.

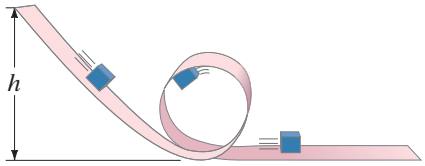


FIGURE 8–47
Problem 89.

90. Some electric power companies use placement of water to store energy. Water is pumped from a low reservoir to a high reservoir. To store the energy produced in 1.0 hour by a 180-MW electric power plant, how many cubic meters of water will have to be pumped from the lower to the upper reservoir? Assume the upper reservoir is 380 m above the lower one. Water has a mass of 1.00×10^3 kg for every 1.0 m^3 $(1\,\text{MW} = 10^6\,\text{W})$.

91. A film of Jesse Owens's famous long jump (Fig. 8–48) in the 1936 Olympics shows that his center of mass rose 1.1 m from launch point to the top of the arc. What minimum speed did he need at launch if he was traveling at 6.5 m/s at the top of the arc?



FIGURE 8–48
Problem 91.

92. The nuclear force between two neutrons in a nucleus is described roughly by the Yukawa potential

$$U(r) = -U_0 \frac{r_0}{r} e^{-r/r_0},$$

where r is the distance between the neutrons and U_0 and r_0 ($\approx 10^{-15}$ m) are constants. (*a*) Determine the force $F(r)$. (*b*) What is the ratio $F(3r_0)/F(r_0)$? (*c*) Calculate this same ratio for the force between two electrically charged particles where $U(r) = -C/r$, with C a constant. (*d*) Why is the Yukawa force referred to as a "short-range" force?

93. A 56-kg student runs at 6.0 m/s, grabs a hanging 10.0-m-long rope, and swings out over a lake (Fig. 8–49). He releases the rope when his velocity is zero. (*a*) What is the angle θ when he releases the rope? (*b*) What is the tension in the rope just before he releases it? (*c*) What is the maximum tension in the rope during the swing?

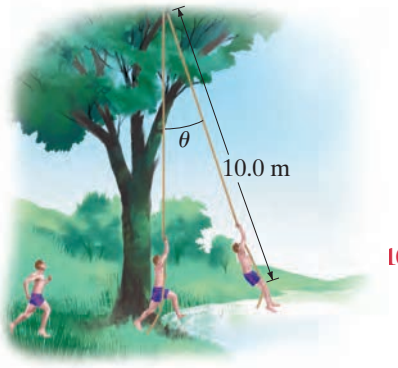


FIGURE 8–49
Problem 93.

94. A fire hose for use in urban areas must be able to shoot a stream of water to a maximum height of 33 m. The water leaves the hose at ground level as a circular stream 3.0 cm in diameter. What minimum power is required to create such a stream of water? Every cubic meter of water has a mass of 1.00×10^3 kg.

95. Proper design of automobile braking systems must account for heat buildup under heavy braking. Calculate the thermal energy dissipated from brakes in a 1500-kg car that descends a 17° hill. The car begins braking when its speed is 95 km/h and slows to a speed of 35 km/h in a distance of 0.30 km measured along the road.

96. (*a*) Suppose we have three masses, m_1, m_2, and m_3, that initially are extremely ($\approx$ infinitely) far apart from each other. Show that the work needed to bring them to the positions shown in Fig. 8–50 is

$$W = -G\left(\frac{m_1 m_2}{r_{12}} + \frac{m_1 m_3}{r_{13}} + \frac{m_2 m_3}{r_{23}}\right).$$

(*b*) Can we say that this formula also gives the potential energy of the system, or the potential energy of one or two of the objects? (*c*) Is W equal to the binding energy of the system—that is, is W equal to the energy required to separate the components by an infinite distance? Explain.

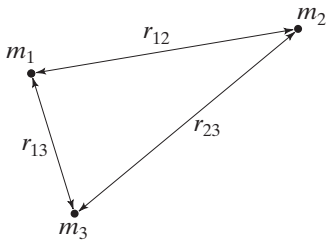


FIGURE 8–50 Problem 96.

97. How much work can a 3.0-hp motor do in 1.0h?

98. Estimate the energy required from fuel to launch a 1225-kg satellite into orbit 1375 km above the Earth's surface. Consider two cases: (*a*) the satellite is launched into an equatorial orbit from a point on the Earth's equator, and (*b*) it is launched from the North Pole into a polar orbit.

99. A satellite is in an elliptic orbit around the Earth (Fig. 8–51). Its speed at the perigee A is 8650 m/s. (*a*) Use conservation of energy to determine its speed at B. (*b*) Use conservation of energy to determine the speed at the apogee C.

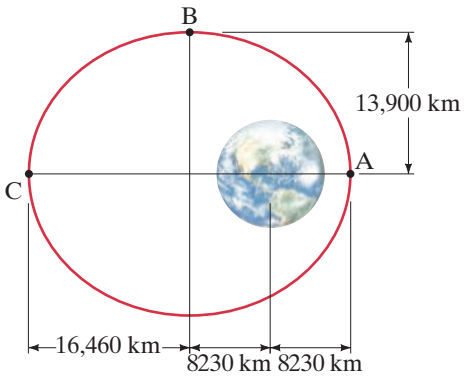


FIGURE 8–51
Problem 99.

100. Suppose the gravitational potential energy of an object of mass m at a distance r from the center of the Earth is given by

$$U(r) = -\frac{GMm}{r} e^{-\alpha r}$$

where α is a positive constant. (Newton's law of universal gravitation has $\alpha = 0$). (*a*) What would be the force on the object as a function of r? (*b*) What would be the object's escape velocity in terms of the Earth's radius R_E?

101. A particle of mass m moves under the influence of a potential energy

$$U(x) = \frac{a}{x} + bx$$

where a and b are positive constants and the particle is restricted to the region $x > 0$. Find a point of equilibrium for the particle and demonstrate that it is stable.

102. An elevator cable breaks when a 975-kg elevator is 28.5 m above a huge spring ($k = 8.00 \times 10^4$ N/m) at the bottom of the shaft. Calculate (*a*) the work done by gravity on the elevator before it hits the spring; (*b*) the speed of the elevator just before striking the spring; and (*c*) the amount the spring compresses (use conservation of energy). Ignore any energy transformed to heat.

103. A 1600-kg spacecraft is launched from Earth and lands on the Moon. How does its total potential energy before launch compare with its total potential energy after landing?

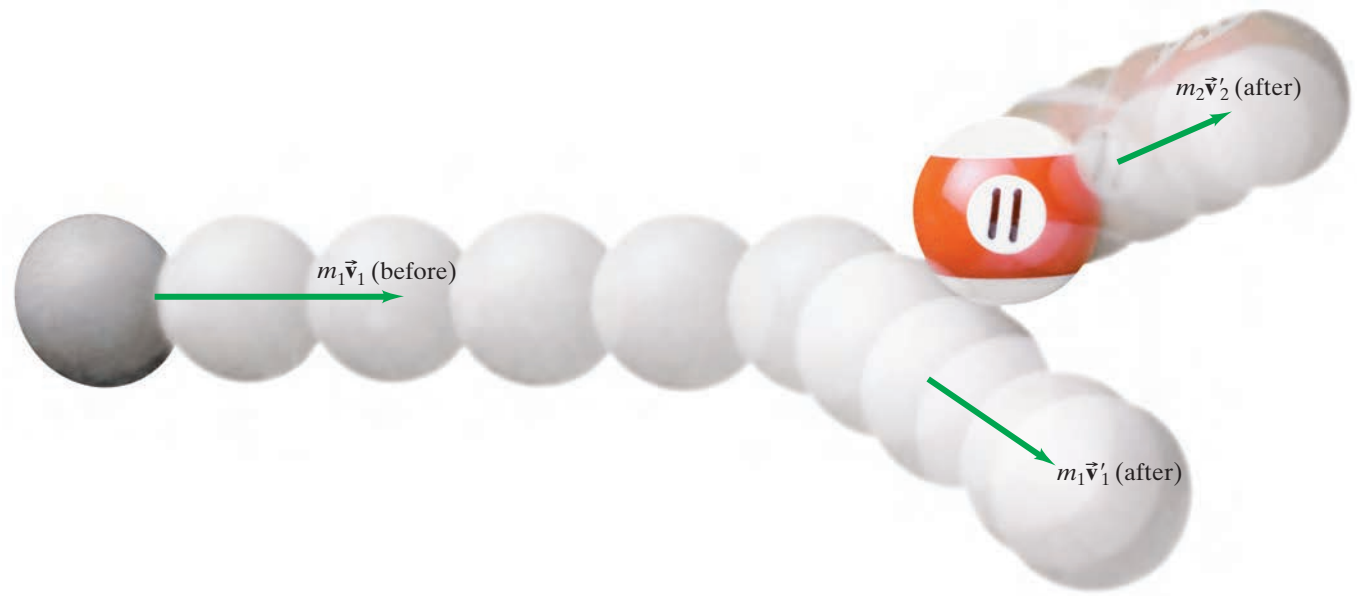
104. The two atoms in a diatomic molecule exert an attractive force on each other at large distances and a repulsive force at short distances. The magnitude of the force between two atoms in a diatomic molecule can be approximated by the Lennard-Jones force, or $F(r) = F_0\left[2(\sigma/r)^{13} - (\sigma/r)^7\right]$, where r is the separation between the two atoms, and σ and F_0 are constants. For an oxygen molecule (which is diatomic) $F_0 = 9.60 \times 10^{-11}$ N and $\sigma = 3.50 \times 10^{-11}$ m. (*a*) Integrate the equation for $F(r)$ to determine the potential energy $U(r)$ of the oxygen molecule. (*b*) Find the equilibrium distance r_0 between the two atoms. (*c*) Graph $F(r)$ and $U(r)$ between $0.9\,r_0$ and $2.5\,r_0$.

ANSWERS TO EXERCISES

A: (*b*).
B: (*c*).
C: Equal speeds.
D: (*e*), (*e*); (*e*), (*c*).

Conservation of linear momentum is another great conservation law of physics. Collisions, such as between billiard or pool balls, illustrate this law very nicely: the total vector momentum just before the collision equals the total vector momentum just after the collision. In this photo, the moving cue ball makes a glancing collision with the 11 ball which is initially at rest. After the collision, both balls move at angles, but the sum of their vector momenta equals the initial vector momentum of the incoming cue ball.

We will consider both elastic collisions (where kinetic energy is also conserved) and inelastic collisions. We also examine the concept of center of mass, and how it helps us in the study of complex motion.



CHAPTER 9

Linear Momentum

CHAPTER-OPENING QUESTIONS—Guess now!

1. A railroad car loaded with rocks coasts on a level track without friction. A worker at the back of the car starts throwing the rocks horizontally backward from the car. Then what happens?

(a) The car slows down.
(b) The car speeds up.
(c) First the car speeds up and then it slows down.
(d) The car's speed remains constant.
(e) None of these.

2. Which answer would you choose if the rocks fall out through a hole in the floor of the car, one at a time?

The law of conservation of energy, which we discussed in the previous Chapter, is one of several great conservation laws in physics. Among the other quantities found to be conserved are linear momentum, angular momentum, and electric charge. We will eventually discuss all of these because the conservation laws are among the most important ideas in science. In this Chapter we discuss linear momentum and its conservation. The law of conservation of momentum is essentially a reworking of Newton's laws that gives us tremendous physical insight and problem-solving power.

CONTENTS

9 – 1 Momentum and Its Relation to Force
9 – 2 Conservation of Momentum
9 – 3 Collisions and Impulse
9 – 4 Conservation of Energy and Momentum in Collisions
9 – 5 Elastic Collisions in One Dimension
9 – 6 Inelastic Collisions
9 – 7 Collisions in 2 or 3 Dimensions
9 – 8 Center of Mass (CM)
9 – 9 CM and Translational Motion
*9–10 Systems of Variable Mass; Rocket Propulsion

The law of conservation of momentum is particularly useful when dealing with a system of two or more objects that interact with each other, such as in collisions of ordinary objects or nuclear particles.

Our focus up to now has been mainly on the motion of a single object, often thought of as a "particle" in the sense that we have ignored any rotation or internal motion. In this Chapter we will deal with systems of two or more objects, and—toward the end of the Chapter—the concept of center of mass.

9–1 Momentum and Its Relation to Force

The **linear momentum** (or "momentum" for short) of an object is defined as the product of its mass and its velocity. Momentum (plural is *momenta*) is represented by the symbol $\vec{\mathbf{p}}$. If we let m represent the mass of an object and $\vec{\mathbf{v}}$ represent its velocity, then its momentum $\vec{\mathbf{p}}$ is defined as

$$\vec{\mathbf{p}} = m\vec{\mathbf{v}}. \tag{9–1}$$

Velocity is a vector, so momentum too is a vector. The direction of the momentum is the direction of the velocity, and the magnitude of the momentum is $p = mv$. Because velocity depends on the reference frame, so does momentum; thus the reference frame must be specified. The unit of momentum is that of mass $\times$ velocity, which in SI units is kg · m/s. There is no special name for this unit.

Everyday usage of the term *momentum* is in accord with the definition above. According to Eq. 9–1, a fast-moving car has more momentum than a slow-moving car of the same mass; a heavy truck has more momentum than a small car moving with the same speed. The more momentum an object has, the harder it is to stop it, and the greater effect it will have on another object if it is brought to rest by striking that object. A football player is more likely to be stunned if tackled by a heavy opponent running at top speed than by a lighter or slower-moving tackler. A heavy, fast-moving truck can do more damage than a slow-moving motorcycle.

EXERCISE A Can a small sports car ever have the same momentum as a large sport-utility vehicle with three times the sports car's mass? Explain.

A force is required to change the momentum of an object, whether it is to increase the momentum, to decrease it, or to change its direction. Newton originally stated his second law of motion in terms of momentum (although he called the product mv the "quantity of motion"). Newton's statement of the **second law of motion**, translated into modern language, is as follows:

> **The rate of change of momentum of an object is equal to the net force applied to it.**

We can write this as an equation,

NEWTON'S SECOND LAW

$$\Sigma\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt}, \tag{9–2}$$

⚠ CAUTION

The change in the momentum vector is in the direction of the net force

where $\Sigma\vec{\mathbf{F}}$ is the net force applied to the object (the vector sum of all forces acting on it). We can readily derive the familiar form of the second law, $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$, from Eq. 9–2 for the case of constant mass. If $\vec{\mathbf{v}}$ is the velocity of an object at any moment, then Eq. 9–2 gives

$$\Sigma\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt} = \frac{d(m\vec{\mathbf{v}})}{dt} = m\frac{d\vec{\mathbf{v}}}{dt} = m\vec{\mathbf{a}} \qquad \text{[constant mass]}$$

because, by definition, $\vec{\mathbf{a}} = d\vec{\mathbf{v}}/dt$ and we assume m = constant. Newton's statement, Eq. 9–2, is actually more general than simply $\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ because Eq. 9–2 is valid also in situations when the mass may change. For example, rockets lose mass as they burn fuel (Section 9–10).

EXAMPLE 9–1 **ESTIMATE** **Force of a tennis serve.** For a top player, a tennis ball may leave the racket on the serve with a speed of 55 m/s (about 120 mi/h), Fig. 9–1. If the ball has a mass of 0.060 kg and is in contact with the racket for about 4 ms $(4 \times 10^{-3}\,\text{s})$, estimate the average force on the ball. Would this force be large enough to lift a 70-kg person?

APPROACH We write Newton's second law, Eq. 9–2, for the average force as

$$F_{\text{avg}} = \frac{\Delta p}{\Delta t} = \frac{mv_2 - mv_1}{\Delta t},$$

where mv_1 and mv_2 are the initial and final momenta. The tennis ball is hit when its initial velocity v_1 is very nearly zero at the top of the throw, so we set $v_1 = 0$, and we assume $v_2 = 55\,\text{m/s}$ is in the horizontal direction. We ignore all other forces on the ball during this brief time interval, such as gravity, in comparison to the force exerted by the tennis racket.

SOLUTION The force exerted on the ball by the racket is

$$F_{\text{avg}} = \frac{\Delta p}{\Delta t} = \frac{mv_2 - mv_1}{\Delta t} = \frac{(0.060\,\text{kg})(55\,\text{m/s}) - 0}{0.004\,\text{s}} \approx 800\,\text{N}.$$

This is a large force, larger than the weight of a 70-kg person, which would require a force $mg = (70\,\text{kg})(9.8\,\text{m/s}^2) \approx 700\,\text{N}$ to lift.

NOTE The force of gravity acting on the tennis ball is $mg = (0.060\,\text{kg})(9.8\,\text{m/s}^2) = 0.59\,\text{N}$, which justifies our ignoring it compared to the enormous force the racket exerts.

NOTE High-speed photography and radar can give us an estimate of the contact time and the velocity of the ball leaving the racket. But a direct measurement of the force is not practical. Our calculation shows a handy technique for determining an unknown force in the real world.



FIGURE 9–1 Example 9–1.

EXAMPLE 9–2 **Washing a car: momentum change and force.** Water leaves a hose at a rate of 1.5 kg/s with a speed of 20 m/s and is aimed at the side of a car, which stops it, Fig. 9–2. (That is, we ignore any splashing back.) What is the force exerted by the water on the car?

APPROACH The water leaving the hose has mass and velocity, so it has a momentum p_{initial} in the horizontal (x) direction, and we assume gravity doesn't pull the water down significantly before it hits the car. When the water hits the car, the water loses this momentum $(p_{\text{final}} = 0)$. We use Newton's second law in the momentum form, Eq. 9–2, to find the force that the car exerts on the water to stop it. By Newton's third law, the force exerted by the water on the car is equal and opposite. We have a continuing process: 1.5 kg of water leaves the hose in each 1.0-s time interval. So let us write $F = \Delta p/\Delta t$ where $\Delta t = 1.0\,\text{s}$, and $mv_{\text{initial}} = (1.5\,\text{kg})(20\,\text{m/s}) = 30\,\text{kg}\cdot\text{m/s}$.

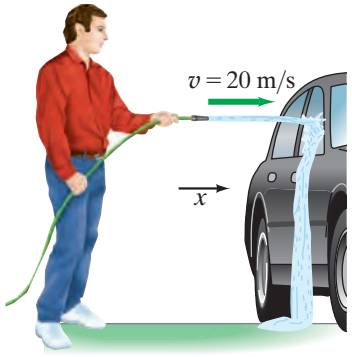
SOLUTION The force (assumed constant) that the car must exert to change the momentum of the water is

$$F = \frac{\Delta p}{\Delta t} = \frac{p_{\text{final}} - p_{\text{initial}}}{\Delta t} = \frac{0 - 30\,\text{kg}\cdot\text{m/s}}{1.0\,\text{s}} = -30\,\text{N}.$$

The minus sign indicates that the force exerted by the car on the water is opposite to the water's original velocity. The car exerts a force of 30 N to the left to stop the water, so by Newton's third law, the water exerts a force of 30 N to the right on the car.

NOTE Keep track of signs, although common sense helps too. The water is moving to the right, so common sense tells us the force on the car must be to the right.

FIGURE 9–2 Example 9–2.



EXERCISE B If the water splashes back from the car in Example 9–2, would the force on the car be larger or smaller?

9–2 Conservation of Momentum

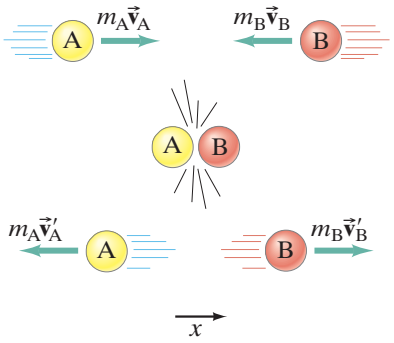


FIGURE 9–3 Momentum is conserved in a collision of two balls, labeled A and B.

The concept of momentum is particularly important because, if no net external force acts on a system, the total momentum of the system is a conserved quantity. This was expressed in Eq. 9–2 for a single object, but it holds also for a system of objects, as we shall discuss.

Consider the head-on collision of two billiard balls, as shown in Fig. 9–3. We assume the net external force on this system of two balls is zero—that is, the only significant forces during the collision are the forces that each ball exerts on the other. Although the momentum of each of the two balls changes as a result of the collision, the *sum* of their momenta is found to be the same before as after the collision. If $m_A\vec{\mathbf{v}}_A$ is the momentum of ball A and $m_B\vec{\mathbf{v}}_B$ the momentum of ball B, both measured just before the collision, then the total momentum of the two balls before the collision is the vector sum $m_A\vec{\mathbf{v}}_A + m_B\vec{\mathbf{v}}_B$. Immediately after the collision, the balls each have a different velocity and momentum, which we designate by a "prime" on the velocity: $m_A\vec{\mathbf{v}}'_A$ and $m_B\vec{\mathbf{v}}'_B$. The total momentum after the collision is the vector sum $m_A\vec{\mathbf{v}}'_A + m_B\vec{\mathbf{v}}'_B$. No matter what the velocities and masses are, experiments show that the total momentum before the collision is the same as afterward, whether the collision is head-on or not, as long as no net external force acts:

CONSERVATION OF MOMENTUM (two objects colliding)

$$\text{momentum before} = \text{momentum after}$$

$$m_A\vec{\mathbf{v}}_A + m_B\vec{\mathbf{v}}_B = m_A\vec{\mathbf{v}}'_A + m_B\vec{\mathbf{v}}'_B. \qquad [\Sigma\vec{\mathbf{F}}_{\text{ext}} = 0] \quad \textbf{(9–3)}$$

That is, the total vector momentum of the system of two colliding balls is conserved: it stays constant. (We saw this result in this Chapter's opening photograph.)

Although the law of conservation of momentum was discovered experimentally, it can be derived from Newton's laws of motion, which we now show.

Let us consider two objects of mass m_A and m_B that have momenta $\vec{\mathbf{p}}_A$ $(= m_A\vec{\mathbf{v}}_A)$ and $\vec{\mathbf{p}}_B$ $(= m_B\vec{\mathbf{v}}_B)$ before they collide and $\vec{\mathbf{p}}'_A$ and $\vec{\mathbf{p}}'_B$ after they collide, as in Fig. 9–4. During the collision, suppose that the force exerted by object A on object B at any instant is $\vec{\mathbf{F}}$. Then, by Newton's third law, the force exerted by object B on object A is $-\vec{\mathbf{F}}$. During the brief collision time, we assume no other (external) forces are acting (or that $\vec{\mathbf{F}}$ is much greater than any other external forces acting). Thus we have

$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}_B}{dt}$$

and

$$-\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}_A}{dt}.$$

We add these two equations together and find

$$0 = \frac{d(\vec{\mathbf{p}}_A + \vec{\mathbf{p}}_B)}{dt}$$

which tells us that

$$\vec{\mathbf{p}}_A + \vec{\mathbf{p}}_B = \text{constant}.$$

This is Eq. 9–3. The total momentum is conserved.

We have put this derivation in the context of a collision. As long as no external forces act, it is valid over any time interval, and conservation of momentum is always valid as long as no external forces act on the chosen system. In the real world, external forces do act: friction on billiard balls, gravity acting on a tennis ball, and so on. So we often want our "observation time" (before and after) to be small. When a racket hits a tennis ball or a bat hits a baseball, both before and after the "collision" the ball moves as a projectile under the action of gravity and air resistance.

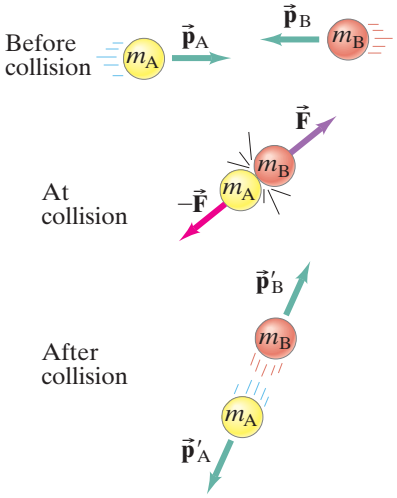


FIGURE 9–4 Collision of two objects. Their momenta before collision are $\vec{\mathbf{p}}_A$ and $\vec{\mathbf{p}}_B$, and after collision are $\vec{\mathbf{p}}'_A$ and $\vec{\mathbf{p}}'_B$. At any moment during the collision each exerts a force on the other of equal magnitude but opposite direction.

However, when the bat or racket hits the ball, during the brief time of the collision those external forces are insignificant compared to the collision force the bat or racket exerts on the ball. Momentum is conserved (or very nearly so) as long as we measure $\vec{\mathbf{p}}_A$ and $\vec{\mathbf{p}}_B$ just before the collision and $\vec{\mathbf{p}}'_A$ and $\vec{\mathbf{p}}'_B$ immediately after the collision (Eq. 9–3). We can not wait for external forces to produce their effect before measuring $\vec{\mathbf{p}}'_A$ and $\vec{\mathbf{p}}'_B$.

Our derivation of the conservation of momentum can be extended to include any number of interacting objects. Let $\vec{\mathbf{P}}$ represent the total momentum of a system of n interacting objects which we number from 1 to n:

$$\vec{\mathbf{P}} = m_1\vec{\mathbf{v}}_1 + m_2\vec{\mathbf{v}}_2 + \cdots + m_n\vec{\mathbf{v}}_n = \Sigma\vec{\mathbf{p}}_i.$$

We differentiate with respect to time:

$$\frac{d\vec{\mathbf{P}}}{dt} = \Sigma\frac{d\vec{\mathbf{p}}_i}{dt} = \Sigma\vec{\mathbf{F}}_i \qquad \textbf{(9–4)}$$

where $\vec{\mathbf{F}}_i$ represents the *net* force on the i^{th} object. The forces can be of two types: (1) *external forces* on objects of the system, exerted by objects outside the system, and (2) *internal forces* that objects within the system exert on other objects in the system. By Newton's third law, the internal forces occur in pairs: if one object exerts a force on a second object, the second exerts an equal and opposite force on the first object. Thus, in the sum over all the forces in Eq. 9–4, all the internal forces cancel each other in pairs. So only external forces are left in Eq. 9–4:

$$\frac{d\vec{\mathbf{P}}}{dt} = \Sigma\vec{\mathbf{F}}_{\text{ext}}, \qquad \textbf{(9–5)}$$

NEWTON'S SECOND LAW (for a system of objects)

where $\Sigma\vec{\mathbf{F}}_{\text{ext}}$ is the sum of all external forces acting on our system. If the net external force is zero, then $d\vec{\mathbf{P}}/dt = 0$, so $\Delta\vec{\mathbf{P}} = 0$ or $\vec{\mathbf{P}} = \text{constant}$. Thus we see that

when the net external force on a system of objects is zero, the total momentum of the system remains constant.

LAW OF CONSERVATION OF LINEAR MOMENTUM

This is the **law of conservation of momentum**. It can also be stated as

the total momentum of an isolated system of objects remains constant.

By **system**, we simply mean a set of objects that we choose, and which may interact with each other. An **isolated system** is one in which the only (significant) forces are those between the objects in the system. The sum of all these "internal" forces within the system will be zero because of Newton's third law. If there are *external forces*—by which we mean forces exerted by objects *outside* the system—and they don't add up to zero, then the system is not isolated and the total momentum of the system won't be conserved. However, if the system can be redefined so as to include the other objects exerting these forces, then the conservation of momentum principle can apply. For example, if we take as our system a falling rock, it does not conserve momentum because an external force, the force of gravity exerted by the Earth, accelerates the rock and changes its momentum. However, if we include the Earth in the system, the total momentum of rock plus Earth is conserved. (This means that the Earth comes up to meet the rock. But the Earth's mass is so great, its upward velocity is tiny and unnoticeable.)

Although the law of conservation of momentum follows from Newton's second law, as we have seen, it is in fact more general than Newton's laws. In the tiny world of the atom, Newton's laws fail, but the great conservation laws—those of energy, momentum, angular momentum, and electric charge—have been found to hold in every experimental situation tested. It is for this reason that the conservation laws are considered more basic than Newton's laws.

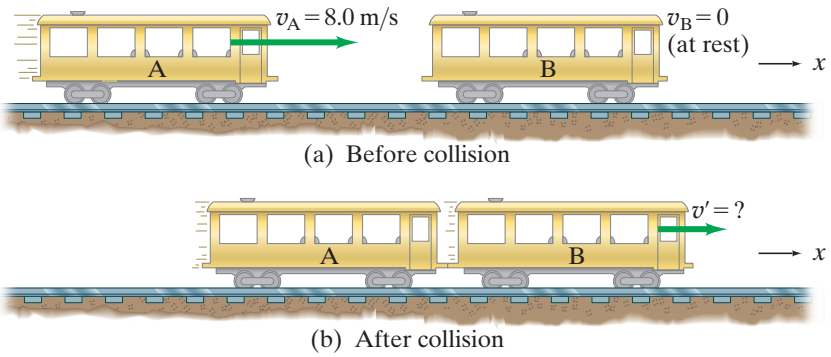


FIGURE 9–5 Example 9–3.

EXAMPLE 9–3 **Railroad cars collide: momentum conserved.** A 10,000-kg railroad car, A, traveling at a speed of 8.0 m/s strikes an identical car, B, at rest. If the cars lock together as a result of the collision, what is their common speed just afterward? See Fig. 9–5.

APPROACH We choose our system to be the two railroad cars. We consider a very brief time interval, from just before the collision until just after, so that external forces such as friction can be ignored. Then we apply conservation of momentum:

$$P_{\text{initial}} = P_{\text{final}}.$$

SOLUTION The initial total momentum is

$$P_{\text{initial}} = m_A v_A + m_B v_B = m_A v_A$$

because car B is at rest initially $(v_B = 0)$. The direction is to the right in the $+x$ direction. After the collision, the two cars become attached, so they will have the same speed, call it v'. Then the total momentum after the collision is

$$P_{\text{final}} = (m_A + m_B)\, v'.$$

We have assumed there are no external forces, so momentum is conserved:

$$P_{\text{initial}} = P_{\text{final}}$$

$$m_A v_A = (m_A + m_B)\, v'.$$

Solving for v', we obtain

$$v' = \frac{m_A}{m_A + m_B} v_A = \left(\frac{10{,}000\text{ kg}}{10{,}000\text{ kg} + 10{,}000\text{ kg}}\right)(8.0\text{ m/s}) = 4.0\text{ m/s},$$

to the right. Their mutual speed after collision is half the initial speed of car A because their masses are equal.

NOTE We kept symbols until the very end, so we have an equation we can use in other (related) situations.

NOTE We haven't mentioned friction here. Why? Because we are examining speeds just before and just after the very brief time interval of the collision, and during that brief time friction can't do much—it is ignorable (but not for long: the cars will slow down because of friction).

EXERCISE C In Example 9–3, what result would you get if (*a*) $m_B = 3m_A$, (*b*) m_B is much larger than m_A $(m_B \gg m_A)$, and (*c*) $m_B \ll m_A$?

EXERCISE D A 50-kg child runs off a dock at 2.0 m/s (horizontally) and lands in a waiting rowboat of mass 150 kg. At what speed does the rowboat move away from the dock?

The law of conservation of momentum is particularly useful when we are dealing with fairly simple systems such as colliding objects and certain types of "explosions." For example, *rocket propulsion*, which we saw in Chapter 4 can be understood on the basis of action and reaction, can also be explained on the basis of the conservation of momentum. We can consider the rocket and fuel as an isolated system if it is far out in space (no external forces). In the reference frame of the rocket before any fuel is ejected, the total momentum of rocket plus fuel is zero. When the fuel burns, the total momentum remains unchanged: the backward momentum of the expelled gases is just balanced by the forward momentum gained by the rocket itself (see Fig. 9–6). Thus, a rocket can accelerate in empty space. There is no need for the expelled gases to push against the Earth or the air (as is sometimes erroneously thought). Similar examples of (nearly) isolated systems where momentum is conserved are the recoil of a gun when a bullet is fired (Example 9–4), and the movement of a rowboat just after a package is thrown from it.

PHYSICS APPLIED
Rocket propulsion

CAUTION
Rockets accelerate by pushing out fuel gas they burn, not by pushing on Earth

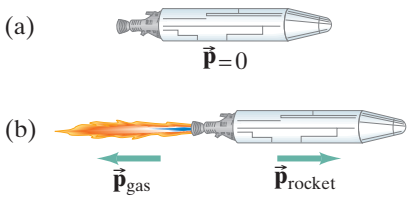


FIGURE 9–6 (a) A rocket, containing fuel, at rest in some reference frame. (b) In the same reference frame, the rocket fires, and gases are expelled at high speed out the rear. The total vector momentum, $\vec{\mathbf{P}} = \vec{\mathbf{p}}_{\text{gas}} + \vec{\mathbf{p}}_{\text{rocket}}$, remains zero.

EXAMPLE 9–4 **Rifle recoil.** Calculate the recoil velocity of a 5.0-kg rifle that shoots a 0.020-kg bullet at a speed of 620 m/s, Fig. 9–7.

APPROACH Our system is the rifle and the bullet, both at rest initially, just before the trigger is pulled. The trigger is pulled, an explosion occurs inside the bullet's shell, and we look at the rifle and bullet just as the bullet leaves the barrel (Fig. 9–7b). The bullet moves to the right $(+x)$, and the gun recoils to the left. During the very short time interval of the explosion, we can assume the external forces are small compared to the forces exerted by the exploding gunpowder. Thus we can apply conservation of momentum, at least approximately.

SOLUTION Let subscript B represent the bullet and R the rifle; the final velocities are indicated by primes. Then momentum conservation in the x direction gives

$$\begin{aligned} \text{momentum before} &= \text{momentum after} \\ m_{\text{B}}v_{\text{B}} + m_{\text{R}}v_{\text{R}} &= m_{\text{B}}v'_{\text{B}} + m_{\text{R}}v'_{\text{R}} \\ 0 + 0 &= m_{\text{B}}v'_{\text{B}} + m_{\text{R}}v'_{\text{R}}. \end{aligned}$$

We solve for the unknown v'_{R}, and find

$$v'_{\text{R}} = -\frac{m_{\text{B}}v'_{\text{B}}}{m_{\text{R}}} = -\frac{(0.020\,\text{kg})\,(620\,\text{m/s})}{(5.0\,\text{kg})} = -2.5\,\text{m/s}.$$

Since the rifle has a much larger mass, its (recoil) velocity is much less than that of the bullet. The minus sign indicates that the velocity (and momentum) of the rifle is in the negative x direction, opposite to that of the bullet.

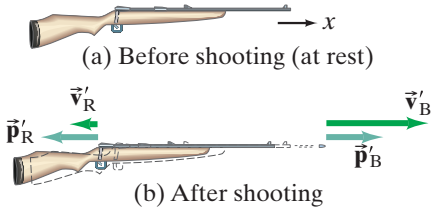


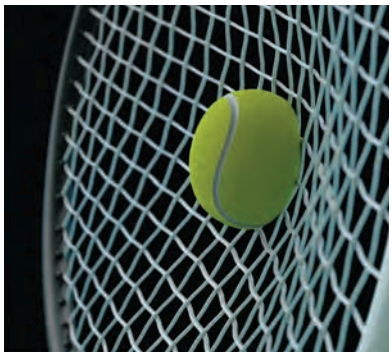
FIGURE 9–7 Example 9–4.

CONCEPTUAL EXAMPLE 9–5 **Falling on or off a sled.** (*a*) An empty sled is sliding on frictionless ice when Jeanne drops vertically from a tree down onto the sled. When she lands, does the sled speed up, slow down, or keep the same speed? (*b*) Later: Jeanne falls sideways off the sled. When she drops off, does the sled speed up, slow down, or keep the same speed?

RESPONSE (*a*) Because Jeanne falls vertically onto the sled, she has no initial horizontal momentum. Thus the total horizontal momentum afterward equals the momentum of the sled initially. Since the mass of the system (sled + person) has increased, the speed must decrease.

(*b*) At the instant Jeanne falls off, she is moving with the same horizontal speed as she was while on the sled. At the moment she leaves the sled (before hitting the ground), she has the same momentum she had an instant before. Because her momentum does not change, neither does the sled's (total momentum conserved); the sled keeps the same speed.

EXERCISE E Return to the Chapter-Opening Questions, page 227, and answer them again now. Try to explain why you may have answered differently the first time.



(a)



(b)

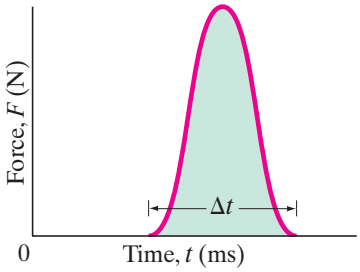
FIGURE 9–8 (a) Tennis racket striking a ball. Both the ball and the racket strings are deformed due to the large force each exerts on the other. (b) Also for a baseball and bat.

9–3 Collisions and Impulse

Collisions are a common occurrence in everyday life: a tennis racket or a baseball bat striking a ball, billiard balls colliding, a hammer hitting a nail. When a collision occurs, the force that each object exerts on the other one is usually far stronger than any external forces. We can then ignore the effects of any other forces during the brief time interval of the collision.

During a collision of two ordinary objects, both objects are deformed, often considerably, because of the large forces involved (Fig. 9–8). When the collision occurs, the force each exerts on the other usually jumps from zero at the moment of contact to a very large force within a very short time, and then rapidly returns to zero again. A graph of the magnitude of the force that one object exerts on the other during a collision, as a function of time, is something like the red curve in Fig. 9–9. The time interval Δt is usually very distinct and usually very small, typically milliseconds for a macroscopic collision.

FIGURE 9–9 Force as a function of time during a typical collision. F can become very large; Δt is typically milliseconds for macroscopic collisions.



From Newton's second law, Eq. 9–2, the *net* force on an object is equal to the rate of change of its momentum:

$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt}.$$

(We have written $\vec{\mathbf{F}}$ instead of $\Sigma\vec{\mathbf{F}}$ for the net force, which we assume is entirely due to the brief but large force that acts during the collision.) This equation applies to *each* of the objects in a collision. During the infinitesimal time interval dt, the momentum changes by

$$d\vec{\mathbf{p}} = \vec{\mathbf{F}}\,dt. \qquad \textbf{(9–6a)}$$

If we integrate this over the duration of a collision, we have

$$\int_{\text{i}}^{\text{f}} d\vec{\mathbf{p}} = \vec{\mathbf{p}}_{\text{f}} - \vec{\mathbf{p}}_{\text{i}} = \int_{t_{\text{i}}}^{t_{\text{f}}} \vec{\mathbf{F}}\,dt,$$

where $\vec{\mathbf{p}}_{\text{i}}$ and $\vec{\mathbf{p}}_{\text{f}}$ are the initial and final momenta of the object, just before and just after the collision. The integral of the net force over the time interval during which it acts is called the **impulse**, $\vec{\mathbf{J}}$:

$$\vec{\mathbf{J}} = \int_{t_{\text{i}}}^{t_{\text{f}}} \vec{\mathbf{F}}\,dt.$$

Thus the change in momentum of an object, $\Delta\vec{\mathbf{p}} = \vec{\mathbf{p}}_{\text{f}} - \vec{\mathbf{p}}_{\text{i}}$, is equal to the impulse acting on it:

$$\Delta\vec{\mathbf{p}} = \vec{\mathbf{p}}_{\text{f}} - \vec{\mathbf{p}}_{\text{i}} = \int_{t_{\text{i}}}^{t_{\text{f}}} \vec{\mathbf{F}}\,dt = \vec{\mathbf{J}}. \qquad \textbf{(9–6b)}$$

The units for impulse are the same as for momentum, kg · m/s (or N · s) in SI. Since $\vec{\mathbf{J}} = \int \vec{\mathbf{F}}\,dt$, we can state that the impulse $\vec{\mathbf{J}}$ of a force is equal to the area under the F versus t curve, as indicated by the shading in Fig. 9–9.

Equation 9–6b (also Eq. 9–6a) is true only if $\vec{\mathbf{F}}$ is the *net* force on the object. It is valid for *any* net force $\vec{\mathbf{F}}$ where $\vec{\mathbf{p}}_{\text{i}}$ and $\vec{\mathbf{p}}_{\text{f}}$ correspond precisely to the times t_{i} and t_{f}.

But the impulse concept is really most useful for so-called *impulsive forces*—that is, for a force like that shown in Fig. 9–9, which has a very large magnitude over a very short time interval and is essentially zero outside this time interval. For most collision processes, the impulsive force is much larger than any other force acting, and the others can be neglected. For such an impulsive force, the time interval over which we take the integral in Eq. 9–6b is not critical as long as we start before t_i and end after t_f, since $\vec{\mathbf{F}}$ is essentially zero outside this time interval. (Of course, if the chosen time interval is too large, the effect of the other (external) forces does become significant—such as the flight of a tennis ball which, after the impulsive force administered by the racket, begins to fall under gravity.)

It is sometimes useful to speak of the average force, $\vec{\mathbf{F}}_{\text{avg}}$, during a collision, defined as that constant force which, if acting over the same time interval $\Delta t = t_f - t_i$ as the actual force, would produce the same impulse and change in momentum. Thus

$$\vec{\mathbf{F}}_{\text{avg}}\,\Delta t = \int_{t_i}^{t_f} \vec{\mathbf{F}}\,dt.$$

Figure 9–10 shows the magnitude of the average force, F_{avg}, for the impulsive force of Fig. 9–9. The rectangular area $F_{\text{avg}}\,\Delta t$ equals the area under the impulsive force curve.

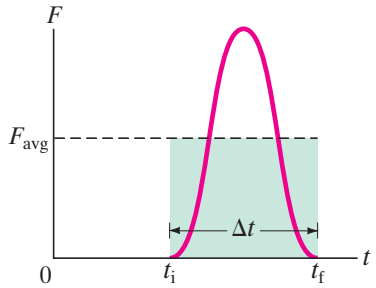


FIGURE 9–10 The average force F_{avg} acting over a very brief time interval Δt gives the same impulse $(F_{\text{avg}}\,\Delta t)$ as the actual force.

EXAMPLE 9–6 **ESTIMATE** **Karate blow.** Estimate the impulse and the average force delivered by a karate blow (Fig. 9–11) that breaks a board a few cm thick. Assume the hand moves at roughly 10 m/s when it hits the board.

APPROACH We use the momentum-impulse relation, Eq. 9–6b. The hand's speed changes from 10 m/s to zero over a distance of perhaps one cm (roughly how much your hand and the board compress before your hand comes to a stop, and the board begins to give way). The hand's mass should probably include part of the arm, and we take it to be roughly $m \approx 1\,\text{kg}$.

SOLUTION The impulse J equals the change in momentum

$$J = \Delta p = (1\,\text{kg})(10\,\text{m/s} - 0) = 10\,\text{kg}\cdot\text{m/s}.$$

We obtain the force from the definition of impulse $F_{\text{avg}} = J/\Delta t$; but what is Δt? The hand is brought to rest over the distance of roughly a centimeter: $\Delta x \approx 1\,\text{cm}$. The average speed during the impact is $\bar{v} = (10\,\text{m/s} + 0)/2 = 5\,\text{m/s}$ and equals $\Delta x/\Delta t$. Thus $\Delta t = \Delta x/\bar{v} \approx (10^{-2}\,\text{m})/(5\,\text{m/s}) = 2\times10^{-3}\,\text{s}$ or about 2 ms. The force is thus (Eq. 9–6b) about

$$F_{\text{avg}} = \frac{J}{\Delta t} = \frac{10\,\text{kg}\cdot\text{m/s}}{2\times10^{-3}\,\text{s}} \approx 5000\,\text{N} = 5\,\text{kN}.$$

FIGURE 9–11 Example 9–6.



9–4 Conservation of Energy and Momentum in Collisions

During most collisions, we usually don't know how the collision force varies over time, and so analysis using Newton's second law becomes difficult or impossible. But by making use of the conservation laws for momentum and energy, we can still determine a lot about the motion after a collision, given the motion before the collision. We saw in Section 9–2 that in the collision of two objects such as billiard balls, the total momentum is conserved. If the two objects are very hard and no heat or other form of energy is produced in the collision, then the total kinetic energy (K) of the two objects is the same after the collision as before. For the brief moment during which the two objects are in contact, some (or all) of the energy is stored momentarily in the form of elastic potential energy. But if we compare the total kinetic energy just before the collision with the total kinetic energy just after the collision, and they are found to be the same, then we say that the total kinetic energy is conserved. Such a collision is called an **elastic collision**.

CAUTION
Elastic collision: total K conserved

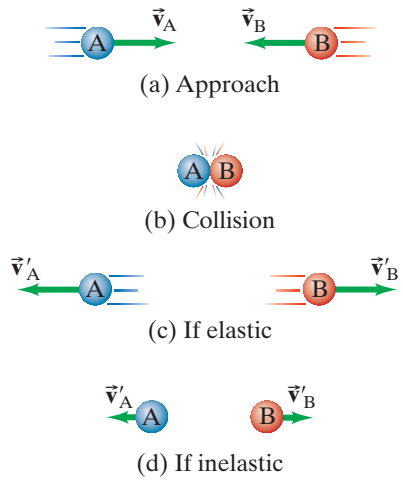


FIGURE 9–12 Two equal-mass objects (a) approach each other with equal speeds, (b) collide, and then (c) bounce off with equal speeds in the opposite directions if the collision is elastic, or (d) bounce back much less or not at all if the collision is inelastic (some of the kinetic energy is transformed to other forms of energy such as sound and heat).

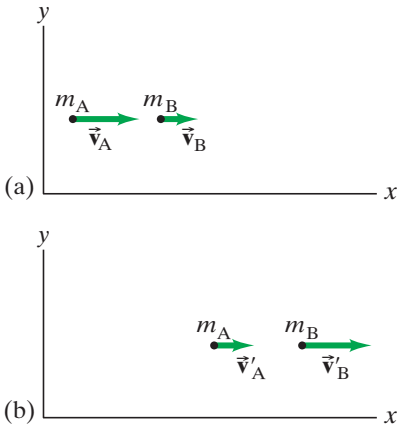


FIGURE 9–13 Two small objects of masses m_A and m_B, (a) before the collision and (b) after the collision.

If we use the subscripts A and B to represent the two objects, we can write the equation for conservation of total kinetic energy as

$$\text{total kinetic energy before} = \text{total kinetic energy after}$$

$$\tfrac{1}{2}m_A v_A^2 + \tfrac{1}{2}m_B v_B^2 = \tfrac{1}{2}m_A v_A'^2 + \tfrac{1}{2}m_B v_B'^2. \qquad \text{[elastic collision]} \quad \textbf{(9–7)}$$

Primed quantities ($'$) mean after the collision, and unprimed mean before the collision, just as in Eq. 9–3 for conservation of momentum.

At the atomic level the collisions of atoms and molecules are often elastic. But in the "macroscopic" world of ordinary objects, an elastic collision is an ideal that is never quite reached, because at least a little thermal energy (and perhaps sound and other forms of energy) is always produced during a collision. The collision of two hard elastic balls, such as billiard balls, however, is very close to being perfectly elastic, and we often treat it as such.

We do need to remember that even when the kinetic energy is not conserved, the *total* energy is always conserved.

Collisions in which kinetic energy is not conserved are said to be **inelastic collisions**. The kinetic energy that is lost is changed into other forms of energy, often thermal energy, so that the total energy (as always) is conserved. In this case,

$$K_A + K_B = K'_A + K'_B + \text{thermal and other forms of energy.}$$

See Fig. 9–12, and the details in its caption. We discuss inelastic collisions in Section 9–6.

9–5 Elastic Collisions in One Dimension

We now apply the conservation laws for momentum and kinetic energy to an elastic collision between two small objects that collide head-on, so all the motion is along a line. To be general, we assume that the two objects are moving, and their velocities v_A and v_B are along the x axis before the collision, Fig. 9–13a. After the collision, their velocities are v'_A and v'_B, Fig. 9–13b. For any $v > 0$, the object is moving to the right (increasing x), whereas for $v < 0$, the object is moving to the left (toward decreasing values of x).

From conservation of momentum, we have

$$m_A v_A + m_B v_B = m_A v'_A + m_B v'_B.$$

Because the collision is assumed to be elastic, kinetic energy is also conserved:

$$\tfrac{1}{2}m_A v_A^2 + \tfrac{1}{2}m_B v_B^2 = \tfrac{1}{2}m_A v_A'^2 + \tfrac{1}{2}m_B v_B'^2.$$

We have two equations, so we can solve for two unknowns. If we know the masses and velocities before the collision, then we can solve these two equations for the velocities after the collision, v'_A and v'_B. We derive a helpful result by rewriting the momentum equation as

$$m_A(v_A - v'_A) = m_B(v'_B - v_B), \qquad \textbf{(i)}$$

and we rewrite the kinetic energy equation as

$$m_A(v_A^2 - v_A'^2) = m_B(v_B'^2 - v_B^2).$$

Noting that algebraically $a^2 - b^2 = (a - b)(a + b)$, we write this last equation as

$$m_A(v_A - v'_A)(v_A + v'_A) = m_B(v'_B - v_B)(v'_B + v_B). \qquad \textbf{(ii)}$$

We divide Eq. (ii) by Eq. (i), and (assuming $v_A \neq v'_A$ and $v_B \neq v'_B$)[†] obtain

$$v_A + v'_A = v'_B + v_B.$$

[†]Note that Eqs. (i) and (ii), which are the conservation laws for momentum and kinetic energy, are both satisfied by the solution $v'_A = v_A$ and $v'_B = v_B$. This is a valid solution, but not very interesting. It corresponds to no collision at all—when the two objects miss each other.

We can rewrite this equation as

$$v_A - v_B = v'_B - v'_A$$

or

$$v_A - v_B = -(v'_A - v'_B). \quad [\text{head-on (1-D) elastic collision}] \quad \textbf{(9–8)}$$

CAUTION
Relative speeds (1-D only) (independent of masses)

This is an interesting result: it tells us that for any elastic head-on collision, the relative speed of the two objects after the collision $(v'_A - v'_B)$ has the same magnitude (but opposite direction) as before the collision, no matter what the masses are.

Equation 9–8 was derived from conservation of kinetic energy for elastic collisions, and can be used in place of it. Because the v's are not squared in Eq. 9–8, it is simpler to use in calculations than the conservation of kinetic energy equation (Eq. 9–7) directly.

EXAMPLE 9–7 **Equal masses.** Billiard ball A of mass m moving with speed v_A collides head-on with ball B of equal mass. What are the speeds of the two balls after the collision, assuming it is elastic? Assume (*a*) both balls are moving initially (v_A and v_B), (*b*) ball B is initially at rest $(v_B = 0)$.

APPROACH There are two unknowns, v'_A and v'_B, so we need two independent equations. We focus on the time interval from just before the collision until just after. No net external force acts on our system of two balls (mg and the normal force cancel), so momentum is conserved. Conservation of kinetic energy applies as well because we are told the collision is elastic.

SOLUTION (*a*) The masses are equal $(m_A = m_B = m)$ so conservation of momentum gives

$$v_A + v_B = v'_A + v'_B.$$

We need a second equation, because there are two unknowns. We could use the conservation of kinetic energy equation, or the simpler Eq. 9–8 derived from it:

$$v_A - v_B = v'_B - v'_A.$$

We add these two equations and obtain

$$v'_B = v_A$$

and then subtract the two equations to obtain

$$v'_A = v_B.$$

That is, the balls exchange velocities as a result of the collision: ball B acquires the velocity that ball A had before the collision, and vice versa.

(*b*) If ball B is at rest initially, so that $v_B = 0$, we have

$$v'_B = v_A \quad \text{and} \quad v'_A = 0.$$

That is, ball A is brought to rest by the collision, whereas ball B acquires the original velocity of ball A. See Fig. 9–14.

NOTE The result in part (*b*) is often observed by billiard and pool players, and is valid only if the two balls have equal masses (and no spin is given to the balls).

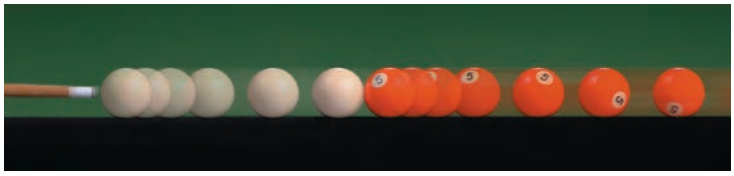


FIGURE 9–14 In this multi-flash photo of a head-on collision between two balls of equal mass, the white cue ball is accelerated from rest by the cue stick and then strikes the red ball, initially at rest. The white ball stops in its tracks and the (equal mass) red ball moves off with the same speed as the white ball had before the collision. See Example 9–7. (Actually if you measure the distance between the equally-timed strobe photos of the red ball, they are slightly less than for the white cue ball; the red ball's speed is slightly less: a little kinetic energy was transformed to heat or sound energy.)

EXAMPLE 9–8 **Unequal masses, target at rest.** A very common practical situation is for a moving object (m_A) to strike a second object (m_B, the "target") at rest $(v_B = 0)$. Assume the objects have unequal masses, and that the collision is elastic and occurs along a line (head-on). (*a*) Derive equations for v'_B and v'_A in terms of the initial velocity v_A of mass m_A and the masses m_A and m_B. (*b*) Determine the final velocities if the moving object m_A is much more massive than the target $(m_A \gg m_B)$. (*c*) Determine the final velocities if the moving object is much less massive than the target $(m_A \ll m_B)$.

APPROACH The momentum equation (with $v_B = 0$) is

$$m_B v'_B = m_A(v_A - v'_A).$$

Kinetic energy is also conserved, and to use it we use Eq. 9–8 and rewrite it as

$$v'_A = v'_B - v_A.$$

SOLUTION (*a*) We substitute the above v'_A equation into the momentum equation $(m_B v'_B = m_A v_A - m_A v'_B + m_A v_A)$ and rearrange to find

$$v'_B = v_A\left(\frac{2m_A}{m_A + m_B}\right).$$

We substitute this value for v'_B back into the equation $v'_A = v'_B - v_A$ to obtain

$$v'_A = v_A\left(\frac{2m_A}{m_A + m_B}\right) - v_A = v_A\left(\frac{2m_A}{m_A + m_B} - 1\right),$$

or

$$v'_A = v_A\left(\frac{m_A - m_B}{m_A + m_B}\right).$$

To check these two equations we have derived, we let $m_A = m_B$, and we obtain

$$v'_B = v_A \quad \text{and} \quad v'_A = 0.$$

This is the same case treated in Example 9–7, and we get the same result: for objects of equal mass, one of which is initially at rest, the velocity of the one moving initially is completely transferred to the object originally at rest.

(*b*) We are given $v_B = 0$ and $m_A \gg m_B$. A very heavy moving object strikes a light object at rest, and we have, using the relations for v'_B and v'_A above,

$$v'_B \approx 2v_A \quad \text{and} \quad v'_A \approx v_A.$$

Thus the velocity of the heavy incoming object is practically unchanged, whereas the light object, originally at rest, takes off with twice the velocity of the heavy one. The velocity of a heavy bowling ball, for example, is hardly affected by striking a much lighter bowling pin.

(*c*) This time we have $v_B = 0$ and $m_A \ll m_B$. A moving light object strikes a very massive object at rest. In this case, using the equations in part (*a*)

$$v'_B \approx 0 \quad \text{and} \quad v'_A \approx -v_A.$$

The massive object at rest remains essentially at rest, and the very light incoming object rebounds with essentially its same speed but in the opposite direction. For example, a tennis ball colliding head-on with a stationary bowling ball will hardly affect the bowling ball, but will rebound with nearly the same speed it had initially, just as if it had struck a hard wall.

NOTE In nuclear reactors, neutrons are crucial to create a chain reaction of uranium fissions for producing energy release, as discussed in Chapter 42, Section 42–3. The neutrons (released in the fission reactions) must be slowed down in order to be effective in producing further fissions, and also not escape. The material that slows down neutrons by collisions, called the *moderator*, slows neutrons down faster if made up of material whose atoms have mass closer to that of a neutron, as we saw in Example 9–7 (also part (*a*) above): one collision with an equal mass particle can slow a neutron to zero. But if it collides with a much lighter or heavier particle (parts (*b*) and (*c*) above) its speed doesn't change much.

PHYSICS APPLIED
Nuclear reactors: neutrons and moderator

It can be shown (it is given as Problem 34) that for any elastic head-on collision the speeds after the collision are

$$v'_B = v_A\left(\frac{2m_A}{m_A + m_B}\right) + v_B\left(\frac{m_B - m_A}{m_A + m_B}\right)$$

and

$$v'_A = v_A\left(\frac{m_A - m_B}{m_A + m_B}\right) + v_B\left(\frac{2m_B}{m_A + m_B}\right).$$

These general equations should not be memorized. They can always be derived from the conservation laws: it is simplest to start a Problem from scratch, as we did above in Example 9–8, and as shown in the next Example.

EXAMPLE 9–9 **A nuclear collision.** A proton (p) of mass 1.01 u (unified atomic mass units) traveling with a speed of 3.60×10^4 m/s has an elastic head-on collision with a helium (He) nucleus $(m_{He} = 4.00\,\text{u})$ initially at rest. What are the velocities of the proton and helium nucleus after the collision? (As mentioned in Chapter 1, $1\,\text{u} = 1.66 \times 10^{-27}$ kg, but we won't need this fact.) Assume this nuclear collision takes place in nearly empty space.

APPROACH Like Examples 9–7 and 9–8, this is an elastic head-on collision. The masses of our two particles are not equal. The only external force could be Earth's gravity, but it is insignificant compared to the powerful forces between the two particles at the moment of collision. So again we use the conservation laws of momentum and of kinetic energy, and apply them to our system of two particles.

SOLUTION We use the subscripts p for the proton and He for the helium nucleus. We are given $v_{He} = 0$ and $v_p = 3.60 \times 10^4$ m/s. We want to find the velocities v'_p and v'_{He} after the collision. From conservation of momentum,

$$m_p v_p + 0 = m_p v'_p + m_{He} v'_{He}.$$

Because the collision is elastic, the kinetic energy of our system of two particles is conserved and we can use Eq. 9–8, which becomes

$$v_p - 0 = v'_{He} - v'_p.$$

Thus

$$v'_p = v'_{He} - v_p,$$

and substituting this into our momentum equation displayed above, we get

$$m_p v_p = m_p v'_{He} - m_p v_p + m_{He} v'_{He}.$$

Solving for v'_{He}, the only unknown in this equation, we obtain

$$v'_{He} = \frac{2m_p v_p}{m_p + m_{He}} = \frac{2(1.01\,\text{u})(3.60 \times 10^4\,\text{m/s})}{(4.00\,\text{u} + 1.01\,\text{u})} = 1.45 \times 10^4\,\text{m/s}.$$

The other unknown is v'_p, which we can now obtain from

$$v'_p = v'_{He} - v_p = (1.45 \times 10^4\,\text{m/s}) - (3.60 \times 10^4\,\text{m/s}) = -2.15 \times 10^4\,\text{m/s}.$$

The minus sign for v'_p tells us that the proton reverses direction upon collision, and we see that its speed is less than its initial speed (see Fig. 9–15).

NOTE This result makes sense: the lighter proton would be expected to "bounce back" from the more massive helium nucleus, but not with its full original velocity as from a rigid wall (which corresponds to extremely large, or infinite, mass).

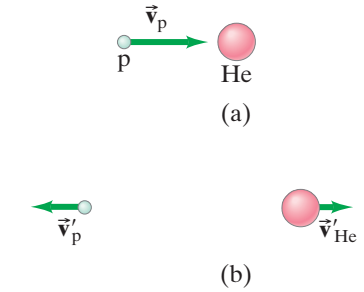


FIGURE 9–15 Example 9–9: (a) before collision, (b) after collision.

9–6 Inelastic Collisions

Collisions in which kinetic energy is not conserved are called **inelastic collisions**. Some of the initial kinetic energy is transformed into other types of energy, such as thermal or potential energy, so the total kinetic energy after the collision is less than the total kinetic energy before the collision. The inverse can also happen when potential energy (such as chemical or nuclear) is released, in which case the total kinetic energy after the interaction can be greater than the initial kinetic energy. Explosions are examples of this type.

Typical macroscopic collisions are inelastic, at least to some extent, and often to a large extent. If two objects stick together as a result of a collision, the collision is said to be **completely inelastic**. Two colliding balls of putty that stick together or two railroad cars that couple together when they collide are examples of completely inelastic collisions. The kinetic energy in some cases is all transformed to other forms of energy in an inelastic collision, but in other cases only part of it is. In Example 9–3, for instance, we saw that when a traveling railroad car collided with a stationary one, the coupled cars traveled off with some kinetic energy. In a completely inelastic collision, the maximum amount of kinetic energy is transformed to other forms consistent with conservation of momentum. Even though kinetic energy is not conserved in inelastic collisions, the total energy is always conserved, and the total vector momentum is also conserved.

PHYSICS APPLIED
Ballistic pendulum

EXAMPLE 9–10 **Ballistic pendulum.** The *ballistic pendulum* is a device used to measure the speed of a projectile, such as a bullet. The projectile, of mass m, is fired into a large block (of wood or other material) of mass M, which is suspended like a pendulum. (Usually, M is somewhat greater than m.) As a result of the collision, the pendulum and projectile together swing up to a maximum height h, Fig. 9–16. Determine the relationship between the initial horizontal speed of the projectile, v, and the maximum height h.

APPROACH We can analyze the process by dividing it into two parts or two time intervals: (1) the time interval from just before to just after the collision itself, and (2) the subsequent time interval in which the pendulum moves from the vertical hanging position to the maximum height h.

In part (1), Fig. 9–16a, we assume the collision time is very short, so that the projectile is embedded in the block before the block has moved significantly from its rest position directly below its support. Thus there is effectively no net external force, and we can apply conservation of momentum to this completely inelastic collision.

In part (2), Fig. 9–16b, the pendulum begins to move, although gravity is acting, tending to pull it back toward the vertical position.† So for part (2), we cannot use conservation of momentum. But we can use conservation of mechanical energy because gravity is a conservative force (Chapter 8). The kinetic energy immediately after the collision is changed entirely to gravitational potential energy when the pendulum reaches its maximum height, h.

SOLUTION In part (1) momentum is conserved:

$$\begin{aligned} \text{total } P \text{ before} &= \text{total } P \text{ after} \\ mv &= (m + M)\,v', \end{aligned} \qquad \textbf{(i)}$$

where v' is the speed of the block and embedded projectile just after the collision, before they have moved significantly.

In part (2), mechanical energy is conserved. We choose $y = 0$ when the pendulum hangs vertically, and then $y = h$ when the pendulum–projectile system reaches its maximum height. Thus we write

$$(K + U) \text{ just after collision} = (K + U) \text{ at pendulum's maximum height}$$

or

$$\tfrac{1}{2}(m + M)\,v'^2 + 0 = 0 + (m + M)\,gh. \qquad \textbf{(ii)}$$

We solve for v':

$$v' = \sqrt{2gh}.$$

Inserting this result for v' into Eq. (i) above, and solving for v, gives

$$v = \frac{m + M}{m} v' = \frac{m + M}{m} \sqrt{2gh},$$

which is our final result.

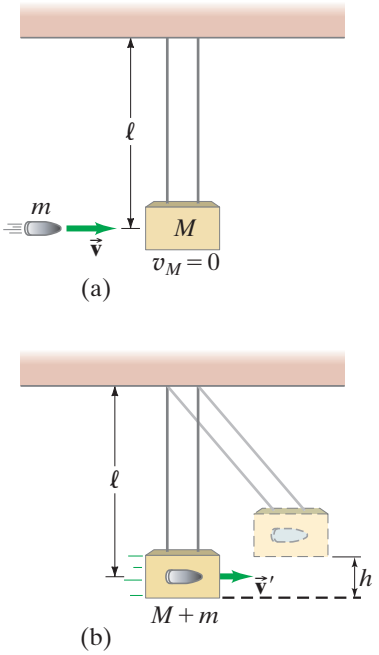


FIGURE 9–16 Ballistic pendulum. Example 9–10.

†The force due to tension in the cords is perpendicular to the motion and so does no work—see Fig. 8–14.

NOTE The separation of the process into two parts was crucial. Such an analysis is a powerful problem-solving tool. But how do you decide how to make such a division? Think about the conservation laws. They are your *tools*. Start a problem by asking yourself whether the conservation laws apply in the given situation. Here, we determined that momentum is conserved only during the brief collision, which we called part (1). But in part (1), because the collision is inelastic, the conservation of mechanical energy is not valid. Then in part (2), conservation of mechanical energy is valid, but not conservation of momentum.

Note, however, that if there had been significant motion of the pendulum during the deceleration of the projectile in the block, then there *would* have been an external force (gravity) during the collision, so conservation of momentum would not have been valid in part (1).

PROBLEM SOLVING
Use the conservation laws to analyze a problem

EXAMPLE 9–11 **Railroad cars again.** For the completely inelastic collision of the two railroad cars that we considered in Example 9–3, calculate how much of the initial kinetic energy is transformed to thermal or other forms of energy.

APPROACH The railroad cars stick together after the collision, so this is a completely inelastic collision. By subtracting the total kinetic energy after the collision from the total initial kinetic energy, we can find how much energy is transformed to other types of energy.

SOLUTION Before the collision, only car A is moving, so the total initial kinetic energy is $\frac{1}{2}m_A v_A^2 = \frac{1}{2}(10{,}000\,\text{kg})\,(8.0\,\text{m/s})^2 = 3.20 \times 10^5\,\text{J}$. After the collision, both cars are moving with half the speed, $v' = 4.0\,\text{m/s}$, by conservation of momentum (Example 9–3). So the total kinetic energy afterward is $K' = \frac{1}{2}(m_A + m_B)v'^2 = \frac{1}{2}(20{,}000\,\text{kg})\,(4.0\,\text{m/s})^2 = 1.60 \times 10^5\,\text{J}$. Hence the energy transformed to other forms is

$$(3.20 \times 10^5\,\text{J}) - (1.60 \times 10^5\,\text{J}) \;=\; 1.60 \times 10^5\,\text{J},$$

which is half the original kinetic energy.

9–7 Collisions in 2 or 3 Dimensions

Conservation of momentum and energy can also be applied to collisions in two or three dimensions, where the vector nature of momentum is especially important. One common type of non-head-on collision is that in which a moving object (called the "projectile") strikes a second object initially at rest (the "target"). This is the common situation in games such as billiards and pool, and for experiments in atomic and nuclear physics (the projectiles, from radioactive decay or a high-energy accelerator, strike a stationary target nucleus, Fig. 9–17).

Figure 9–18 shows the incoming projectile, m_A, heading along the x axis toward the target object, m_B, which is initially at rest. If these are billiard balls, m_A strikes m_B not quite head-on and they go off at the angles θ'_A and θ'_B, respectively, which are measured relative to m_A's initial direction (the x axis).†

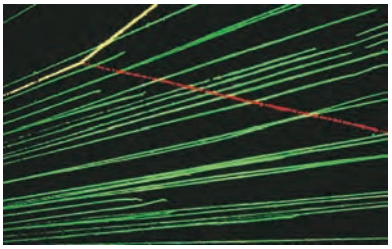


FIGURE 9–17 A recent color-enhanced version of a cloud-chamber photograph made in the early days (1920s) of nuclear physics. Green lines are paths of helium nuclei (He) coming from the left. One He, highlighted in yellow, strikes a proton of the hydrogen gas in the chamber, and both scatter at an angle; the scattered proton's path is shown in red.

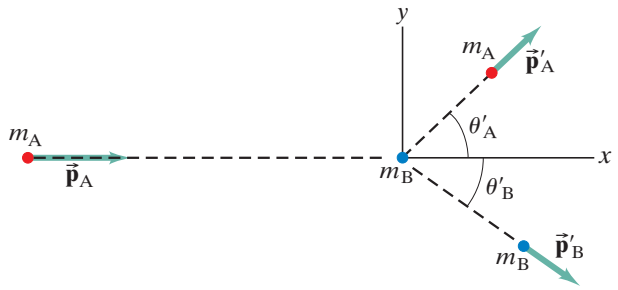


FIGURE 9–18 Object A, the projectile, collides with object B, the target. After the collision, they move off with momenta $\vec{p}'_A$ and $\vec{p}'_B$ at angles θ'_A and θ'_B.

†The objects may begin to deflect even before they touch if electric, magnetic, or nuclear forces act between them. You might think, for example, of two magnets oriented so that they repel each other: when one moves toward the other, the second moves away before the first one touches it.

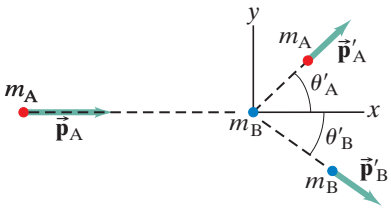


FIGURE 9–18 (Repeated.) Projectile A hits target B; A and B move off at angles θ'_A, θ'_B.

Let us apply the law of conservation of momentum to a collision like that of Fig. 9–18. We choose the xy plane to be the plane in which the initial and final momenta lie. Momentum is a vector, and because the total momentum is conserved, its components in the x and y directions also are conserved. The x component of momentum conservation gives

$$p_{Ax} + p_{Bx} = p'_{Ax} + p'_{Bx}$$

or, with $p_{Bx} = m_B v_{Bx} = 0$,

p_x conserved

$$m_A v_A = m_A v'_A \cos\theta'_A + m_B v'_B \cos\theta'_B, \qquad \textbf{(9–9a)}$$

where the primes ($'$) refer to quantities *after* the collision. There is no motion in the y direction initially, so the y component of the total momentum is zero before the collision. The y component equation of momentum conservation is then

$$p_{Ay} + p_{By} = p'_{Ay} + p'_{By}$$

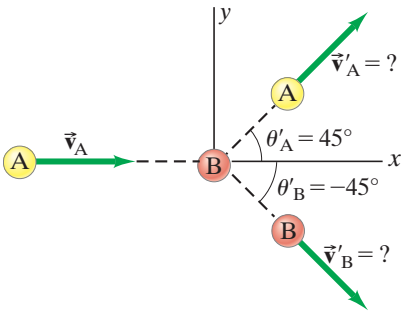
or

p_y conserved

$$0 = m_A v'_A \sin\theta'_A + m_B v'_B \sin\theta'_B. \qquad \textbf{(9–9b)}$$

When we have two independent equations, we can solve for two unknowns at most.

FIGURE 9–19 Example 9–12.



EXAMPLE 9–12 Billiard ball collision in 2-D. Billiard ball A moving with speed $v_A = 3.0\,\text{m/s}$ in the $+x$ direction (Fig. 9–19) strikes an equal-mass ball B initially at rest. The two balls are observed to move off at 45° to the x axis, ball A above the x axis and ball B below. That is, $\theta'_A = 45°$ and $\theta'_B = -45°$ in Fig. 9–19. What are the speeds of the two balls after the collision?

APPROACH There is no net external force on our system of two balls, assuming the table is level (the normal force balances gravity). Thus momentum conservation applies, and we apply it to both the x and y components using the xy coordinate system shown in Fig. 9–19. We get two equations, and we have two unknowns, v'_A and v'_B. From *symmetry* we might guess that the two balls have the same speed. But let us not assume that now. Even though we are not told whether the collision is elastic or inelastic, we can still use conservation of momentum.

SOLUTION We apply conservation of momentum for the x and y components, Eqs. 9–9a and b, and we solve for v'_A and v'_B. We are given $m_A = m_B (= m)$, so

$$\text{(for } x\text{)} \qquad mv_A = mv'_A \cos(45°) + mv'_B \cos(-45°)$$

and

$$\text{(for } y\text{)} \qquad 0 = mv'_A \sin(45°) + mv'_B \sin(-45°).$$

The m's cancel out in both equations (the masses are equal). The second equation yields [recall from trigonometry that $\sin(-\theta) = -\sin\theta$]:

$$v'_B = -v'_A \frac{\sin(45°)}{\sin(-45°)} = -v'_A\left(\frac{\sin 45°}{-\sin 45°}\right) = v'_A.$$

So they do have equal speeds as we guessed at first. The x component equation gives [recall that $\cos(-\theta) = \cos\theta$]:

$$v_A = v'_A \cos(45°) + v'_B \cos(45°) = 2v'_A \cos(45°);$$

solving for v'_A (which also equals v'_B) gives

$$v'_A = \frac{v_A}{2\cos(45°)} = \frac{3.0\,\text{m/s}}{2(0.707)} = 2.1\,\text{m/s}.$$

If we know that a collision is elastic, we can also apply conservation of kinetic energy and obtain a third equation in addition to Eqs. 9–9a and b:

$$K_A + K_B = K'_A + K'_B$$

or, for the collision shown in Fig. 9–18 or 9–19 (where $K_B = 0$),

$$\tfrac{1}{2}m_A v_A^2 = \tfrac{1}{2}m_A v_A'^2 + \tfrac{1}{2}m_B v_B'^2. \quad \text{[elastic collision]} \quad \textbf{(9–9c)}$$

Total K conserved, if elastic

If the collision is elastic, we have three independent equations and can solve for three unknowns. If we are given m_A, m_B, v_A (and v_B, if it is not zero), we cannot, for example, predict the final variables, v'_A, v'_B, θ'_A, and θ'_B, because there are four of them. However, if we measure one of these variables, say θ'_A, then the other three variables (v'_A, v'_B, and θ'_B) are uniquely determined, and we can determine them using Eqs. 9–9a, b, and c.

A note of caution: Eq. 9–8 (page 237) does *not* apply for two-dimensional collisions. It works only when a collision occurs along a line.

⚠ CAUTION

Equation 9–8 applies only in 1-D

EXAMPLE 9–13 **Proton–proton collision.** A proton traveling with speed 8.2×10^5 m/s collides elastically with a stationary proton in a hydrogen target as in Fig. 9–18. One of the protons is observed to be scattered at a 60° angle. At what angle will the second proton be observed, and what will be the velocities of the two protons after the collision?

APPROACH We saw a two-dimensional collision in Example 9–12, where we needed to use only conservation of momentum. Now we are given less information: we have three unknowns instead of two. Because the collision is elastic, we can use the kinetic energy equation as well as the two momentum equations.

SOLUTION Since $m_A = m_B$, Eqs. 9–9a, b, and c become

$$v_A = v'_A \cos\theta'_A + v'_B \cos\theta'_B \quad \textbf{(i)}$$

$$0 = v'_A \sin\theta'_A + v'_B \sin\theta'_B \quad \textbf{(ii)}$$

$$v_A^2 = v_A'^2 + v_B'^2, \quad \textbf{(iii)}$$

where $v_A = 8.2 \times 10^5$ m/s and $\theta'_A = 60°$ are given. In the first and second equations, we move the v'_A terms to the left side and square both sides of the equations:

$$v_A^2 - 2v_A v'_A \cos\theta'_A + v_A'^2 \cos^2\theta'_A = v_B'^2 \cos^2\theta'_B$$

$$v_A'^2 \sin^2\theta'_A = v_B'^2 \sin^2\theta'_B.$$

We add these two equations and use $\sin^2\theta + \cos^2\theta = 1$ to get:

$$v_A^2 - 2v_A v'_A \cos\theta'_A + v_A'^2 = v_B'^2.$$

Into this equation we substitute $v_B'^2 = v_A^2 - v_A'^2$, from equation (iii) above, and get

$$2v_A'^2 = 2v_A v'_A \cos\theta'_A$$

or

$$v'_A = v_A \cos\theta'_A = (8.2 \times 10^5 \text{ m/s})(\cos 60°) = 4.1 \times 10^5 \text{ m/s}.$$

To obtain v'_B, we use equation (iii) above (conservation of kinetic energy):

$$v'_B = \sqrt{v_A^2 - v_A'^2} = 7.1 \times 10^5 \text{ m/s}.$$

Finally, from equation (ii), we have

$$\sin\theta'_B = -\frac{v'_A}{v'_B}\sin\theta'_A = -\left(\frac{4.1 \times 10^5 \text{ m/s}}{7.1 \times 10^5 \text{ m/s}}\right)(0.866) = -0.50,$$

so $\theta'_B = -30°$. The minus sign means particle B moves at an angle below the x axis if particle A is above the axis, as in Fig. 9–18. (See Fig. 9–20 for an example of such a collision.) The two trajectories are at right angles to each other after the collision. This can be shown to be true in general for 2-D collisions of equal mass particles, one of which is at rest initially: see Problem 51.

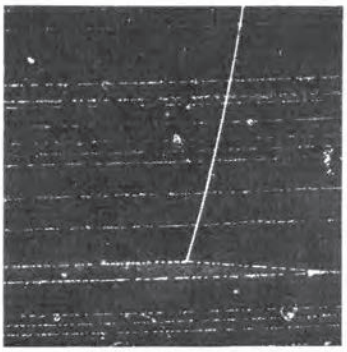


FIGURE 9–20 Photo of a proton–proton collision in a hydrogen bubble chamber (a device that makes visible the paths of elementary particles). The many horizontal lines represent incoming protons which can strike the protons of the hydrogen in the chamber.

PROBLEM SOLVING

Momentum Conservation and Collisions

1. Choose your **system**. If the situation is complex, think about how you might break it up into separate parts when one or more conservation laws apply.
2. If a significant **net external force** acts on your chosen system, be sure the time interval Δt is so short that the effect on momentum is negligible. That is, the forces that act between the interacting objects must be the only significant ones if momentum conservation is to be used. [Note: If this is valid for a portion of the problem, you can use momentum conservation only for that portion.]
3. Draw a **diagram** of the initial situation, just before the interaction (collision, explosion) takes place, and represent the momentum of each object with an arrow and a label. Do the same for the final situation, just after the interaction.
4. Choose a **coordinate system** and "+" and "−" directions. (For a head-on collision, you will need only an x axis.) It is often convenient to choose the $+x$ axis in the direction of one object's initial velocity.
5. Apply the **momentum conservation** equation(s):

   total initial momentum = total final momentum.

   You have one equation for each component (x, y, z): only one equation for a head-on collision.
6. If the collision is elastic, you can also write down a **conservation of kinetic energy** equation:

   total initial kinetic energy = total final kinetic energy.

   Alternately, for 1-D collisions, you can use Eq. 9–8:

   $$v_A - v_B = v'_B - v'_A.$$
7. Solve for the **unknown(s)**.
8. **Check** your work, check the units, and ask yourself whether the results are reasonable.



(a) (b)

FIGURE 9–21 The motion of the diver is pure translation in (a), but is translation plus rotation in (b). The black dots represent the diver's CM at each moment.

9–8 Center of Mass (CM)

Momentum is a powerful concept not only for analyzing collisions but also for analyzing the translational motion of real extended objects. Until now, whenever we have dealt with the motion of an extended object (that is, an object that has size), we have assumed that it could be approximated as a point particle or that it undergoes only translational motion. Real extended objects, however, can undergo rotational and other types of motion as well. For example, the diver in Fig. 9–21a undergoes only translational motion (all parts of the object follow the same path), whereas the diver in Fig. 9–21b undergoes both translational and rotational motion. We will refer to motion that is not pure translation as *general motion*.

Observations indicate that even if an object rotates, or several parts of a system of objects move relative to one another, there is one point that moves in the same path that a particle would move if subjected to the same net force. This point is called the **center of mass** (abbreviated CM). The general motion of an extended object (or system of objects) can be considered as *the sum of the translational motion of the CM, plus rotational, vibrational, or other types of motion about the CM*.

As an example, consider the motion of the center of mass of the diver in Fig. 9–21; the CM follows a parabolic path even when the diver rotates, as shown in Fig. 9–21b. This is the same parabolic path that a projected particle follows when acted on only by the force of gravity (projectile motion, Section 3–7). Other points in the rotating diver's body, such as her feet or head, follow more complicated paths.

Figure 9–22 shows a wrench acted on by zero net force, translating and rotating along a horizontal surface. Note that its CM, marked by a red cross, moves in a straight line, as shown by the dashed white line.

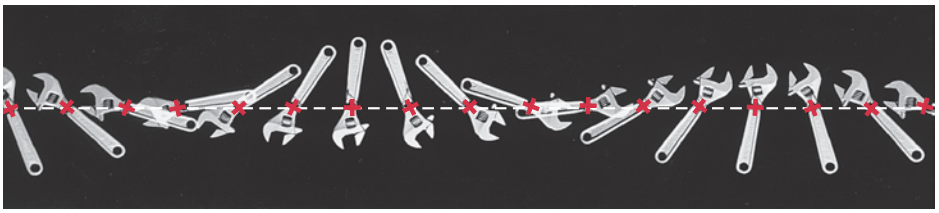


FIGURE 9–22 Translation plus rotation: a wrench moving over a smooth horizontal surface. The CM, marked with a red cross, moves in a straight line because no net force acts on the wrench.

We will show in Section 9–9 that the important properties of the CM follow from Newton's laws if the CM is defined in the following way. We can consider any extended object as being made up of many tiny particles. But first we consider a system made up of only two particles (or small objects), of masses m_A and m_B. We choose a coordinate system so that both particles lie on the x axis at positions x_A and x_B, Fig. 9–23. The center of mass of this system is defined to be at the position x_{CM}, given by

$$x_{CM} = \frac{m_A x_A + m_B x_B}{m_A + m_B} = \frac{m_A x_A + m_B x_B}{M},$$

where $M = m_A + m_B$ is the total mass of the system. The center of mass lies on the line joining m_A and m_B. If the two masses are equal $(m_A = m_B = m)$, then x_{CM} is midway between them, since in this case

$$x_{CM} = \frac{m(x_A + x_B)}{2m} = \frac{(x_A + x_B)}{2}.$$

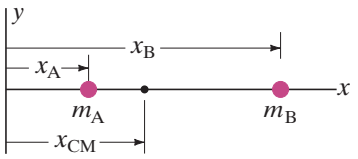


FIGURE 9–23 The center of mass of a two-particle system lies on the line joining the two masses. Here $m_A > m_B$, so the CM is closer to m_A than to m_B.

If one mass is greater than the other, say, $m_A > m_B$, then the CM is closer to the larger mass. If all the mass is concentrated at x_B, so $m_A = 0$, then $x_{CM} = (0x_A + m_B x_B)/(0 + m_B) = x_B$, as we would expect.

Now let us consider a system consisting of n particles, where n could be very large. This system could be an extended object which we consider as being made up of n tiny particles. If these n particles are all along a straight line (call it the x axis), we define the CM of the system to be located at

$$x_{CM} = \frac{m_1 x_1 + m_2 x_2 + \cdots + m_n x_n}{m_1 + m_2 + \cdots + m_n} = \frac{\sum_{i=1}^{n} m_i x_i}{M}, \qquad \textbf{(9–10)}$$

where $m_1, m_2, \ldots m_n$ are the masses of each particle and $x_1, x_2, \ldots x_n$ are their positions. The symbol $\sum_{i=1}^{n}$ is the summation sign meaning to sum over all the particles, where i takes on integer values from 1 to n. (Often we simply write $\Sigma m_i x_i$, leaving out the $i = 1$ to n.) The total mass of the system is $M = \Sigma m_i$.

EXAMPLE 9–14 **CM of three guys on a raft.** Three people of roughly equal masses m on a lightweight (air-filled) banana boat sit along the x axis at positions $x_A = 1.0\,\text{m}$, $x_B = 5.0\,\text{m}$, and $x_C = 6.0\,\text{m}$, measured from the left-hand end as shown in Fig. 9–24. Find the position of the CM. Ignore the boat's mass.

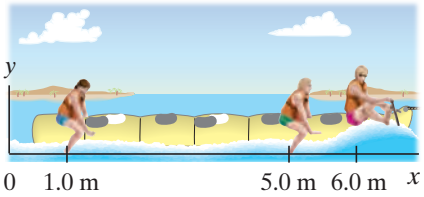
APPROACH We are given the mass and location of the three people, so we use three terms in Eq. 9–10. We approximate each person as a point particle. Equivalently, the location of each person is the position of that person's own CM.

SOLUTION We use Eq. 9–10 with three terms:

$$x_{CM} = \frac{mx_A + mx_B + mx_C}{m + m + m} = \frac{m(x_A + x_B + x_C)}{3m}$$
$$= \frac{(1.0\,\text{m} + 5.0\,\text{m} + 6.0\,\text{m})}{3} = \frac{12.0\,\text{m}}{3} = 4.0\,\text{m}.$$

The CM is 4.0 m from the left-hand end of the boat.

FIGURE 9–24 Example 9–14.



Note that the coordinates of the CM depend on the reference frame or coordinate system chosen. But the physical location of the CM is independent of that choice.

EXERCISE F Calculate the CM of the three people in Example 9–14 taking the origin at the driver $(x_C = 0)$ on the right. Is the physical location of the CM the same?

If the particles are spread out in two or three dimensions, as for a typical extended object, then we define the coordinates of the CM as

$$x_{CM} = \frac{\Sigma m_i x_i}{M}, \quad y_{CM} = \frac{\Sigma m_i y_i}{M}, \quad z_{CM} = \frac{\Sigma m_i z_i}{M}, \tag{9–11}$$

where x_i, y_i, z_i are the coordinates of the particle of mass m_i and again $M = \Sigma m_i$ is the total mass.

Although we usually calculate the components of the CM (Eq. 9–11), it is sometimes convenient (for example, for derivations) to write Eq. 9–11 in vector form. If $\vec{\mathbf{r}}_i = x_i\hat{\mathbf{i}} + y_i\hat{\mathbf{j}} + z_i\hat{\mathbf{k}}$ is the position vector of the i^{th} particle, and $\vec{\mathbf{r}}_{CM} = x_{CM}\hat{\mathbf{i}} + y_{CM}\hat{\mathbf{j}} + z_{CM}\hat{\mathbf{k}}$ is the position vector of the center of mass, then

$$\vec{\mathbf{r}}_{CM} = \frac{\Sigma m_i \vec{\mathbf{r}}_i}{M}. \tag{9–12}$$

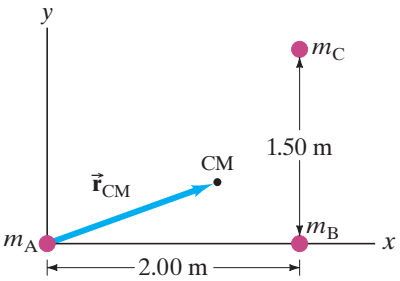


FIGURE 9–25 Example 9–15.

EXAMPLE 9–15 **Three particles in 2-D.** Three particles, each of mass 2.50 kg, are located at the corners of a right triangle whose sides are 2.00 m and 1.50 m long, as shown in Fig. 9–25. Locate the center of mass.

APPROACH We choose our coordinate system as shown (to simplify calculations) with m_A at the origin and m_B on the x axis. Then m_A has coordinates $x_A = y_A = 0$; m_B has coordinates $x_B = 2.0\,\text{m}$, $y_B = 0$; and m_C has coordinates $x_C = 2.0\,\text{m}$, $y_C = 1.5\,\text{m}$.

SOLUTION From Eqs. 9–11,

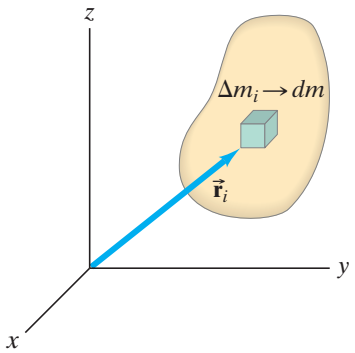
$$x_{CM} = \frac{(2.50\,\text{kg})(0) + (2.50\,\text{kg})(2.00\,\text{m}) + (2.50\,\text{kg})(2.00\,\text{m})}{3(2.50\,\text{kg})} = 1.33\,\text{m}$$

$$y_{CM} = \frac{(2.50\,\text{kg})(0) + (2.50\,\text{kg})(0) + (2.50\,\text{kg})(1.50\,\text{m})}{7.50\,\text{kg}} = 0.50\,\text{m}.$$

The CM and the position vector $\vec{\mathbf{r}}_{CM}$ are shown in Fig. 9–25, inside the "triangle" as we should expect.

EXERCISE G A diver does a high dive involving a flip and a half-pike (legs and arms straight, but body bent in half). What can you say about the diver's center of mass? (*a*) It accelerates with a magnitude of $9.8\,\text{m/s}^2$ (ignoring air friction). (*b*) It moves in a circular path because of the rotation of the diver. (*c*) It must always be roughly located inside the diver's body, somewhere in the geometric center. (*d*) All of the above are true.

FIGURE 9–26 An extended object, here shown in only two dimensions, can be considered to be made up of many tiny particles (n), each having a mass Δm_i. One such particle is shown located at a point $\vec{\mathbf{r}}_i = x_i\hat{\mathbf{i}} + y_i\hat{\mathbf{j}} + z_i\hat{\mathbf{k}}$. We take the limit of $n \to \infty$ so Δm_i becomes the infinitesimal dm.



It is often convenient to think of an extended object as made up of a continuous distribution of matter. In other words, we consider the object to be made up of n particles, each of mass Δm_i in a tiny volume around a point x_i, y_i, z_i, and we take the limit of n approaching infinity (Fig. 9–26). Then Δm_i becomes the infinitesimal mass dm at points x, y, z. The summations in Eqs. 9–11 and 9–12 become integrals:

$$x_{CM} = \frac{1}{M}\int x\,dm, \quad y_{CM} = \frac{1}{M}\int y\,dm, \quad z_{CM} = \frac{1}{M}\int z\,dm, \tag{9–13}$$

where the sum over all the mass elements is $\int dm = M$, the total mass of the object. In vector notation, this becomes

$$\vec{\mathbf{r}}_{CM} = \frac{1}{M}\int \vec{\mathbf{r}}\,dm. \tag{9–14}$$

A concept similar to *center of mass* is **center of gravity** (CG). The CG of an object is that point at which the force of gravity can be considered to act. The force of gravity actually acts on *all* the different parts or particles of an object, but for purposes of the translational motion of an object as a whole, we can assume that the entire weight of the object (the sum of the weights of all its parts) acts at the CG. There is a conceptual difference between the center of gravity and the center of mass, but for nearly all practical purposes, they are at the same point.†

†There would be a difference between the CM and CG only in the unusual case of an object so large that the acceleration due to gravity, g, was different at different parts of the object.

EXAMPLE 9–16 **CM of a thin rod.** (*a*) Show that the CM of a uniform thin rod of length ℓ and mass M is at its center. (*b*) Determine the CM of the rod assuming its linear mass density λ (its mass per unit length) varies linearly from $\lambda = \lambda_0$ at the left end to double that value, $\lambda = 2\lambda_0$, at the right end.

APPROACH We choose a coordinate system so that the rod lies on the x axis with the left end at $x = 0$, Fig. 9–27. Then $y_{CM} = 0$ and $z_{CM} = 0$.

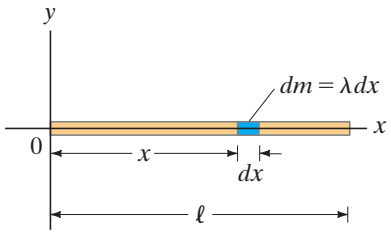


FIGURE 9–27 Example 9–16.

SOLUTION (*a*) The rod is uniform, so its mass per unit length (linear mass density λ) is constant and we write it as $\lambda = M/\ell$. We now imagine the rod as divided into infinitesimal elements of length dx, each of which has mass $dm = \lambda\, dx$. We use Eq. 9–13:

$$x_{CM} = \frac{1}{M}\int_{x=0}^{\ell} x\, dm = \frac{1}{M}\int_0^{\ell} \lambda x\, dx = \frac{\lambda}{M}\frac{x^2}{2}\bigg|_0^{\ell} = \frac{\lambda \ell^2}{2M} = \frac{\ell}{2}$$

where we used $\lambda = M/\ell$. This result, x_{CM} at the center, is what we expected.
(*b*) Now we have $\lambda = \lambda_0$ at $x = 0$ and we are told that λ increases linearly to $\lambda = 2\lambda_0$ at $x = \ell$. So we write (let α be a constant)

$$\lambda = \lambda_0(1 + \alpha x)$$

which satisfies $\lambda = \lambda_0$ at $x = 0$, increases linearly, and gives $\lambda = 2\lambda_0$ at $x = \ell$ if $(1 + \alpha\ell) = 2$. In other words, $\alpha = 1/\ell$. Again we use Eq. 9–13, with $\lambda = \lambda_0(1 + x/\ell)$:

$$x_{CM} = \frac{1}{M}\int_{x=0}^{\ell} \lambda x\, dx = \frac{1}{M}\lambda_0\int_0^{\ell}\left(1 + \frac{x}{\ell}\right)x\, dx = \frac{\lambda_0}{M}\left(\frac{x^2}{2} + \frac{x^3}{3\ell}\right)\bigg|_0^{\ell} = \frac{5}{6}\frac{\lambda_0}{M}\ell^2.$$

Now let us write M in terms of λ_0 and ℓ. We can write

$$M = \int_{x=0}^{\ell} dm = \int_0^{\ell} \lambda\, dx = \lambda_0\int_0^{\ell}\left(1 + \frac{x}{\ell}\right)dx = \lambda_0\left(x + \frac{x^2}{2\ell}\right)\bigg|_0^{\ell} = \frac{3}{2}\lambda_0\ell.$$

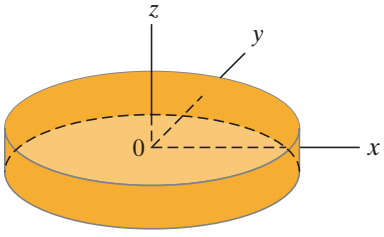
Then

$$x_{CM} = \frac{5}{6}\frac{\lambda_0}{M}\ell^2 = \frac{5}{9}\ell,$$

which is more than halfway along the rod, as we would expect since there is more mass to the right.

For symmetrically shaped objects of uniform composition, such as spheres, cylinders, and rectangular solids, the CM is located at the geometric center of the object. Consider a uniform circular cylinder, such as a solid circular disk. We expect the CM to be at the center of the circle. To show that it is, we first choose a coordinate system whose origin is at the center of the circle with the z axis perpendicular to the disk (Fig. 9–28). When we take the sum $\Sigma m_i x_i$ in Eqs. 9–11, there is as much mass at any $+x_i$ as there is at $-x_i$. So all terms cancel out in pairs and $x_{CM} = 0$. The same is true for y_{CM}. In the vertical (z) direction, the CM must lie halfway between the circular faces: if we choose our origin of coordinates at that point, there is as much mass at any $+z_i$ as at $-z_i$, so $z_{CM} = 0$. For other uniform, symmetrically shaped objects, we can make similar arguments to show that the CM must lie on a line of *symmetry*. If a symmetric body is *not* uniform, then these arguments do not hold. For example, the CM of a wheel or disk weighted on one side is not at the geometric center but closer to the weighted side.

FIGURE 9–28 Cylindrical disk with origin of coordinates at geometric center.



To locate the center of mass of a group of extended objects, we can use Eqs. 9–11, where the m_i are the masses of these objects and x_i, y_i, and z_i are the coordinates of the CM of each of the objects.

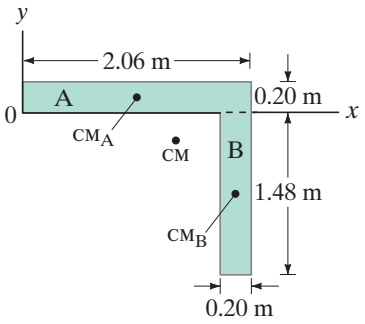


FIGURE 9–29 Example 9–17. This L-shaped object has thickness t (not shown on diagram).

EXAMPLE 9–17 **CM of L-shaped flat object.** Determine the CM of the uniform thin L-shaped construction brace shown in Fig. 9–29.

APPROACH We can consider the object as two rectangles: rectangle A, which is $2.06\,\text{m} \times 0.20\,\text{m}$, and rectangle B, which is $1.48\,\text{m} \times 0.20\,\text{m}$. We choose the origin at 0 as shown. We assume a uniform thickness t.

SOLUTION The CM of rectangle A is at

$$x_A = 1.03\,\text{m},$$
$$y_A = 0.10\,\text{m}.$$

The CM of B is at

$$x_B = 2.06\,\text{m} - \tfrac{1}{2}(0.20\,\text{m}) = 1.96\,\text{m},$$
$$y_B = -0.74\,\text{m}.$$

The mass of A, whose thickness is t, is ρV where volume $V = (\text{length}) \times (\text{width}) \times (\text{thickness})$ is

$$M_A = (2.06\,\text{m})(0.20\,\text{m})(t)(\rho) = (0.412\,\text{m}^2)(\rho t),$$

where ρ is the uniform density (mass per unit volume). The mass of B is

$$M_B = (1.48\,\text{m})(0.20\,\text{m})(\rho t) = (0.296\,\text{m}^2)(\rho t),$$

and the total mass is $M = M_A + M_B = (0.708\,\text{m}^2)(\rho t)$. Thus

$$x_{CM} = \frac{M_A x_A + M_B x_B}{M} = \frac{(0.412\,\text{m}^2)(1.03\,\text{m}) + (0.296\,\text{m}^2)(1.96\,\text{m})}{(0.708\,\text{m}^2)} = 1.42\,\text{m},$$

where ρt was canceled out in numerator and denominator. Similarly,

$$y_{CM} = \frac{(0.412\,\text{m}^2)(0.10\,\text{m}) + (0.296\,\text{m}^2)(-0.74\,\text{m})}{(0.708\,\text{m}^2)} = -0.25\,\text{m},$$

which puts the CM approximately at the point so labeled in Fig. 9–29. In thickness, $z_{CM} = t/2$, since the object is assumed to be uniform.

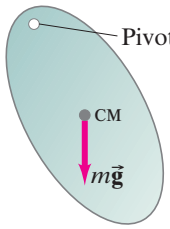
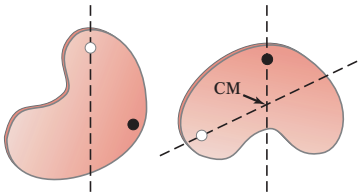


FIGURE 9–30 Determining the CM of a flat uniform body.

FIGURE 9–31 Finding the CM.



Note in this last Example that the CM can actually lie *outside* the object. Another example is a doughnut whose CM is at the center of the hole.

It is often easier to determine the CM or CG of an extended object experimentally rather than analytically. If an object is suspended from any point, it will swing (Fig. 9–30) due to the force of gravity on it, unless it is placed so its CM lies on a vertical line directly below the point from which it is suspended. If the object is two-dimensional, or has a plane of symmetry, it need only be hung from two different pivot points and the respective vertical (plumb) lines drawn. Then the center of mass will be at the intersection of the two lines, as in Fig. 9–31. If the object doesn't have a plane of symmetry, the CM with respect to the third dimension is found by suspending the object from at least three points whose plumb lines do not lie in the same plane. For symmetrically shaped objects, the CM is located at the geometric center of the object.

9–9 Center of Mass and Translational Motion

As mentioned in Section 9–8, a major reason for the importance of the concept of center of mass is that the translational motion of the CM for a system of particles (or an extended object) is directly related to the net force acting on the system as a whole. We now show this, by examining the motion of a system of n particles

of total mass M, and we assume all the masses m_i remain constant. We begin by rewriting Eq. 9–12 as

$$M\vec{\mathbf{r}}_{\text{CM}} = \Sigma m_i \vec{\mathbf{r}}_i.$$

We differentiate this equation with respect to time:

$$M\frac{d\vec{\mathbf{r}}_{\text{CM}}}{dt} = \Sigma m_i \frac{d\vec{\mathbf{r}}_i}{dt}$$

or

$$M\vec{\mathbf{v}}_{\text{CM}} = \Sigma m_i \vec{\mathbf{v}}_i, \qquad \textbf{(9–15)}$$

where $\vec{\mathbf{v}}_i = d\vec{\mathbf{r}}_i/dt$ is the velocity of the i^{th} particle of mass m_i, and $\vec{\mathbf{v}}_{\text{CM}}$ is the velocity of the CM. We take the derivative with respect to time again and obtain

$$M\frac{d\vec{\mathbf{v}}_{\text{CM}}}{dt} = \Sigma m_i \vec{\mathbf{a}}_i,$$

where $\vec{\mathbf{a}}_i = d\vec{\mathbf{v}}_i/dt$ is the acceleration of the i^{th} particle. Now $d\vec{\mathbf{v}}_{\text{CM}}/dt$ is the acceleration of the CM, $\vec{\mathbf{a}}_{\text{CM}}$. By Newton's second law, $m_i\vec{\mathbf{a}}_i = \vec{\mathbf{F}}_i$ where $\vec{\mathbf{F}}_i$ is the net force on the i^{th} particle. Therefore

$$M\vec{\mathbf{a}}_{\text{CM}} = \vec{\mathbf{F}}_1 + \vec{\mathbf{F}}_2 + \cdots + \vec{\mathbf{F}}_n = \Sigma\vec{\mathbf{F}}_i. \qquad \textbf{(9–16)}$$

That is, the vector sum of all the forces acting on the system is equal to the total mass of the system times the acceleration of its center of mass. Note that our system of n particles could be the n particles that make up one or more extended objects.

The forces $\vec{\mathbf{F}}_i$ exerted on the particles of the system can be divided into two types: (1) *external forces* exerted by objects outside the system and (2) *internal forces* that particles within the system exert on one another. By Newton's third law, the internal forces occur in pairs: if one particle exerts a force on a second particle in our system, the second must exert an equal and opposite force on the first. Thus, in the sum over all the forces in Eq. 9–16, these internal forces cancel each other in pairs. We are left, then, with only the external forces on the right side of Eq. 9–16:

$$M\vec{\mathbf{a}}_{\text{CM}} = \Sigma\vec{\mathbf{F}}_{\text{ext}}, \qquad [\text{constant } M] \quad \textbf{(9–17)}$$

NEWTON'S SECOND LAW (for a system)

where $\Sigma\vec{\mathbf{F}}_{\text{ext}}$ is the sum of all the external forces acting on our system, which is the *net force* acting on the system. Thus

the sum of all the forces acting on the system is equal to the total mass of the system times the acceleration of its center of mass.

This is **Newton's second law** for a system of particles. It also applies to an extended object (which can be thought of as a collection of particles), and to a system of objects. Thus we conclude that

the center of mass of a system of particles (or objects) with total mass M moves like a single particle of mass M acted upon by the same net external force.

TRANSLATIONAL MOTION OF CM

That is, the system translates as if all its mass were concentrated at the CM and all the external forces acted at that point. We can thus treat the *translational motion* of any object or system of objects as the motion of a particle (see Figs. 9–21 and 9–22) at the position of the CM. This result simplifies our analysis of the motion of complex systems and extended objects. Although the motion of various parts of the system may be complicated, we may often be satisfied with knowing the motion of the center of mass. This result also allows us to solve certain types of problems very easily, as illustrated by the following Example.

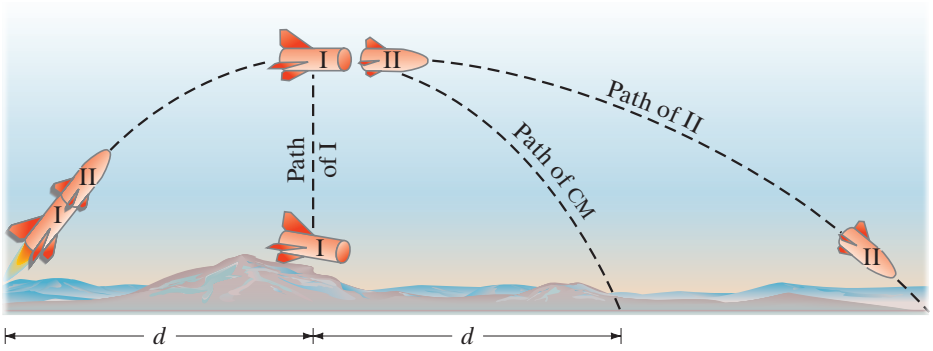


FIGURE 9–32 Example 9–18.

CONCEPTUAL EXAMPLE 9–18 **A two-stage rocket.** A rocket is shot into the air as shown in Fig. 9–32. At the moment the rocket reaches its highest point, a horizontal distance d from its starting point, a prearranged explosion separates it into two parts of equal mass. Part I is stopped in midair by the explosion and falls vertically to Earth. Where does part II land? Assume $\vec{\mathbf{g}}$ = constant.

RESPONSE After the rocket is fired, the path of the CM of the system continues to follow the parabolic trajectory of a projectile acted on only by a constant gravitational force. The CM will thus land at a point $2d$ from the starting point. Since the masses of I and II are equal, the CM must be midway between them at any time. Therefore, part II lands a distance $3d$ from the starting point.

NOTE If part I had been given a kick up or down, instead of merely falling, the solution would have been somewhat more complicated.

EXERCISE H A woman stands up in a rowboat and walks from one end of the boat to the other. How does the boat move, as seen from the shore?

We can write Eq. 9–17, $M\vec{\mathbf{a}}_{\text{CM}} = \Sigma\vec{\mathbf{F}}_{\text{ext}}$, in terms of the total momentum $\vec{\mathbf{P}}$ of a system of particles: $\vec{\mathbf{P}} = m_1\vec{\mathbf{v}}_1 + m_2\vec{\mathbf{v}}_2 + \cdots + m_n\vec{\mathbf{v}}_n = \Sigma\vec{\mathbf{p}}_i$, as we saw in Section 9–2. From Eq. 9–15 $\left(M\vec{\mathbf{v}}_{\text{CM}} = \Sigma m_i\vec{\mathbf{v}}_i\right)$ we have

$$\vec{\mathbf{P}} = M\vec{\mathbf{v}}_{\text{CM}}. \qquad \textbf{(9–18)}$$

Thus, *the total linear momentum of a system of particles is equal to the product of the total mass* M *and the velocity of the center of mass of the system.* Or, *the linear momentum of an extended object is the product of the object's mass and the velocity of its* CM.

If we differentiate Eq. 9–18 with respect to time, we obtain (assuming the total mass M is constant)

$$\frac{d\vec{\mathbf{P}}}{dt} = M\frac{d\vec{\mathbf{v}}_{\text{CM}}}{dt} = M\vec{\mathbf{a}}_{\text{CM}}.$$

From Eq. 9–17, we see that

NEWTON'S SECOND LAW (for a system)

$$\frac{d\vec{\mathbf{P}}}{dt} = \Sigma\vec{\mathbf{F}}_{\text{ext}}, \qquad \text{[same as Eq. 9–5]}$$

where $\Sigma\vec{\mathbf{F}}_{\text{ext}}$ is the net external force on the system. This is just Eq. 9–5 obtained earlier: **Newton's second law for a system of objects**. It is valid for any definite fixed system of particles or objects. If we know $\Sigma\vec{\mathbf{F}}_{\text{ext}}$, we can determine how the total momentum changes.

PHYSICS APPLIED
Distant planets discovered

An interesting application is the discovery of nearby stars (see Section 6–5) that seem to "wobble." What could cause such a wobble? It could be that a planet orbits the star, and each exerts a gravitational force on the other. The planets may be too small and too far away to be observed directly by telescopes. But the slight wobble in the motion of the star suggests that both the planet and the star (its sun) orbit about their mutual center of mass, and hence the star appears to have a wobble. Irregularities in the star's motion can be measured to high accuracy, yielding information on the size of the planets' orbits and their masses. See Fig. 6–19 in Chapter 6.

*9–10 Systems of Variable Mass; Rocket Propulsion

We now treat objects or systems whose mass varies. Such systems could be treated as a type of inelastic collision, but it is simpler to use Eq. 9–5, $d\vec{\mathbf{P}}/dt = \Sigma\vec{\mathbf{F}}_{\text{ext}}$, where $\vec{\mathbf{P}}$ is the total momentum of the system and $\Sigma\vec{\mathbf{F}}_{\text{ext}}$ is the net external force exerted on it. Great care must be taken to define the *system*, and to include all changes in momentum. An important application is to rockets, which propel themselves forward by the ejection of burned gases: the force exerted by the gases on the rocket accelerates the rocket. The mass M of the rocket decreases as it ejects gas, so for the rocket $dM/dt < 0$. Another application of changing mass is the dropping of material (gravel, packaged goods) onto a conveyor belt. In this situation, the mass M of the loaded conveyor belt increases and $dM/dt > 0$.

To treat the general case of variable mass, let us consider the system shown in Fig. 9–33. At some time t, we have a mass M with momentum $M\vec{\mathbf{v}}$. We also have a tiny (infinitesimal) mass dM traveling with velocity $\vec{\mathbf{u}}$ which is about to enter mass M (Fig. 9–33a). An infinitesimal time dt later, the mass dM combines with mass M (Fig. 9–33b). For simplicity we will refer to this as a "collision." So our large mass has changed in mass from M to $M + dM$ in the time dt. Note that dM can be less than zero, as for a rocket propelled by ejected gases whose mass M thus decreases.

In order to apply Eq. 9–5, $d\vec{\mathbf{P}}/dt = \Sigma\vec{\mathbf{F}}_{\text{ext}}$, we must consider a definite fixed system of particles. That is, in considering the change in momentum, $d\vec{\mathbf{P}}$, we must consider the momentum of the same particles initially and finally. We will thus define our *system* as including M plus dM. Then initially, at time t, the total momentum is $M\vec{\mathbf{v}} + \vec{\mathbf{u}}\,dM$ (Fig. 9–33). At time $t + dt$, after dM has combined with M, the velocity of the whole is now $\vec{\mathbf{v}} + d\vec{\mathbf{v}}$ and the total momentum is $(M + dM)(\vec{\mathbf{v}} + d\vec{\mathbf{v}})$. So the change in momentum $d\vec{\mathbf{P}}$ is

$$\begin{aligned} d\vec{\mathbf{P}} &= (M + dM)(\vec{\mathbf{v}} + d\vec{\mathbf{v}}) - (M\vec{\mathbf{v}} + \vec{\mathbf{u}}\,dM) \\ &= M\,d\vec{\mathbf{v}} + \vec{\mathbf{v}}\,dM + dM\,d\vec{\mathbf{v}} - \vec{\mathbf{u}}\,dM. \end{aligned}$$

The term $dM\,d\vec{\mathbf{v}}$ is the product of two differentials and is zero even after we "divide by dt," which we do, and apply Eq. 9–5 to obtain

$$\Sigma\vec{\mathbf{F}}_{\text{ext}} = \frac{d\vec{\mathbf{P}}}{dt} = \frac{M\,d\vec{\mathbf{v}} + \vec{\mathbf{v}}\,dM - \vec{\mathbf{u}}\,dM}{dt}.$$

Thus we get

$$\Sigma\vec{\mathbf{F}}_{\text{ext}} = M\frac{d\vec{\mathbf{v}}}{dt} - (\vec{\mathbf{u}} - \vec{\mathbf{v}})\frac{dM}{dt}. \qquad \textbf{(9–19a)}$$

Note that the quantity $(\vec{\mathbf{u}} - \vec{\mathbf{v}})$ is the relative velocity, $\vec{\mathbf{v}}_{\text{rel}}$, of dM with respect to M. That is,

$$\vec{\mathbf{v}}_{\text{rel}} = \vec{\mathbf{u}} - \vec{\mathbf{v}}$$

is the velocity of the entering mass dM as seen by an observer on M. We can rearrange Eq. 9–19a:

$$M\frac{d\vec{\mathbf{v}}}{dt} = \Sigma\vec{\mathbf{F}}_{\text{ext}} + \vec{\mathbf{v}}_{\text{rel}}\frac{dM}{dt}. \qquad \textbf{(9–19b)}$$

We can interpret this equation as follows. $Md\vec{\mathbf{v}}/dt$ is the mass times the acceleration of M. The first term on the right, $\Sigma\vec{\mathbf{F}}_{\text{ext}}$, refers to the external force on the mass M (for a rocket, it would include the force of gravity and air resistance). It does *not* include the force that dM exerts on M as a result of their collision. This is taken care of by the second term on the right, $\vec{\mathbf{v}}_{\text{rel}}\,(dM/dt)$, which represents the rate at which momentum is being transferred into (or out of) the mass M because of the mass that is added to (or leaves) it. It can thus be interpreted as the force exerted on the mass M due to the addition (or ejection) of mass.

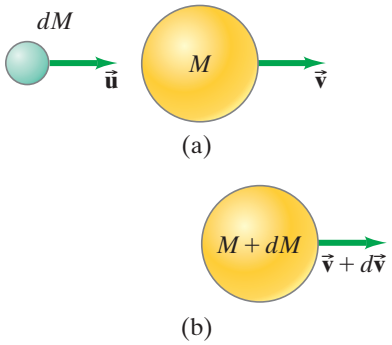


FIGURE 9–33 (a) At time t, a mass dM is about to be added to an object of mass M. (b) At time $t + dt$, the mass dM has been added to our mass which now has mass $M + dM$.

For a rocket this term $\vec{\mathbf{v}}_{\text{rel}}\,(dM/dt)$ is called the **thrust**, since it represents the force exerted on the rocket by the expelled gases. For a rocket ejecting burned fuel, $dM/dt < 0$, but so is $\vec{\mathbf{v}}_{\text{rel}}$ (gases are forced out the back), so the second term in Eq. 9–19b acts to increase $\vec{\mathbf{v}}$.

PHYSICS APPLIED
Rocket propulsion

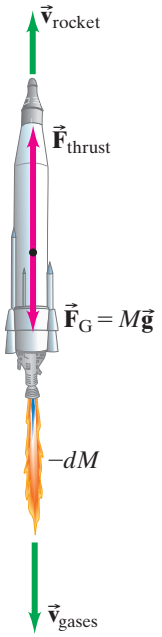


FIGURE 9–34 Example 9–19; $\vec{\mathbf{v}}_{\text{rel}} = \vec{\mathbf{v}}_{\text{gases}} - \vec{\mathbf{v}}_{\text{rocket}}$. M is the mass of the rocket at any instant and is decreasing until burnout.

EXAMPLE 9–19 **Rocket propulsion.** A fully fueled rocket has a mass of 21,000 kg, of which 15,000 kg is fuel. The burned fuel is spewed out the rear at a rate of 190 kg/s with a speed of 2800 m/s relative to the rocket. If the rocket is fired vertically upward (Fig. 9–34) calculate: (*a*) the thrust of the rocket; (*b*) the net force on the rocket at blastoff, and just before burnout (when all the fuel has been used up); (*c*) the rocket's velocity as a function of time, and (*d*) its final velocity at burnout. Ignore air resistance and assume the acceleration due to gravity is constant at $g = 9.80\,\text{m/s}^2$.

APPROACH In part (*a*), the thrust is defined (see above) as the last term in Eq. 9–19b, $v_{\text{rel}}(dM/dt)$. The net force [for (*b*)] is the vector sum of the thrust and gravity. The velocity is found from Eq. 9–19b. The mass M of the rocket varies in time, decreasing as burnt fuel is forced out (most of the rocket's initial mass is fuel).

SOLUTION (*a*) The thrust is:

$$F_{\text{thrust}} = v_{\text{rel}}\frac{dM}{dt} = (-2800\,\text{m/s})(-190\,\text{kg/s}) = 5.3\times10^5\,\text{N},$$

where we have taken upward as positive so v_{rel} is negative because it is downward, and dM/dt is negative because the rocket's mass is diminishing.

(*b*) $F_{\text{ext}} = Mg = (2.1\times10^4\,\text{kg})(9.80\,\text{m/s}^2) = 2.1\times10^5\,\text{N}$ initially, when $M = 21{,}000\,\text{kg}$; then at burnout, when $M = 21{,}000\,\text{kg} - 15{,}000\,\text{kg} = 6{,}000\,\text{kg}$, $F_{\text{ext}} = (6.0\times10^3\,\text{kg})(9.80\,\text{m/s}^2) = 5.9\times10^4\,\text{N}$. The net force on the rocket at blastoff is the thrust minus the rocket's weight (the "external force") as shown in Fig. 9–34:

$$F_{\text{net}} = 5.3\times10^5\,\text{N} - 2.1\times10^5\,\text{N} = 3.2\times10^5\,\text{N}, \quad [\text{blastoff}]$$

and just before burnout it is

$$F_{\text{net}} = 5.3\times10^5\,\text{N} - 5.9\times10^4\,\text{N} = 4.7\times10^5\,\text{N}. \quad [\text{burnout}]$$

After burnout, the net force on the rocket is only that of gravity, $-5.9\times10^4\,\text{N}$, slowing it down.

(*c*) From Eq. 9–19b we have

$$dv = \frac{F_{\text{ext}}}{M}\,dt + v_{\text{rel}}\frac{dM}{M},$$

where $F_{\text{ext}} = -Mg$, and M is the mass of the rocket and is a function of time. Since v_{rel} is constant, we can integrate this easily:

$$\int_{v_0}^{v} dv = -\int_0^t g\,dt + v_{\text{rel}}\int_{M_0}^{M}\frac{dM}{M}$$

or

$$v(t) = v_0 - gt + v_{\text{rel}}\ln\frac{M}{M_0},$$

where $v(t)$ is the rocket's velocity and M its mass at any time t. Note that v_{rel} is negative ($-2800\,\text{m/s}$ in our case) because it is opposite to the motion, and that $\ln(M/M_0)$ is also negative because $M_0 > M$. Hence the last term, which represents the thrust, is positive and acts to increase the velocity.

(*d*) The time required to reach burnout is the time needed to use up all the fuel (15,000 kg) at a rate of 190 kg/s; so at burnout,

$$t = \frac{1.50\times10^4\,\text{kg}}{190\,\text{kg/s}} = 79\,\text{s}.$$

If we take $v_0 = 0$, then using the result of part (*c*):

$$v = 0 - (9.80\,\text{m/s}^2)(79\,\text{s}) + (-2800\,\text{m/s})\left(\ln\frac{6000\,\text{kg}}{21{,}000\,\text{kg}}\right) = 2700\,\text{m/s}.$$

EXAMPLE 9–20 **Conveyor belt.** You are designing a conveyor system for a gravel yard. A hopper drops gravel gently at a rate of 75.0 kg/s onto a conveyor belt that moves at a constant speed $v = 2.20\,\text{m/s}$ (Fig. 9–35). (*a*) Determine the additional force (over and above its own internal friction) needed to keep the conveyor belt moving as gravel falls on it. (*b*) What power output would be needed from the motor that drives the conveyor belt?

PHYSICS APPLIED
Moving conveyor belt

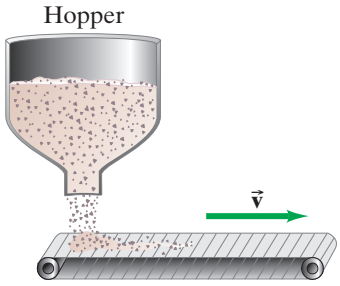


FIGURE 9–35 Example 9–20. Gravel dropped from hopper onto conveyor belt.

APPROACH We assume that the hopper is at rest so $u = 0$ in Eq. 9–19a, and that the hopper has just begun dropping gravel so $dM/dt = 75.0\,\text{kg/s}$.

SOLUTION (*a*) The belt needs to move at a constant speed $(dv/dt = 0)$, so Eq. 9–19a, written for one dimension, gives:

$$\begin{aligned} F_{\text{ext}} &= M\frac{dv}{dt} - (u - v)\frac{dM}{dt} \\ &= 0 - (0 - v)\frac{dM}{dt} \\ &= v\frac{dM}{dt} = (2.20\,\text{m/s})(75.0\,\text{kg/s}) = 165\,\text{N}. \end{aligned}$$

(*b*) This force does work at the rate (Eq. 8–21)

$$\begin{aligned} \frac{dW}{dt} &= \vec{\mathbf{F}}_{\text{ext}} \cdot \vec{\mathbf{v}} = v^2\frac{dM}{dt} \\ &= 363\,\text{W}, \end{aligned}$$

which is the minimum power output required of the motor.

NOTE This work does not all go into kinetic energy of the gravel, since

$$\frac{dK}{dt} = \frac{d}{dt}\left(\frac{1}{2}Mv^2\right) = \frac{1}{2}\frac{dM}{dt}v^2,$$

which is only half the rate at which work is done by $\vec{\mathbf{F}}_{\text{ext}}$. The other half of the external work done goes into thermal energy produced by friction between the gravel and the belt (the same friction force that accelerates the gravel).

NOTE When gravel reaches the far end of the conveyor belt and falls off, there is no effect or force on the conveyor belt. The difference here is that the gravel falling *onto* the belt from the hopper at the left must be accelerated.

CAUTION
Note difference between gravel falling on *and* off *belt*

Summary

The **linear momentum**, $\vec{\mathbf{p}}$, of an object is defined as the product of its mass times its velocity,

$$\vec{\mathbf{p}} = m\vec{\mathbf{v}}. \qquad \textbf{(9–1)}$$

In terms of momentum, **Newton's second law** can be written as

$$\Sigma\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt}. \qquad \textbf{(9–2)}$$

That is, the rate of change of momentum of an object equals the net force exerted on it.

When the net external force on a system of objects is zero, the total momentum remains constant. This is the **law of conservation of momentum**. Stated another way, the total momentum of an isolated system of objects remains constant.

The law of conservation of momentum is very useful in dealing with **collisions**. In a collision, two (or more) objects interact with each other over a very short time interval, and the force each exerts on the other during this time interval is very large compared to any other forces acting.

The **impulse** delivered by a force on an object is defined as

$$\vec{\mathbf{J}} = \int \vec{\mathbf{F}}\,dt$$

and is equal to the change in momentum of the object as long as $\vec{\mathbf{F}}$ is the net force on the object:

$$\Delta\vec{\mathbf{p}} = \vec{\mathbf{p}}_{\text{f}} - \vec{\mathbf{p}}_{\text{i}} = \int_{t_{\text{i}}}^{t_{\text{f}}} \vec{\mathbf{F}}\,dt = \vec{\mathbf{J}}. \qquad \textbf{(9–6b)}$$

Total momentum is conserved in *any* collision as long as any net external force is zero or negligible. If $m_{\text{A}}\vec{\mathbf{v}}_{\text{A}}$ and $m_{\text{B}}\vec{\mathbf{v}}_{\text{B}}$ are the momenta of two objects before the collision and $m_{\text{A}}\vec{\mathbf{v}}'_{\text{A}}$ and $m_{\text{B}}\vec{\mathbf{v}}'_{\text{B}}$ are their momenta after, then momentum conservation tells us that

$$m_{\text{A}}\vec{\mathbf{v}}_{\text{A}} + m_{\text{B}}\vec{\mathbf{v}}_{\text{B}} = m_{\text{A}}\vec{\mathbf{v}}'_{\text{A}} + m_{\text{B}}\vec{\mathbf{v}}'_{\text{B}} \qquad \textbf{(9–3)}$$

for this two-object system.

Total energy is also conserved. But this may not be helpful unless kinetic energy is conserved, in which case the collision is called an **elastic collision** and we can write

$$\tfrac{1}{2}m_A v_A^2 + \tfrac{1}{2}m_B v_B^2 = \tfrac{1}{2}m_A v_A'^2 + \tfrac{1}{2}m_B v_B'^2. \quad \textbf{(9–7)}$$

If kinetic energy is not conserved, the collision is called **inelastic**. Macroscopic collisions are generally inelastic. If two colliding objects stick together as the result of a collision, the collision is said to be **completely inelastic**.

For a system of particles, or for an extended object that can be considered as having a continuous distribution of matter, the **center of mass** (CM) is defined as

$$x_{CM} = \frac{\Sigma m_i x_i}{M}, \quad y_{CM} = \frac{\Sigma m_i y_i}{M}, \quad z_{CM} = \frac{\Sigma m_i z_i}{M} \quad \textbf{(9–11)}$$

or

$$x_{CM} = \frac{1}{M}\int x\,dm, \quad y_{CM} = \frac{1}{M}\int y\,dm, \quad z_{CM} = \frac{1}{M}\int z\,dm, \quad \textbf{(9–13)}$$

where M is the total mass of the system.

The center of mass of a system is important because this point moves like a single particle of mass M acted on by the same net external force, $\Sigma\vec{\mathbf{F}}_{ext}$. In equation form, this is just Newton's second law for a system of particles (or extended objects):

$$M\vec{\mathbf{a}}_{CM} = \Sigma\vec{\mathbf{F}}_{ext}, \quad \textbf{(9–17)}$$

where M is the total mass of the system, $\vec{\mathbf{a}}_{CM}$ is the acceleration of the CM of the system, and $\Sigma\vec{\mathbf{F}}_{ext}$ is the total (net) external force acting on all parts of the system.

For a system of particles of total linear momentum $\vec{\mathbf{P}} = \Sigma m_i\vec{\mathbf{v}}_i = M\vec{\mathbf{v}}_{CM}$, Newton's second law is

$$\frac{d\vec{\mathbf{P}}}{dt} = \Sigma\vec{\mathbf{F}}_{ext}. \quad \textbf{(9–5)}$$

[*If the mass M of an object is not constant, then

$$M\frac{d\vec{\mathbf{v}}}{dt} = \Sigma\vec{\mathbf{F}}_{ext} + \vec{\mathbf{v}}_{rel}\frac{dM}{dt} \quad \textbf{(9–19b)}$$

where $\vec{\mathbf{v}}$ is the velocity of the object at any instant and $\vec{\mathbf{v}}_{rel}$ is the relative velocity at which mass enters (or leaves) the object.]

Questions

1. We claim that momentum is conserved. Yet most moving objects eventually slow down and stop. Explain.
2. A light object and a heavy object have the same kinetic energy. Which has the greater momentum? Explain.
3. When a person jumps from a tree to the ground, what happens to the momentum of the person upon striking the ground?
4. When you release an inflated but untied balloon, why does it fly across the room?
5. Explain, on the basis of conservation of momentum, how a fish propels itself forward by swishing its tail back and forth.
6. According to Eq. 9–6a (also Eq. 9–6b), the longer the impact time t of an impulse, the smaller the force $\vec{\mathbf{F}}$ needs to be for the same momentum change, and hence the smaller the deformation of the object on which the force acts. On this basis, explain the value of *air bags*, which are intended to inflate during an automobile collision and reduce the possibility of fracture or death.
7. If a falling ball made a perfectly elastic collision with the floor, would it rebound to its original height? Explain.
8. A boy stands on the back of a rowboat and dives into the water. What happens to the boat as the boy leaves it? Explain.
9. It is said that in ancient times a rich man with a bag of gold coins was stranded on the surface of a frozen lake. Because the ice was frictionless, he could not push himself to shore and froze to death. What could he have done to save himself had he not been so miserly?
10. The speed of a tennis ball on the return of a serve can be just as fast as the serve, even though the racket isn't swung very fast. How can this be?
11. Is it possible for an object to receive a larger impulse from a small force than from a large force? Explain.
12. How could a force give zero impulse over a nonzero time interval even though the force is not zero for at least a part of that time interval?
13. In a collision between two cars, which would you expect to be more damaging to the occupants: if the cars collide and remain together, or if the two cars collide and rebound backward? Explain.
14. A very elastic ball is dropped from a height h onto a hard steel plate (fixed to the Earth), from which it rebounds at very nearly the same speed. (*a*) Is the momentum of the ball conserved during any part of this process? (*b*) If we consider the ball and the Earth as our system, during what parts of the process is momentum conserved? (*c*) Answer part (*b*) for a piece of putty that falls and sticks to the steel plate.
15. Cars used to be built as rigid as possible to withstand collisions. Today, though, cars are designed to have "crumple zones" that collapse upon impact. What is the advantage of this new design?
16. At a hydroelectric power plant, water is directed at high speed against turbine blades on an axle that turns an electric generator. For maximum power generation, should the turbine blades be designed so that the water is brought to a dead stop, or so that the water rebounds? Explain.
17. A squash ball hits a wall at a 45° angle as shown in Fig. 9–36. What is the direction (*a*) of the change in momentum of the ball, (*b*) of the force on the wall?

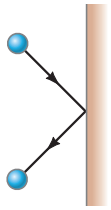


FIGURE 9–36
Question 17.

18. Two blocks of mass m_1 and m_2 rest on a frictionless table and are connected by a spring. The blocks are pulled apart, stretching the spring, and then released. Describe the subsequent motion of the two blocks.
19. Why can a batter hit a pitched baseball farther than a ball he himself has tossed up in the air?
20. If a small 20-passenger plane is not full, sometimes passengers are told they must sit in certain seats and may not move to empty seats. Why might this be?
21. Why do you tend to lean backward when carrying a heavy load in your arms?
22. Why is the CM of a 1-m length of pipe at its midpoint, whereas this is not true for your arm or leg?

23. Describe an analytic way of determining the CM of any thin, triangular-shaped, uniform plate.

24. How can a rocket change direction when it is far out in space and essentially in a vacuum?

25. Bob and Jim decide to play *tug-of-war* on a frictionless (icy) surface. Jim is considerably stronger than Bob, but Bob weighs 160 lb whereas Jim weighs 145 lb. Who loses by crossing over the midline first? Explain.

26. In one type of nuclear radioactive decay, an electron and a recoil nucleus are emitted but often do not separate along the same line. Use conservation of momentum in two dimensions to explain why this implies the emission of at least one other particle (it came to be called a "neutrino").

27. Show on a diagram how your CM shifts when you move from a lying position to a sitting position.

28. Place yourself facing the edge of an open door. Position your feet astride the door with your nose and abdomen touching the door's edge. Try to rise on your tiptoes. Why can't this be done?

29. If only an external force can change the momentum of the center of mass of an object, how can the internal force of the engine accelerate a car?

30. A rocket following a parabolic path through the air suddenly explodes into many pieces. What can you say about the motion of this system of pieces?

31. At a carnival game you try to knock over a heavy cylinder by throwing a small ball at it. You have a choice of throwing either a ball that will stick to the cylinder, or a second ball of equal mass and speed that will bounce backward off the cylinder. Which ball is more likely to make the cylinder move, and why?

MisConceptual Questions

1. Two children float motionlessly in a space station. The 20-kg girl pushes on the 40-kg boy and he sails away at 1.0 m/s. The girl
(*a*) remains motionless.
(*b*) moves in the same direction at 1.0 m/s.
(*c*) moves in the opposite direction at 1.0 m/s.
(*d*) moves in the opposite direction at 2.0 m/s.
(*e*) none of these.

2. A small boat coasts at constant speed under a bridge. A heavy sack of sand is dropped from the bridge onto the boat. The speed of the boat
(*a*) increases.
(*b*) decreases.
(*c*) does not change.
(*d*) Without knowing the mass of the boat and the sand, we can't tell.

3. An astronaut is a short distance away from her space station without a tether rope. She has a large wrench. What should she do with the wrench to move toward the space station?
(*a*) Throw it directly away from the space station.
(*b*) Throw it directly toward the space station.
(*c*) Throw it toward the station without letting go of it.
(*d*) Throw it parallel to the direction of the station's orbit.
(*e*) Throw it opposite to the direction of the station's orbit.

4. A space vehicle, in circular orbit around the Earth, collides with a small asteroid which ends up in the vehicle's storage bay. For this collision,
(*a*) only momentum is conserved.
(*b*) only kinetic energy is conserved.
(*c*) both momentum and kinetic energy are conserved.
(*d*) neither momentum nor kinetic energy is conserved.

5. A baseball is pitched horizontally toward home plate with a velocity of 110 km/h. In which of the following scenarios does the change in momentum of the baseball have the largest magnitude?
(*a*) The catcher catches the ball.
(*b*) The ball is popped straight up at a speed of 110 km/h.
(*c*) The baseball is hit straight back to the pitcher at a speed of 110 km/h.
(*d*) Scenarios (*a*) and (*b*) have the same change in momentum.
(*e*) Scenarios (*a*), (*b*), and (*c*) have the same change in momentum.

6. You are lying in bed and want to shut your bedroom door. You have a bouncy ball and a blob of clay, both with the same mass. Which one would be more effective to throw at your door to close it?
(*a*) The ball. (*c*) Both the same.
(*b*) The blob of clay. (*d*) Neither will work.

7. A small car and a heavy pickup truck are both out of gas. The truck has twice the mass of the car. After you push first the car and then the truck for the *same amount of time* with the same force, what can you say about the momentum and kinetic energy of the car and the truck? Ignore friction.
(*a*) They have the same momentum and the same kinetic energy.
(*b*) The car has more momentum and more kinetic energy than the truck.
(*c*) The truck has more momentum and more kinetic energy than the car.
(*d*) They have the same momentum, but the car has more kinetic energy than the truck.
(*e*) They have the same kinetic energy, but the truck has more momentum than the car.

8. Answer the previous Question (#7) but now assume that you push both the car and the truck for the *same distance* with the same force. [*Hint*: See also Chapter 7.]

9. A railroad tank car contains milk and moves at a constant speed along a level track. The milk begins to leak out the bottom. The car then
(*a*) slows down.
(*b*) speeds up.
(*c*) maintains a constant speed.
(*d*) Need more information about the rate of the leak.

10. A truck going 15 km/h has a head-on collision with a small car going 30 km/h. Which statement best describes the situation?
(*a*) The truck has the greater change of momentum because it has the greater mass.
(*b*) The car has the greater change of momentum because it has the greater speed.
(*c*) Neither the car nor the truck changes its momentum in the collision because momentum is conserved.
(*d*) They both have the same change in magnitude of momentum because momentum is conserved.
(*e*) None of the above is necessarily true.

11. A bowling ball hangs from a 1.0-m-long cord, Fig. 9–37: (i) A 200-gram putty ball moving 5.0 m/s hits the bowling ball and sticks to it, causing the bowling ball to swing up; (ii) a 200-gram rubber ball moving 5.0 m/s hits the bowling ball and bounces straight back at nearly 5.0 m/s, causing the bowling ball to swing up. Describe what happens.
(*a*) The bowling ball swings up by the same amount in both (i) and (ii).
(*b*) The ball swings up farther in (i) than in (ii).
(*c*) The ball swings up farther in (ii) than in (i).
(*d*) Not enough information is given; we need the contact time between the rubber ball and the bowling ball.

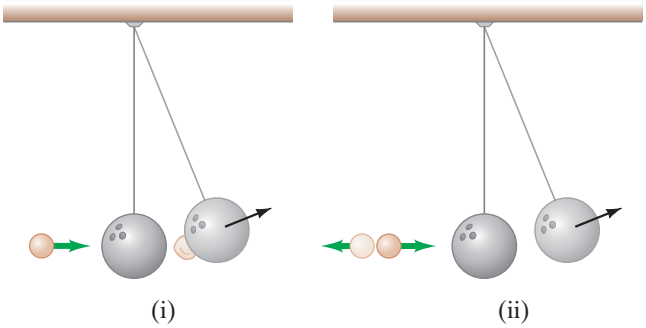


FIGURE 9–37 MisConceptual Question 11.

12. Inelastic and elastic collisions are similar in that
(*a*) momentum and kinetic energy are conserved in both.
(*b*) momentum is conserved in both.
(*c*) momentum and potential energy are conserved in both.
(*d*) kinetic energy is conserved in both.

13. Figure 9–38 shows the paths of two colliding steel balls, *A* and *B*. Which set of arrows below best represents the direction of the *change* in momentum of each ball?

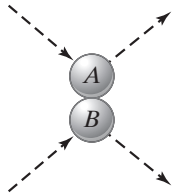
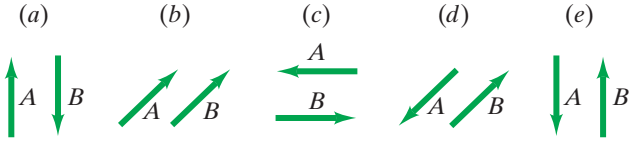


FIGURE 9–38 (Above.)



Problems

9–1 and 9–2 Momentum and Its Conservation

1. (I) What is the magnitude of the momentum of a 32-g sparrow flying with a speed of 8.4 m/s?

2. (I) A 7150-kg railroad car travels alone on a level frictionless track with a constant speed of 15.0 m/s. A 3650-kg load, initially at rest, is dropped onto the car. What will be the car's new speed?

3. (I) A 110-kg tackler moving at 2.5 m/s meets head-on (and holds on to) an 82-kg halfback moving at 4.4 m/s. What will be their mutual speed immediately after the collision?

4. (I) Calculate the force exerted on a rocket when the propelling gases are being expelled at a rate of 1200 kg/s with a speed of 4.5×10^4 m/s.

5. (II) The momentum of a particle, in SI units, is given by $\vec{\mathbf{p}} = 4.8t^2\hat{\mathbf{i}} - 8.0\hat{\mathbf{j}} - 9.4t\hat{\mathbf{k}}$. What is the force exerted on the particle as a function of time?

6. (II) A child in a boat throws a 5.30-kg package out horizontally with a speed of 10.0 m/s, Fig. 9–39. Calculate the velocity of the boat immediately after, assuming it was initially at rest. The mass of the child is 24.0 kg and that of the boat is 42.0 kg.



FIGURE 9–39 Problem 6.

7. (II) An atomic nucleus initially moving at 320 m/s emits an alpha particle in the direction of its velocity, and the remaining nucleus slows to 280 m/s. If the alpha particle has a mass of 4.0 u and the original nucleus has a mass of 222 u, what speed does the alpha particle have when it is emitted?

8. (II) The force on a particle of mass m is given by $\vec{\mathbf{F}} = 26\hat{\mathbf{i}} - 12t^2\hat{\mathbf{j}}$ where F is in N and t in s. What will be the change in the particle's momentum between $t = 1.0$ s and $t = 2.0$ s?

9. (II) An object at rest is suddenly broken apart into two fragments by an explosion. One fragment acquires twice the kinetic energy of the other. What is the ratio of their masses?

10. (II) A 22-g bullet traveling 240 m/s penetrates a 2.0-kg block of wood and emerges going 130 m/s. If the block is stationary on a frictionless surface when hit, how fast does it move after the bullet emerges?

11. (II) A 0.145-kg baseball pitched horizontally at 35.0 m/s strikes a bat and pops straight up to a height of 31.5 m. If the contact time between bat and ball is 2.5 ms, calculate the average force between the ball and bat during contact.

12. (II) A rocket of total mass 3180 kg is traveling in outer space with a velocity of 125 m/s. To alter its course by 35.0°, its rockets can be fired briefly in a direction perpendicular to its original motion. If the rocket gases are expelled at a speed of 1750 m/s, how much mass must be expelled?

13. (II) A rocket of mass m traveling with speed v_0 along the x axis suddenly shoots out fuel equal to one-third its mass, perpendicular to the x axis (along the y axis) with speed $2v_0$. Express the final velocity of the rocket in $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ notation.

14. (II) The *radioactive decay* of a neutron into a proton, an electron, and a neutrino is an example of a three-particle decay process. Use the vector nature of momentum to show that if the neutron is initially at rest, the velocity vectors of the three must be coplanar (that is, all in the same plane). The result is not true for numbers greater than three.

15. (II) A mass $m_A = 2.0\,\text{kg}$, moving with velocity $\vec{\mathbf{v}}_A = (4.0\,\hat{\mathbf{i}} + 5.0\,\hat{\mathbf{j}} - 2.0\,\hat{\mathbf{k}})\,\text{m/s}$, collides with mass $m_B = 3.0\,\text{kg}$, which is initially at rest. Immediately after the collision, mass m_A is observed traveling at velocity $\vec{\mathbf{v}}'_A = (-2.0\,\hat{\mathbf{i}} + 3.0\,\hat{\mathbf{k}})\,\text{m/s}$. Find the velocity of mass m_B after the collision. Assume no outside force acts on the two masses during the collision.

16. (III) Air in a 120-km/h wind strikes head-on the face of a building 45 m wide by 75 m high and is brought to rest. If air has a mass of 1.3 kg per cubic meter, determine the average force of the wind on the building.

17. (III) A 195-kg projectile, fired with a speed of 116 m/s at a 60.0° angle, breaks into three pieces of equal mass at the highest point of its arc (where its velocity is horizontal). Two of the fragments move with the same speed right after the explosion as the entire projectile had just before the explosion; one of these moves vertically downward and the other horizontally. Determine (*a*) the velocity of the third fragment immediately after the explosion and (*b*) the energy released in the explosion.

9–3 Collisions and Impulse

18. (I) A 0.145-kg baseball pitched at 31.0 m/s is hit on a horizontal line drive straight back at the pitcher at 46.0 m/s. If the contact time between bat and ball is $5.00 \times 10^{-3}\,\text{s}$, calculate the force (assumed to be constant) between the ball and bat.

19. (II) A golf ball of mass 0.045 kg is hit off the tee at a speed of 38 m/s. The golf club was in contact with the ball for $3.5 \times 10^{-3}\,\text{s}$. Find (*a*) the impulse imparted to the golf ball, and (*b*) the average force exerted on the ball by the golf club.

20. (II) A tennis ball of mass $m = 0.060\,\text{kg}$ and speed $v = 28\,\text{m/s}$ strikes a wall at a 45° angle and rebounds with the same speed at 45° (Fig. 9–40). What is the impulse (magnitude and direction) given to the ball?

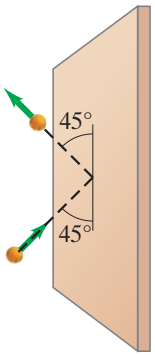


FIGURE 9–40
Problem 20.

21. (II) A 125-kg astronaut (including space suit) acquires a speed of 2.50 m/s by pushing off with her legs from a 2200-kg space capsule. (*a*) What is the change in speed of the space capsule? (*b*) If the push lasts 0.600 s, what is the average force exerted by each on the other? As the reference frame, use the position of the capsule before the push. (*c*) What is the kinetic energy of each after the push?

22. (II) A 95-kg fullback is running at 3.0 m/s to the east and is stopped in 0.75 s by a head-on tackle by a tackler running due west. Calculate (*a*) the original momentum of the fullback, (*b*) the impulse exerted on the fullback, (*c*) the impulse exerted on the tackler, and (*d*) the average force exerted on the tackler.

23. (II) Suppose the force acting on a tennis ball (mass 0.060 kg) points in the $+x$ direction and is given by the graph of Fig. 9–41 as a function of time. (*a*) Use graphical methods (count squares) to estimate the total impulse given the ball. (*b*) Estimate the velocity of the ball after being struck, assuming the tennis ball is being served so it is nearly at rest initially.

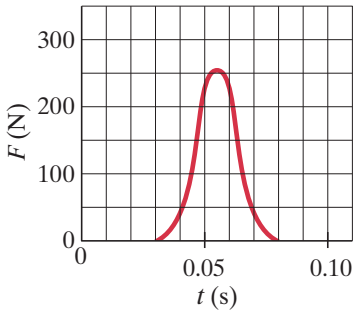


FIGURE 9–41
Problem 23.

24. (II) The force on a bullet along the barrel of a firearm is given by the formula $F = [740 - (2.3 \times 10^5\,\text{s}^{-1})\,t]\,\text{N}$ over the time interval $t = 0$ to $t = 3.0 \times 10^{-3}\,\text{s}$. (*a*) Plot a graph of F versus t for $t = 0$ to $t = 3.0\,\text{ms}$. (*b*) Use the graph to estimate the impulse given the bullet. (*c*) Determine the impulse by integration. (*d*) If the bullet achieves a speed of 260 m/s as a result of this impulse, given to it in the barrel of a gun, what must the bullet's mass be? (*e*) What is the recoil speed of the 4.5-kg gun?

25. (II) (*a*) A molecule of mass m and speed v strikes a wall at right angles and rebounds back with the same speed. If the collision time is Δt, what is the average force on the wall during the collision? (*b*) If molecules, all of this type, strike the wall at intervals a time t apart (on the average) what is the average force on the wall averaged over a long time?

26. (III) (*a*) Calculate the impulse experienced when a 68-kg person lands on firm ground after jumping from a height of 2.8 m. (*b*) Estimate the average force exerted on the person's feet by the ground if the landing is stiff-legged, and again (*c*) with bent legs. With stiff legs, assume the body moves 1.0 cm during impact, and when the legs are bent, about 50 cm. [*Hint*: The average net force on him, which is related to impulse, is the vector sum of gravity and the force exerted by the ground. See Fig. 9–42.] We will see in Chapter 12 that the force in (*b*) exceeds the ultimate strength of bone (Table 12–2).



FIGURE 9–42
Problem 26.

27. (III) A scale is adjusted so that when a large, shallow pan is placed on it, it reads zero. A water faucet at height $h = 2.35\,\text{m}$ above is turned on and water falls into the pan at a rate $R = 0.142\,\text{kg/s}$. Determine (*a*) a formula for the scale reading as a function of time t and (*b*) the reading for $t = 11.8\,\text{s}$. (*c*) Repeat (*a*) and (*b*), but replace the shallow pan with a tall, narrow cylindrical container of area $A = 18.5\,\text{cm}^2$ (the level rises in this case).

9–4 and 9–5 Elastic Collisions

28. (II) A 0.450-kg hockey puck, moving east (+ x direction) with a speed of 7.40 m/s, has a head-on collision with a 0.900-kg puck initially at rest. Assuming a perfectly elastic collision, what will be the speed and direction of each puck after the collision?

29. (II) Two billiard balls of equal mass undergo a perfectly elastic head-on collision. If one ball's initial speed was 2.00 m/s, and the other's was 3.60 m/s in the opposite direction, what will be their speeds and directions after the collision?

30. (II) A 0.280-kg croquet ball makes an elastic head-on collision with a second ball initially at rest. The second ball moves off with half the original speed of the first ball. (*a*) What is the mass of the second ball? (*b*) What fraction of the original kinetic energy ($\Delta K/K$) gets transferred to the second ball?

31. (II) A block of mass m_2 is at rest on a long frictionless table. A block of mass m_1 is projected toward the first block, and makes a perfectly elastic head-on collision with block m_2. Determine the value of m_2 (in terms of m_1) for which both blocks move with the same speed but in opposite directions after the collision.

32. (II) A ball of mass 0.320 kg that is moving with a speed of 5.5 m/s collides head-on and elastically with another ball initially at rest. Immediately after the collision, the incoming ball bounces backward with a speed of 3.8 m/s. Calculate (*a*) the velocity of the target ball after the collision, and (*b*) the mass of the target ball.

33. (II) Determine the fraction of kinetic energy lost by a neutron $(m_1 = 1.01\text{ u})$ when it collides head-on and elastically with a target particle at rest which is (*a*) ${}^{1}_{1}\text{H}$ ($m = 1.01$ u); (*b*) ${}^{2}_{1}\text{H}$ (heavy hydrogen, $m = 2.01$ u); (*c*) ${}^{12}_{6}\text{C}$ ($m = 12.00$ u); (*d*) ${}^{208}_{82}\text{Pb}$ (lead, $m = 208$ u).

34. (II) Show that, in general, for any head-on one-dimensional elastic collision, the speeds after collision are

$$v'_B = v_A\left(\frac{2m_A}{m_A + m_B}\right) + v_B\left(\frac{m_B - m_A}{m_A + m_B}\right)$$

and

$$v'_A = v_A\left(\frac{m_A - m_B}{m_A + m_B}\right) + v_B\left(\frac{2m_B}{m_A + m_B}\right),$$

where v_A and v_B are the initial speeds of the two objects of mass m_A and m_B.

35. (III) Two objects, one initially at rest, undergo a 1-D elastic collision. If one-fourth of the initial kinetic energy is transferred to the target, what is the ratio of their masses?

36. (III) A 3.0-kg block slides along a frictionless tabletop at 8.0 m/s toward a second block (at rest) of mass 4.5 kg. A coil spring, which obeys Hooke's law and has spring constant $k = 850$ N/m, is attached to the second block in such a way that it will be compressed when struck by the moving block, Fig. 9–43. (*a*) What will be the maximum compression of the spring? (*b*) What will be the final velocities of the blocks after the collision? (*c*) Is the collision elastic? Ignore the mass of the spring.

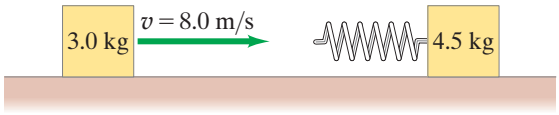


FIGURE 9–43 Problem 36.

9–6 Inelastic Collisions

37. (I) In a ballistic pendulum experiment, projectile 1 results in a maximum height h of the pendulum equal to 2.6 cm. A second projectile (of the same mass) causes the pendulum to swing twice as high, $h_2 = 5.2$ cm. The second projectile was how many times faster than the first?

38. (II) A 28-g rifle bullet traveling 190 m/s embeds itself in a 3.1-kg pendulum hanging on a 2.4-m-long string, which makes the pendulum swing upward in an arc. Determine the vertical and horizontal components of the pendulum's maximum displacement.

39. (II) An internal explosion breaks an object, initially at rest, into two pieces, one of which has 1.5 times the mass of the other. If 4600 J is released in the explosion, how much kinetic energy does each piece acquire?

40. (II) You drop a 21-g ball from a height of 1.5 m and it only bounces back to a height of 0.85 m. What was the total impulse on the ball when it hit the floor? (Ignore air resistance.)

41. (II) Croquet ball A moving at 4.3 m/s makes a head-on collision with ball B of equal mass initially at rest. Immediately after the collision, ball B moves forward at 3.0 m/s. What fraction of the initial kinetic energy is lost in the collision?

42. (II) A 144-g baseball moving 28.0 m/s strikes a stationary 4.85-kg brick resting on small rollers so it moves without significant friction. After hitting the brick, the baseball bounces straight back, and the brick moves forward at 1.10 m/s. (*a*) What is the baseball's speed after the collision? (*b*) Find the total kinetic energy before and after the collision.

43. (II) A 2.0-kg ball moving 3.5 m/s strikes a 3.0-kg ball at rest head-on. If the total kinetic energy after the collision is 75% of that before the collision, what are the speeds and directions of both balls after the collision?

44. (II) A pendulum consists of a mass M hanging at the bottom end of a massless rod of length ℓ, which has a frictionless pivot at its top end. A mass m, moving horizontally as shown in Fig. 9–44 with velocity v, impacts M and becomes embedded. What is the smallest value of v sufficient to cause the pendulum (with embedded mass m) to swing clear over the top of its arc?

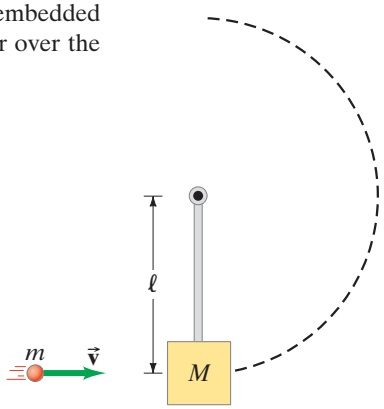


FIGURE 9–44
Problem 44.

45. (II) A bullet of mass $m = 0.0010$ kg embeds itself in a wooden block with mass $M = 0.999$ kg, which then compresses a spring ($k = 140$ N/m) by a distance $x = 0.050$ m before coming to rest. The coefficient of kinetic friction between the block and table is $\mu = 0.50$. (*a*) What is the initial velocity (assumed horizontal) of the bullet? (*b*) What fraction of the bullet's initial kinetic energy is dissipated (in damage to the wooden block, rising temperature, etc.) in the collision between the bullet and the block?

46. (II) A measure of inelasticity in a head-on collision of two objects is the **coefficient of restitution**, defined as

$$e = \frac{v'_A - v'_B}{v_B - v_A},$$

where $v'_A - v'_B$ is the relative velocity of the two objects after the collision and $v_B - v_A$ is their relative velocity before it. (*a*) Show that $e = 1$ for a perfectly elastic collision, and $e = 0$ for a completely inelastic collision. (*b*) A simple method for measuring the coefficient of restitution for an object colliding with a very hard surface like steel is to drop the object onto a heavy steel plate, as shown in Fig. 9–45. Determine a formula for e in terms of the original height h and the maximum height h' reached after collision.

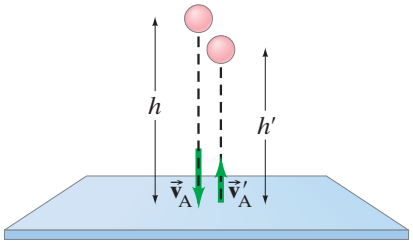


FIGURE 9–45 Problem 46.
Measurement of coefficient of restitution.

9–7 Collisions in 2 or 3 Dimensions

47. (II) Billiard ball A of mass $m_A = 0.120\,\text{kg}$ moving with speed $v_A = 2.80\,\text{m/s}$ strikes ball B, initially at rest, of mass $m_B = 0.140\,\text{kg}$. As a result of the collision, ball A is deflected off at an angle of 35.0° with a speed $v'_A = 2.10\,\text{m/s}$. (*a*) Taking the x axis to be the original direction of motion of ball A, write down the equations expressing the conservation of momentum for the components in the x and y directions separately. (*b*) Solve these equations for the speed, v'_B, and angle, θ'_B, of ball B after the collision. Do not assume the collision is elastic.

48. (II) A radioactive nucleus at rest decays into a second nucleus, an electron, and a neutrino. The electron and neutrino are emitted at right angles and have momenta of $9.6 \times 10^{-23}\,\text{kg}\cdot\text{m/s}$ and $6.2 \times 10^{-23}\,\text{kg}\cdot\text{m/s}$, respectively. Determine the magnitude and the direction of the momentum of the second (recoiling) nucleus.

49. (II) An atomic nucleus of mass m traveling with speed v collides elastically with a target particle of mass $2m$ (initially at rest) and is scattered at 90°. (*a*) At what angle does the target particle move after the collision? (*b*) What are the final speeds of the two particles? (*c*) What fraction of the initial kinetic energy is transferred to the target particle?

50. (II) A neutron collides elastically with a helium nucleus (at rest initially) whose mass is 4.0 times that of the neutron. The helium nucleus is observed to move off at an angle $\theta'_{He} = 45°$. Determine the angle of the neutron, θ'_n, and the speeds of the two particles, v'_n and v'_{He}, after the collision. The neutron's initial speed is $6.2 \times 10^5\,\text{m/s}$.

51. (III) Prove that in a non-head-on elastic collision of two objects of identical mass, with one being a target initially at rest, the angle between their velocity vectors after collision is always 90°.

52. (III) A neon atom ($m = 20.0\,\text{u}$) makes a perfectly elastic collision with another atom at rest. After the impact, the neon atom travels away at a 55.6° angle from its original direction and the unknown atom travels away at a −50.0° angle. What is the mass (in u) of the unknown atom? [*Hint*: You could use the law of sines.]

53. (III) Two billiard balls in contact with each other are struck by a third ball moving at 5.5 m/s as shown in Fig. 9–46a. If both of the first two balls move after the collision at an angle of 32° to the initial path of the third ball (Fig. 9–46b), find the velocities of all three balls after the collision. Assume all balls have the same mass, and the collision is elastic.

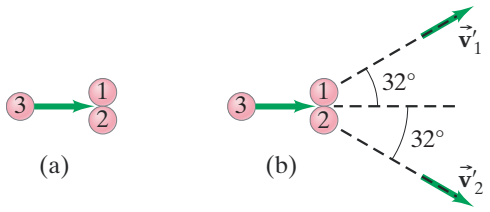


FIGURE 9–46 Problem 53.

9–8 Center of Mass (CM)

54. (I) The distance between a carbon atom ($m = 12\,\text{u}$) and an oxygen atom ($m = 16\,\text{u}$) in the CO molecule is $1.13 \times 10^{-10}\,\text{m}$. How far from the carbon atom is the center of mass of the molecule?

55. (I) Find the center of mass of the three-mass system shown in Fig. 9–47 relative to the 1.00-kg mass.

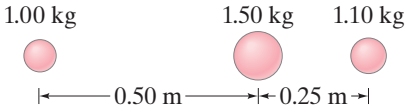


FIGURE 9–47 Problem 55.

56. (I) The CM of an empty 1750-kg Jeep is 2.40 m behind the front of the Jeep. How far from the front of the Jeep will the CM be when two people sit in the front seat 2.80 m from the front of the Jeep, and three people sit in the back seat 3.90 m from the front? Assume that each person has a mass of 65.0 kg.

57. (II) Three cubes, of side ℓ_0, $2\ell_0$, and $3\ell_0$, are placed next to one another (in contact) with their centers along a straight line as shown in Fig. 9–48. What is the position, along this line, of the CM of this system? Assume the cubes are made of the same uniform material.

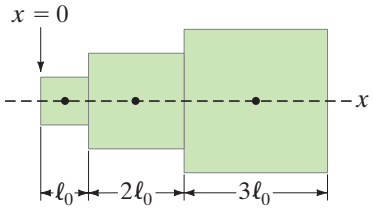


FIGURE 9–48 Problem 57

58. (II) Find the center of mass of the ammonia molecule. The chemical formula is NH_3. The hydrogens are at the corners of an equilateral triangle (with sides 0.16 nm) that forms the base of a pyramid, with nitrogen at the apex (0.037 nm vertically above the plane of the triangle).

59. (II) A (lightweight) pallet has a load of ten identical cases of tomato paste (see Fig. 9–49), each of which is a cube of length ℓ. Find the center of gravity in the horizontal plane, so that the crane operator can pick up the load without tipping it.

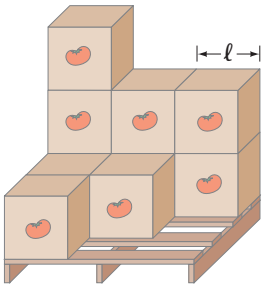


FIGURE 9–49
Problem 59.

60. (II) A uniform circular plate of radius $2R$ has a circular hole of radius R cut out of it. The center C′ of the smaller circle is a distance $0.80R$ from the center C of the larger circle, Fig. 9–50. What is the position of the center of mass of the plate? [*Hint*: Try subtraction.]

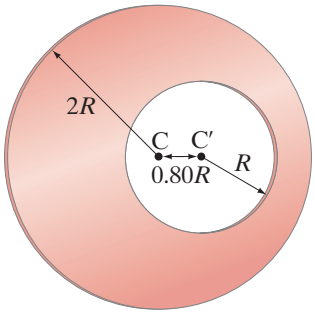


FIGURE 9–50 Problem 60.

61. (II) A square uniform raft, 18 m by 18 m, of mass 5800 kg, is used as a ferryboat. If three cars, each of mass 1350 kg and pointing east, occupy the NE, SE, and SW corners, determine the CM of the loaded ferryboat relative to the center of the raft. Each car is 5.0 m long and 2.0 m wide, with the CM at the center.

62. (II) A uniform thin wire is bent into a semicircle of radius R. Determine the coordinates of its center of mass with respect to an origin of coordinates at the center of the "full" circle.

63. (III) Determine the CM of a machine part that is a uniform cone of height h and radius R, Fig. 9–51. [*Hint*: Divide the cone into an infinite number of circular disks of thickness dz, one of which is shown.]

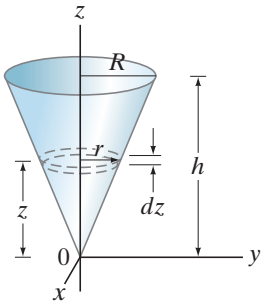


FIGURE 9–51 Problem 63.

64. (III) Determine the CM of a uniform pyramid that has four triangular faces and a square base with equal sides all of length s. [*Hint*: See Problem 63.]

9–9 CM and Translational Motion

65. (II) The masses of the Earth and Moon are 5.98×10^{24} kg and 7.35×10^{22} kg, respectively, and their centers are separated by 3.84×10^{8} m. (*a*) Where is the CM of the Earth–Moon system located? (*b*) What can you say about the motion of the Earth–Moon system about the Sun, and of the Earth and Moon separately about the Sun?

66. (II) A mallet consists of a uniform cylindrical head of mass 2.60 kg and diameter 0.0800 m attached at its center to a uniform cylindrical handle of mass 0.500 kg and length 0.240 m, as shown in Fig. 9–52. If this mallet is tossed, spinning, into the air, how far above the bottom of the handle is the point that will follow a parabolic trajectory?

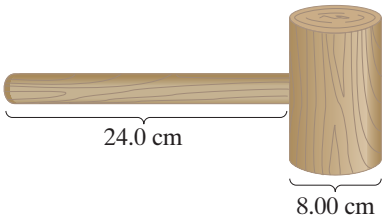


FIGURE 9–52
Problem 66.

67. (II) A 52-kg woman and a 72-kg man stand 8.6 m apart on nearly frictionless ice. (*a*) How far from the woman is their CM? (*b*) If each holds one end of a rope, and the man pulls on the rope so that he moves 2.5 m, how far from the woman will he be now? (*c*) How far will the man have moved from his original position when he collides with the woman?

68. (II) Suppose that in Example 9–18 (Fig. 9–32), $m_{II} = 3m_I$. (*a*) Where then would m_{II} land? (*b*) What if $m_I = 3m_{II}$?

69. (II) Mass $M_A = 35$ kg and mass $M_B = 25$ kg. They have velocities (in m/s) $\vec{\mathbf{v}}_A = 14\hat{\mathbf{i}} - 16\hat{\mathbf{j}}$ and $\vec{\mathbf{v}}_B = -25\hat{\mathbf{i}} + 14\hat{\mathbf{j}}$. Determine the velocity of the center of mass of the system.

70. (III) A huge balloon and its gondola, of mass M, are in the air and stationary with respect to the ground. A passenger, of mass m, then climbs out and slides down a rope with speed v, measured with respect to the balloon. With what speed and direction (relative to Earth) does the balloon then move? What happens if the passenger stops?

71. (III) A 360-kg four-wheel flat cart 15 m long is moving horizontally with a speed of 6.0 m/s. A 110-kg worker starts walking from one end of the cart to the other in the direction of motion, with speed 2.0 m/s with respect to the cart. In the time it takes for him to reach the other end, how far has the cart moved?

*9–10 Variable Mass

***72.** (II) A 3500-kg rocket is to be accelerated at $3.5\,g$ at take-off from the Earth. If the gases can be ejected at a rate of 27 kg/s, what must be their exhaust speed?

***73.** (II) A rocket traveling 1950 m/s away from the Earth at an altitude of 6400 km fires its rockets, which eject gas at a speed of 1200 m/s (relative to the rocket). If the mass of the rocket at this moment is 25,000 kg and an acceleration of $1.5\ \text{m/s}^2$ is desired, at what rate must the gases be ejected?

***74.** (II) The jet engine of an airplane takes in 120 kg of air per second, which is burned with 4.2 kg of fuel per second. The burned gases leave the plane at a speed of 550 m/s (relative to the plane). If the plane is traveling 270 m/s (600 mi/h), determine: (*a*) the thrust due to ejected fuel; (*b*) the thrust due to accelerated air passing through the engine; and (*c*) the power (hp) delivered.

*75. (II) Suppose the conveyor belt of Example 9–20 is retarded by a friction force of 150 N. Determine the required output power (hp) of the motor as a function of time from the moment gravel first starts falling ($t = 0$) until 3.0 s after the gravel begins to be dumped off the end of the 22-m-long conveyor belt.

*76. (III) A sled filled with sand slides without friction down a 32° slope. Sand leaks out a hole in the sled at a rate of 2.0 kg/s. If the sled starts from rest with an initial total mass of 45.0 kg, how long does it take the sled to travel 120 m along the slope?

General Problems

77. Two astronauts, one of mass 195 kg and the other 235 kg, including their space suits, are initially at rest together in outer space. They then push each other apart. How far apart are they when the lighter astronaut has moved 12 m from their starting point?

78. A 4800-kg open truck coasts at a constant speed of 6.70 m/s. Snow begins to fall vertically and fills the truck at a rate of 3.80 kg/min. Ignoring friction, what is the truck's speed after 60.0 min?

79. Two bumper cars in an amusement park ride collide elastically as one approaches the other directly from the rear (Fig. 9–53). Car A has a mass of 435 kg and car B 495 kg, owing to differences in passenger mass. If car A approaches at 4.50 m/s and car B is moving at 3.70 m/s, calculate (*a*) their velocities after the collision, and (*b*) the change in momentum of each.

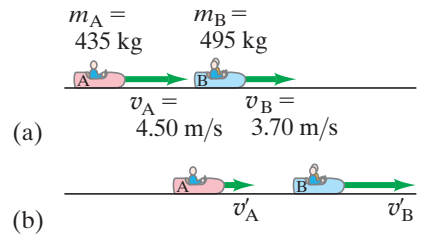


FIGURE 9–53 Problem 79: (a) before collision, (b) after collision.

80. A gun fires a bullet vertically into a 1.40-kg block of wood at rest on a thin horizontal sheet, Fig. 9–54. If the bullet has a mass of 15.0 g and a speed of 230 m/s, how high will the block rise into the air after the bullet becomes embedded in it?

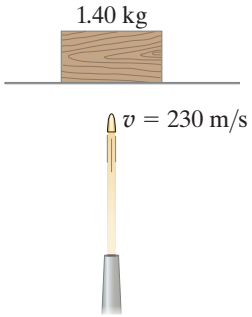


FIGURE 9–54 Problem 80.

81. A 145-g baseball, moving along the x axis with speed 30.0 m/s, strikes a fence at a 45° angle and rebounds along the y axis with unchanged speed. Give its change in momentum using unit vector notation.

82. A meteor whose mass was about 1.5×10^8 kg struck the Earth ($m_E = 6.0 \times 10^{24}$ kg) with a speed of about 25 km/s and came to rest in the Earth. (*a*) What was the Earth's recoil speed (relative to Earth at rest before the collision)? (*b*) What fraction of the meteor's kinetic energy was transformed to kinetic energy of the Earth? (*c*) By how much did the Earth's kinetic energy change as a result of this collision?

83. You are the design engineer in charge of the *crashworthiness* of new automobile models. Cars are tested by smashing them into fixed, massive barriers at 45 km/h. A new model of mass 1500 kg takes 0.15 s from the time of impact until it is brought to rest. (*a*) Calculate the average force exerted on the car by the barrier. (*b*) Calculate the average deceleration of the car in g's.

84. After a completely inelastic collision between two objects of equal mass, each having initial speed, v, the two move off together with speed $v/3$. What was the angle between their initial directions?

85. A 5.5-kg object moving in the $+x$ direction at 6.5 m/s collides head-on with an 8.0-kg object moving in the $-x$ direction at 4.0 m/s. Determine the final velocity of each object if: (*a*) the objects stick together; (*b*) the collision is elastic; (*c*) the 5.5-kg object is at rest after the collision; (*d*) the 8.0-kg object is at rest after the collision; (*e*) the 5.5-kg object has a velocity of 4.0 m/s in the $-x$ direction after the collision. Finally, (*f*) are the results in (*c*), (*d*), and (*e*) "reasonable"? Explain.

86. Two balls, of masses $m_A = 42$ g and $m_B = 96$ g, are suspended as shown in Fig. 9–55. The lighter ball is pulled away to a 66° angle with the vertical and released. (*a*) What is the velocity of the lighter ball before impact? (*b*) What is the velocity of each ball after the elastic collision? (*c*) What will be the maximum height of each ball after the elastic collision?

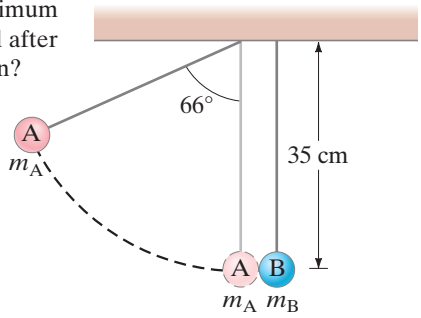


FIGURE 9–55 Problem 86.

87. A block of mass $m = 2.20$ kg slides down a 30.0° incline which is 3.60 m high. At the bottom, it strikes a block of mass $M = 7.70$ kg which is at rest on a horizontal surface, Fig. 9–56. (Assume a smooth transition at the bottom of the incline.) If the collision is elastic, and friction can be ignored, determine (*a*) the speeds of the two blocks after the collision, and (*b*) how far back up along the incline will the smaller mass go.

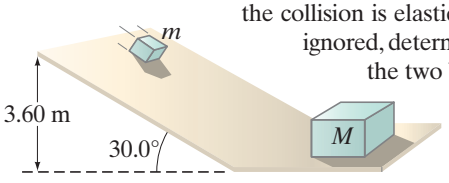


FIGURE 9–56 Problems 87 and 88.

88. In Problem 87 (Fig. 9–56), what is the upper limit on mass m if it is to rebound from M, slide up the incline, stop, slide down the incline, and collide with M again?

89. An *extrasolar planet* can be detected by observing the *wobble* it produces on the star around which it orbits. Suppose an extrasolar planet of mass m_B orbits around its star of mass m_A. If no external force acts on this simple two-object system, then its CM is stationary. Assume m_A and m_B are in circular orbits with radii r_A and r_B about the system's CM. (*a*) Show that

$$r_A = \frac{m_B}{m_A} r_B.$$

(*b*) Now consider a Sun-like star and a single planet with the same characteristics as Jupiter. That is, $m_B = 1.0 \times 10^{-3} m_A$ and the planet has an orbital radius of 8.0×10^{11} m. Determine the radius r_A of the star's orbit about the system's CM. (*c*) When viewed from Earth, the distant system appears to wobble over a distance of $2r_A$. If astronomers are able to detect angular displacements θ of about 1 *milliarcsec* (1 arcsec $= \frac{1}{3600}$ of a degree), from what distance d (in light-years) can the star's wobble be detected (1 ly $= 9.46 \times 10^{15}$ m)? (*d*) The star nearest to our Sun is about 4 ly away. Assuming stars are uniformly distributed throughout our region of the Milky Way Galaxy, about how many stars can this technique be applied to in the search for extrasolar planetary systems?

90. In a physics lab, a cube slides down a frictionless incline as shown in Fig. 9–57 and elastically strikes another cube at the bottom that is only one-half its mass. If the incline is 45 cm high and the table is 95 cm off the floor, where does each cube land? [*Hint*: Both leave the incline moving horizontally.]

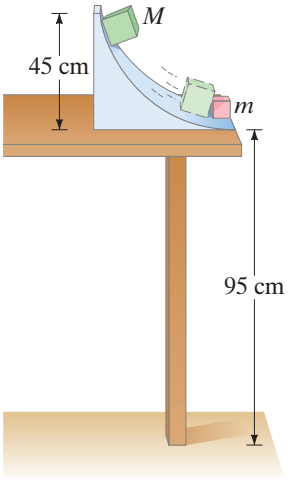
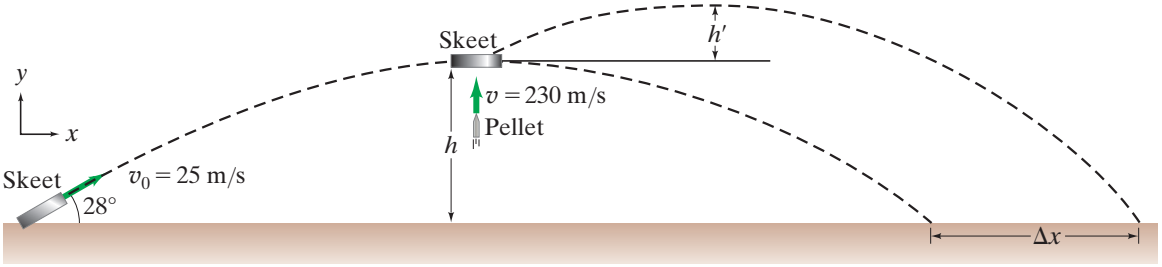


FIGURE 9–57 Problem 90.

91. A 0.20-kg *skeet* (clay target) is fired at an angle of 28° to the horizontal with a speed of 25 m/s (Fig. 9–58). When it reaches the maximum height, h, it is hit from below by a 15-g pellet traveling vertically upward at a speed of 230 m/s. The pellet is embedded in the skeet. (*a*) How much higher, h', does the skeet go up? (*b*) How much extra distance, Δx, does the skeet travel because of the collision?

FIGURE 9–58 Problem 91.



92. A rifle is aimed at a 2.0-kg block of wood along an inclined plane making an angle of 25°, as shown in Fig. 9–59. A 9.5-g bullet is fired at 760 m/s and becomes embedded in the block. How far up the incline does the block/bullet slide? (*a*) Ignore the friction. (*b*) Assume $\mu_k = 0.33$.

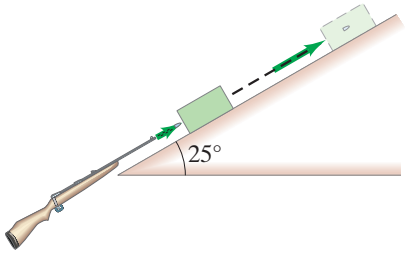


FIGURE 9–59 Problem 92.

93. Astronomers estimate that a 2.0-km-diameter *asteroid* collides with the Earth once every million years. The collision could pose a threat to life on Earth. (*a*) Assume a spherical asteroid has a mass of 3200 kg for each cubic meter of volume and moves toward the Earth at 15 km/s. How much destructive energy could be released when it embeds itself in the Earth? (*b*) For comparison, a nuclear bomb could release about 5.0×10^{16} J. How many such bombs would have to explode simultaneously to release the destructive energy of the asteroid collision with the Earth?

94. An astronaut of mass 210 kg including his suit and jet pack wants to acquire a velocity of 2.0 m/s to move back toward his space shuttle. Assuming the jet pack can eject gas with a velocity of 35 m/s, what mass of gas will need to be ejected?

95. Two blocks of mass m_A and m_B, resting on a frictionless table, are connected by a stretched spring and then released (Fig. 9–60). (*a*) Is there a net external force on the system before release? (*b*) Determine the ratio of their speeds, v_A/v_B. (*c*) What is the ratio of their kinetic energies? (*d*) Describe the motion of the CM of this system. Ignore mass of spring.

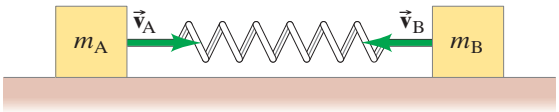


FIGURE 9–60 Problem 95.

96. A fake hockey puck of mass $4m$ has been rigged to explode. Initially the puck is at rest on a frictionless ice rink. Then it bursts into three pieces. One chunk, of mass m, slides across the ice at velocity $v\hat{\mathbf{i}}$. Another chunk, of mass $2m$, slides across the ice at velocity $2v\hat{\mathbf{j}}$. Determine the velocity of the third chunk.

97. Determine the CM of a thin, uniform, semicircular plate, Fig. 9–61. [*Hint*: Integrate; see Problem 63.]

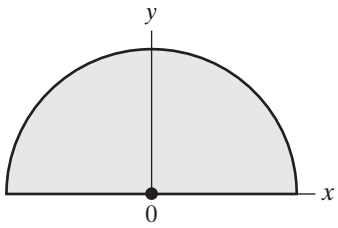


FIGURE 9–61 Problem 97.

98. A massless spring with spring constant k is placed between a block of mass m and a block of mass $3m$. Initially the blocks are at rest on a frictionless surface and they are held together so that the spring between them is compressed by an amount D from its equilibrium length. The blocks are then released and the spring pushes them off in opposite directions. Find the speeds of the two blocks when they detach from the spring.

***99.** Consider the truck of Problem 78, which is slowly filling with snow. (*a*) Determine the speed of the truck as a function of time using Eqs. 9–19. (*b*) What is the speed of the truck after 60.0 min? Does this agree with the simpler calculation (Problem 78)?

100. **The gravitational slingshot effect**. Figure 9–62 shows the planet Saturn moving in the negative x direction at its orbital speed (with respect to the Sun) of 9.6 km/s. The mass of Saturn is 5.69×10^{26} kg. A spacecraft with mass 825 kg approaches Saturn. When far from Saturn, it moves in the $+x$ direction at 10.4 km/s. The gravitational attraction of Saturn (a conservative force) acting on the spacecraft causes it to swing around the planet (orbit shown as dashed line) and head off in the opposite direction. Using momentum conservation in one dimension, estimate the final speed of the spacecraft after it is far enough away to be considered free of Saturn's gravitational pull. Assume the spacecraft does not affect the orbit of Saturn whose mass is so much larger.

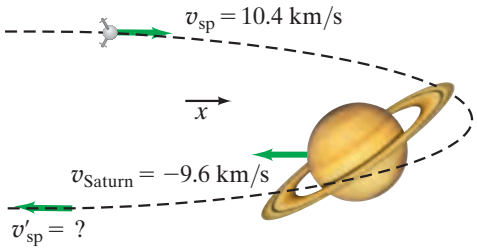


FIGURE 9–62 Problem 100.

101. During a Chicago storm, winds can whip at speeds of 110 km/h. (*a*) Show that the air strikes a person horizontally at the rate of 40 kg/s per square meter and assume it is brought to rest. (Air has a mass of 1.29 kg per m^3 of volume.) (*b*) Calculate the *force of the wind* on a person. Assume the person is 1.60 m high and 0.50 m wide. Compare to the typical maximum force of friction ($\mu \approx 1.0$) between the person and the ground, if the person has a mass of 70 kg.

102. A novice pool player is faced with the corner pocket shot shown in Fig. 9–63. Relative dimensions are also shown. Should the player worry that this might be a "*scratch shot*," in which the cue ball will also fall into a pocket? Give details. Assume equal-mass balls and an elastic collision. Ignore spin.

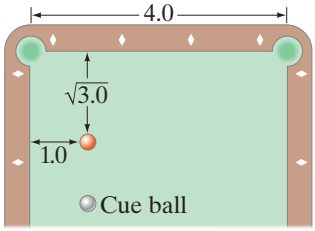


FIGURE 9–63 Problem 102.

103. In order to convert a tough split in *bowling*, it is necessary to strike the pin a glancing blow as shown in Fig. 9–64. Assume that the bowling ball, traveling at 14.0 m/s just before it strikes the pin, has five times the mass of a pin and that the pin goes off at 75° from the original direction of the ball. Calculate the speed (*a*) of the pin and (*b*) of the ball just after collision, and (*c*) calculate the angle θ through which the ball was deflected. Assume the collision is elastic and ignore any spin of the ball.

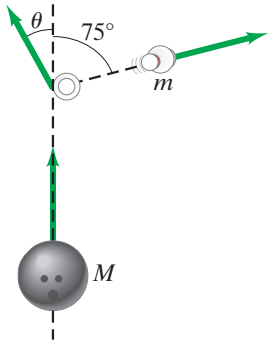


FIGURE 9–64 Problem 103.

104. A small ball of mass m and a large ball of mass M, with $M \gg m$, are dropped from a height h above the floor. Ball m is dropped a very short time after ball M, and directly above it. Ball M collides elastically with the floor, rebounds, and then collides elastically with ball m (which is still falling) just after it rebounds off the floor. Show that the smaller ball rebounds to a height $9h$.

ANSWERS TO EXERCISES

A: Yes, if the sports car's speed is three times greater.
B: Larger (Δp is greater).
C: (*a*) 2.0 m/s; (*b*) almost zero; (*c*) almost 8.0 m/s.
D: 0.50 m/s.
E: (*b*); (*d*).
F: $x_{CM} = -2.0$ m; yes.
G: (*a*).
H: The boat moves in the opposite direction.

You too can experience rapid rotation—if your stomach can take the high angular velocity and centripetal acceleration of some of the faster amusement park rides. If not, try the slower merry-go-round or Ferris wheel. Rotating carnival rides have rotational kinetic energy as well as angular momentum. Angular acceleration is produced by a net torque, and rotating objects have rotational kinetic energy.



CHAPTER 10

Rotational Motion

CONTENTS

10 – 1 Angular Quantities
10 – 2 Vector Nature of Angular Quantities
10 – 3 Constant Angular Acceleration
10 – 4 Torque
10 – 5 Rotational Dynamics; Torque and Rotational Inertia
10 – 6 Solving Problems in Rotational Dynamics
10 – 7 Determining Moments of Inertia
10 – 8 Rotational Kinetic Energy
10 – 9 Rotational Plus Translational Motion; Rolling
*10–10 Why Does a Rolling Sphere Slow Down?

CHAPTER-OPENING QUESTION—Guess now!

A solid ball and a solid cylinder roll down a ramp. They both start from rest at the same time. Which gets to the bottom first?

(a) They get there at the same time.
(b) They get there at almost exactly the same time except for frictional differences.
(c) The ball gets there first.
(d) The cylinder gets there first.
(e) Can't tell without knowing the mass of each.
(f) Can't tell without knowing the radius of each.

Until now, we have been concerned mainly with translational motion. We discussed the kinematics and dynamics of translational motion (the role of force). We also discussed energy and momentum for translational motion. In this Chapter and the next we will deal with rotational motion. We first discuss the kinematics of rotational motion and then rotational dynamics (involving torque), as well as rotational kinetic energy and angular momentum (the rotational analog of linear momentum). Our understanding of the world around us will be increased significantly—from rotating bicycle wheels and CDs (or vinyl) to amusement park rides, a spinning skater, the rotating Earth, and a centrifuge—and there may be a few surprises.

We will consider mainly the rotation of rigid objects about a fixed axis. A **rigid object** is an object with a definite shape that doesn't change, so that the particles composing it stay in fixed positions relative to one another. Any real object is capable of vibrating or deforming when a force is exerted on it. But these effects are often very small, so the concept of an ideal rigid object is very useful as a good approximation.

10–1 Angular Quantities

The motion of a rigid object can be analyzed as the translational motion of its center of mass plus rotational motion *about* its center of mass (Sections 9–8 and 9–9). We have already discussed translational motion in detail, so now we focus our attention on purely rotational motion. By *purely rotational motion* of an object about a fixed axis, we mean that all points in the object move in circles, such as the point P on the rotating wheel of Fig. 10–1, and that the centers of these circles all lie on one line called the **axis of rotation**. In Fig. 10–1 the axis of rotation is perpendicular to the page and passes through point O. We assume the axis is fixed in an inertial reference frame, but we will not always insist that the axis pass through the center of mass.

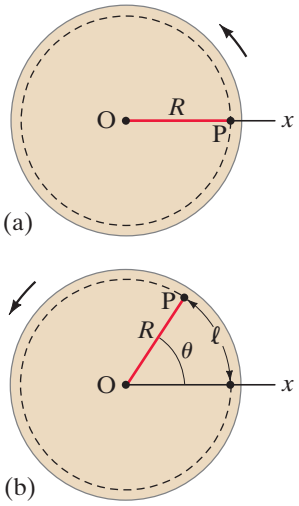


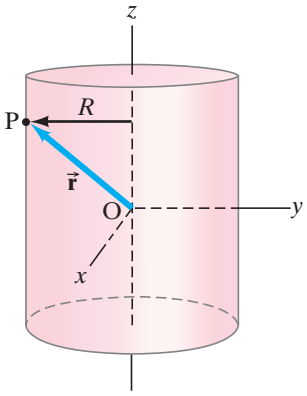
FIGURE 10–1 Looking at a wheel that is rotating counterclockwise about an axis through the wheel's center at O (axis perpendicular to the page). Each point, such as point P, moves in a circular path; ℓ is the distance P travels as the wheel rotates through the angle θ.

For a three-dimensional rigid object rotating about a fixed axis, we will use the symbol R to represent the perpendicular distance of a point or particle from the axis of rotation. We do this to distinguish R from r, which will continue to represent the position of a particle with reference to the origin (point) of some coordinate system. This distinction is illustrated in Fig. 10–2. This distinction may seem like a small point, but not being fully aware of it can cause huge errors when working with rotational motion.

CAUTION
Distinguish R from r

Every point in an object rotating about a fixed axis moves in a circle (shown dashed in Fig. 10–1 for point P) whose center is on the axis of rotation and whose radius is R, the distance of that point from the axis of rotation. A straight line drawn from the axis to any point in the object sweeps out the same angle θ in the same time interval.

FIGURE 10–2 Showing the distinction between $\vec{\mathbf{r}}$ (the position vector) and R (the distance from the rotation axis) for a point P on the edge of a cylinder rotating about the z axis.



To indicate the **angular position** of a rotating object, or how far it has rotated, we specify the angle θ of some particular line in the object (red in Fig. 10–1) with respect to some reference line, such as the x axis in Fig. 10–1. A point in the object, such as P in Fig. 10–1b, moves through an angle θ when it travels the distance ℓ measured along the circumference of its circular path. Angles are commonly specified in degrees, but the mathematics of circular motion is much simpler if we use the *radian* for angular measure. One **radian** (abbreviated rad) is defined as the angle subtended by an arc whose length is equal to the radius. For example, in Fig. 10–1, point P is a distance R from the axis of rotation, and it has moved a distance ℓ along the arc of a circle. The arc length ℓ is said to "subtend" the angle θ. In general, any angle θ, in radians, is given by

$$\theta = \frac{\ell}{R}, \qquad [\theta \text{ in radians}] \quad \textbf{(10–1a)}$$

If $\ell = R$, then $\theta = 1$ rad. Equation 10–1a does not hold if θ is in degrees.

CAUTION
Use radians in calculating, not degrees

The radian is dimensionless since it is the ratio of two lengths. Nonetheless when giving an angle in radians, we always write "rad" to remind us it is not degrees. It is often useful to rewrite Eq. 10–1a in terms of arc length ℓ:

$$\ell = R\theta. \qquad \textbf{(10–1b)}$$

Radians can be related to degrees in the following way. In a complete circle there are 360°, which must correspond to an arc length equal to the circumference of the circle, $\ell = 2\pi R$. For a full circle, $\theta = \ell/R = 2\pi R/R = 2\pi$ rad. Thus

$$360^\circ = 2\pi \text{ rad}.$$

One radian is then $360^\circ/2\pi \approx 360^\circ/6.28 \approx 57.3^\circ$. An object that makes one complete revolution (rev) has rotated through 360°, or 2π radians:

$$1 \text{ rev} = 360^\circ = 2\pi \text{ rad}.$$

1 rad ≈ 57.3°

EXAMPLE 10–1 **Bike wheel.** A bike wheel rotates 4.50 revolutions. How many radians has it rotated?

APPROACH All we need is a conversion of units using

$$1\text{ revolution} = 360^\circ = 2\pi\text{ rad} = 6.28\text{ rad}.$$

SOLUTION

$$4.50\text{ revolutions} = (4.50\text{ rev})\left(2\pi\frac{\text{rad}}{\text{rev}}\right) = 9.00\pi\text{ rad} = 28.3\text{ rad}.$$

EXAMPLE 10–2 **Birds of prey—in radians.** A particular bird's eye can just distinguish objects that subtend an angle no smaller than about 3×10^{-4} rad. (*a*) How many degrees is this? (*b*) How small an object can the bird just distinguish when flying at a height of 100 m (Fig. 10–3a)?

APPROACH For (*a*) we use the relation $360^\circ = 2\pi$ rad. For (*b*) we use Eq. 10–1b, $\ell = R\theta$, to find the arc length.

SOLUTION (*a*) We convert 3×10^{-4} rad to degrees:

$$(3\times10^{-4}\text{ rad})\left(\frac{360^\circ}{2\pi\text{ rad}}\right) = 0.017^\circ,$$

or about 0.02°.

(*b*) We use Eq. 10–1b, $\ell = R\theta$. For small angles, the arc length ℓ and the chord length are approximately the same (Fig. 10–3b). Since $R = 100$ m and $\theta = 3\times10^{-4}$ rad, we find

$$\ell = (100\text{ m})(3\times10^{-4}\text{ rad}) = 3\times10^{-2}\text{ m} = 3\text{ cm}.$$

A bird can distinguish a small mouse (about 3 cm long) from a height of 100 m. That is good eyesight.

NOTE Had the angle been given in degrees, we would first have had to convert it to radians to make this calculation. Equation 10–1 is valid *only* if the angle is specified in radians. Degrees (or revolutions) won't work.

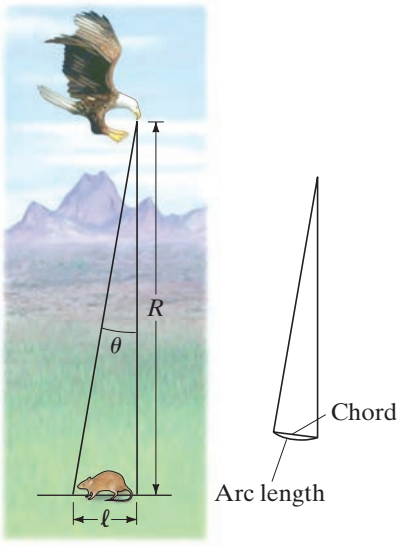


FIGURE 10–3 (a) Example 10–2. (b) For small angles, arc length and the chord length (straight line) are nearly equal. For an angle as large as 15°, the error in making this estimate is only 1%. For larger angles the error increases rapidly. (The **chord** is the straight-line distance between the ends of the arc.)

To describe rotational motion, we use angular quantities, such as angular velocity and angular acceleration. These quantities are defined in analogy to the corresponding quantities in linear motion; they describe the rotating object as a whole, so they each have the same value for every point in a rotating rigid object. Each point in a rotating object also has translational velocity and acceleration, but they have different values for different points in the object.

When an object, such as the bicycle wheel in Fig. 10–4, rotates from some initial position, specified by θ_1, to some final position, θ_2, its **angular displacement** is

$$\Delta\theta = \theta_2 - \theta_1.$$

The *angular velocity* (denoted by ω, the Greek lowercase letter omega) is defined in analogy with linear (translational) velocity that was discussed in Chapter 2. Instead of linear displacement, we use the angular displacement. Thus the **average angular velocity** of an object rotating about a fixed axis is defined as the time rate of change of angular position:

$$\overline{\omega} = \frac{\Delta\theta}{\Delta t}, \qquad \textbf{(10–2a)}$$

where $\Delta\theta$ is the angle through which the object has rotated in the time interval Δt. The **instantaneous angular velocity** is the limit of this ratio as Δt approaches zero:

$$\omega = \lim_{\Delta t\to0}\frac{\Delta\theta}{\Delta t} = \frac{d\theta}{dt}. \qquad \textbf{(10–2b)}$$

Angular velocity has units of radians per second (rad/s). Note that *all points in a rigid object rotate with the same angular velocity*, since every position in the object moves through the same angle in the same time interval.

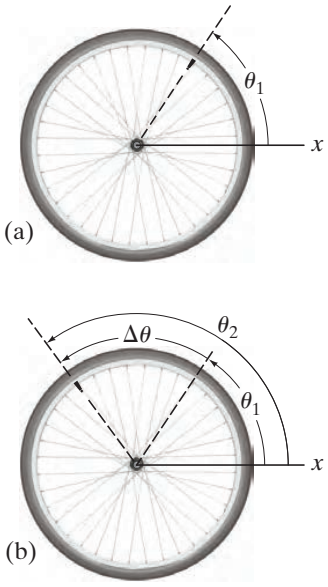


FIGURE 10–4 A wheel rotates from (a) initial position θ_1 to (b) final position θ_2. The angular displacement is $\Delta\theta = \theta_2 - \theta_1$.

An object such as the wheel in Fig. 10–4 can rotate about a fixed axis either clockwise or counterclockwise. The direction can be specified with a + or − sign. The usual convention is to choose the angular displacement $\Delta\theta$ and angular velocity ω as positive when the wheel rotates counterclockwise. If the rotation is clockwise, then θ would decrease, so $\Delta\theta$ and ω would be negative.

Angular acceleration (denoted by α, the Greek lowercase letter alpha), in analogy to linear acceleration, is defined as the change in angular velocity divided by the time required to make this change. The **average angular acceleration** is defined as

$$\bar{\alpha} = \frac{\omega_2 - \omega_1}{\Delta t} = \frac{\Delta\omega}{\Delta t}, \qquad \textbf{(10–3a)}$$

where ω_1 is the angular velocity initially, and ω_2 is the angular velocity after a time interval Δt. **Instantaneous angular acceleration** is defined as the limit of this ratio as Δt approaches zero:

$$\alpha = \lim_{\Delta t \to 0} \frac{\Delta\omega}{\Delta t} = \frac{d\omega}{dt}. \qquad \textbf{(10–3b)}$$

Since ω is the same for all points of a rotating object, Eq. 10–3 tells us that α also will be the same for all points. Thus, ω and α are properties of the rotating object as a whole. With ω measured in radians per second and t in seconds, α has units of radians per second squared (rad/s^2).

Each point or particle of a rotating object has, at any instant, a linear velocity v and a linear acceleration a. We can relate the linear quantities at each point, v and a, to the angular quantities, ω and α, for a rigid object rotating around a fixed axis. Consider a point P located a distance R from the axis of rotation, as in Fig. 10–5. If the object rotates with angular velocity ω, any point will have a linear velocity whose direction is tangent to its circular path. The magnitude of that point's linear velocity is $v = d\ell/dt$. From Eq. 10–1b, a change in rotation angle $d\theta$ (in radians) is related to the linear distance traveled by $d\ell = R\,d\theta$. Hence

$$v = \frac{d\ell}{dt} = R\frac{d\theta}{dt}$$

or, since $d\theta/dt = \omega$,

$$v = R\omega. \qquad \textbf{(10–4)}$$

In this very useful Eq. 10–4, R is the distance of a point from the rotation axis and ω is given in rad/s. We see that although ω is the same for every point in the rotating object at any instant, the linear velocity v is greater for points farther from the axis (Fig. 10–6). Note that Eq. 10–4 is valid both instantaneously and on the average.

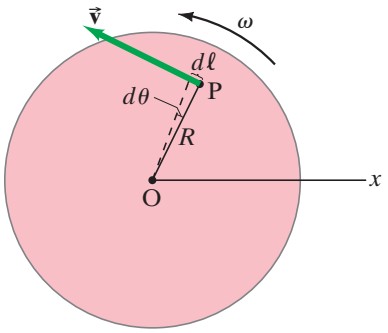


FIGURE 10–5 A point P on a rotating wheel has a linear velocity $\vec{\mathbf{v}}$ at any moment.

CONCEPTUAL EXAMPLE 10–3 **Is the lion faster than the horse?** On a rotating carousel or merry-go-round, one child sits on a horse near the outer edge and another child sits on a lion halfway out from the center. (*a*) Which child has the greater linear velocity? (*b*) Which child has the greater angular velocity?

RESPONSE (*a*) The *linear* velocity is the distance traveled divided by the time interval. In one rotation the child on the outer edge travels a longer distance than the child near the center, but the time interval is the same for both. Thus the child at the outer edge, on the horse, has the greater linear velocity.

(*b*) The *angular* velocity is the angle of rotation of the carousel as a whole divided by the time interval. For example, in one rotation both children rotate through the same angle. The two children have the same angular velocity.

If the angular velocity of a rotating object changes, the object as a whole—and each point in it—has an angular acceleration. Each point then also has a component of linear acceleration tangent to that point's circular path.

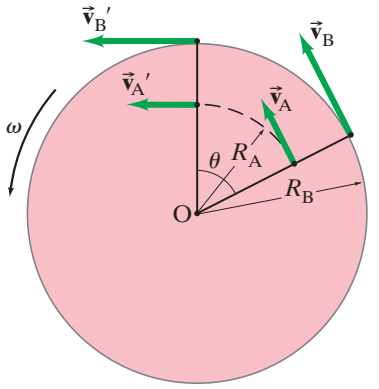


FIGURE 10–6 A wheel rotating uniformly counterclockwise. Two points on the wheel, at distances R_A and R_B from the center, have the same angular velocity ω because they travel through the same angle θ in the same time interval. But the two points have different linear velocities because they travel different distances in the same time interval. Since $R_B > R_A$, then $v_B > v_A$ (because $v = R\omega$).

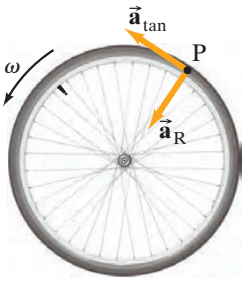


FIGURE 10–7 On a rotating wheel whose angular speed is increasing, a point P has both tangential and radial (centripetal) components of linear acceleration. (See also Chapter 5.)

We use Eq. 10–4 ($v = R\omega$) to see that the angular acceleration α is related to the tangential linear acceleration $a_{\tan}$ of a point in the rotating object by

$$a_{\tan} = \frac{dv}{dt} = R\frac{d\omega}{dt}$$

or

$$a_{\tan} = R\alpha. \qquad \textbf{(10–5)}$$

In this equation, R is the radius of the circle in which the particle is moving (its distance from the rotation axis), and the subscript "tan" in $a_{\tan}$ stands for "tangential."

The total linear acceleration of a point in the rotating object at any instant is the vector sum of two components:

$$\vec{\mathbf{a}} = \vec{\mathbf{a}}_{\tan} + \vec{\mathbf{a}}_{\mathrm{R}},$$

where the radial component, $\vec{\mathbf{a}}_{\mathrm{R}}$, is the radial or "centripetal" acceleration and its direction is toward the center of the point's circular path; see Fig. 10–7. We saw in Chapter 5 (Eq. 5–1) that a particle moving in a circle of radius R with linear speed v has a radial acceleration $a_{\mathrm{R}} = v^2/R$; we can rewrite this in terms of ω using Eq. 10–4:

$$a_{\mathrm{R}} = \frac{v^2}{R} = \frac{(R\omega)^2}{R} = \omega^2 R. \qquad \textbf{(10–6)}$$

Thus the centripetal acceleration is greater the farther you are from the axis of rotation: the children farthest out on a carousel experience the greatest radial acceleration.

Table 10–1 summarizes the relationships between the angular quantities describing the rotation of an object to the linear quantities for each point of the object.

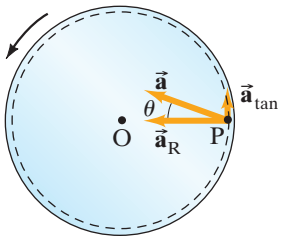
TABLE 10–1
Linear and Rotational Quantities

Linear	Type	Rotational	Relation (θ in radians)
x	displacement	θ	$x = R\theta$
v	velocity	ω	$v = R\omega$
$a_{\tan}$	acceleration	α	$a_{\tan} = R\alpha$

FIGURE 10–8 Example 10–4. The total acceleration vector $\vec{\mathbf{a}} = \vec{\mathbf{a}}_{\tan} + \vec{\mathbf{a}}_{\mathrm{R}}$, at $t = 8.0$ s.



(a)



(b)

EXAMPLE 10–4 **Angular and linear velocities and accelerations.** A carousel is initially at rest. At $t = 0$ it is given a constant angular acceleration $\alpha = 0.060\ \mathrm{rad/s^2}$, which increases its angular velocity for 8.0 s. At $t = 8.0$ s, determine the magnitude of the following quantities: (*a*) the angular velocity of the carousel; (*b*) the linear velocity of a child (Fig. 10–8a) located 5.5 m from the center, point P in Fig. 10–8b; (*c*) the tangential (linear) acceleration of that child; (*d*) the centripetal acceleration of the child; and (*e*) the total linear acceleration of the child.

APPROACH The angular acceleration α is constant, so we can use Eq. 10–3a, $\alpha = \bar{\alpha} = \Delta\omega/\Delta t$ to solve for ω after a time $t = 8.0$ s. With this ω and the given α, we determine the other quantities using the relations we just developed, Eqs. 10–4, 10–5, and 10–6.

SOLUTION (*a*) In Equation 10–3a, $\bar{\alpha} = (\omega_2 - \omega_1)/\Delta t$, we put $\Delta t = 8.0$ s, $\bar{\alpha} = 0.060\ \mathrm{rad/s^2}$, and $\omega_1 = 0$. Solving for ω_2, we get

$$\omega_2 = \omega_1 + \bar{\alpha}\,\Delta t = 0 + (0.060\ \mathrm{rad/s^2})(8.0\ \mathrm{s}) = 0.48\ \mathrm{rad/s}.$$

During the 8.0-s interval, the carousel has accelerated from $\omega_1 = 0$ (rest) to $\omega_2 = 0.48$ rad/s.

(*b*) The linear velocity of the child, with $R = 5.5$ m at time $t = 8.0$ s, is found using Eq. 10–4:

$$v = R\omega = (5.5\ \mathrm{m})(0.48\ \mathrm{rad/s}) = 2.6\ \mathrm{m/s}.$$

Note that the "rad" has been dropped here because it is dimensionless (and only a reminder)—it is a ratio of two distances, Eq. 10–1a.

(*c*) The child's tangential acceleration is given by Eq. 10–5:

$$a_{\tan} = R\alpha = (5.5\ \mathrm{m})(0.060\ \mathrm{rad/s^2}) = 0.33\ \mathrm{m/s^2},$$

and it is the same throughout the 8.0-s acceleration interval.

(*d*) The child's centripetal acceleration at $t = 8.0$ s is given by Eq. 10–6:

$$a_{\mathrm{R}} = \omega^2 \mathrm{R} = (0.48\ \mathrm{s^{-1}})^2(5.5\ \mathrm{m}) = 1.27\ \mathrm{m/s^2}.$$

(*e*) The two components of linear acceleration calculated in parts (*c*) and (*d*) are perpendicular to each other. Thus the total linear acceleration at $t = 8.0\,\text{s}$ (just before the acceleration stops) has magnitude

$$a = \sqrt{a_{\text{tan}}^2 + a_{\text{R}}^2} = \sqrt{(0.33\,\text{m/s}^2)^2 + (1.27\,\text{m/s}^2)^2} = 1.31\,\text{m/s}^2.$$

Its direction (Fig. 10–8b) is

$$\theta = \tan^{-1}\left(\frac{a_{\text{tan}}}{a_{\text{R}}}\right) = \tan^{-1}\left(\frac{0.33\,\text{m/s}^2}{1.27\,\text{m/s}^2}\right) = 0.25\,\text{rad},$$

so $\theta \approx 15°$.

NOTE The linear acceleration at this chosen instant is mostly centripetal, and keeps the child moving in a circle with the carousel. The tangential component that speeds up the circular motion is smaller. (After $t = 8.0\,\text{s}$ it is zero.)

We can relate the angular velocity ω to the frequency of rotation, f. The **frequency** is the number of complete revolutions (rev) per second, as we saw in Chapter 5. One revolution (of a wheel, say) corresponds to an angle of 2π radians, and thus $1\,\text{rev/s} = 2\pi\,\text{rad/s}$. Hence, in general, the frequency f is related to the angular velocity ω by

$$f = \frac{\omega}{2\pi}$$

or

$$\omega = 2\pi f. \qquad \textbf{(10–7)}$$

The unit for frequency, revolutions per second (rev/s), is given the special name the **hertz** (Hz). That is

$$1\,\text{Hz} = 1\,\text{rev/s}.$$

Note that "revolution" has no dimensions, so we can also write $1\,\text{Hz} = 1\,\text{s}^{-1}$.

The time required for one complete revolution is the **period** T, and it is related to the frequency by

$$T = \frac{1}{f}. \qquad \textbf{(10–8)}$$

If a particle rotates at a frequency of three revolutions per second, then the period of each revolution is $\frac{1}{3}\,\text{s}$.

EXERCISE A In Example 10–4, we found that the carousel, after 8.0 s, rotates at an angular velocity $\omega = 0.48\,\text{rad/s}$, and continues to do so after $t = 8.0\,\text{s}$ because the acceleration ceased. Determine the frequency and period of the carousel when rotating at this constant angular velocity $\omega = 0.48\,\text{rad/s}$.

EXAMPLE 10–5 **Given ω as function of time.** A disk of radius $R = 3.0\,\text{m}$ rotates at an angular velocity $\omega = (1.6 + 1.2\,t)\,\text{rad/s}$, where t is in seconds. At the instant $t = 2.0\,\text{s}$, determine (*a*) the angular acceleration, and (*b*) the speed v and the components of the acceleration a of a point on the edge of the disk.

APPROACH We use $\alpha = d\omega/dt$, $v = R\omega$, $a_{\text{tan}} = R\alpha$, and $a_{\text{R}} = \omega^2 R$, which are Eqs. 10–3b, 10–4, 10–5, and 10–6. We can write ω explicitly showing units of the constants (in case we want to check later): $\omega = [1.6\,\text{s}^{-1} + (1.2\,\text{s}^{-2})t]$ which will give us s^{-1} (= rad/s) for each term.

SOLUTION (*a*) The angular acceleration is

$$\alpha = \frac{d\omega}{dt} = \frac{d}{dt}[1.6\,\text{s}^{-1} + (1.2\,\text{s}^{-2})t] = 1.2\,\text{rad/s}^2.$$

(*b*) The speed v of a point 3.0 m from the center of the rotating disk at $t = 2.0\,\text{s}$ is, using Eq. 10–4,

$$v = R\omega = (3.0\,\text{m})(1.6 + 1.2\,t)\,\text{s}^{-1} = (3.0\,\text{m})(4.0\,\text{s}^{-1}) = 12.0\,\text{m/s}.$$

The components of the linear acceleration of this point at $t = 2.0\,\text{s}$ are

$$a_{\text{tan}} = R\alpha = (3.0\,\text{m})(1.2\,\text{rad/s}^2) = 3.6\,\text{m/s}^2$$

$$a_{\text{R}} = \omega^2 R = [(1.6 + 1.2\,t)\,\text{s}^{-1}]^2(3.0\,\text{m}) = (4.0\,\text{s}^{-1})^2(3.0\,\text{m}) = 48\,\text{m/s}^2.$$

10–2 Vector Nature of Angular Quantities

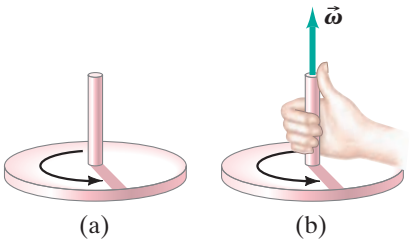


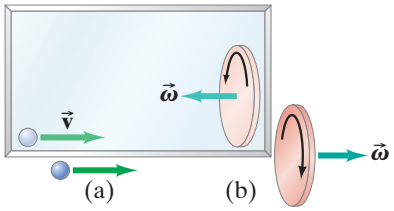
FIGURE 10–9 (a) Rotating wheel. (b) Right-hand rule for obtaining direction of $\vec{\omega}$.

Both $\vec{\omega}$ and $\vec{\alpha}$ can be treated as vectors, and we define their directions in the following way. Consider the rotating wheel shown in Fig. 10–9a. The linear velocities of different particles of the wheel point in all different directions. The only unique direction in space associated with the rotation is along the axis of rotation, perpendicular to the actual motion. We therefore choose the axis of rotation to be the direction of the angular velocity vector, $\vec{\omega}$. Actually, there is still an ambiguity since $\vec{\omega}$ could point in either direction along the axis of rotation (up or down in Fig. 10–9a). The convention we use, called the **right-hand rule**, is this: when the fingers of the right hand are curled around the rotation axis and point in the direction of the rotation, then the thumb points in the direction of $\vec{\omega}$. This is shown in Fig. 10–9b. Note that $\vec{\omega}$ points in the direction a normal (right-handed) screw would move when turned in the direction of rotation. Thus, if the rotation of the wheel in Fig. 10–9a is counterclockwise, the direction of $\vec{\omega}$ is upward as shown in Fig. 10–9b. If the wheel rotates clockwise, then $\vec{\omega}$ points in the opposite direction, downward. Note that no part of the rotating object moves in the direction of $\vec{\omega}$.

If the axis of rotation is fixed in direction, then $\vec{\omega}$ can change only in magnitude. Then $\vec{\alpha} = d\vec{\omega}/dt$ must also point along the axis of rotation. If the rotation is counterclockwise as in Fig. 10–9a and the magnitude of ω is increasing, then $\vec{\alpha}$ points upward; but if ω is decreasing (the wheel is slowing down), $\vec{\alpha}$ points downward. If the rotation is clockwise, $\vec{\alpha}$ points downward if ω is increasing, and $\vec{\alpha}$ points upward if ω is decreasing.

*Pseudovectors (Axial Vectors)

FIGURE 10–10 (a) Velocity is a true vector. The reflection of $\vec{\mathbf{v}}$ points in the same direction. (b) Angular velocity is a pseudovector since it does not follow this rule. As can be seen, the reflection of the wheel rotates in the opposite direction, so the direction of $\vec{\omega}$ is opposite for the reflection.



Strictly speaking, $\vec{\omega}$ and $\vec{\alpha}$ are not quite vectors. The problem is that they do not behave like vectors under reflection. Suppose, as we are looking directly into a mirror, a particle moving with velocity $\vec{\mathbf{v}}$ to the right passes in front of and parallel to the mirror. In the reflection of the mirror, $\vec{\mathbf{v}}$ still points to the right, Fig. 10–10a. Thus a true vector, like velocity, when pointing parallel to the face of the mirror has the same direction in the reflection as in actuality. Now consider a wheel rotating in front of the mirror, so $\vec{\omega}$ points to the right. (We will be looking at the edge of the wheel.) As viewed in the mirror, Fig. 10–10b, the wheel will be rotating in the opposite direction. So $\vec{\omega}$ will point in the *opposite* direction (to the left) in the mirror. Because $\vec{\omega}$ is different under reflection than a true vector, $\vec{\omega}$ is called a **pseudovector** or **axial vector**. The angular acceleration $\vec{\alpha}$ is also a pseudovector, as are all cross products of true vectors (Section 11–2). The difference between true vectors and pseudovectors is important in elementary particle physics, but will not concern us in this book.

10–3 Constant Angular Acceleration

In Chapter 2, we derived the useful kinematic equations (Eqs. 2–12) that relate acceleration, velocity, distance, and time for the special case of uniform linear acceleration. Those equations were derived from the definitions of linear velocity and acceleration, assuming constant acceleration. The definitions of angular velocity and angular acceleration are just like those for their linear counterparts, except that θ replaces the linear displacement x, ω replaces v, and α replaces a. Therefore, the angular equations for **constant angular acceleration** will be analogous to Eqs. 2–12 with x replaced by θ, v by ω, and a by α, and they can be derived in exactly the same way. We summarize the angular equations here, opposite their linear equivalents, Eqs. 2–12 (for simplicity we chose $x_0 = 0$ and $\theta_0 = 0$ at the initial time $t_0 = 0$):

Angular	Linear			
$\omega = \omega_0 + \alpha t$	$v = v_0 + at$	[constant α, a]	**(10–9a)**	*Kinematic equations*
$\theta = \omega_0 t + \frac{1}{2}\alpha t^2$	$x = v_0 t + \frac{1}{2}at^2$	[constant α, a]	**(10–9b)**	*for constant*
$\omega^2 = \omega_0^2 + 2\alpha\theta$	$v^2 = v_0^2 + 2ax$	[constant α, a]	**(10–9c)**	*angular acceleration*
$\overline{\omega} = \frac{\omega + \omega_0}{2}$	$\overline{v} = \frac{v + v_0}{2}$	[constant α, a]	**(10–9d)**	*$(x_0 = 0,\ \theta_0 = 0)$*

Note that ω_0 represents the angular velocity at $t_0 = 0$, whereas θ and ω represent the angular position and angular velocity, respectively, at time t. Since the angular acceleration is constant, $\alpha = \overline{\alpha}$.

EXAMPLE 10–6 **Centrifuge acceleration.** A centrifuge rotor is accelerated for 30 s from rest to 20,000 rpm (revolutions per minute). (*a*) What is its average angular acceleration (rad/s^2)? (*b*) Through how many revolutions has the centrifuge rotor turned during its acceleration period, assuming constant angular acceleration?

PHYSICS APPLIED *Centrifuge*

APPROACH To determine $\overline{\alpha} = \Delta\omega/\Delta t$, we need the initial and final angular velocities. For (*b*), we use Eqs. 10–9 (recall that one revolution corresponds to $\theta = 2\pi$ rad).

SOLUTION (*a*) The initial angular velocity is $\omega = 0$. The final angular velocity is

$$\omega = 2\pi f = (2\pi \text{ rad/rev})\frac{(20{,}000 \text{ rev/min})}{(60 \text{ s/min})} = 2100 \text{ rad/s}.$$

Then, since $\overline{\alpha} = \Delta\omega/\Delta t$ and $\Delta t = 30$ s, we have

$$\overline{\alpha} = \frac{\omega - \omega_0}{\Delta t} = \frac{2100 \text{ rad/s} - 0}{30 \text{ s}} = 70 \text{ rad/s}^2.$$

That is, every second the rotor's angular velocity increases by 70 rad/s, or by (70 rad/s)(1 rev/2π rad) = 11 revolutions per second.

(*b*) To find θ we could use either Eq. 10–9b or 10–9c, or both to check our answer. The former gives

$$\theta = \omega_0 t + \tfrac{1}{2}\alpha t^2 = 0 + \tfrac{1}{2}(70 \text{ rad/s}^2)(30 \text{ s})^2 = 3.15 \times 10^4 \text{ rad},$$

where we have kept an extra digit because this is an intermediate result. To find the total number of revolutions, we divide by 2π rad/rev and obtain

$$\frac{3.15 \times 10^4 \text{ rad}}{2\pi \text{ rad/rev}} = 5.0 \times 10^3 \text{ rev} = 5{,}000 \text{ revolutions}.$$

NOTE Let us calculate θ using Eq. 10–9c:

$$\theta = \frac{\omega^2 - \omega_0^2}{2\alpha} = \frac{(2100 \text{ rad/s})^2 - 0}{2(70 \text{ rad/s}^2)} = 3.15 \times 10^4 \text{ rad}$$

which checks our answer above using Eq. 10–9b perfectly.

10–4 Torque

We have so far discussed rotational kinematics—the description of rotational motion in terms of angular position, angular velocity, and angular acceleration. Now we discuss the dynamics, or causes, of rotational motion. Just as we found analogies between linear and rotational motion for the description of motion, so rotational equivalents for dynamics exist as well.

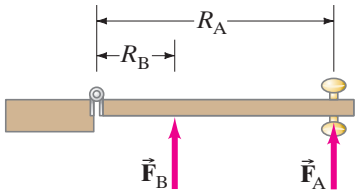


FIGURE 10–11 Top view of a door. Applying the same force with different lever arms, R_A and R_B. If $R_A = 3R_B$, then to create the same effect (angular acceleration), F_B needs to be three times F_A.

To make an object start rotating about an axis clearly requires a force. But the direction of this force, and where it is applied, are also important. Take, for example, an ordinary situation such as the overhead view of the door in Fig. 10–11. If you apply a force $\vec{\mathbf{F}}_A$ perpendicular to the door as shown, you will find that the greater the magnitude, F_A, the more quickly the door opens. But now if you apply the same force at a point closer to the hinge—say, $\vec{\mathbf{F}}_B$ in Fig. 10–11—the door will not open so quickly. The effect of the force is less: *where* the force acts, as well as its magnitude and direction, affects how quickly the door opens. Indeed, if only this one force acts, the angular acceleration of the door is proportional not only to the magnitude of the force, but is also directly proportional to *the perpendicular distance from the axis of rotation to the line along which the force acts*. This distance is called the **lever arm**, or **moment arm**, of the force, and is labeled R_A and R_B for the two forces in Fig. 10–11. Thus, if R_A in Fig. 10–11 is three times larger than R_B, then the angular acceleration of the door will be three times as great, assuming that the magnitudes of the forces are the same. To say it another way, if $R_A = 3R_B$, then F_B must be three times as large as F_A to give the same angular acceleration. (Figure 10–12 shows two examples of tools whose long lever arms are very effective.)

FIGURE 10–12 (a) A plumber can exert greater torque using a wrench with a long lever arm. (b) A tire iron too can have a long lever arm.



(a) (b)

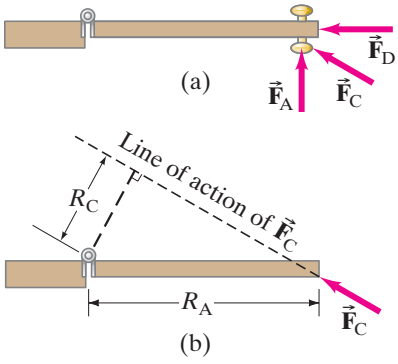
The angular acceleration, then, is proportional to the product of the *force times the lever arm*. This product is called the *moment of the force* about the axis, or, more commonly, it is called the **torque**, and is represented by τ (Greek lowercase letter tau). Thus, the angular acceleration α of an object is directly proportional to the net applied torque τ:

$$\alpha \propto \tau,$$

and we see that it is torque that gives rise to angular acceleration. This is the rotational analog of Newton's second law for linear motion $(a \propto F)$.

We defined the lever arm as the *perpendicular* distance from the axis of rotation to the line of action of the force—that is, the distance which is perpendicular both to the axis of rotation and to an imaginary line drawn along the direction of the force. We do this to take into account the effect of forces acting at an angle. It is clear that a force applied at an angle, such as $\vec{\mathbf{F}}_C$ in Fig. 10–13, will be less effective than the same magnitude force applied perpendicular to the door, such as $\vec{\mathbf{F}}_A$ (Fig. 10–13a). And if you push on the end of the door so that the force is directed toward the hinge (the axis of rotation), as indicated by $\vec{\mathbf{F}}_D$, the door will not rotate at all.

FIGURE 10–13 (a) Forces acting at different angles at the doorknob. (b) The lever arm for $\vec{\mathbf{F}}_C$ is defined as the perpendicular distance from the axis of rotation (the hinge) to the line of action of the force (R_C for the force $\vec{\mathbf{F}}_C$).



The lever arm for a force such as $\vec{\mathbf{F}}_C$ is found by drawing a line along the direction of $\vec{\mathbf{F}}_C$ (this is the "line of action" of $\vec{\mathbf{F}}_C$). Then we draw another line, perpendicular to this line of action, that goes to the axis of rotation and is perpendicular also to it. The length of this second line is the lever arm for $\vec{\mathbf{F}}_C$ and is labeled R_C in Fig. 10–13b. The lever arm for $\vec{\mathbf{F}}_A$ is the full distance from the hinge to the doorknob, R_A (just as in Fig. 10–11). Thus R_C is much smaller than R_A.

The magnitude of the torque associated with $\vec{\mathbf{F}}_C$ is then $R_C F_C$. This short lever arm R_C and the corresponding smaller torque associated with $\vec{\mathbf{F}}_C$ are consistent with the observation that $\vec{\mathbf{F}}_C$ is less effective in accelerating the door than is $\vec{\mathbf{F}}_A$ with its larger lever arm. When the lever arm is defined in this way, experiment shows that the relation $\alpha \propto \tau$ is valid in general. Notice in Fig. 10–13a that the line of action of the force $\vec{\mathbf{F}}_D$ passes through the hinge, and hence its lever arm is zero. Consequently, zero torque is associated with $\vec{\mathbf{F}}_D$ and it gives rise to no angular acceleration, in accord with everyday experience (you can't get a door to start moving by pushing directly at the hinge).

In general, then, we can write the magnitude of the torque about a given axis as

$$\tau = R_\perp F, \qquad \textbf{(10–10a)}$$

where $R_\perp$ is the lever arm, and the perpendicular symbol ($\perp$) reminds us that we must use the distance from the axis of rotation that is perpendicular to the line of action of the force (Fig. 10–14a).

An equivalent way of determining the torque associated with a force is to resolve the force into components parallel and perpendicular to the line that connects the axis to the point of application of the force, as shown in Fig. 10–14b. The component $F_\parallel$ exerts no torque since it is directed at the rotation axis (its lever arm is zero). Hence the torque will be equal to $F_\perp$ times the distance R from the axis to the point of application of the force:

$$\tau = RF_\perp. \qquad \textbf{(10–10b)}$$

This gives the same result as Eq. 10–10a because $F_\perp = F\sin\theta$ and $r_\perp = r\sin\theta$. Thus

$$\tau = RF\sin\theta \qquad \textbf{(10–10c)}$$

in either case. [Note that θ is the angle between the directions of $\vec{\mathbf{F}}$ and R (radial line from the axis to the point where $\vec{\mathbf{F}}$ acts)]. We can use any of Eqs. 10–10 to calculate the torque, whichever is easiest.

Because torque is a distance times a force, it is measured in units of m · N in SI units,† cm · dyne in the cgs system, and ft · lb in the English system.

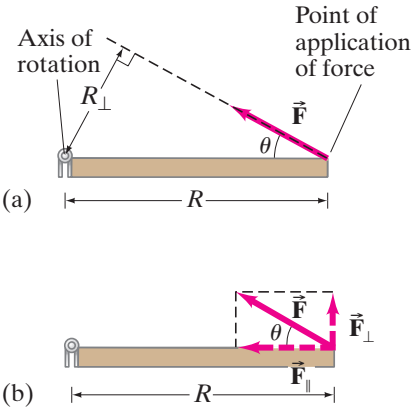
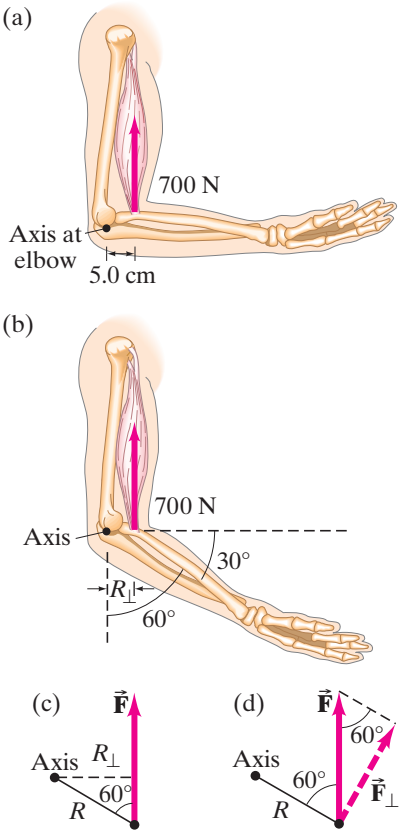


FIGURE 10–14
Torque = $R_\perp F = RF_\perp$.

FIGURE 10–15 Example 10–7.



EXAMPLE 10–7 **Biceps torque.** The biceps muscle exerts a vertical force on the lower arm, bent as shown in Figs. 10–15a and b. For each case, calculate the torque about the axis of rotation through the elbow joint, assuming the muscle is attached 5.0 cm from the elbow as shown.

APPROACH The force is given, and the lever arm in (*a*) is given. In (*b*) we have to take into account the angle to get the lever arm.

SOLUTION (*a*) $F = 700\,\text{N}$ and $R_\perp = 0.050\,\text{m}$, so

$$\tau = R_\perp F = (0.050\,\text{m})(700\,\text{N}) = 35\,\text{m}\cdot\text{N}.$$

(*b*) Because the arm is at an angle below the horizontal, the lever arm is shorter (Fig. 10–15c) than in part (*a*): $R_\perp = (0.050\,\text{m})(\sin 60°)$, where $\theta = 60°$ is the angle between $\vec{\mathbf{F}}$ and R. F is still 700 N, so

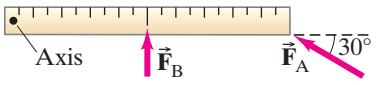
$$\tau = (0.050\,\text{m})(\sin 60°)(700\,\text{N}) = 30\,\text{m}\cdot\text{N}.$$

The arm can exert less torque at this angle than when it is at 90°. Weight machines at gyms are often designed to take this variation with angle into account.

NOTE In (*b*), we could instead have used $\tau = RF_\perp$. As shown in Fig. 10–15d, $F_\perp = F\sin 60°$. Then $\tau = RF_\perp = RF\sin\theta = (0.050\,\text{m})(700\,\text{N})(\sin 60°)$ gives the same result.

EXERCISE B Two forces ($F_\text{B} = 20\,\text{N}$ and $F_\text{A} = 30\,\text{N}$) are applied to a meter stick which can rotate about its left end, Fig. 10–16. Force $\vec{\mathbf{F}}_\text{B}$ is applied perpendicularly at the midpoint. Which force exerts the greater torque: F_A, F_B, or both the same?

FIGURE 10–16 Exercise B.



When more than one torque acts on an object, the angular acceleration α is found to be proportional to the *net* torque. If all the torques acting on an object tend to rotate it in the same direction about a fixed axis of rotation, the net torque is the sum of the torques. But if, let's say, one torque acts to rotate an object in one direction, and a second torque acts to rotate the object in the opposite direction, the net torque is the difference of the two torques. We normally assign a positive sign to torques that act to rotate the object counterclockwise (just as θ is usually positive counterclockwise), and a negative sign to torques that act to rotate the object clockwise.

†Note that the units for torque are the same as those for energy. We write the unit for torque here as m · N (in SI) to distinguish it from energy (N · m) because the two quantities are very different. The special name *joule* (1 J = 1 N · m) is used only for energy (and for work), *never* for torque.

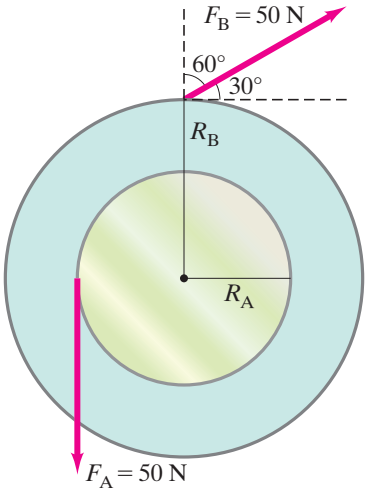


FIGURE 10–17 Example 10–8. The torque due to $\vec{\mathbf{F}}_A$ tends to accelerate the wheel counterclockwise, whereas the torque due to $\vec{\mathbf{F}}_B$ tends to accelerate the wheel clockwise.

EXAMPLE 10–8 **Torque on a compound wheel.** Two thin disk-shaped wheels, of radii $R_A = 30$ cm and $R_B = 50$ cm, are attached to each other on an axle that passes through the center of each, as shown in Fig. 10–17. Calculate the net torque on this compound wheel due to the two forces shown, each of magnitude 50 N.

APPROACH The force $\vec{\mathbf{F}}_A$ acts to rotate the system counterclockwise, whereas $\vec{\mathbf{F}}_B$ acts to rotate it clockwise. So the two forces act in opposition to each other. We must choose one direction of rotation to be positive—say, counterclockwise. Then $\vec{\mathbf{F}}_A$ exerts a positive torque, $\tau_A = R_A F_A$, since the lever arm is R_A. On the other hand, $\vec{\mathbf{F}}_B$ produces a negative (clockwise) torque and does not act perpendicular to R_B, so we must use the perpendicular component of $\vec{\mathbf{F}}_B$ to calculate the torque it produces: $\tau_B = -R_B F_{B\perp} = -R_B F_B \sin\theta$, where $\theta = 60°$. (Recall that θ must be the angle between $\vec{\mathbf{F}}_B$ and a radial line from the axis.)

SOLUTION The net torque is

$$\begin{aligned}\tau &= R_A F_A - R_B F_B \sin 60° \\ &= (0.30\text{ m})(50\text{ N}) - (0.50\text{ m})(50\text{ N})(\sin 60°) = -6.7\text{ m}\cdot\text{N}.\end{aligned}$$

This net torque acts to accelerate the rotation of the wheel in the clockwise direction.

10–5 Rotational Dynamics; Torque and Rotational Inertia

We discussed in Section 10–4 that the angular acceleration α of a rotating object is proportional to the net torque τ applied to it:

$$\alpha \propto \Sigma\tau.$$

We write $\Sigma\tau$ to remind us that it is the *net* torque (sum of all torques acting on the object) that is proportional to α. This corresponds to Newton's second law for translational motion, $a \propto \Sigma F$. In the translational case, the acceleration is not only proportional to the net force, but it is also inversely proportional to the inertia of the object, which we call its mass, m. Thus we wrote $a = \Sigma F/m$. But what plays the role of mass for the rotational case? That is what we now set out to determine. At the same time, we will see that the relation $\alpha \propto \Sigma\tau$ follows directly from Newton's second law, $\Sigma F = ma$.

We first examine a very simple case: a particle of mass m revolving in a circle of radius R at the end of a string or rod whose mass we can ignore compared to m (Fig. 10–18). Consider a force F that acts on the mass m tangent to the circle as shown. The torque that gives rise to an angular acceleration is $\tau = RF$. If we use Newton's second law for linear quantities, $\Sigma F = ma$, and Eq. 10–5 relating the angular acceleration to the tangential linear acceleration, $a_{\tan} = R\alpha$, then we have

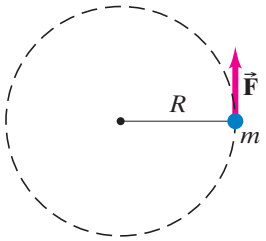
$$\begin{aligned}F &= ma \\ &= mR\alpha,\end{aligned}$$

where α is given in rad/s^2. When we multiply both sides of this equation by R, we find that the torque $\tau = RF = R(mR\alpha)$, or

$$\tau = mR^2\alpha. \qquad \text{[single particle]} \quad \textbf{(10–11)}$$

Here at last we have a direct relation between the angular acceleration and the applied torque τ. The quantity mR^2 represents the *rotational inertia* of the particle and is called its **moment of inertia**.

FIGURE 10–18 A mass m revolving in a circle of radius R about a fixed point.



Now let us consider a rotating rigid object, such as a wheel rotating about a fixed axis through its center, such as an axle. We can think of the wheel as consisting of many particles located at various distances from the axis of rotation. We can apply Eq. 10–11 to each particle of the object; that is, we write $\tau_i = m_i R_i^2 \alpha$ for the i^{th} particle of the object. Then we sum over all the particles. The sum of the various torques is just the total torque, $\Sigma\tau$, so we obtain:

$$\Sigma\tau_i = \left(\Sigma m_i R_i^2\right)\alpha \qquad \text{[axis fixed]} \quad \textbf{(10–12)}$$

where we factored out α since it is the same for all the particles of a rigid object.

The sum $\Sigma m_i R_i^2$ in Eq. 10–12 represents the sum of the masses of each particle in the object multiplied by the square of the distance of that particle from the axis of rotation. If we assign each particle a number $(1, 2, 3, \ldots)$, then

$$\Sigma m_i R_i^2 = m_1 R_1^2 + m_2 R_2^2 + m_3 R_3^2 + \cdots.$$

This sum is called the **moment of inertia** (also **rotational inertia**) I of the object:

$$I = \Sigma m_i R_i^2 = m_1 R_1^2 + m_2 R_2^2 + \cdots. \qquad \textbf{(10–13)}$$

Combining Eqs. 10–12 and 10–13, we can write

$$\Sigma\tau = I\alpha. \qquad \left[\begin{array}{c}\text{axis fixed in}\\ \text{inertial reference frame}\end{array}\right] \qquad \textbf{(10–14)}$$

NEWTON'S SECOND LAW FOR ROTATION

This is the rotational equivalent of Newton's second law. It is valid for the rotation of a rigid object about a fixed axis.[†] Equation 10–14 is valid even when the object is translating with acceleration, but only if we calculate I and α about the center of mass of the object, and the rotation axis through the CM doesn't change direction. A ball rolling down a ramp is an example. (We show this in Chapter 11.) Thus

$$(\Sigma\tau)_{\text{CM}} = I_{\text{CM}}\alpha_{\text{CM}}, \qquad \left[\begin{array}{c}\text{axis fixed in direction,}\\ \text{but may accelerate}\end{array}\right] \qquad \textbf{(10–15)}$$

where the subscript CM means "calculated about the center of mass." (The axis fixed in direction means a rolling wheel makes no left or right turns).

We see that the moment of inertia, I, which is a measure of the rotational inertia of an object, plays the same role for rotational motion that mass does for translational motion. As can be seen from Eq. 10–13, the rotational inertia of an object depends not only on its mass, but also on how that mass is distributed with respect to the axis. For example, a large-diameter cylinder will have greater rotational inertia than one of equal mass but smaller diameter, Fig. 10–19. The former will be harder to start rotating, and harder to stop. When the mass is concentrated farther from the axis of rotation, the rotational inertia is greater. For rotational motion, the mass of an object can *not* be considered as concentrated at its center of mass.

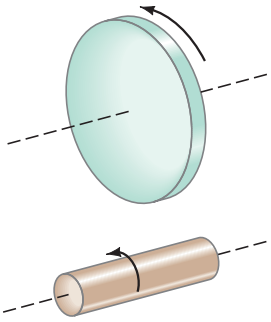


FIGURE 10–19 A large-diameter cylinder has greater rotational inertia than one of equal mass but smaller diameter.

CAUTION

Mass can not *be considered concentrated at* CM *for rotational motion*

EXAMPLE 10–9 **Two weights on a bar: different axis, different I.** Two small "weights," of mass 5.0 kg and 7.0 kg, are mounted 4.0 m apart on a light rod (whose mass can be ignored), as shown in Fig. 10–20. Calculate the moment of inertia of the system (*a*) when rotated about an axis halfway between the weights, Fig. 10–20a, and (*b*) when rotated about an axis 0.50 m to the left of the 5.0-kg mass (Fig. 10–20b).

APPROACH In each case, the moment of inertia of the system is found by summing over the two parts using Eq. 10–13.

SOLUTION (*a*) Both weights are the same distance, 2.0 m, from the axis of rotation. Thus

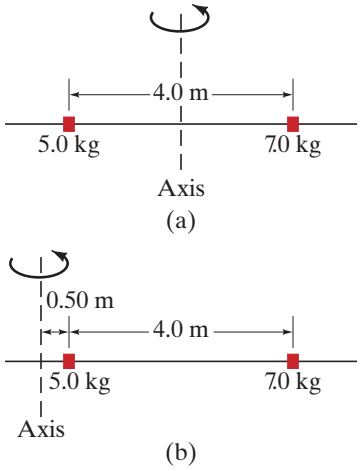
$$\begin{aligned} I = \Sigma mR^2 &= (5.0\,\text{kg})(2.0\,\text{m})^2 + (7.0\,\text{kg})(2.0\,\text{m})^2 \\ &= 20\,\text{kg}\cdot\text{m}^2 + 28\,\text{kg}\cdot\text{m}^2 = 48\,\text{kg}\cdot\text{m}^2. \end{aligned}$$

(*b*) The 5.0-kg mass is now 0.50 m from the axis, and the 7.0-kg mass is 4.50 m from the axis. Then

$$\begin{aligned} I = \Sigma mR^2 &= (5.0\,\text{kg})(0.50\,\text{m})^2 + (7.0\,\text{kg})(4.5\,\text{m})^2 \\ &= 1.3\,\text{kg}\cdot\text{m}^2 + 142\,\text{kg}\cdot\text{m}^2 = 143\,\text{kg}\cdot\text{m}^2. \end{aligned}$$

NOTE This Example illustrates two important points. First, the moment of inertia of a given system is different for different axes of rotation. Second, we see in part (*b*) that mass close to the axis of rotation contributes little to the total moment of inertia; here, the 5.0-kg object contributed less than 1% to the total.

FIGURE 10–20 Example 10–9. Calculating the moment of inertia.



CAUTION

I depends on axis of rotation and on distribution of mass

[†]That is, the axis is fixed relative to the object and is fixed in an inertial reference frame. This includes an axis moving at uniform velocity in an inertial frame, since the axis can be considered fixed in a second inertial frame that moves with respect to the first.

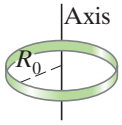
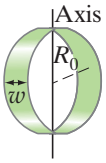
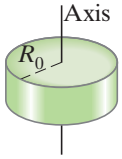
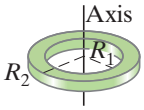
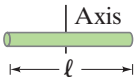
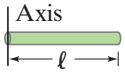
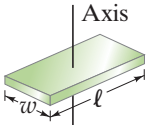
	Object	Location of axis		Moment of inertia
(a)	**Thin hoop**, radius R_0	Through center		MR_0^2
(b)	**Thin hoop**, radius R_0 width w	Through central diameter		$\frac{1}{2}MR_0^2 + \frac{1}{12}Mw^2$
(c)	**Solid cylinder**, radius R_0	Through center		$\frac{1}{2}MR_0^2$
(d)	**Hollow cylinder**, inner radius R_1 outer radius R_2	Through center		$\frac{1}{2}M(R_1^2 + R_2^2)$
(e)	**Uniform sphere**, radius r_0	Through center		$\frac{2}{5}Mr_0^2$
(f)	**Long uniform rod**, length ℓ	Through center		$\frac{1}{12}M\ell^2$
(g)	**Long uniform rod**, length ℓ	Through end		$\frac{1}{3}M\ell^2$
(h)	**Rectangular thin plate**, length ℓ, width w	Through center		$\frac{1}{12}M(\ell^2 + w^2)$

FIGURE 10–21 Moments of inertia for various objects of uniform composition, each with mass M. [We use R for radial distance from an axis, and r for distance from a point (only in e, the sphere), as discussed in Fig. 10–2.]

For most ordinary objects, the mass is distributed continuously, and the calculation of the moment of inertia, ΣmR^2, can be difficult. Expressions can, however, be worked out (using calculus) for the moments of inertia of regularly shaped objects in terms of the dimensions of the objects, as we will discuss in Section 10–7. Figure 10–21 gives these expressions for a number of solids rotated about the axes specified. The only one for which the result is obvious is that for the thin hoop or ring rotated about an axis passing through its center perpendicular to the plane of the hoop (Fig. 10–21a). For this hoop, all the mass is concentrated at the same distance from the axis, R_0. Thus $\Sigma mR^2 = (\Sigma m)R_0^2 = MR_0^2$, where M is the total mass of the hoop.

When calculation is difficult, I can be determined experimentally by measuring the angular acceleration α about a fixed axis due to a known net torque, $\Sigma\tau$, and applying Newton's second law, $I = \Sigma\tau/\alpha$, Eq. 10–14.

10–6 Solving Problems in Rotational Dynamics

When working with torque and angular acceleration (Eq. 10–14), it is important to use a consistent set of units, which in SI is: α in rad/s^2; τ in m · N; and the moment of inertia, I, in kg · m^2.

PROBLEM SOLVING

Rotational Motion

1. As always, draw a clear and complete **diagram**.
2. Choose the object or objects that will be the **system** to be studied.
3. Draw a **free-body diagram** for the object under consideration (or for each object separately, if more than one), showing all (and only) the forces acting on that object and exactly where they act, so you can determine the torque due to each. Gravity acts at the CM of the object (Section 9–8).
4. Identify the axis of rotation and determine the **torques** about it. Choose positive and negative directions of rotation (counterclockwise and clockwise), and assign the correct sign to each torque.
5. Apply **Newton's second law for rotation**, $\Sigma\tau = I\alpha$. If the moment of inertia is not given, and it is not the unknown sought, you need to determine it first. Use consistent units, which in SI are: α in rad/s^2; τ in m · N; and I in kg · m^2.
6. Also apply **Newton's second law for translation**, $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$, and **other** laws or principles as needed.
7. **Solve** the resulting equation(s) for the unknown(s).
8. Do a rough **estimate** to determine if your answer is reasonable.

EXAMPLE 10–10 **A heavy pulley.** A 15.0-N force (represented by $\vec{\mathbf{F}}_T$) is applied to a cord wrapped around a pulley of mass $M = 4.00\,\text{kg}$ and radius $R_0 = 33.0\,\text{cm}$, Fig. 10–22. The pulley accelerates uniformly from rest to an angular speed of 30.0 rad/s in 3.00 s. Assume there is a frictional torque $\tau_{fr} = 1.10\,\text{m} \cdot \text{N}$ at the axle. Determine the moment of inertia of the pulley rotating about its center.

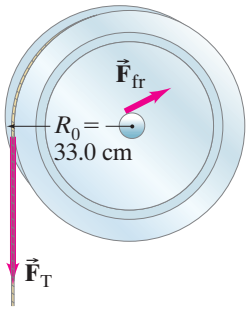


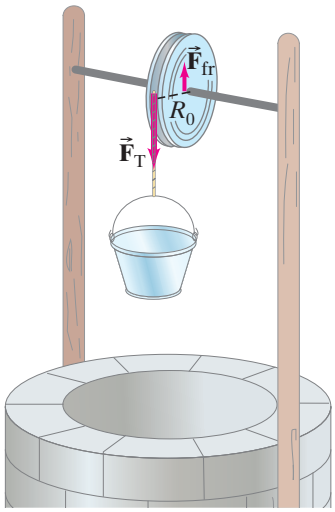
FIGURE 10–22 Example 10–10.

APPROACH We follow the steps of the Problem Solving Strategy above.

SOLUTION

1. **Draw a diagram**. The pulley and the attached cord are shown in Fig. 10–22.
2. **Choose the system**: the pulley.
3. **Draw a free-body diagram**. The force that the cord exerts on the pulley is shown as $\vec{\mathbf{F}}_T$ in Fig. 10–22. The friction force acts all around the axle, retarding the motion, as suggested by $\vec{\mathbf{F}}_{fr}$ in Fig. 10–22. We are given only its torque, which is what we need. Two other forces could be included in the diagram: the force of gravity mg down and whatever force keeps the axle in place (they balance each other). They do not contribute to the torque (their lever arms are zero) and so we omit them to keep our diagram simple.
4. **Determine the torques**. The cord exerts a force $\vec{\mathbf{F}}_T$ that acts at the edge of the pulley, so its lever arm is R_0. The torque exerted by the cord equals R_0F_T and is counterclockwise, which we choose to be positive. The frictional torque is given as $\tau_{fr} = 1.10\,\text{m} \cdot \text{N}$; it opposes the motion and is negative.
5. **Apply Newton's second law for rotation**. The net torque is
$$\Sigma\tau = R_0F_T - \tau_{fr} = (0.330\,\text{m})(15.0\,\text{N}) - 1.10\,\text{m} \cdot \text{N} = 3.85\,\text{m} \cdot \text{N}.$$
The angular acceleration α is found from the given data that it takes 3.00 s to accelerate the pulley from rest to $\omega = 30.0\,\text{rad/s}$:
$$\alpha = \frac{\Delta\omega}{\Delta t} = \frac{30.0\,\text{rad/s} - 0}{3.00\,\text{s}} = 10.0\,\text{rad/s}^2.$$
Newton's second law, $\Sigma\tau = I\alpha$, can be solved for I which is the unknown: $I = \Sigma\tau/\alpha$.
6. **Other calculations**: None needed.
7. **Solve for unknowns**. From Newton's second law (check units; rad has none):
$$I = \frac{\Sigma\tau}{\alpha} = \frac{3.85\,\text{m} \cdot \text{N}}{10.0\,\text{rad/s}^2} = 0.385\,\text{kg} \cdot \text{m}^2.$$
8. **Do a rough estimate**. We can do a rough estimate of the moment of inertia by assuming the pulley is a uniform cylinder and using Fig. 10–21c:
$$I \approx \tfrac{1}{2}MR_0^2 = \tfrac{1}{2}(4.00\,\text{kg})(0.330\,\text{m})^2 = 0.218\,\text{kg} \cdot \text{m}^2.$$
This is the same order of magnitude as our result, but numerically is less. This makes sense because a pulley is not usually uniform, but instead has more of its mass concentrated toward the outside edge, and would be expected to have a greater moment of inertia. A thin hoop, Fig. 10–21a, ought to have a greater I than our pulley, and indeed it does: $I = MR_0^2 = 0.436\,\text{kg} \cdot \text{m}^2$.

PROBLEM SOLVING
Usefulness and power of rough estimates



(a)



(b)

FIGURE 10–23 Example 10–11. (a) Pulley and falling bucket of mass m. This is also the free-body diagram for the pulley. (b) Free-body diagram for the bucket.

EXAMPLE 10–11 **Pulley and bucket.** Consider again the pulley in Example 10–10. But instead of a constant 15.0-N force being exerted on the cord, we now have a bucket of weight $w = 15.0\,\text{N}$ (mass $m = w/g = 1.53\,\text{kg}$) hanging from the cord. See Fig. 10–23a. We assume the cord has negligible mass and does not stretch or slip on the pulley. Calculate the angular acceleration α of the pulley and the linear acceleration a of the bucket. Assume the same frictional torque $\tau_{fr} = 1.10\,\text{m}\cdot\text{N}$ acts on the pulley.

APPROACH This situation looks a lot like Example 10–10, Fig. 10–22. But there is a big difference: the tension in the cord is now an unknown, and it is no longer equal to the weight of the bucket if the bucket accelerates. Our system has two parts: the bucket, which can undergo translational motion (Fig. 10–23b is its free-body diagram); and the pulley. The pulley does not translate, but it can rotate. We apply the rotational version of Newton's second law to the pulley, $\Sigma\tau = I\alpha$, and the linear version to the bucket, $\Sigma F = ma$.

SOLUTION Let F_T be the tension in the cord. Then a force F_T acts at the edge of the pulley, and we apply Newton's second law, Eq. 10–14, for the rotation of the pulley:

$$I\alpha = \Sigma\tau = R_0F_T - \tau_{fr}. \qquad \text{[pulley]}$$

Next we look at the (linear) motion of the bucket of mass m. Figure 10–23b, the free-body diagram for the bucket, shows that two forces act on the bucket: the force of gravity mg acts downward, and the tension of the cord F_T pulls upward. Applying Newton's second law, $\Sigma F = ma$, for the bucket, we have (taking downward as positive):

$$mg - F_T = ma. \qquad \text{[bucket]}$$

Note that the tension F_T, which is the force exerted on the edge of the pulley, is *not* equal to the weight of the bucket $(= mg = 15.0\,\text{N})$. There must be a net force on the bucket if it is accelerating, so $F_T < mg$. We can also see this from the last equation above, $F_T = mg - ma$.

To obtain α, we note that the tangential acceleration of a point on the edge of the pulley is the same as the acceleration of the bucket if the cord doesn't stretch or slip. Hence we can use Eq. 10–5, $a_{tan} = a = R_0\alpha$. Substituting $F_T = mg - ma = mg - mR_0\alpha$ into the first equation above (Newton's second law for rotation of the pulley), we obtain

$$I\alpha = \Sigma\tau = R_0F_T - \tau_{fr} = R_0(mg - mR_0\alpha) - \tau_{fr} = mgR_0 - mR_0^2\alpha - \tau_{fr}.$$

The unknown α appears on the left and in the second term on the far right, so we bring that term to the left side and solve for α:

$$\alpha = \frac{mgR_0 - \tau_{fr}}{I + mR_0^2}.$$

The numerator $(mgR_0 - \tau_{fr})$ is the net torque, and the denominator $(I + mR_0^2)$ is the total rotational inertia of the system. With $mg = 15.0\,\text{N}$ $(m = 1.53\,\text{kg})$ and, from Example 10–10, $I = 0.385\,\text{kg}\cdot\text{m}^2$ and $\tau_{fr} = 1.10\,\text{m}\cdot\text{N}$, then

$$\alpha = \frac{(15.0\,\text{N})(0.330\,\text{m}) - 1.10\,\text{m}\cdot\text{N}}{0.385\,\text{kg}\cdot\text{m}^2 + (1.53\,\text{kg})(0.330\,\text{m})^2} = 6.98\,\text{rad/s}^2.$$

The angular acceleration is somewhat less in this case than the $10.0\,\text{rad/s}^2$ of Example 10–10. Why? Because $F_T\,(= mg - ma = 15.0\,\text{N} - ma)$ is less than the 15.0-N force in Example 10–10. The linear acceleration of the bucket is

$$a = R_0\alpha = (0.330\,\text{m})(6.98\,\text{rad/s}^2) = 2.30\,\text{m/s}^2.$$

NOTE The tension in the cord F_T is less than mg because the bucket accelerates.

EXAMPLE 10–12 **Rotating rod.** A uniform rod of mass M and length ℓ can pivot freely (i.e., we ignore friction) about a hinge or pin attached to the case of a large machine, as in Fig. 10–24. The rod is held horizontally and then released. At the moment of release (when you are no longer exerting a force holding it up), determine (*a*) the angular acceleration of the rod and (*b*) the linear acceleration of the tip of the rod. Assume the force of gravity acts at the center of mass of the rod, as shown.

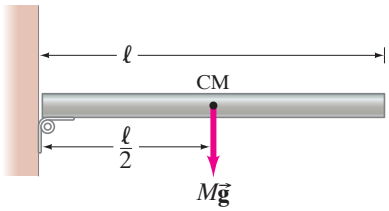


FIGURE 10–24 Example 10–12.

APPROACH (*a*) The only torque on the rod about the hinge is that due to gravity, which acts with a force $F = Mg$ downward with a lever arm $\ell/2$ at the moment of release (the CM is at the center of a uniform rod). There is also a force on the rod at the hinge, but with the hinge as axis of rotation, the lever arm of this force is zero. The moment of inertia of a uniform rod pivoted about its end is (Fig. 10–21g) $I = \frac{1}{3}M\ell^2$. In part (*b*) we use $a_{\tan} = R\alpha$.

SOLUTION (*a*) We use Eq. 10–14, solving for α to obtain the initial angular acceleration of the rod:

$$\alpha = \frac{\tau}{I} = \frac{Mg\frac{\ell}{2}}{\frac{1}{3}M\ell^2} = \frac{3}{2}\frac{g}{\ell}.$$

As the rod descends, the force of gravity on it is constant but the torque due to this force is not constant because the lever arm decreases. Hence the rod's angular acceleration is *not* constant.

(*b*) The initial linear acceleration of the tip of the rod is found from the relation $a_{\tan} = R\alpha$ (Eq. 10–5) with $R = \ell$:

$$a_{\tan} = \ell\alpha = \tfrac{3}{2}g.$$

NOTE The tip of the rod initially falls with an acceleration greater than g! A small object balanced on the tip of the rod would be left behind when the rod is released. In contrast, the CM of the rod, at a distance $\ell/2$ from the pivot, has acceleration $a_{\tan} = (\ell/2)\alpha = \frac{3}{4}g$. (Because the angular acceleration decreases in time, so does the linear acceleration of the rod tip.)

10–7 Determining Moments of Inertia

By Experiment

The moment of inertia of any object about any axis can be determined experimentally, such as by measuring the net torque $\Sigma\tau$ required to give the object an angular acceleration α. Then, from Eq. 10–14, $I = \Sigma\tau/\alpha$. See Example 10–10.

Using Calculus

For simple systems of masses or particles, the moment of inertia can be calculated directly, as in Example 10–9. Many objects can be considered as a continuous distribution of mass. In this case, Eq. 10–13 defining moment of inertia becomes

$$I = \int R^2\, dm, \qquad \textbf{(10–16)}$$

where dm represents the mass of any infinitesimal particle of the object and R is the perpendicular distance of this particle from the axis of rotation. The integral is taken over the whole object. This is easily done only for objects of simple geometric shape.

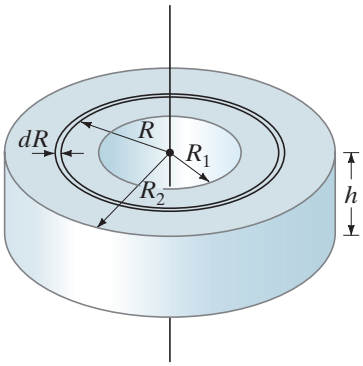


FIGURE 10–25 Determining the moment of inertia of a hollow cylinder (Example 10–13).

EXAMPLE 10–13 **Cylinder, solid or hollow.** (*a*) Show that the moment of inertia of a uniform hollow cylinder of inner radius R_1, outer radius R_2, and mass M, is $I = \frac{1}{2}M(R_1^2 + R_2^2)$, as stated in Fig. 10–21d, if the rotation axis is through the center along the axis of symmetry. (*b*) Obtain the moment of inertia for a solid cylinder.

APPROACH We know that the moment of inertia of a thin ring of radius R is mR^2 (Fig. 10–21a). So we divide the cylinder into thin concentric cylindrical rings or hoops of thickness dR, one of which is indicated in Fig. 10–25. The cylinder is uniform, which means its density ρ (mass per unit volume $= M/V$) is constant. So we write the mass dm of one thin ring as

$$dm = \rho\, dV,$$

where dV is the volume of the thin ring of radius R, thickness dR, and height h. Since $dV = (2\pi R)(dR)(h)$, we have

$$dm = 2\pi\rho h R\, dR.$$

SOLUTION (*a*) The moment of inertia is obtained by integrating (summing) over all these rings:

$$I = \int R^2\, dm = \int_{R_1}^{R_2} 2\pi\rho h R^3\, dR = 2\pi\rho h\left[\frac{R_2^4 - R_1^4}{4}\right] = \frac{\pi\rho h}{2}(R_2^4 - R_1^4),$$

where we are given that the cylinder has uniform density, $\rho =$ constant. (If this were not so, we would have to know ρ as a function of R before the integration could be carried out.) The volume V of this hollow cylinder is $V = (\pi R_2^2 - \pi R_1^2)h$, so its mass M is

$$M = \rho V = \rho\pi(R_2^2 - R_1^2)h.$$

Since $(R_2^4 - R_1^4) = (R_2^2 - R_1^2)(R_2^2 + R_1^2)$, we have

$$I = \frac{\rho\pi h}{2}(R_2^2 - R_1^2)(R_2^2 + R_1^2) = \tfrac{1}{2}M(R_1^2 + R_2^2),$$

as stated in Fig. 10–21d.

(*b*) For a solid cylinder, $R_1 = 0$ and if we set $R_2 = R_0$, then

$$I = \tfrac{1}{2}MR_0^2,$$

which is that given in Fig. 10–21c for a solid cylinder of mass M and radius R_0.

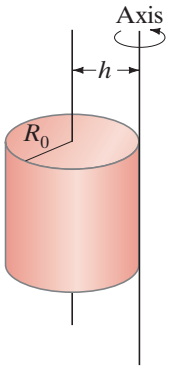
The Parallel-Axis Theorem

There are two simple theorems that are helpful in obtaining moments of inertia. The first is called the **parallel-axis theorem**. It relates the moment of inertia I of an object of total mass M about any axis, and its moment of inertia I_{CM} about an axis passing through the center of mass and parallel to the first axis. If the two axes are a distance h apart, then

$$I = I_{CM} + Mh^2. \qquad \text{[parallel axis]} \quad \textbf{(10–17)}$$

Thus, for example, if the moment of inertia about an axis through the CM is known, the moment of inertia about any axis parallel to this axis is easily obtained.

FIGURE 10–26 Example 10–14.



EXAMPLE 10–14 **Parallel axis.** Determine the moment of inertia of a solid cylinder of radius R_0 and mass M about an axis tangent to its edge and parallel to its symmetry axis, Fig. 10–26.

APPROACH We use the parallel-axis theorem with $I_{CM} = \frac{1}{2}MR_0^2$ (Fig. 10–21c).

SOLUTION Since $h = R_0$, Eq. 10–17 gives

$$I = I_{CM} + Mh^2 = \tfrac{3}{2}MR_0^2.$$

EXERCISE C In Figs. 10–21f and g, the moments of inertia for a thin rod about two different axes are given. Are they related by the parallel-axis theorem? Please show how.

*Proof of the Parallel-Axis Theorem

The proof of the parallel-axis theorem is as follows. We choose our coordinate system so the origin is at the CM, and I_{CM} is the moment of inertia about the z axis, perpendicular to the page at the point O. Figure 10–27 shows a cross-section of an object of arbitrary shape in the xy plane. We let I represent the moment of inertia of the object about an axis parallel to the z axis that passes through the point A in Fig. 10–27 where the point A has coordinates x_A and y_A. Let x_i, y_i, and m_i represent the coordinates and mass of an arbitrary particle of the object (shown as a tiny cube). The square of the distance from this particle to A is $[(x_i - x_A)^2 + (y_i - y_A)^2]$. So the moment of inertia, I, about the axis through A is

$$\begin{aligned} I &= \Sigma m_i[(x_i - x_A)^2 + (y_i - y_A)^2] \\ &= \Sigma m_i(x_i^2 + y_i^2) - 2x_A\Sigma m_i x_i - 2y_A\Sigma m_i y_i + (\Sigma m_i)(x_A^2 + y_A^2). \end{aligned}$$

The first term on the right is just $I_{CM} = \Sigma m_i(x_i^2 + y_i^2)$ since the CM is at the origin. The second and third terms are zero since, by definition of the CM, $\Sigma m_i x_i = \Sigma m_i y_i = 0$ because $x_{CM} = y_{CM} = 0$. The last term is Mh^2 since $\Sigma m_i = M$ and $(x_A^2 + y_A^2) = h^2$ where h is the distance of A from the CM. Thus, $I = I_{CM} + Mh^2$ for this cross-section, and will be for all other cross-sections of the object. Hence we have proved $I = I_{CM} + Mh^2$, which is Eq. 10–17.

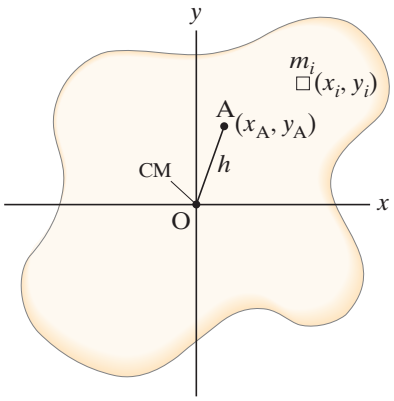


FIGURE 10–27 Derivation of the parallel-axis theorem.

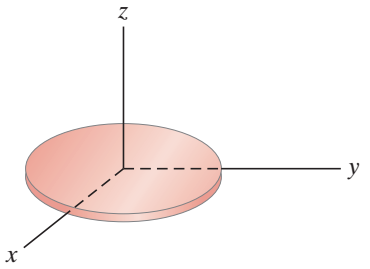
*The Perpendicular-Axis Theorem

The parallel-axis theorem can be applied to any object. The second theorem, the **perpendicular-axis theorem**, can be applied only to plane (flat) objects—that is, to two-dimensional objects, or objects of uniform thickness whose thickness can be neglected compared to the other dimensions. This theorem states that the sum of the moments of inertia of a thin plane object about any two perpendicular axes in the plane of the object, is equal to the moment of inertia about an axis through their point of intersection perpendicular to the plane of the object. That is, if the object is in the xy plane (Fig. 10–28),

$$I_z = I_x + I_y. \qquad \text{[object in } xy \text{ plane]} \quad \textbf{(10–18)}$$

Here I_z, I_x, I_y are moments of inertia about the z, x, and y axes. The proof is straightforward: $I_x = \Sigma m_i y_i^2$, because y_i is the perpendicular distance of any particle i of the object from the x axis; similarly, $I_y = \Sigma m_i x_i^2$, and $I_z = \Sigma m_i(x_i^2 + y_i^2)$, and so Eq. 10–18 follows directly.

FIGURE 10–28 The perpendicular-axis theorem.



10–8 Rotational Kinetic Energy

The quantity $\frac{1}{2}mv^2$ is the kinetic energy of an object undergoing translational motion. An object rotating about an axis is said to have **rotational kinetic energy**. By analogy with translational kinetic energy, we might expect this to be given by the expression $\frac{1}{2}I\omega^2$ where I is the moment of inertia of the object and ω is its angular velocity. We can indeed show this.

Consider any rigid rotating object as made up of many tiny particles, each of mass m_i. If we let R_i represent the distance of any one particle from the axis of rotation, then its linear velocity is $v_i = R_i\omega$, where ω is the same for all particles of the object. The total kinetic energy of the whole object will be the sum of the kinetic energies of all its particles:

$$\begin{aligned} K &= \Sigma(\tfrac{1}{2}m_i v_i^2) = \Sigma(\tfrac{1}{2}m_i R_i^2\omega^2) \\ &= \tfrac{1}{2}\Sigma(m_i R_i^2)\,\omega^2. \end{aligned}$$

We have factored out the $\frac{1}{2}$ and the ω^2 since they are the same for every particle of a rigid object. Since $\Sigma m_i R_i^2 = I$, the moment of inertia, we see that the kinetic energy, K, of an object rotating about a fixed axis is, as expected,

$$K = \tfrac{1}{2}I\omega^2. \qquad \text{[rotation about a fixed axis]} \quad \textbf{(10–19)}$$

If the axis is not fixed in space, the rotational kinetic energy can take on a more complicated form.

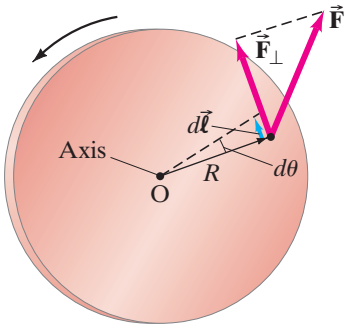


FIGURE 10–29 Calculating the work done by a torque acting on a rigid object rotating about a fixed axis.

The work done on an object rotating about a fixed axis can be written in terms of angular quantities. Suppose a force $\vec{\mathbf{F}}$ is exerted at a point whose distance from the axis of rotation is R, as in Fig. 10–29. The work done by this force is

$$W = \int \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}} = \int F_\perp R \, d\theta,$$

where $d\vec{\boldsymbol{\ell}}$ is an infinitesimal distance perpendicular to R with magnitude $d\ell = R\,d\theta$, and $F_\perp$ is the component of $\vec{\mathbf{F}}$ perpendicular to R and parallel to $d\vec{\boldsymbol{\ell}}$ (Fig. 10–29). But $F_\perp R$ is the torque about the axis, so

$$W = \int_{\theta_1}^{\theta_2} \tau \, d\theta \qquad \textbf{(10–20)}$$

is the work done by a torque τ to rotate an object through the angle $\theta_2 - \theta_1$. The rate of work done, or power P, at any instant is

$$P = \frac{dW}{dt} = \tau \frac{d\theta}{dt} = \tau\omega. \qquad \textbf{(10–21)}$$

The work-energy principle (Section 7–4) holds for rotation of a rigid object about a fixed axis: from Eq. 10–14 we have

$$\tau = I\alpha = I\frac{d\omega}{dt} = I\frac{d\omega}{d\theta}\frac{d\theta}{dt} = I\omega\frac{d\omega}{d\theta},$$

where we used the chain rule and $\omega = d\theta/dt$. Then $\tau\,d\theta = I\omega\,d\omega$ and

$$W = \int_{\theta_1}^{\theta_2} \tau \, d\theta = \int_{\omega_1}^{\omega_2} I\omega \, d\omega = \tfrac{1}{2}I\omega_2^2 - \tfrac{1}{2}I\omega_1^2. \qquad \textbf{(10–22)}$$

This is the work-energy principle for a rigid object rotating about a fixed axis. It states that the work done in rotating an object through an angle $\theta_2 - \theta_1$ is equal to the change in rotational kinetic energy of the object.

PHYSICS APPLIED
Energy from a flywheel

EXAMPLE 10–15 ESTIMATE **Flywheel.** Flywheels, which are simply large rotating disks, have been suggested as a means of storing energy for solar-powered generating systems. Estimate the kinetic energy that can be stored in an 80,000-kg (80-ton) flywheel with a diameter of 10 m (a three-story building). Assume it could hold together (without flying apart due to internal stresses) at 100 rpm.

APPROACH We use Eq. 10–19, $K = \frac{1}{2}I\omega^2$, but only after changing 100 rpm to ω in rad/s.

SOLUTION We are given

$$\omega = 100\,\text{rpm} = \left(100\,\frac{\text{rev}}{\text{min}}\right)\left(\frac{1\,\text{min}}{60\,\text{s}}\right)\left(\frac{2\pi\,\text{rad}}{\text{rev}}\right) = 10.5\,\text{rad/s}.$$

The kinetic energy stored in the disk $\left(\text{for which } I = \frac{1}{2}MR_0^2\right)$ is

$$K = \tfrac{1}{2}I\omega^2 = \tfrac{1}{2}\left(\tfrac{1}{2}MR_0^2\right)\omega^2$$

$$= \tfrac{1}{4}\left(8.0\times10^4\,\text{kg}\right)(5\,\text{m})^2(10.5\,\text{rad/s})^2 = 5.5\times10^7\,\text{J}.$$

NOTE In terms of kilowatt-hours $\left[1\,\text{kWh} = (1000\,\text{J/s})(3600\,\text{s/h})(1\,\text{h}) = 3.6\times10^6\,\text{J}\right]$, this energy is only about 15 kWh, which is not a lot of energy (one 3-kW oven would use it all in 5 h). Thus flywheels seem unlikely for this application.

EXAMPLE 10–16 **Rotating rod.** A rod of mass M is pivoted on a frictionless hinge at one end, as shown in Fig. 10–30. The rod is held at rest horizontally and then released. Determine the angular velocity of the rod when it reaches the vertical position, and the speed of the rod's tip at this moment.

APPROACH We can use the work-energy principle here. The work done is due to gravity, and is equal to the change in gravitational potential energy of the rod.

SOLUTION Since the CM of the rod drops a vertical distance $\ell/2$, the work done by gravity is

$$W = Mg\frac{\ell}{2}.$$

The initial kinetic energy is zero. Hence, from the work-energy principle,

$$\tfrac{1}{2}I\omega^2 = Mg\frac{\ell}{2}.$$

Since $I = \frac{1}{3}M\ell^2$ for a rod pivoted about its end (Fig. 10–21g), we can solve for ω:

$$\omega = \sqrt{\frac{3g}{\ell}}.$$

The tip of the rod will have a linear speed (see Eq. 10–4)

$$v = \ell\omega = \sqrt{3g\ell}.$$

NOTE The speed of the rod's tip is greater than the speed of an object that falls vertically a height ℓ, whose speed is only $v = \sqrt{2g\ell}$.

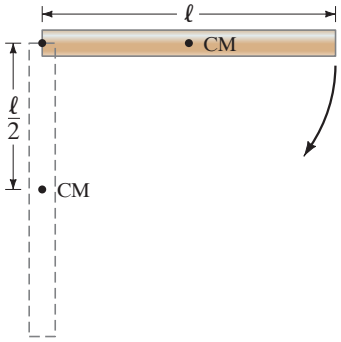


FIGURE 10–30 Example 10–16.

EXERCISE D Estimate the energy stored in the rotational motion of a hurricane. Model the hurricane as a uniform cylinder 300 km in diameter and 5 km high, made of air whose mass is 1.3 kg per m^3. Estimate the outer edge of the hurricane to move at a speed of 200 km/h.

10–9 Rotational Plus Translational Motion; Rolling

Rolling Without Slipping

The rolling motion of a ball or wheel is familiar in everyday life: a ball rolling across the floor, or the wheels and tires of a car or bicycle rolling along the pavement. Rolling *without slipping* depends on static friction between the rolling object and the ground. The friction is static because the rolling object's point of contact with the ground is at rest at each moment.

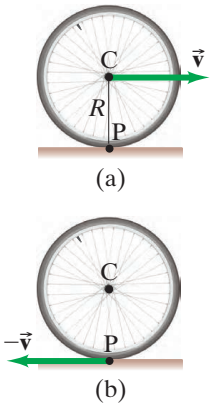
Rolling without slipping involves both rotation and translation. There is a simple relation between the linear speed v of the axle and the angular velocity ω of the rotating wheel or sphere: namely, $v = R\omega$, where R is the radius, as we now show. Figure 10–31a shows a wheel rolling to the right without slipping. At the instant shown, point P on the wheel is in contact with the ground and is momentarily at rest. (If P was not at rest, the wheel would be slipping.) The velocity of the axle at the wheel's center C is $\vec{\mathbf{v}}$. In Fig. 10–31b we have put ourselves in the reference frame of the wheel—that is, we are moving to the right with velocity $\vec{\mathbf{v}}$ relative to the ground. In this reference frame the axle C is at rest, whereas the ground and point P are moving to the left with velocity $-\vec{\mathbf{v}}$ as shown. In Fig. 10–31b, we are seeing pure rotation. We can then use Eq. 10–4 to obtain $v = R\omega$ in Fig. 10–31b, where R is the radius of the wheel. This is the same v as in Fig. 10–31a, so we see that the linear speed v of the axle relative to the ground is related to the angular velocity ω by

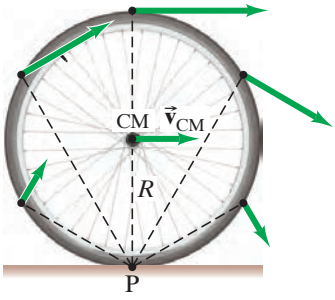
$$v = R\omega. \qquad \text{[rolling without slipping]}$$

This is valid only if there is no slipping. If the rotating wheel or sphere is uniform, or at least the CM is at the geometric center C (the axle), then we can write

$$v_{CM} = R\omega. \qquad \text{[rolling without slipping; CM at center]} \quad \textbf{(10–23)}$$

FIGURE 10–31 (a) A wheel rolling to the right. Its center C moves with velocity $\vec{\mathbf{v}}$. Point P, the point of contact with the ground, is at rest at this instant. (b) The same wheel as seen from a reference frame in which the axle of the wheel C is at rest—that is, we are moving to the right with velocity $\vec{\mathbf{v}}$ relative to the ground. Point P, which was at rest in (a), here in (b) is moving to the left with velocity $-\vec{\mathbf{v}}$ as shown. (See also Section 3–9 on relative velocity.)





(a)



(b)

FIGURE 10–32 (a) A rolling wheel rotates about the instantaneous axis (perpendicular to the page) passing through the point of contact with the ground, P. The arrows represent the instantaneous velocity of each point. (b) Photograph of a rolling wheel. The spokes are more blurred where the speed is greater.

Instantaneous Axis

When a wheel rolls without slipping, the point of contact of the wheel with the ground is instantaneously at rest. It is sometimes useful to think of the motion of the wheel as pure rotation about this "instantaneous axis" passing through that point P (Fig. 10–32a). Points close to the ground have a small linear speed, as they are close to this instantaneous axis, whereas points farther away have a greater linear speed. This can be seen in a photograph of a real rolling wheel (Fig. 10–32b): spokes near the top of the wheel appear more blurry because they are moving faster than those near the bottom of the wheel.

Total Kinetic Energy = $K_{CM} + K_{rot}$

An object that rotates while its center of mass (CM) undergoes translational motion will have both translational and rotational kinetic energy. Equation 10–19, $K = \frac{1}{2}I\omega^2$, gives the rotational kinetic energy if the rotation axis is fixed. If the object is moving, such as a wheel rolling along the ground, Fig. 10–33, there will also be translational kinetic energy. To obtain the total kinetic energy, we note that the rolling wheel undergoes pure rotation about its instantaneous point of contact P, Fig. 10–32. As we saw in the discussion of Fig. 10–31, the speed v of the center relative to the ground equals the speed of a point on the edge of the wheel relative to its center. Both of these speeds are related to the radius R by $v = R\omega$. Thus the angular velocity ω about point P is the same ω for the wheel rotating about its center. So the total kinetic energy, which about point P is pure rotation, is

$$K_{tot} = \tfrac{1}{2}I_P\omega^2,$$

where I_P is the rolling object's moment of inertia about the instantaneous axis at P. We can write K_{tot} in terms of the center of mass using the parallel-axis theorem: $I_P = I_{CM} + MR^2$, where we have substituted $h = R$ in Eq. 10–17, the parallel axis theorem. Thus

$$K_{tot} = \tfrac{1}{2}I_{CM}\omega^2 + \tfrac{1}{2}MR^2\omega^2.$$

But $R\omega = v_{CM}$, the speed of the center of mass (Eq. 10–23). So the total kinetic energy of a rolling object is

$$K_{tot} = \tfrac{1}{2}I_{CM}\omega^2 + \tfrac{1}{2}Mv_{CM}^2, \qquad \textbf{(10–24)}$$

where v_{CM} is the linear velocity of the CM, I_{CM} is the moment of inertia about an axis through the CM, ω is the angular velocity about this axis, and M is the total mass of the object. Equation 10–24 is also valid in general, not just for a rolling wheel or sphere.

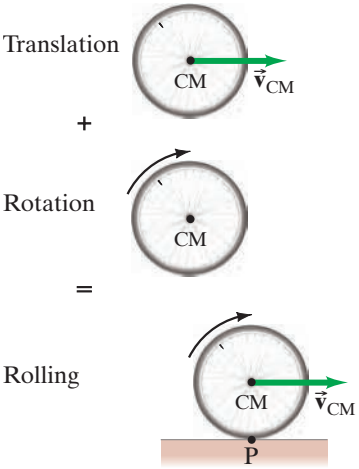


FIGURE 10–33 A wheel rolling without slipping can be considered as translation of the wheel as a whole with velocity $\vec{\mathbf{v}}_{CM}$ plus rotation about the CM.

EXAMPLE 10–17 **Sphere rolling down an incline.** What will be the speed of the CM of a uniform solid sphere of mass M and radius r_0 when it reaches the bottom of an incline if it starts from rest at a vertical height H and rolls without slipping? See Fig. 10–34. (Assume no slipping occurs because there is plenty of static friction.) Compare your result to that for an object *sliding* down a frictionless incline.

APPROACH We use the law of conservation of energy with gravitational potential energy, now including rotational as well as translational kinetic energy.

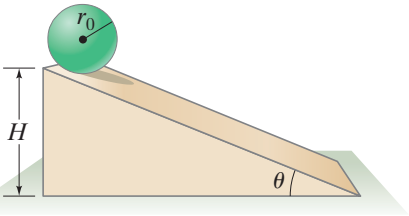


FIGURE 10–34 A uniform sphere rolling down a hill has both translational and rotational kinetic energy. Example 10–17.

SOLUTION The total energy of the sphere when its CM is a vertical distance y above the base of the incline is

$$E = \tfrac{1}{2}Mv_{CM}^2 + \tfrac{1}{2}I_{CM}\omega^2 + Mgy,$$

where v_{CM} is the speed of the center of mass, and Mgy is the gravitational potential energy. Applying conservation of energy, we equate the total energy at the top $(y = H + r_0,\ v_{CM} = 0,\ \omega = 0)$ to the total energy at the bottom $(y = r_0)$:

$$E_{top} = E_{bottom}$$
$$0 + 0 + Mg(H + r_0) = \tfrac{1}{2}Mv_{CM}^2 + \tfrac{1}{2}I_{CM}\omega^2 + Mgr_0. \quad \text{[energy conservation]}$$

The moment of inertia of a solid sphere about an axis through its center of mass is $I_{CM} = \frac{2}{5}Mr_0^2$, Fig. 10–21e. Since the sphere rolls without slipping, we have $\omega = v_{CM}/r_0$ (recall Fig. 10–31). Hence

$$MgH = \tfrac{1}{2}Mv_{CM}^2 + \tfrac{1}{2}\left(\tfrac{2}{5}Mr_0^2\right)\left(\frac{v_{CM}^2}{r_0^2}\right).$$

Canceling the M's and r_0's we obtain

$$\left(\tfrac{1}{2} + \tfrac{1}{5}\right)v_{CM}^2 = gH$$

or

$$v_{CM} = \sqrt{\tfrac{10}{7}gH}. \quad \text{[rolling sphere]}$$

We can compare this result for the speed of the CM of a rolling sphere to the speed of an object sliding down a plane without rotating and without friction, $\frac{1}{2}mv^2 = mgH$. For the sliding object, $v = \sqrt{2gH}$, which is greater than for a rolling sphere $\left(2 > \frac{10}{7}\right)$. An object sliding without friction or rotation transforms its initial potential energy entirely into translational kinetic energy (none into rotational kinetic energy), so the speed of its center of mass is greater.

CAUTION
Rolling objects go slower than sliding objects because of rotational kinetic energy

NOTE Our result for the rolling sphere shows (perhaps surprisingly) that v_{CM} is independent of both the mass M and the radius r_0 of the sphere.

CONCEPTUAL EXAMPLE 10–18 Which is fastest? Several objects roll without slipping down an incline of vertical height H, all starting from rest at the same moment. The objects are a thin hoop (or a plain wedding band), a spherical marble, a solid cylinder (a D-cell battery), and an empty soup can. In addition, a greased box slides down without friction. In what order do they reach the bottom of the incline? All the objects have the same mass M and radius R (if rolling).

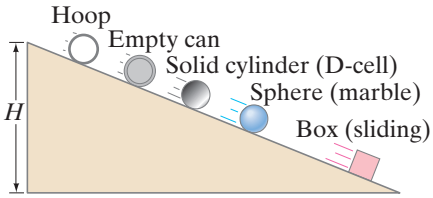
RESPONSE We use conservation of energy with gravitational potential energy plus rotational and translational kinetic energy. The sliding box would be fastest because the potential energy loss (MgH) is transformed completely into translational kinetic energy of the box, whereas for rolling objects the initial potential energy is shared between translational and rotational kinetic energies, and so the speed of the CM is less. For each of the rolling objects we can state that the decrease in potential energy equals the increase in translational plus rotational kinetic energy:

$$MgH = \tfrac{1}{2}Mv_{CM}^2 + \tfrac{1}{2}I_{CM}\omega^2.$$

For all our rolling objects, the moment of inertia I_{CM} is a numerical factor times MR^2 (Fig. 10–21). The mass M is in each term of the equation above so it cancels out. Thus the translational speed v_{CM} at the bottom does not depend on M. The CM speed also does not depend on the radius R because $\omega = v_{CM}/R$, so R^2 cancels out in the term $\frac{1}{2}I_{CM}\omega^2$. Thus the speed v_{CM} at the bottom of the incline depends only on that numerical factor in I_{CM} which expresses how the mass is distributed. The hoop, with all its mass concentrated at radius R $\left(I_{CM} = MR^2\right)$, has the largest moment of inertia. Hence the hoop has the most rotational kinetic energy and the least transational kinetic energy, and lowest speed at the bottom. The hoop arrives behind the D-cell $\left(I_{CM} = \frac{1}{2}MR^2\right)$, which will be behind the marble $\left(I_{CM} = \frac{2}{5}MR^2\right)$. The empty can, which is a hoop plus a thin disk, has most of its mass concentrated at R; so it will be a bit faster than the pure hoop but slower than the D-cell. See Fig. 10–35.

NOTE The rolling objects do not even have to have the same radius and mass: the speed at the bottom does not depend on the object's mass M or radius R, but only on its shape (and the height of the hill H).

FIGURE 10–35 Example 10–18.



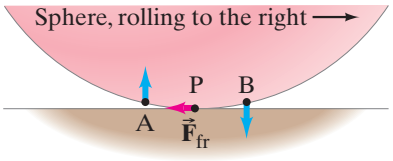


FIGURE 10–36 A sphere rolling to the right on a plane surface. The point in contact with the ground at any moment, point P, is momentarily at rest. Point A to the left of P is moving nearly vertically upward at the instant shown, and point B to the right is moving nearly vertically downward. An instant later, point B will touch the plane and be at rest momentarily. Thus the force $\vec{\mathbf{F}}_{fr}$ is perpendicular to the motion, so no work is done by the force of static friction.

If there had been little or no static friction between the rolling objects and the plane in these Examples, the round objects would have slid rather than rolled, or a combination of both. Static friction must be present to make a round object roll. We did not need to take friction into account in the energy equation for the rolling objects because it is *static* friction and does no work—the point of contact of a sphere at each instant does not slide, but moves perpendicular to the plane (first down and then up as shown in Fig. 10–36) as it rolls. Thus, no work is done by the static friction force because the force and the motion (displacement) are perpendicular. The reason the rolling objects in Examples 10–17 and 10–18 move down the slope more slowly than if they were sliding is *not* because friction slows them down. Rather, it is because some of the gravitional potential energy is converted to rotational kinetic energy, leaving less for the translational kinetic energy.

EXERCISE E Return to the Chapter-Opening Question, page 264, and answer it again now. Try to explain why you may have answered differently the first time.

Using $\Sigma\tau_{CM} = I_{CM}\alpha_{CM}$

We can examine objects rolling down a plane not only from the point of view of kinetic energy, as we did in Examples 10–17 and 10–18, but also in terms of forces and torques. If we calculate torques about an axis fixed in direction (even if the axis is accelerating) which passes through the center of mass of the rolling sphere, then

$$\Sigma\tau_{CM} = I_{CM}\alpha_{CM}$$

is valid, as we discussed in Section 10–5. See Eq. 10–15, whose validity we will show in Section 11–4. Be careful, however: do not assume $\Sigma\tau = I\alpha$ is always valid. You cannot just calculate τ, I, and α about any axis unless the axis is

(1) fixed in an inertial reference frame, or
(2) fixed in direction but passes through the CM of the object.

CAUTION
When is $\Sigma\tau = I\alpha$ valid?

EXAMPLE 10–19 **Analysis of a sphere on an incline using forces.** Analyze the rolling sphere of Example 10–17, Fig. 10–34, in terms of forces and torques, and determine the velocity v_{CM} of the sphere's CM at the base of the incline, and the magnitude of the friction force, F_{fr}, Fig. 10–37.

APPROACH We analyze the motion as translation of the CM plus rotation about the CM. F_{fr} is due to static friction and we cannot assume $F_{fr} = \mu_s F_N$, only $F_{fr} \leq \mu_s F_N$.

SOLUTION For translation in the x direction we have from $\Sigma F = ma$,

$$Mg\sin\theta - F_{fr} = Ma_{CM}.$$

In the y direction

$$F_N - Mg\cos\theta = 0$$

since there is no acceleration perpendicular to the plane. This last equation merely tells us the magnitude of the normal force,

$$F_N = Mg\cos\theta.$$

For the rotational motion about the CM, we use Newton's second law for rotation $\Sigma\tau_{CM} = I_{CM}\alpha_{CM}$ (Eq. 10–15), calculating about an axis passing through the CM of the sphere (radius r_0) but fixed in direction:

$$F_{fr}r_0 = \left(\tfrac{2}{5}Mr_0^2\right)\alpha_{CM}.$$

The other forces, $\vec{\mathbf{F}}_N$ and $M\vec{\mathbf{g}}$, point through the axis of rotation (CM), so have lever arms equal to zero and do not appear here. As we saw on page 283 and Fig. 10–31, $\omega = v/r_0$ where $v = v_{CM}$ is the speed of the CM. Taking derivatives of $\omega = v_{CM}/r_0$ with respect to time we have $\alpha_{CM} = a_{CM}/r_0$; substituting into the last equation we find

$$F_{fr} = \tfrac{2}{5}Ma_{CM}.$$

When we substitute this into the first equation in this solution, we get

$$Mg\sin\theta - \tfrac{2}{5}Ma_{CM} = Ma_{CM},$$

or

$$a_{CM} = \tfrac{5}{7}g\sin\theta.$$

We thus see that the acceleration of the CM of a rolling sphere is less than that for an object sliding without friction $(a = g\sin\theta)$. The sphere started from

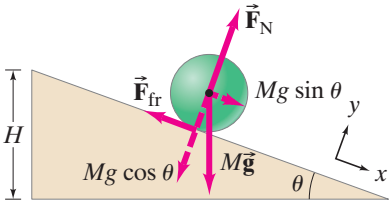


FIGURE 10–37 Example 10–19.

rest $(v_0 = 0)$ at the top of the incline (height H). To find the speed v_{CM} of the CM at the bottom we use Eq. 2–12c for constant linear acceleration $(v^2 = v_0^2 + 2ax)$, where the total distance traveled along the plane is $x = H/\sin\theta$ (see Fig. 10–37). Thus

$$v_{CM} = \sqrt{2a_{CM}x} = \sqrt{2\left(\frac{5}{7}g\sin\theta\right)\left(\frac{H}{\sin\theta}\right)} = \sqrt{\frac{10}{7}gH}.$$

This is the same result obtained in Example 10–17, although less effort was needed there. To get the magnitude of the force of friction, we use the equations obtained above:

$$F_{fr} = \tfrac{2}{5}Ma_{CM} = \tfrac{2}{5}M\left(\tfrac{5}{7}g\sin\theta\right) = \tfrac{2}{7}Mg\sin\theta.$$

NOTE If the coefficient of static friction is sufficiently small, or θ sufficiently large so this equation gives $F_{fr} > \mu_s F_N$, the sphere will not simply roll but will also slip as it moves down the plane. Because $F_{fr} = \frac{2}{7}Mg\sin\theta$ and $\mu_s F_N = \mu_s Mg\cos\theta$, then $F_{fr} > \mu_s F_N$ is equivalent to $\tan\theta > \frac{7}{2}\mu_s$.

*More Advanced Examples

Here we do three more Examples, all of them fun and interesting. When they use $\Sigma\tau = I\alpha$, we must remember that this equation is valid only if τ, α, and I are calculated about an axis that either (1) is fixed in an inertial reference frame, or (2) passes through the CM of the object and remains fixed in direction.

EXAMPLE 10–20 A falling yo-yo. String is wrapped around a uniform solid cylinder (something like a yo-yo) of mass M and radius R, and the cylinder starts falling from rest, Fig. 10–38a. As the cylinder falls, find (a) the acceleration of its CM and (b) the tension in the string.

APPROACH As always we begin with a free-body diagram, Fig. 10–38b, which shows the weight of the cylinder acting at the CM and the tension of the string $\vec{\mathbf{F}}_T$ acting at the edge of the cylinder. We write Newton's second law for the linear motion of the CM, choosing *down* as positive:

$$Ma_{CM} = \Sigma F = Mg - F_T.$$

Since we do not know the tension in the string, we cannot immediately solve for a_{CM}. So we try Newton's second law for the rotational motion, calculated about the center of mass (let clockwise be +):

$$\Sigma\tau_{CM} = I_{CM}\alpha_{CM}$$
$$F_T R = \tfrac{1}{2}MR^2\alpha_{CM}.$$

Because the cylinder "rolls without slipping" down the string, we have the additional relation that $a_{CM} = R\alpha_{CM}$. (Because $\frac{d}{dt}(v_{CM} = R\omega)$; Eq. 10–23 and Fig. 10–31.)

SOLUTION (a) The torque equation becomes

$$F_T R = \tfrac{1}{2}MR^2\left(\frac{a_{CM}}{R}\right) = \tfrac{1}{2}MRa_{CM}$$

so

$$F_T = \tfrac{1}{2}Ma_{CM}.$$

Substituting this into the force equation above, we obtain

$$Ma_{CM} = Mg - F_T = Mg - \tfrac{1}{2}Ma_{CM}.$$

Solving for a_{CM} we find that $a_{CM} = \frac{2}{3}g$. That is, the linear acceleration of the CM of the cylinder is less than what it would be if the cylinder is simply dropped. This makes sense because gravity is not the only vertical force acting; the tension in the string is acting as well.

(b) Since $a_{CM} = \frac{2}{3}g$, $F_T = \frac{1}{2}Ma_{CM} = \frac{1}{3}Mg$.

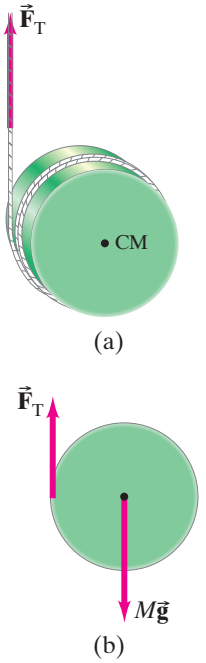


FIGURE 10–38 Example 10–20.

EXERCISE F Find the acceleration a of a yo-yo whose spindle has radius $\frac{1}{2}R$. Assume the moment of inertia is still $\frac{1}{2}MR^2$ (ignore the mass of the spindle).

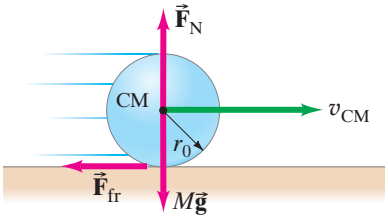


FIGURE 10–39 Example 10–21. A spherical bowling ball, rolling and slipping.

EXAMPLE 10–21 **What if a rolling ball slips?** A bowling ball of mass M and radius r_0 is thrown along a level surface so that initially $(t = 0)$ it slides with a linear speed v_0 but does not rotate $(\omega_0 = 0)$. As it slides, it begins to spin, and eventually rolls without slipping. How long does it take to begin rolling without slipping?

APPROACH The free-body diagram is shown in Fig. 10–39, with the ball moving to the right. The friction force does two things: it acts to slow down the translational motion of the CM; and it immediately acts to start the ball rotating clockwise. This may seem like a hard problem, but just watch.

SOLUTION Newton's second law for translation of the ball's CM is

$$Ma_{CM} = \Sigma F = -F_{fr} = -\mu_k F_N = -\mu_k Mg,$$

where μ_k is the coefficient of kinetic friction because the ball is sliding. Thus the ball's acceleration is $a_{CM} = -\mu_k g$, which means the ball slows down. The velocity of the CM is

$$v_{CM} = v_0 + a_{CM}t = v_0 - \mu_k gt. \qquad \textbf{(i)}$$

Next we apply Newton's second law for rotation about the CM, $I_{CM}\alpha_{CM} = \Sigma\tau_{CM}$:

$$\tfrac{2}{5}Mr_0^2\alpha_{CM} = F_{fr}r_0 = \mu_k F_N r_0 = \mu_k Mgr_0.$$

The angular acceleration of the ball while it is slipping is thus $\alpha_{CM} = 5\mu_k g/2r_0$, which is constant. Then the angular velocity of the slipping ball is (Eq. 10–9a)

$$\omega_{CM} = \omega_0 + \alpha_{CM}t = 0 + \frac{5\mu_k gt}{2r_0}. \qquad \textbf{(ii)}$$

The ball starts rolling immediately after it touches the ground, but it rolls and slips at the same time to begin with. It eventually stops slipping, and then rolls without slipping. The condition for *rolling without slipping* is that

$$v_{CM} = \omega_{CM}r_0,$$

which is Eq. 10–23, and is *not* valid if there is slipping. Thus rolling without slipping begins at a time $t = t_1$ when $v_{CM} = \omega_{CM}r_0$, which we substitute into Eq. (i) $[\omega_{CM}r_0 = v_0 - \mu_k gt_1]$ and then use Eq. (ii) to find

$$\frac{5\mu_k gt_1}{2r_0}r_0 = v_0 - \mu_k gt_1$$

so

$$t_1 = \frac{2v_0}{7\mu_k g}.$$

NOTE Kinetic friction slows the ball until t_1, when v_{CM} becomes equal to $\omega_{CM}r_0$, after which static friction acts.

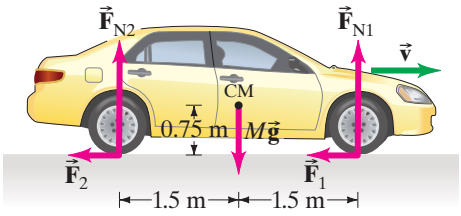
CAUTION
Rolling with slipping

CAUTION
Rolling without slipping

PHYSICS APPLIED
Braking distribution of a car

EXAMPLE 10–22 **ESTIMATE** **Braking a car.** When the brakes of a car are applied, the front of the car dips down a bit; and the force on the front tires is greater than on the rear tires. To see why, estimate the magnitude of the normal forces, F_{N1} and F_{N2}, on the front and rear tires of the car shown in Fig. 10–40 when the car brakes and decelerates at a rate $a = 0.50\,g$. The car has mass $M = 1200$ kg, the distance between the front and rear axles is 3.0 m, and its CM (where the force of gravity acts) is midway between the axles 75 cm above the ground.

FIGURE 10–40 Forces on a braking car (Example 10–22).



APPROACH Figure 10–40 is the free-body diagram showing all the forces on the car. F_1 and F_2 are the frictional forces that decelerate the car. We let F_1 be the sum of the forces on both front tires, and F_2 likewise for the two rear tires. F_{N1} and F_{N2} are the normal forces the road exerts on the tires and, for our estimate, we assume the static friction force acts the same for all the tires, so that F_1 and F_2 are proportional respectively to F_{N1} and F_{N2}:

$$F_1 = \mu F_{N1} \quad \text{and} \quad F_2 = \mu F_{N2}.$$

The forces F_1 and F_2 are due to static friction. But μ is less than the static coefficient of friction μ_s $(F_{fr} \le \mu_s F_N)$, unless the car is just about to skid.

SOLUTION The friction forces F_1 and F_2 decelerate the car, so Newton's second law gives (we are given $a = 0.50g$)

$$F_1 + F_2 = Ma = 0.50Mg = (1200\,\text{kg})(0.50)(9.8\,\text{m/s}^2) = 5900\,\text{N}. \quad \textbf{(i)}$$

While the car is braking its motion is only translational, so the net torque on the car is zero. If we calculate the torques about the CM as axis, the forces F_1, F_2, and F_{N2} all act to rotate the car clockwise, and only F_{N1} acts to rotate it counterclockwise; so the torque due to F_{N1} must balance the other three torques. Hence, F_{N1} must be significantly greater than F_{N2}. Mathematically, we have for the torques calculated about the CM:

$$(1.5\,\text{m})F_{N1} - (1.5\,\text{m})F_{N2} - (0.75\,\text{m})F_1 - (0.75\,\text{m})F_2 = 0.$$

Since F_1 and F_2 are proportional to F_{N1} and F_{N2} $(F_1 = \mu F_{N1},\ F_2 = \mu F_{N2})$, we can write this as

$$(1.5\,\text{m})(F_{N1} - F_{N2}) - (0.75\,\text{m})(\mu)(F_{N1} + F_{N2}) = 0. \quad \textbf{(ii)}$$

Also, since the car does not accelerate vertically, we have

$$Mg = F_{N1} + F_{N2} = \frac{F_1 + F_2}{\mu}. \quad \textbf{(iii)}$$

Comparing Eq. (iii) to Eq. (i), we see that $\mu = (F_1 + F_2)/Mg = a/g = 0.50$. Now we solve Eq. (ii) for F_{N1} and use $\mu = 0.50$ to obtain

$$F_{N1} = F_{N2}\left(\frac{2 + \mu}{2 - \mu}\right) = \frac{5}{3}F_{N2}. \quad \textbf{(iv)}$$

Thus F_{N1} is $1\frac{2}{3}$ times greater than F_{N2}. (The arrows in Fig. 10–40 representing $\vec{\mathbf{F}}_{N1}$ and $\vec{\mathbf{F}}_{N2}$ should thus be of different lengths.) Actual magnitudes are determined from Eqs. (iii) and (i):

$$F_{N1} + F_{N2} = \frac{F_1 + F_2}{\mu} = \frac{5900\,\text{N}}{0.50} = 11{,}800\,\text{N},$$

and from (iv)

$$F_{N1} + F_{N2} = \tfrac{5}{3}F_{N2} + F_{N2} = (\tfrac{5}{3} + 1)F_{N2} = \tfrac{8}{3}F_{N2}.$$

So $F_{N2} = \frac{3}{8}(F_{N1} + F_{N2}) = \frac{3}{8}(11{,}800\,\text{N}) = 4400\,\text{N}$ and $F_{N1} = 11{,}800\,\text{N} - F_{N2} = 11{,}800\,\text{N} - 4400\,\text{N} = 7400\,\text{N}$.

NOTE Because the force on the front tires is generally greater than on the rear tires, cars are often designed with larger brake pads on the front wheels than on the rear. Or, to say it another way, if the brake pads are equal, the front ones wear out a lot faster.

*10–10 Why Does a Rolling Sphere Slow Down? [Neglected Aspects]

A sphere of mass M and radius r_0 rolling on a horizontal flat surface eventually comes to rest. What force causes it to come to rest? You might think it is friction, but when you examine the problem, a paradox seems to arise.

Suppose a sphere is rolling to the right as shown in Fig. 10–41, and is slowing down. By Newton's second law, $\Sigma\vec{\mathbf{F}} = M\vec{\mathbf{a}}$, there must be a force $\vec{\mathbf{F}}$ (presumably frictional) acting to the left as shown, so that the acceleration $\vec{\mathbf{a}}$ will also point to the left and v will decrease. Curiously enough, though, if we now look at the torque equation (calculated about the center of mass), $\Sigma\tau_{CM} = I_{CM}\alpha$, we see that the force $\vec{\mathbf{F}}$ acts to increase the angular acceleration α, and thus to *increase* the velocity of the sphere. Thus the paradox. The force $\vec{\mathbf{F}}$ acts to decelerate the sphere if we look at the translational motion, but speeds it up if we look at the rotational motion.

The resolution of this apparent paradox is that other forces are acting on the sphere: gravity, $M\vec{\mathbf{g}}$, and the normal force $\vec{\mathbf{F}}_N$ $(= -M\vec{\mathbf{g}})$. These act vertically and hence do not affect the horizontal translational motion.

CAUTION
Paradox?

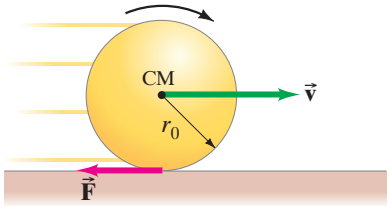


FIGURE 10–41 Sphere rolling to the right.

CAUTION

Real objects are not perfectly rigid

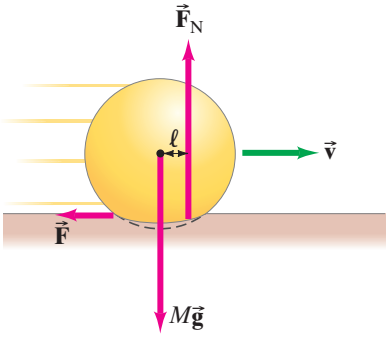


FIGURE 10–42 The normal force, $\vec{\mathbf{F}}_N$, exerts a torque that slows down the sphere. The deformation of the sphere and the surface it moves on have been exaggerated for detail.

If we assume the sphere and plane are rigid, so the sphere is in contact at only one point, these forces $M\vec{\mathbf{g}}$ and $\vec{\mathbf{F}}_N$ could give rise to no torques about the CM either, since they would act through the CM.

The only recourse we have to resolve the paradox is to give up our *idealization* that the objects are *rigid*. In fact, all objects are deformable to some extent. Our sphere flattens slightly and the level surface also acquires a slight depression where the two are in contact. There is an *area* of contact, not a point. Hence there can be a torque acting in this area of contact in the opposite direction to the torque associated with $\vec{\mathbf{F}}$, and thus acts to slow down the rotation of the sphere. This torque is associated with the normal force $\vec{\mathbf{F}}_N$ that the table exerts on the sphere over the whole area of contact. The net effect is that we can consider $\vec{\mathbf{F}}_N$ acting vertically a distance ℓ in front of the CM as shown in Fig. 10–42 (where the deformation is greatly exaggerated).

Is it reasonable that the normal force $\vec{\mathbf{F}}_N$ should effectively act in *front* of the CM as shown in Fig. 10–42? Yes. The sphere is rolling, and the leading edge strikes the surface with a slight impulse. The table therefore pushes upward a bit more strongly on the front part of the sphere than it would if the sphere is at rest. At the back part of the area of contact, the rolling sphere is starting to move away (upward) and so the table is not in as good contact and pushes upward on it less strongly than if the sphere is at rest. The table pushing up more strongly on the front part of the area of contact gives rise to the necessary torque and justifies the effective acting point of $\vec{\mathbf{F}}_N$ being in front of the CM.

When other forces are present, the tiny torque τ_N due to $\vec{\mathbf{F}}_N$ can usually be ignored. For example, when a sphere or cylinder rolls down an incline, the force of gravity has far more influence than τ_N, so the latter can be ignored. For many purposes (but not all), we can assume a hard sphere is in contact with a hard surface at essentially one point.

Summary

When a rigid object rotates about a fixed axis, each point of the object moves in a circular path. Lines drawn perpendicularly from the rotation axis to different points in the object all sweep out the same angle θ in any given time interval.

Angles are conveniently measured in **radians**. One radian is the angle subtended by an arc whose length is equal to the radius R, or (since the circumference of a circle is $2\pi R$)

$$2\pi \text{ rad} = 360^\circ \quad \text{so} \quad 1 \text{ rad} \approx 57.3^\circ.$$

All parts of a rigid object rotating about a fixed axis have the same **angular velocity** ω and the same **angular acceleration** α at any instant, where

$$\omega = \frac{d\theta}{dt} \qquad \textbf{(10–2b)}$$

and

$$\alpha = \frac{d\omega}{dt}. \qquad \textbf{(10–3b)}$$

The units of ω and α are rad/s and rad/s^2.

The linear velocity v and acceleration a of a point located a distance R from a fixed axis of rotation are related to the angular quantities by

$$v = R\omega \qquad \textbf{(10–4)}$$
$$a_{\tan} = R\alpha \qquad \textbf{(10–5)}$$
$$a_R = \omega^2 R = v^2/R \qquad \textbf{(10–6)}$$

where $a_{\tan}$ and a_R are the tangential and radial (centripetal) components of the linear acceleration. The frequency f and period T are related to ω (rad/s) by

$$\omega = 2\pi f \qquad \textbf{(10–7)}$$
$$T = 1/f. \qquad \textbf{(10–8)}$$

Angular velocity and angular acceleration are vectors. For a rigid object rotating about a fixed axis, both $\vec{\boldsymbol{\omega}}$ and $\vec{\boldsymbol{\alpha}}$ point along the rotation axis. The direction of $\vec{\boldsymbol{\omega}}$ is given by the **right-hand rule**.

If a rigid object undergoes uniformly accelerated rotational motion (α = constant), equations analogous to those for linear motion are valid:

$$\omega = \omega_0 + \alpha t, \qquad \theta = \omega_0 t + \tfrac{1}{2}\alpha t^2,$$
$$\omega^2 = \omega_0^2 + 2\alpha\theta, \qquad \bar{\omega} = \frac{\omega + \omega_0}{2}. \qquad \textbf{(10–9)}$$

The **torque** τ due to a force $\vec{\mathbf{F}}$ exerted on a rigid object is equal to

$$\tau = R_\perp F = RF_\perp = RF\sin\theta, \qquad \textbf{(10–10)}$$

where $R_\perp$, called the **lever arm**, is the perpendicular distance from the axis of rotation to the line along which the force acts, and θ is the angle between $\vec{\mathbf{F}}$ and R.

The rotational equivalent of Newton's second law is

$$\Sigma\tau = I\alpha, \qquad \textbf{(10–14)}$$

where $I = \Sigma m_i R_i^2$ is the **moment of inertia** of the object about the axis of rotation. This relation is valid for a rigid object rotating about an axis fixed in an inertial reference frame, or when τ, I, and α are calculated about the center of mass of an object even if the CM is moving. I depends not only on the mass of the object but also on how the mass is distributed relative to the axis of rotation. For a uniform solid cylinder of radius R_0 and mass M, I has the form $I = \frac{1}{2}MR_0^2$ (see Fig. 10–21).

The **rotational kinetic energy** of an object rotating about a fixed axis with angular velocity ω is

$$K = \tfrac{1}{2} I\omega^2. \qquad \textbf{(10–19)}$$

For an object undergoing both translational and rotational motion, the total kinetic energy is the sum of the translational kinetic energy of the object's CM plus the rotational kinetic energy of the object about its CM:

$$K_{\text{tot}} = \tfrac{1}{2} M v_{\text{CM}}^2 + \tfrac{1}{2} I_{\text{CM}}\omega^2 \qquad \textbf{(10–24)}$$

as long as the rotation axis is fixed in direction.

The following Table summarizes angular (or rotational) quantities, comparing them to their translational analogs.

Translation	Rotation	Connection
x	θ	$x = R\theta$
v	ω	$v = R\omega$
a	α	$a = R\alpha$
m	I	$I = \Sigma mR^2$
F	τ	$\tau = RF\sin\theta$
$K = \frac{1}{2}mv^2$	$K = \frac{1}{2}I\omega^2$	
$W = Fd$	$W = \tau\theta$	
$\Sigma F = ma$	$\Sigma\tau = I\alpha$	

Questions

1. A bicycle **odometer** (which counts revolutions and is calibrated to report distance traveled) is attached near the wheel hub and is calibrated for 27-inch wheels. What happens if you use it on a bicycle with 24-inch wheels?
2. Suppose a disk rotates at constant angular velocity. (*a*) Does a point on the rim have radial and/or tangential acceleration? (*b*) If the disk's angular velocity increases uniformly, does the point have radial and/or tangential acceleration? (*c*) For which cases would the magnitude of either component of linear acceleration change?
3. Could a nonrigid object be described by a single value of the angular velocity ω? Explain.
4. Why are Eqs. 10–4 and 10–5 valid for radians but not for revolutions or degrees? (Read Section 10–1 and follow the derivations carefully.)
5. The angular velocity of a wheel rotating on a horizontal axle points west. In what direction is the linear velocity of a point on the top of the wheel? If the angular acceleration points east, describe the tangential linear acceleration of this point at the top of the wheel. Is the angular speed increasing or decreasing?
6. In what direction is the Earth's angular velocity vector as it rotates daily about its axis, north or south?
7. Can a small force ever exert a greater torque than a larger force? Explain.
8. Why is it more difficult to do a *sit-up* with your hands behind your head than when your arms are stretched out in front of you? A diagram may help you to answer this.
9. If the net force on a system is zero, is the net torque also zero? If the net torque on a system is zero, is the net force zero? Explain and give examples.
10. Mammals that depend on being able to run fast have slender lower legs with flesh and muscle concentrated high, close to the body (Fig. 10–43). On the basis of rotational dynamics, explain why this distribution of mass is advantageous.



FIGURE 10–43 Question 10. A gazelle.

11. This book has three symmetry axes through its center, all mutually perpendicular. The book's moment of inertia would be smallest about which of the three? Explain.
12. Can the mass of a rigid object be considered concentrated at its CM for rotational motion? Explain.
13. The moment of inertia of a rotating solid disk about an axis through its CM is $\frac{1}{2}MR_0^2$ (Fig. 10–21c). Suppose instead that a parallel axis of rotation passes through a point on the edge of the disk. Will the moment of inertia be the same, larger, or smaller? Explain why.
14. Why do *tightrope walkers* (Fig. 10–44) carry a long, narrow rod?



FIGURE 10–44 Question 14.

15. Two inclines have the same height but make different angles with the horizontal. The same steel ball rolls without slipping down each incline. On which incline will the speed of the ball at the bottom be greater? Explain.
16. Two spheres look identical and have the same mass. However, one is hollow and the other is solid. Describe an experiment to determine which is which.
17. A sphere and a cylinder have the same radius and the same mass. They start from rest at the top of an incline. (*a*) Which reaches the bottom first? (*b*) Which has the greater speed at the bottom? (*c*) Which has the greater total kinetic energy at the bottom? (*d*) Which has the greater rotational kinetic energy at the bottom? Explain your answers.
18. Two solid spheres simultaneously start rolling (from rest) down an incline. One sphere has twice the radius and twice the mass of the other. Which reaches the bottom of the incline first? Which has the greater speed there? Which has the greater total kinetic energy at the bottom?

MisConceptual Questions

1. Bonnie sits on the outer rim of a merry-go-round, and Jill sits midway between the center and the rim. The merry-go-round makes one complete revolution every 2 seconds. Jill's linear velocity is:
(*a*) the same as Bonnie's.
(*b*) twice Bonnie's.
(*c*) half of Bonnie's.
(*d*) one-quarter of Bonnie's.
(*e*) four times Bonnie's.

2. The rigid disk in Fig. 10–45 rotates about point A. Which statement below is true?
(*a*) Point P has a higher angular velocity than point Q.
(*b*) Point P has a lower angular velocity than point Q.
(*c*) Point P has the same angular velocity as point Q.
(*d*) Point P has the same angular velocity as point Q if the angular velocity is measured in rad/s but not if it is measured in degrees/s.

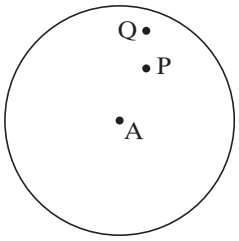


FIGURE 10–45
MisConceptual Question 2.

3. An object at rest begins to rotate with a constant angular acceleration. If this object rotates through an angle θ in time t, through what angle did it rotate in the time $\frac{1}{2}t$?
(*a*) $\frac{1}{2}\theta$. (*b*) $\frac{1}{4}\theta$. (*c*) θ. (*d*) 2θ. (*e*) 4θ.

4. An accurate car speedometer, which is supposed to read the linear speed of the car, uses a device that actually measures the angular speed of the tires. If larger-diameter tires are mounted on the car instead, how will that affect the speedometer reading? The speedometer
(*a*) will still read the speed accurately.
(*b*) will read low.
(*c*) will read high.

5. The solid dot shown in Fig. 10–46 is a pivot point. The board can rotate about the pivot. Which force shown exerts the largest magnitude torque on the board?

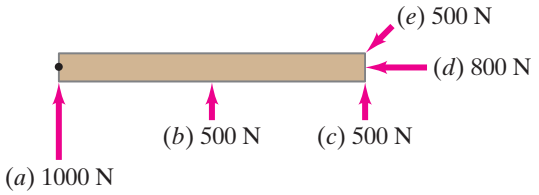


FIGURE 10–46 MisConceptual Questions 5 and 6.

6. Which force (or forces) in Fig. 10–46 exerts the smallest torque on the board?

7. Consider a force $F = 80\,\text{N}$ applied to a beam as shown in Fig. 10–47. The length of the beam is $\ell = 5.0\,\text{m}$, and $\theta = 37°$, so that $x = 3.0\,\text{m}$ and $y = 4.0\,\text{m}$. Of the following expressions, which ones give the correct torque produced by the force $\vec{\mathbf{F}}$ around point P?
(*a*) 80 N.
(*b*) (80 N)(5.0 m).
(*c*) (80 N)(5.0 m)(sin 37°).
(*d*) (80 N)(4.0 m).
(*e*) (80 N)(3.0 m).
(*f*) (48 N)(5.0 m).
(*g*) (48 N)(4.0 m)(sin 37°).

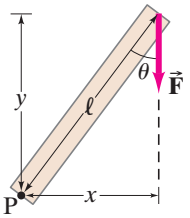


FIGURE 10–47
MisConceptual Question 7.

8. Two spheres have the same radius and equal mass. One sphere is solid, and the other is hollow and made of a denser material. Which one has the bigger moment of inertia about an axis through its center?
(*a*) The solid one.
(*b*) The hollow one.
(*c*) Both the same.

9. Two wheels, having the same radius and mass, rotate at the same angular velocity (Fig. 10–48). One wheel is made with spokes so nearly all the mass is at the rim. The other is a solid disk. How do their rotational kinetic energies compare?
(*a*) They are nearly the same.
(*b*) The wheel with spokes has about twice the kinetic energy.
(*c*) The wheel with spokes has more kinetic energy, but not twice as much.
(*d*) The solid wheel has about twice the kinetic energy.
(*e*) The solid wheel has more kinetic energy, but not twice as much.

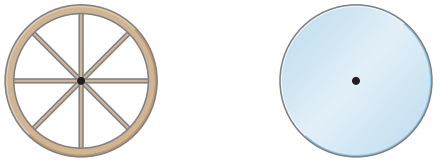


FIGURE 10–48 MisConceptual Question 9.

10. Four identical small masses are connected in a flat perfect square (Fig. 10–49). Rank the relative rotational inertias (I_A, I_B, I_C) about the three axes of rotation shown. Axes A and B are in the plane of the square, and axis C is perpendicular to the plane, through mass m_1.
(*a*) $I_A < I_B < I_C$.
(*b*) $I_C < I_B < I_A$.
(*c*) $I_A = I_B = I_C$.
(*d*) $I_B < I_A < I_C$.
(*e*) $I_C < I_A < I_B$.

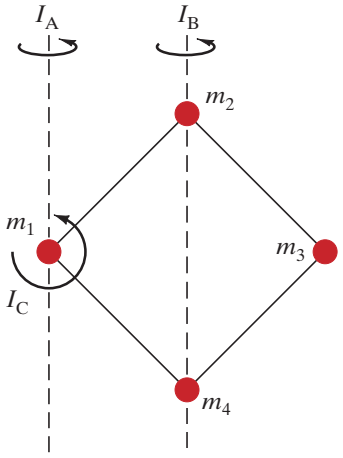


FIGURE 10–49
MisConceptual Question 10.

11. If you used 1000 J of energy to throw a ball, would it travel faster if you threw the ball (ignoring air resistance)
(*a*) so that it was also rotating?
(*b*) so that it wasn't rotating?
(*c*) It makes no difference.

12. A small solid sphere and a small thin hoop are rolling along a horizontal surface with the same translational speed when they encounter a 20° rising slope. If these two objects roll up the slope without slipping, which will rise farther up the slope?
(*a*) The sphere.
(*b*) The hoop.
(*c*) Both the same.
(*d*) More information about the objects' mass and diameter is needed.

13. Two uniform spheres simultaneously start rolling (from rest) down an incline. One sphere has twice the radius and twice the mass of the other. (*i*) Which reaches the bottom of the incline first? (*ii*) Which has the greater total kinetic energy at the bottom?
(*a*) The same; the same.
(*b*) The same; the larger one.
(*c*) The larger one; the larger one.
(*d*) The smaller one; the larger one.

14. Starting from rest, a solid sphere rolls without slipping down an inclined plane. At the bottom of the incline what does the angular velocity depend upon? Choose all that apply.
(*a*) The height of the incline.
(*b*) The length of the incline.
(*c*) The mass of the sphere.
(*d*) The radius of the sphere.

Problems

10–1 Angular Quantities

1. (I) Express the following angles in radians: (*a*) 45.0°, (*b*) 60.0°, (*c*) 90.0°, (*d*) 360.0°, and (*e*) 445°. Give as numerical values and as fractions of π.

2. (I) The Sun subtends an angle of about 0.5° to us on Earth, 150 million km away. Estimate the radius of the Sun.

3. (I) A laser beam is directed at the Moon, 380,000 km from Earth. The beam diverges at an angle θ (Fig. 10–50) of 1.4×10^{-5} rad. What diameter spot will it make on the Moon?

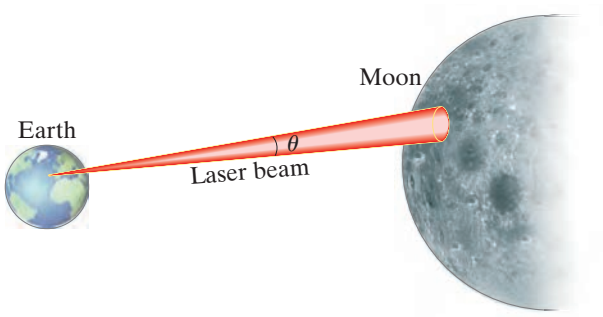


FIGURE 10–50 Problem 3.

4. (I) The blades in a blender rotate at a rate of 8500 rpm (rpm = revolutions per minute = rev/min). When the motor is turned off during operation, the blades slow to rest in 4.0 s. What is the angular acceleration as the blades slow down?

5. (II) The platter of the **hard drive** of a computer rotates at 7200 rpm (rpm = revolutions per minute = rev/min). (*a*) What is the angular velocity (rad/s) of the platter? (*b*) If the reading head of the drive is located 3.00 cm from the rotation axis, what is the linear speed of the point on the platter just below it? (*c*) If a single bit requires 0.50 μm of length along the direction of motion, how many bits per second can the writing head write when it is 3.00 cm from the axis?

6. (II) A child rolls a ball on a level floor 3.1 m to another child. If the ball makes 12.0 revolutions, what is its diameter?

7. (II) (*a*) A grinding wheel 0.35 m in diameter rotates at 2400 rpm. Calculate its angular velocity in rad/s. (*b*) What are the linear speed and radial acceleration of a point on the edge of the grinding wheel?

8. (II) A bicycle with tires 68 cm in diameter travels 8.6 km. How many revolutions do the wheels make?

9. (II) Calculate the average angular velocity (*a*) of a clock's second hand, (*b*) its minute hand, and (*c*) its hour hand. State in rad/s. (*d*) What is the average angular acceleration in each case?

10. (II) A rotating merry-go-round makes one complete revolution in 4.0 s (Fig. 10–51). (*a*) What is the linear speed of a child seated 1.2 m from the center? (*b*) What is her acceleration?



FIGURE 10–51 Problem 10.

11. (II) **Total solar eclipses** can happen on Earth because of amazing coincidences: for one, the rare near-perfect alignment of Earth, Moon, and Sun. Using the information inside the front cover, calculate the angular diameters (in radians) of the Sun and the Moon, as seen from Earth, and then comment about full vs. partial eclipses.

12. (II) What is the linear speed, due to the Earth's rotation, of a point (*a*) on the equator, (*b*) on the Arctic Circle (latitude 66.5° N), and (*c*) at a latitude of 48.0° N?

13. (II) Calculate the angular velocity of the Earth (*a*) in its orbit around the Sun, and (*b*) about its axis.

14. (II) How fast (in rpm) must a centrifuge rotate if a particle 8.0 cm from the axis of rotation is to experience an acceleration of 100,000 *g*'s?

15. (II) A 61-cm-diameter wheel accelerates uniformly about its center from 120 rpm to 280 rpm in 5.0 s. Determine (*a*) its angular acceleration, and (*b*) the radial and tangential components of the linear acceleration of a point on the edge of the wheel 2.0 s after it has started accelerating.

16. (II) In traveling to the Moon, astronauts aboard the *Apollo* spacecraft put the spacecraft into a slow rotation to distribute the Sun's energy evenly (so one side would not become too hot). At the start of their trip, they accelerated for 12 minutes from no rotation to 1.0 revolution per minute which they then maintained. Think of the spacecraft as a cylinder with a diameter of 8.5 m rotating about its cylindrical axis. Determine (*a*) the angular acceleration, and (*b*) the radial and tangential components of the linear acceleration of a point on the skin of the ship 6.0 min after it started this acceleration.

17. (II) A turntable of radius R_1 is turned by a circular rubber roller of radius R_2 in contact with it at their outer edges. What is the ratio of their angular velocities, ω_1/ω_2?

10–2 Vector Nature of $\vec{\omega}$ and $\vec{\alpha}$

18. (II) The axle of a wheel is mounted on supports that rest on a rotating turntable as shown in Fig. 10–52. The wheel has angular velocity $\omega_1 = 48.0\,\text{rad/s}$ about its axle, and the turntable has angular velocity $\omega_2 = 35.0\,\text{rad/s}$ about a vertical axis. (Note arrows showing these motions in the figure.) (*a*) What are the directions of $\vec{\boldsymbol{\omega}}_1$ and $\vec{\boldsymbol{\omega}}_2$ at the instant shown? (*b*) What is the resultant angular velocity of the wheel, as seen by an outside observer, at the instant shown? Give the magnitude and direction. (*c*) What is the magnitude and direction of the angular acceleration of the wheel at the instant shown? Take the z axis vertically upward and the direction of the axle at the moment shown to be the x axis pointing to the right.

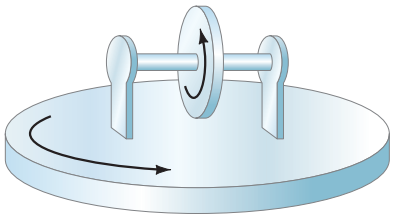


FIGURE 10–52 Problem 18.

10–3 Constant Angular Acceleration

19. (I) A centrifuge accelerates uniformly from rest to 15,000 rpm in 220 s. Through how many revolutions did it turn in this time?

20. (I) Pilots can be tested for the stresses of flying high-speed jets in a whirling "human centrifuge," which takes 1.0 min to turn through 26 complete revolutions before reaching its final speed. (*a*) What was its angular acceleration (assumed constant), and (*b*) what was its final angular speed in rpm?

21. (II) A cooling fan is turned off when it is running at 780 rev/min. It turns 1250 revolutions before it comes to a stop. (*a*) What was the fan's angular acceleration, assumed constant? (*b*) How long did it take the fan to come to a complete stop?

22. (II) A wheel 34 cm in diameter accelerates uniformly from 240 rpm to 360 rpm in 6.8 s. How far will a point on the edge of the wheel have traveled in this time?

23. (II) The tires of a car make 75 revolutions as the car reduces its speed uniformly from 95 km/h to 55 km/h. The tires have a diameter of 0.80 m. (*a*) What was the angular acceleration of the tires? If the car continues to decelerate at this rate, (*b*) how much more time is required for it to stop, and (*c*) how far does it go before stopping?

24. (II) A small rubber wheel is used to drive a large pottery wheel. The two wheels are mounted so that their circular edges touch. The small wheel has a radius of 2.0 cm and accelerates at the rate of $6.8\,\text{rad/s}^2$, and it is in contact with the pottery wheel (radius 27.0 cm) without slipping. Calculate (*a*) the angular acceleration of the pottery wheel, and (*b*) the time it takes the pottery wheel to reach its required speed of 65 rpm.

25. (II) The angle through which a rotating wheel has turned in time t is given by $\theta = 9.5\,t - 13.0\,t^2 + 1.6\,t^4$, where θ is in radians and t in seconds. Determine an expression (*a*) for the instantaneous angular velocity ω and (*b*) for the instantaneous angular acceleration α. (*c*) Evaluate ω and α at $t = 3.0\,\text{s}$. (*d*) What is the average angular velocity, and (*e*) the average angular acceleration between $t = 2.0\,\text{s}$ and $t = 3.0\,\text{s}$?

26. (II) The angular acceleration of a wheel, as a function of time, is $\alpha = 4.2\,t^2 - 9.0\,t$, where α is in rad/s^2 and t in seconds. If the wheel starts from rest ($\theta = 0$, $\omega = 0$, at $t = 0$), determine a formula for (*a*) the angular velocity ω and (*b*) the angular position θ, both as a function of time. (*c*) Evaluate ω and θ at $t = 2.0\,\text{s}$.

10–4 Torque

27. (I) A 58-kg person riding a bike puts all her weight on each pedal when climbing a hill. The pedals rotate in a circle of radius 17 cm. (*a*) What is the maximum torque she exerts? (*b*) How could she exert more torque?

28. (I) Two blocks, each of mass m, are attached to the ends of a massless rod which pivots as shown in Fig. 10–53. Initially the rod is held in the horizontal position and then released. Calculate the magnitude and direction of the net torque on this system when it is first released.

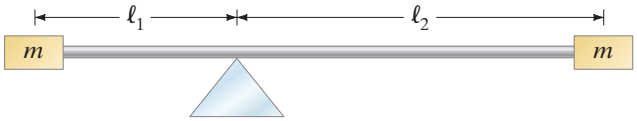


FIGURE 10–53 Problem 28.

29. (II) A person exerts a horizontal force of 37 N on the end of a door 96 cm wide. What is the magnitude of the torque if the force is exerted (*a*) perpendicular to the door and (*b*) at a 60.0° angle to the face of the door?

30. (II) Calculate the net torque about the axle of the wheel shown in Fig. 10–54. Assume that a friction torque of $0.60\,\text{m}\cdot\text{N}$ opposes the motion.

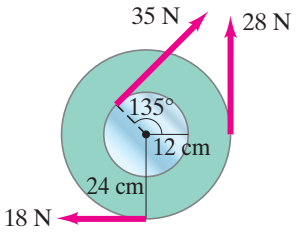


FIGURE 10–54 Problem 30.

31. (II) The bolts on the cylinder head of an engine require tightening to a torque of $95\,\text{m}\cdot\text{N}$. (*a*) If a wrench is 28 cm long, what force perpendicular to the wrench must the mechanic exert at its end? (*b*) If the six-sided bolt head is 15 mm across (Fig. 10–55), estimate the force applied near each of the six points by a wrench.

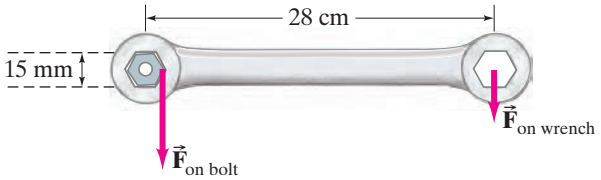


FIGURE 10–55 Problem 31.

32. (II) A wheel of diameter 24.0 cm is constrained to rotate in the xy plane, about the z axis, which passes through its center. A force $\vec{\mathbf{F}} = (-31.0\,\hat{\mathbf{i}} + 43.4\,\hat{\mathbf{j}})$ N acts at a point on the edge of the wheel that lies exactly on the x axis at a particular instant. What is the torque about the rotation axis at this instant?

33. (II) Determine the net torque on the 2.0-m-long uniform beam shown in Fig. 10–56. All forces are shown. Calculate about (*a*) point C, the CM, and (*b*) point P at one end.

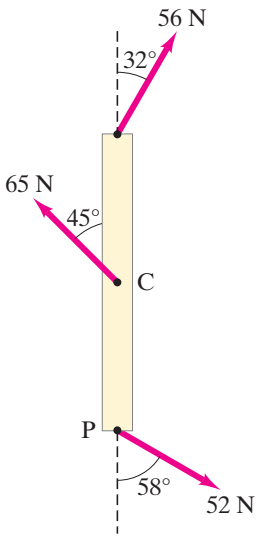


FIGURE 10–56
Problem 33.

10–5 and 10–6 Rotational Dynamics

34. (I) Estimate the moment of inertia of a bicycle wheel 67 cm in diameter. The rim and tire have a combined mass of 1.1 kg. The mass of the hub (at the center) can be ignored (why?).

35. (II) A merry-go-round accelerates from rest to 0.72 rad/s in 34 s. Assuming the merry-go-round is a uniform disk of radius 7.0 m and mass 31,000 kg, calculate the net torque required to accelerate it.

36. (II) An oxygen molecule consists of two oxygen atoms whose total mass is 5.3×10^{-26} kg and whose moment of inertia about an axis perpendicular to the line joining the two atoms, midway between them, is $1.9 \times 10^{-46}\ \text{kg} \cdot \text{m}^2$. From these data, estimate the effective distance between the atoms.

37. (II) A grinding wheel is a uniform cylinder with a radius of 8.50 cm and a mass of 0.380 kg. Calculate (*a*) its moment of inertia about its center, and (*b*) the applied torque needed to accelerate it from rest to 1950 rpm in 5.00 s. Take into account a frictional torque that has been measured to slow down the wheel from 1500 rpm to rest in 55.0 s.

38. (II) The forearm in Fig. 10–57 accelerates a 3.6-kg ball at $7.0\ \text{m/s}^2$ by means of the triceps muscle, as shown. Calculate (*a*) the torque needed, and (*b*) the force that must be exerted by the triceps muscle. Ignore the mass of the arm.

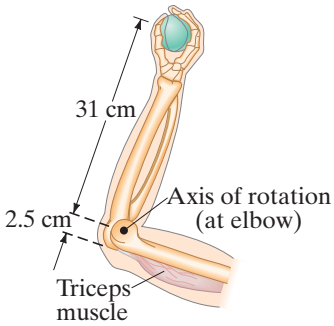


FIGURE 10–57
Problems 38 and 39.

39. (II) Assume that a 1.00-kg ball is thrown solely by the action of the forearm, which rotates about the elbow joint under the action of the triceps muscle, Fig. 10–57. The ball is accelerated uniformly from rest to 8.5 m/s in 0.38 s, at which point it is released. Calculate (*a*) the angular acceleration of the arm, and (*b*) the force required of the triceps muscle. Assume that the forearm has a mass of 3.7 kg and rotates like a uniform rod about an axis at its end.

40. (II) A softball player swings a bat, accelerating it from rest to 2.4 rev/s in a time of 0.20 s. Approximate the bat as a 0.90-kg uniform rod of length 0.95 m, and compute the torque the player applies to one end of it.

41. (II) A small 350-gram ball on the end of a thin, light rod is rotated in a horizontal circle of radius 1.4 m. Calculate (*a*) the moment of inertia of the ball about the center of the circle, and (*b*) the torque needed to keep the ball rotating at constant angular velocity if air resistance exerts a force of 0.020 N on the ball. Ignore air resistance on the rod and its moment of inertia.

42. (II) Calculate the moment of inertia of the array of point objects shown in Fig. 10–58 about (*a*) the *y* axis, and (*b*) the *x* axis. Assume $m = 2.2$ kg, $M = 3.2$ kg, and the objects are wired together by very light, rigid pieces of wire. The array is rectangular and is split through the middle by the *x* axis. (*c*) About which axis would it be harder to accelerate this array?

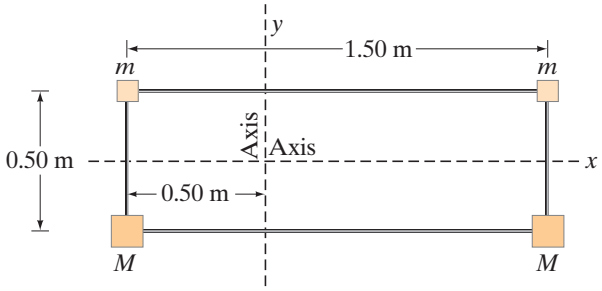


FIGURE 10–58 Problem 42.

43. (II) A potter is shaping a bowl on a potter's wheel rotating at constant angular velocity of 1.6 rev/s (Fig. 10–59). The friction force between her hands and the clay is 1.8 N total. (*a*) How large is her torque on the wheel, if the diameter of the bowl is 9.0 cm? (*b*) How long would it take for the potter's wheel to stop if the only torque acting on it is due to the potter's hands? The moment of inertia of the wheel and the bowl is $0.11\ \text{kg} \cdot \text{m}^2$.



FIGURE 10–59 Problem 43.

44. (II) A dad pushes tangentially on a small hand-driven merry-go-round and is able to accelerate it from rest to a frequency of 15 rpm in 10.0 s. Assume the merry-go-round is a uniform disk of radius 2.5 m and has a mass of 330 kg, and two children (each with a mass of 25 kg) sit opposite each other on the edge. Calculate the torque required to produce the acceleration, neglecting frictional torque. What force is required at the edge?

45. (II) Let us treat a helicopter rotor blade as a long thin rod, as shown in Fig. 10–60. (*a*) If each of the three rotor helicopter blades is 3.75 m long and has a mass of 135 kg, calculate the moment of inertia of the three rotor blades about the axis of rotation. (*b*) How much torque must the motor apply to bring the blades from rest up to a speed of 6.5 rev/s in 8.0 s?

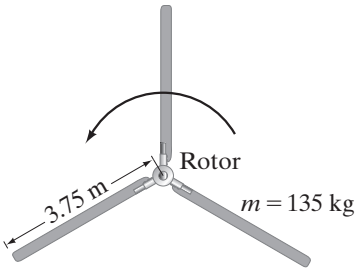


FIGURE 10–60 Problem 45.

46. (II) Suppose the force F_T in the cord hanging from the pulley of Example 10–10, Fig. 10–22, is given by the relation $F_T = 3.00\,t - 0.20\,t^2$ (newtons) where t is in seconds. If the pulley starts from rest, what is the linear speed of a point on its rim 9.0 s later? Ignore friction and use the moment of inertia, calculated in Example 10–10.

47. (II) Four equal masses M are spaced at equal intervals, ℓ, along a horizontal straight rod whose mass can be ignored. The system is to be rotated about a vertical axis passing through the mass at the left end of the rod and perpendicular to it. (*a*) What is the moment of inertia of the system about this axis? (*b*) What minimum force, applied to the farthest mass, will impart an angular acceleration α? (*c*) What is the direction of this force?

48. (II) To get a flat, uniform cylindrical satellite spinning at the correct rate, engineers fire four tangential rockets as shown in Fig. 10–61. Suppose that the satellite has a mass of 3600 kg and a radius of 4.0 m, and that the rockets each add a mass of 250 kg. What is the steady force required of each rocket if the satellite is to reach 28 rpm in 5.0 min, starting from rest?

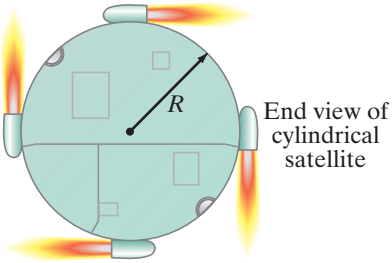


FIGURE 10–61 Problem 48.

49. (II) When discussing moments of inertia, especially for unusual or irregularly shaped objects, it is sometimes convenient to work with the **radius of gyration**, k. This radius is defined so that if all the mass of the object were concentrated at this distance from the axis, the moment of inertia would be the same as that of the original object. Thus, the moment of inertia of any object can be written in terms of its mass M and the radius of gyration as $I = Mk^2$. Determine the radius of gyration for each of the objects (hoop, cylinder, sphere, etc.) shown in Fig. 10–21.

50. (III) Two blocks are connected by a light string passing over a pulley of radius 0.15 m and moment of inertia I. The blocks move (towards the right) with an acceleration of 1.00 m/s^2 along their frictionless inclines (see Fig. 10–62). (*a*) Draw free-body diagrams for each of the two blocks and the pulley. (*b*) Determine F_{TA} and F_{TB}, the tensions in the two parts of the string. (*c*) Find the net torque acting on the pulley, and determine its moment of inertia, I.

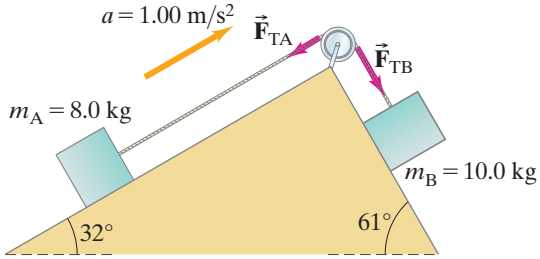


FIGURE 10–62 Problem 50.

51. (III) An *Atwood machine* consists of two masses, $m_A = 65$ kg and $m_B = 75$ kg, connected by a massless inelastic cord that passes over a pulley free to rotate, Fig. 10–63. The pulley is a solid cylinder of radius $R = 0.45$ m and mass 6.0 kg. (*a*) Determine the acceleration of each mass. (*b*) What % error would be made if the moment of inertia of the pulley is ignored? [*Hint*: The tensions F_{TA} and F_{TB} are not equal. We discussed the Atwood machine in Example 4–13, assuming $I = 0$ for the pulley.]

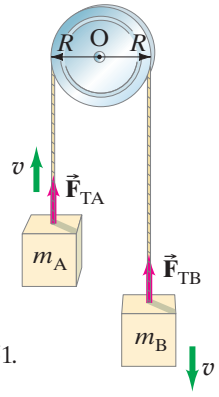


FIGURE 10–63 Problem 51. Atwood machine.

52. (III) A hammer thrower accelerates the hammer of mass 7.30 kg (Fig. 10–64) from rest within four full turns (revolutions) and releases it at a speed of 26.5 m/s. Assuming a uniform rate of increase in angular velocity and a horizontal circular path of radius 1.20 m, calculate (*a*) the angular acceleration, (*b*) the (linear) tangential acceleration, (*c*) the centripetal acceleration just before release, (*d*) the net force being exerted on the hammer by the athlete just before release, and (*e*) the angle of this force with respect to the radius of the circular motion. Ignore gravity.



FIGURE 10–64 Problem 52.

10–7 Determining Moments of Inertia

53. (I) Use the parallel-axis theorem to show that the moment of inertia of a thin rod about an axis perpendicular to the rod at one end is $I = \frac{1}{3}M\ell^2$, given that if the axis passes through the center, $I = \frac{1}{12}M\ell^2$ (Fig. 10–21f and g).

54. (II) Determine the moment of inertia of a 19-kg door around its hinges along one side. The door is 2.5 m high and 1.0 m wide. Ignore the thickness of the door.

55. (II) Two uniform solid spheres of mass M and radius r_0 are connected by a thin (massless) rod of length r_0 so that the centers are $3r_0$ apart. (*a*) Determine the moment of inertia of this system about an axis perpendicular to the rod at its center (parallel axis theorem). (*b*) What is the percentage error if the masses of each sphere are assumed to be concentrated at their centers and a very simple calculation made?

56. (II) A ball of mass M and radius r_1 on the end of a thin massless rod is rotated in a horizontal circle of radius R_0 about an axis of rotation AB, as shown in Fig. 10–65. (*a*) Assuming the mass of the ball to be concentrated at its center of mass, calculate its moment of inertia about AB. (*b*) Using the parallel-axis theorem and considering the finite radius of the ball, calculate the moment of inertia of the ball about AB. (*c*) Calculate the percentage error introduced by the point mass approximation for $r_1 = 9.25$ cm and $R_0 = 1.0$ m. (*d*) Repeat part *c* assuming $r_1 = 8.0$ cm, and comment.

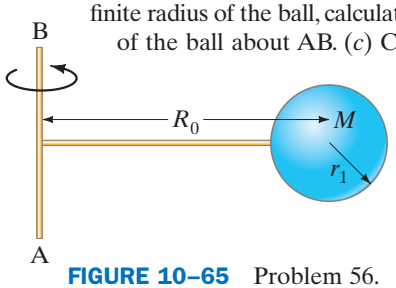


FIGURE 10–65 Problem 56.

57. (II) A thin 7.0-kg wheel of radius 34 cm is a uniform disk and is weighted to one side by a 1.50-kg weight, small in size, placed 24 cm from the center of the wheel. Calculate (*a*) the position of the center of mass of the weighted wheel and (*b*) the moment of inertia about an axis through its CM, perpendicular to its face.

58. (III) Integrate to derive the formula for the moment of inertia of a uniform thin rod of length ℓ about an axis through its center, perpendicular to the rod (see Fig. 10–21f).

59. (III) (*a*) Integrate to obtain the formula given in Fig. 10–21h for the moment of inertia of a uniform, flat, rectangular plate of dimensions $\ell \times w$ about an axis through its center, perpendicular to the plate. (*b*) What is the moment of inertia about each of the axes through the center that are parallel to the edges of the plate?

60. (III) Determine the moment of inertia of a uniform solid cone whose base has radius R_0, height L and mass M. The axis of rotation (z) is the symmetry axis perpendicular to the base, Fig. 10–66. [*Hint*: Think of the cone as a stack of infinitesimally thin disks of mass dm, radius R, and thickness dz.]

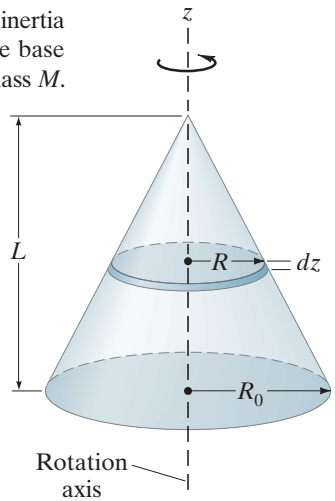


FIGURE 10–66 Problem 60.

10–8 Rotational Kinetic Energy

61. (I) An automobile engine develops a torque of 245 m · N at 3150 rpm. What is the horsepower of the engine?

62. (I) A centrifuge rotor has a moment of inertia of 3.25×10^{-2} kg · m^2. How much energy is required to bring it from rest to 8750 rpm?

63. (II) Estimate the kinetic energy of the Earth with respect to the Sun as the sum of two terms, (*a*) that due to its daily rotation about its axis, and (*b*) that due to its yearly revolution about the Sun. [Assume the Earth is a uniform sphere with mass $= 6.0 \times 10^{24}$ kg, radius $= 6.4 \times 10^6$ m, and is 1.5×10^8 km from the Sun.]

64. (II) A rotating uniform cylindrical platform of mass 220 kg and radius 5.5 m slows down from 3.8 rev/s to rest in 18 s when the driving motor is disconnected. Estimate the power output of the motor (hp) required to maintain a steady speed of 3.8 rev/s.

65. (II) A merry-go-round has a mass of 1240 kg and a radius of 7.50 m. How much net work is required to accelerate it from rest to a rotation rate of 1.00 revolution per 7.00 s? Assume it is a solid cylinder.

66. (II) A uniform thin rod of length ℓ and mass M is suspended freely from one end. It is pulled to the side an angle θ and released. If friction can be ignored, what is its angular velocity, and the speed of its free end, at the lowest point?

67. (II) Two masses, $m_A = 35.0$ kg and $m_B = 41.0$ kg, are connected by a rope that hangs over a pulley (as in Fig. 10–67). The pulley is a uniform cylinder of radius $R = 0.311$ m and mass 3.1 kg. Initially m_A is on the ground and m_B rests 2.5 m above the ground. If the system is released, use conservation of energy to determine the speed of m_B just before it strikes the ground. Assume the pulley bearing is frictionless.

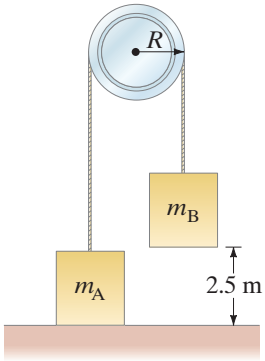


FIGURE 10–67 Problem 67.

68. (III) A 2.30-m-long pole is balanced vertically on its tip. It starts to fall and its lower end does not slip. What will be the speed of the upper end of the pole just before it hits the ground? [*Hint*: Use conservation of energy.]

10–9 Rotational Plus Translational Motion

69. (I) Calculate the translational speed of a uniform solid cylinder when it reaches the bottom of an incline 6.50 m high. Assume it starts from rest and rolls without slipping.

70. (I) A bowling ball of mass 7.3 kg and radius 9.0 cm rolls without slipping down a lane at 3.7 m/s. Calculate its total kinetic energy.

71. (II) A uniform solid sphere of radius $r_0 = 24.5$ cm and mass $m = 1.60$ kg starts from rest and rolls without slipping down a 30.0° incline that is 10.0 m long. (*a*) Calculate its translational and rotational speeds when it reaches the bottom. (*b*) What is the ratio of translational to rotational kinetic energy at the bottom? Avoid putting in numbers until the end so you can answer: (*c*) do your answers in (*a*) and (*b*) depend on the radius of the sphere or its mass?

72. (II) A narrow but solid spool of thread has radius R and mass M. If you pull up on the thread so that the CM of the spool remains suspended in the air at the same place as it unwinds, (*a*) what force must you exert on the thread? (*b*) How much work have you done by the time the spool turns with angular velocity ω?

73. (II) A solid rubber ball rests on the floor of a railroad car when the car begins moving with acceleration a. Assuming the ball rolls without slipping, what is its acceleration relative to (*a*) the car and (*b*) the ground?

74. (II) A thin, hollow 0.545-kg section of pipe of radius 12.0 cm starts rolling (from rest) down a 17.5° incline 5.60 m long. (*a*) If the pipe rolls without slipping, what will be its speed at the base of the incline? (*b*) What will be its total kinetic energy at the base of the incline? (*c*) What minimum value must the coefficient of static friction have if the pipe is not to slip?

***75.** (II) In Example 10–21, (*a*) how far has the ball moved down the lane when it starts rolling without slipping? (*b*) What are its final linear and rotational speeds?

76. (II) A ball of radius r_0 rolls on the inside of a track of radius R_0 (see Fig. 10–68). If the ball starts from rest at the vertical edge of the track, what will be its speed when it reaches the lowest point of the track, rolling without slipping the whole way?

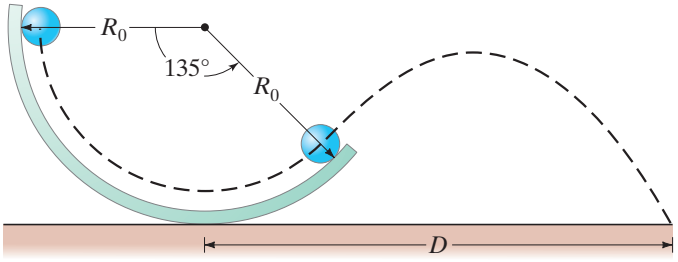


FIGURE 10–68 Problems 76 and 77.

77. (III) A small sphere of radius $r_0 = 1.5$ cm rolls without slipping on the track shown in Fig. 10–68 whose radius is $R_0 = 26.0$ cm. The sphere starts rolling at a height R_0 above the bottom of the track. When it leaves the track after passing through an angle of 135° as shown, (*a*) what will be its speed, and (*b*) at what distance D from the base of the track will the sphere hit the ground? (Ignore thickness of track.)

***78.** (III) A wheel with rotational inertia $I = \frac{1}{2}MR^2$ about its horizontal central axle is set spinning with initial angular speed ω_0. It is then lowered, and at the instant its edge touches the ground the speed of the axle (and CM) is zero. Initially the wheel slips when it touches the ground, but then begins to move forward and eventually rolls without slipping. (*a*) In what direction does friction act on the slipping wheel? (*b*) How long does the wheel slip before it begins to roll without slipping? (*c*) What is the wheel's final translational speed, v_{CM}? [*Hint*: Use $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$, $\Sigma\tau_{CM} = I_{CM}\alpha_{CM}$, and recall that only when there is rolling without slipping is $v_{CM} = \omega R$.]

79. (III) The 1100-kg mass of a car includes four tires, each of mass 35 kg (including wheels) and diameter 0.80 m. Assume each tire and wheel combination acts as a solid cylinder. Determine (*a*) the total kinetic energy of the car when traveling 95 km/h and (*b*) the fraction of the kinetic energy in the tires and wheels. (*c*) If the car is initially at rest and is then pulled by a tow truck with a force of 1500 N, what is the acceleration of the car? Ignore frictional losses. (*d*) What percent error would you make in part (*c*) if you ignored the rotational inertia of the tires and wheels?

*10–10 Rolling Sphere Slows Down

***80.** (II) A rolling ball slows down because the normal force does not pass exactly through the CM of the ball, but passes in front of the CM. Using Fig. 10–42, show that the torque resulting from the normal force ($\tau_N = \ell F_N$ in Fig. 10–42) is $\frac{2}{5}$ of that due to the frictional force, $\tau_{fr} = r_0 F$ where r_0 is the ball's radius; that is, show that $\tau_N = \frac{2}{5}\tau_{fr}$.

General Problems

81. A 1.6-kg grindstone in the shape of a uniform cylinder of radius 0.20 m acquires a rotational rate of 22 rev/s from rest over a 6.0-s interval at constant angular acceleration. Calculate the torque delivered by the motor.

82. On a 12.0-cm-diameter *audio compact* disc (CD), digital bits of information are encoded sequentially along an outward spiraling path. The spiral starts at radius $R_1 = 2.5$ cm and winds its way out to radius $R_2 = 5.8$ cm. To read the digital information, a CD player rotates the CD so that the player's readout laser scans along the spiral's sequence of bits at a constant linear speed of 1.25 m/s. Thus the player must accurately adjust the rotational frequency f of the CD as the laser moves outward. Determine the values for f (in units of rpm) when the laser is located at R_1 and when it is at R_2.

83. (*a*) A yo-yo is made of two solid cylindrical disks, each of mass 0.050 kg and diameter 0.075 m, joined by a (concentric) thin solid cylindrical hub of mass 0.0050 kg and diameter 0.013 m. Use conservation of energy to calculate the linear speed of the yo-yo just before it reaches the end of its 1.0-m-long string, if it is released from rest. (*b*) What fraction of its kinetic energy is rotational?

84. A cyclist accelerates from rest at a rate of 1.00 m/s^2. How fast will a point at the top of the rim of the tire (diameter = 68.0 cm) be moving after 2.75 s? [*Hint*: At any moment, the lowest point on the tire is in contact with the ground and is at rest—see Fig. 10–69.]

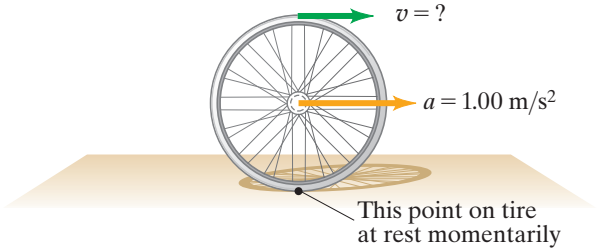


FIGURE 10–69 Problem 84.

85. Suppose David puts a 0.60-kg rock into a sling of length 1.5 m and begins whirling the rock in a nearly horizontal circle, accelerating it from rest to a rate of 75 rpm after 4.5 s. What is the torque required to achieve this feat, and where does the torque come from?

86. A large spool of rope rolls on the ground with the end of the rope lying on the top edge of the spool. A person grabs the end of the rope and walks a distance ℓ, holding onto it, Fig. 10–70. The spool rolls behind the person without slipping. What length of rope unwinds from the spool? How far does the spool's center of mass move?

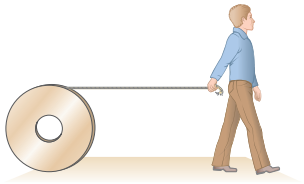


FIGURE 10–70 Problem 86.

87. **Bicycle gears**: (*a*) How is the angular velocity ω_R of the rear wheel of a bicycle related to the angular velocity ω_F of the front sprocket and pedals? Let N_F and N_R be the number of teeth on the front and rear sprockets, respectively, Fig. 10–71, and R_F and R_R their respective radii. The teeth are spaced the same on both sprockets and the rear sprocket is firmly attached to the rear wheel. (*b*) Evaluate the ratio ω_R/ω_F when the front and rear sprockets have 52 and 13 teeth, respectively, and (*c*) when they have 42 and 28 teeth.

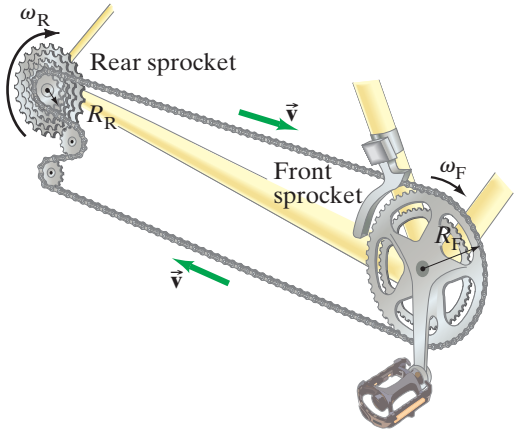


FIGURE 10–71 Problem 87.

88. A string passing over a pulley has a 3.80-kg mass hanging from one end and a 3.05-kg mass hanging from the other end. The pulley is a uniform solid cylinder of radius 4.0 cm and mass 0.60 kg. (*a*) If the bearings of the pulley were frictionless, what would be the acceleration of the two masses? (*b*) In fact, it is found that if the heavier mass is given a downward speed of 0.20 m/s, it comes to rest in 6.2 s. What is the average frictional torque acting on the pulley?

89. Figure 10–72 illustrates an H_2O molecule. The O—H bond length is 0.096 nm and the H—O—H bonds make an angle of 104°. Calculate the moment of inertia of the H_2O molecule (assume the atoms are points) about an axis passing through the center of the oxygen atom (*a*) perpendicular to the plane of the molecule, and (*b*) in the plane of the molecule, bisecting the H—O—H bonds.

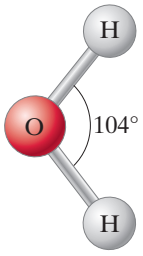


FIGURE 10–72 Problem 89.

90. A hollow cylinder (hoop) is rolling on a horizontal surface at speed $v = 3.0$ m/s when it reaches an 18° incline. (*a*) How far up the incline will it go? (*b*) How long will it be on the incline before it arrives back at the bottom?

91. What minimum horizontal force F is needed to pull a wheel of radius R and mass M over a step of height h as shown in Fig. 10–73 ($R > h$)? (*a*) Assume the force is applied at the top edge as shown. (*b*) Assume the force is applied instead at the wheel's center.

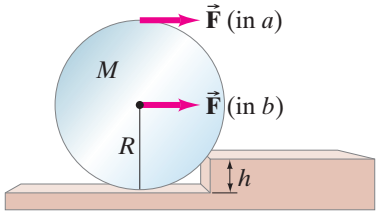


FIGURE 10–73 Problem 91.

92. A 140-g baseball, with a diameter of 7.5 cm, is pitched at 145 km/h. It spins at 1200 rpm. If the baseball is assumed to be a uniform solid sphere, what fraction of its kinetic energy is rotational?

93. If the coefficient of static friction between a car's tires and the pavement is 0.65, calculate the minimum torque that must be applied to the 66-cm-diameter tire of a 1250-kg automobile in order to "lay rubber" (make the wheels spin, slipping as the car accelerates). Assume each wheel supports an equal share of the weight.

94. A 4.00-kg mass and a 3.00-kg mass are attached to opposite ends of a thin 42.0-cm-long horizontal rod (Fig. 10–74). The system is rotating at angular speed $\omega = 5.20$ rad/s about a vertical axle at the center of the rod. Determine (*a*) the kinetic energy K of the system, and (*b*) the net force on each mass. (*c*) Repeat parts (*a*) and (*b*) assuming that the axle passes through the CM of the system.

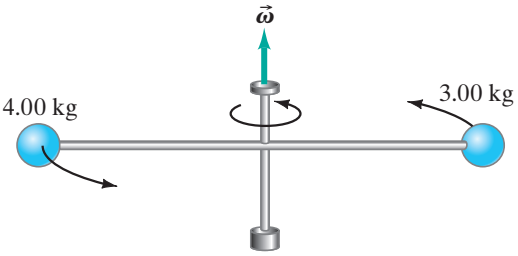


FIGURE 10–74 Problem 94.

95. The radius of the roll of paper shown in Fig. 10–75 is 7.6 cm and its moment of inertia is $I = 3.3 \times 10^{-3}$ kg · m^2. A force of 2.5 N is exerted on the end of the roll for 1.1 s; the paper does not tear so it unrolls. A constant friction torque of 0.11 m · N is exerted on the roll which gradually brings it to a stop. Assuming that the paper's thickness is negligible, calculate (*a*) the length of paper that unrolls during the time that the force is applied (1.1 s) and (*b*) the length of paper that unrolls from the time the force ends to the time when the roll has stopped moving.

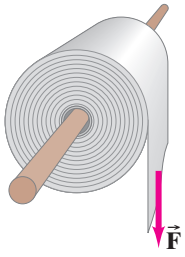


FIGURE 10–75 Problem 95.

96. A solid uniform disk of mass 21.0 kg and radius 85.0 cm is at rest flat on a frictionless surface. Figure 10–76 shows a view from above. A string is wrapped around the rim of the disk and a constant force of 35.0 N is applied to the string. The string does not slip on the rim. (*a*) In what direction does the CM move? When the CM has moved a distance of 5.2 m, determine (*b*) how fast the CM is moving, (*c*) how fast the disk is spinning (in radians per second), and (*d*) how much string has unwrapped from around the rim.

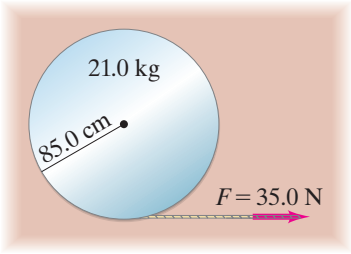


FIGURE 10–76 Problem 96, looking down on the disk.

97. A marble of mass m and radius r rolls along the looped rough track of Fig. 10–77. What is the minimum value of the vertical height h that the marble must drop if it is to reach the highest point of the loop without leaving the track? (*a*) Assume $r \ll R$; (*b*) do not make this assumption. Ignore frictional losses.

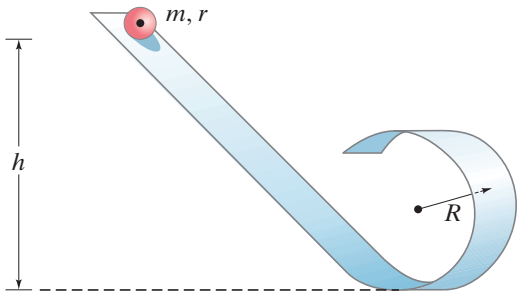


FIGURE 10–77 Problem 97.

98. The density (mass per unit length) of a thin rod of length ℓ increases uniformly from λ_0 at one end to $3\lambda_0$ at the other end. Determine the moment of inertia about an axis perpendicular to the rod through its geometric center.

99. If a billiard ball is hit in just the right way by a cue stick, the ball will roll without slipping immediately after losing contact with the stick. Consider a billiard ball (radius r, mass M) at rest on a horizontal pool table. A cue stick exerts a constant horizontal force F on the ball for a time t at a point that is a height h above the table's surface (see Fig. 10–78). Assume that the coefficient of static friction between the ball and table is μ_s. Determine the value for h so that the ball will roll without slipping immediately after losing contact with the stick.

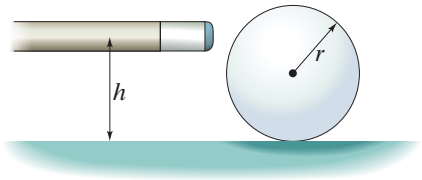


FIGURE 10–78 Problem 99.

100. A 5.0-m-long ladder is leaning against the side of a building making a 35° angle with the building. When a person is about 1/3 of the way up, the ladder slips and falls to the ground in 3.0 s. What is the average angular acceleration of the ladder as it falls?

101. When bicycle and motorcycle riders "pop a wheelie," a large acceleration causes the bike's front wheel to leave the ground. Let M be the total mass of the bike-plus-rider system; let x and y be the horizontal and vertical distance of this system's CM from the rear wheel's point of contact with the ground (Fig. 10–79). (*a*) Determine the horizontal acceleration a required to barely lift the bike's front wheel off of the ground. (*b*) To minimize the acceleration necessary to pop a wheelie, should x be made as small or as large as possible? How about y? How should a rider position his or her body on the bike in order to achieve these optimal values for x and y? (*c*) If $x = 35$ cm and $y = 95$ cm, find a.

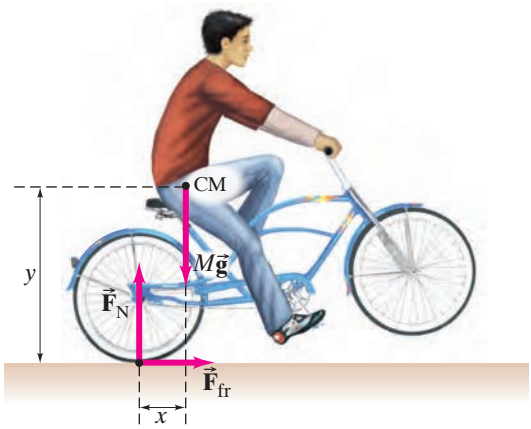


FIGURE 10–79 Problem 101.

102. A crucial part of a piece of machinery starts as a flat uniform cylindrical disk of radius R_0 and mass M. It then has a circular hole of radius R_1 drilled into it (Fig. 10–80). The hole's center is a distance h from the center of the disk. Find the moment of inertia of this disk (with off-center hole) when rotated about its center, C. [*Hint*: Consider a solid disk and "subtract" the hole; use the parallel-axis theorem.]

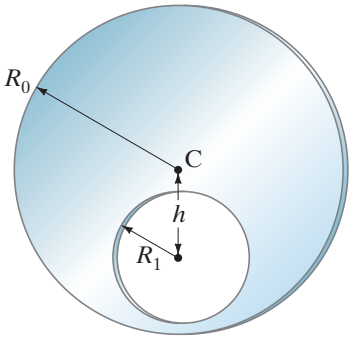


FIGURE 10–80 Problem 102.

103. A thin rod of length ℓ stands vertically on a table. The rod begins to fall, but its lower end does not slide. (*a*) Determine the angular velocity of the rod as a function of the angle ϕ it makes with the tabletop. (*b*) What is the speed of the tip of the rod just before it strikes the table?

104. A cord connected at one end to a block which can slide on an inclined plane has its other end wrapped around a cylinder resting in a depression at the top of the plane as shown in Fig. 10–81. Determine the speed of the block after it has traveled 1.80 m along the plane, starting from rest. Assume (*a*) there is no friction, (*b*) the coefficient of friction between all surfaces is $\mu = 0.035$. [*Hint*: In part (*b*) first determine the normal force on the cylinder, and make any reasonable assumptions needed.]

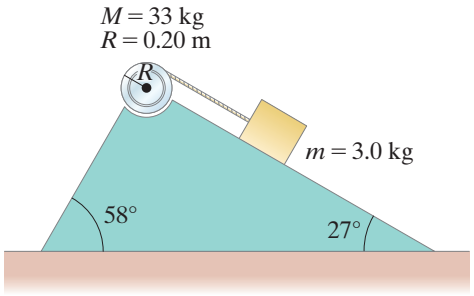


FIGURE 10–81 Problem 104.

***105.** (*a*) For the yo-yo-like cylinder of Example 10–20, we saw that the downward acceleration of its CM was $a = \frac{2}{3}g$. If it starts from rest, what will be the CM velocity after it has fallen a distance h? (*b*) Now use conservation of energy to determine the cylinder's CM velocity after it has fallen a distance h, starting from rest.

106. A thin uniform stick of mass M and length ℓ is positioned vertically, with its tip on a frictionless table. It is released and allowed to fall. Determine the speed of its CM just before it hits the table (Fig. 10–82).

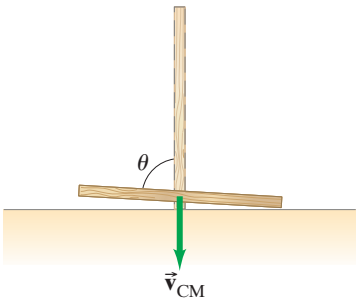
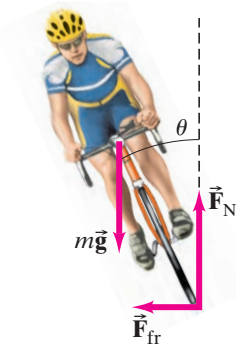


FIGURE 10–82
Problem 106.

107. One possibility for a low-pollution automobile is for it to use energy stored in a heavy rotating **flywheel**. Suppose such a car has a total mass of 1200 kg, uses a uniform cylindrical flywheel of diameter 1.50 m and mass 270 kg, and should be able to travel 350 km without needing a flywheel "spinup." (*a*) Make reasonable assumptions (average frictional retarding force on car = 450 N, thirty acceleration periods from rest to 95 km/h, equal uphill and downhill, and that energy can be put back into the flywheel as the car goes downhill), and estimate what total energy needs to be stored in the flywheel. (*b*) What is the angular velocity of the flywheel when it has a full "energy charge"? (*c*) About how long would it take a 150-hp motor to give the flywheel a full energy charge before a trip? [N.B. Such a car carries dangers, especially when turning. It can flip over. Easier to see after we discuss angular momentum in the next Chapter.]

108. A bicyclist traveling with speed $v = 8.5$ m/s on a flat road is making a turn with a radius $r = 13$ m. There are three forces acting on the cyclist and cycle: the normal force $(\vec{\mathbf{F}}_N)$ and friction force $(\vec{\mathbf{F}}_{fr})$ exerted by the road on the tires; and $m\vec{\mathbf{g}}$, the total weight of the cyclist and cycle. Ignore the small mass of the wheels. (*a*) Explain carefully why the angle θ the bicycle makes with the vertical (Fig. 10–83) must be given by $\tan\theta = F_{fr}/F_N$ if the cyclist is to maintain balance. (*b*) Calculate θ for the values given. [*Hint*: Consider the force needed to provide the centripetal acceleration of the bicycle and rider going around the curve.] (*c*) If the coefficient of static friction between tires and road is $\mu_s = 0.65$, what is the minimum turning radius?





(a) (b)

FIGURE 10–83 Problem 108.

ANSWERS TO EXERCISES

A: $f = 0.076$ Hz; $T = 13$ s.
B: $\vec{\mathbf{F}}_A$.
C: Yes; $\frac{1}{12}M\ell^2 + M(\frac{1}{2}\ell)^2 = \frac{1}{3}M\ell^2$.
D: 4×10^{17} J.
E: (*c*).
F: $a = \frac{1}{3}g$. [*Hint*: $\tau = F_T(R/2)$ and $a = (R/2)\alpha$]

This skater is doing a spin. When her arms are spread outward horizontally, she spins less fast than when her arms are held close to the axis of rotation, as at this moment in the photograph. This is an example of the conservation of angular momentum.

Angular momentum, which we study in this Chapter, is conserved only if no net torque acts on the object or system. Otherwise, the rate of change of angular momentum is proportional to the net applied torque—which, if zero, means the angular momentum is *conserved*. In this Chapter we also examine more complicated aspects of rotational motion.



CHAPTER 11

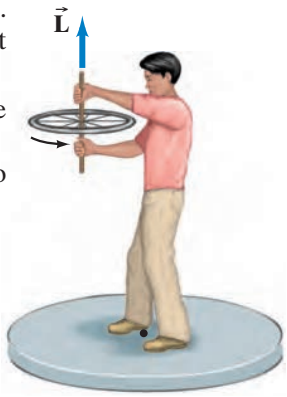
Angular Momentum; General Rotation

CONTENTS

11–1 Angular Momentum—Objects Rotating About a Fixed Axis
11–2 Vector Cross Product; Torque as a Vector
11–3 Angular Momentum of a Particle
11–4 Angular Momentum and Torque for a System of Particles; General Motion
11–5 Angular Momentum and Torque for a Rigid Object
11–6 Conservation of Angular Momentum
*11–7 The Spinning Top and Gyroscope
11–8 Rotating Frames of Reference; Inertial Forces
*11–9 The Coriolis Effect

CHAPTER-OPENING QUESTION—Guess now!

You are standing on a platform at rest, but that is free to rotate. You hold a spinning bicycle wheel by its axle as shown here. You then flip the wheel over so its axle points down. What happens then?



(a) The platform starts rotating in the direction the bicycle wheel was originally rotating.
(b) The platform starts rotating in the direction opposite to the original rotation of the bicycle wheel.
(c) The platform stays at rest.
(d) The platform turns only while you are flipping the wheel.
(e) None of these is correct.

In Chapter 10 we dealt with the kinematics and dynamics of a rigid object rotating about an axis whose direction is fixed in an inertial reference frame. We analyzed the motion in terms of the rotational equivalent of Newton's laws (torque plays the role that force does for translational motion), as well as rotational kinetic energy.

To keep the axis of a rotating object fixed, the object must usually be constrained by external supports (such as bearings at the end of an axle). The motion of objects that are not constrained to rotate about a fixed axis is more difficult to describe and analyze. Indeed, the complete analysis of the general rotational motion of an object (or system of objects) is very complicated, and we will only look at some aspects of general rotational motion in this Chapter.

We start this Chapter by introducing the concept of *angular momentum*, which is the rotational analog of linear momentum. We first treat angular momentum and its conservation for an object rotating about a fixed axis. After that, we examine the vector nature of torque and angular momentum. We will derive some general theorems and apply them to some interesting types of motion, and we will see some examples of axis direction changing.

11–1 Angular Momentum—Objects Rotating About a Fixed Axis

In Chapter 10 we saw that if we use the appropriate angular variables, the kinematic and dynamic equations for rotational motion are analogous to those for ordinary linear motion. In like manner, the linear momentum, $p = mv$, has a rotational analog. It is called **angular momentum**, L. We consider first the case of an object, with moment of inertia I, that is rotating about a fixed axis through the CM. The angular momentum of this object is

$$L = I\omega, \tag{11–1}$$

where ω is the object's angular velocity. The SI units for L are $\text{kg} \cdot \text{m}^2/\text{s}$ which has no special name.

We saw in Chapter 9 (Section 9–1) that Newton's second law can be written not only as $\Sigma F = ma$, but also more generally in terms of momentum (Eq. 9–2), $\Sigma F = dp/dt$. In a similar way, the rotational equivalent of Newton's second law, which we saw in Eqs. 10–14 and 10–15 can be written as $\Sigma\tau = I\alpha$, can also be written in terms of angular momentum. From Eq. 10–3b, the angular acceleration $\alpha = d\omega/dt$. Then $I\alpha = I(d\omega/dt) = d(I\omega)/dt = dL/dt$, so

$$\Sigma\tau = \frac{dL}{dt}. \tag{11–2}$$

NEWTON'S SECOND LAW FOR ROTATION

This derivation assumes that the moment of inertia, I, remains constant. However, Eq. 11–2 is valid even if the moment of inertia changes, and it applies also to a *system* of objects rotating about a fixed axis where $\Sigma\tau$ is the net external torque (discussed in Section 11–4). Equation 11–2 is *Newton's second law for rotational motion* about a fixed axis, and is also valid for a moving object if its rotation is about an axis passing through its center of mass (as for Eq. 10–15).

Conservation of Angular Momentum

Angular momentum is an important concept in physics because, under certain conditions, it is a conserved quantity. What are the conditions for which it is conserved? From Eq. 11–2 we see that if the net external torque $\Sigma\tau$ on an object (or system of objects) is zero, then

$$\frac{dL}{dt} = 0 \quad \text{and} \quad L = I\omega = \text{constant.} \qquad [\Sigma\tau = 0]$$

This is the **law of conservation of angular momentum** for a rotating object:

The total angular momentum of a rotating object remains constant if the net external torque acting on it is zero.

CONSERVATION OF ANGULAR MOMENTUM

The law of conservation of angular momentum is one of the great conservation laws of physics, along with those for energy and linear momentum.

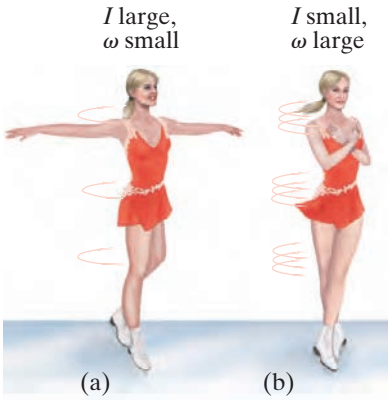


FIGURE 11–1 A skater spinning on ice, illustrating conservation of angular momentum: (a) I is large and ω is small; (b) I is smaller so ω is larger.

When there is zero net torque acting on an object, and the object is rotating about a fixed axis or about an axis through its center of mass whose direction does not change, we can write

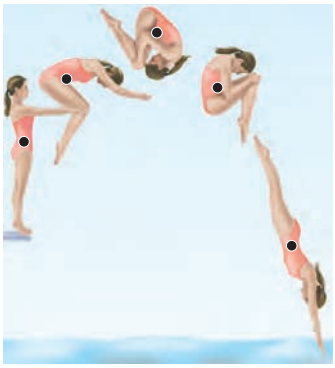
$$I\omega = I_0\omega_0 = \text{constant}.$$

I_0 and ω_0 are the moment of inertia and angular velocity, respectively, about that axis at some initial time $(t = 0)$, and I and ω are their values at some other time. The parts of the object may alter their positions relative to one another, so that I changes. But then ω changes as well, so that the product $I\omega$ remains constant.

Many interesting phenomena can be understood on the basis of conservation of angular momentum. Consider a skater doing a spin on the tips of her skates, Fig. 11–1. She rotates at a relatively low speed when her arms are outstretched, but when she brings her arms in close to her body, she suddenly spins much faster. From the definition of moment of inertia, $I = \Sigma mR^2$, it is clear that when she pulls her arms in closer to the axis of rotation, R is reduced for the arms so her moment of inertia is reduced. Since the angular momentum $I\omega$ remains constant (we ignore the small torque due to friction), if I decreases, then the angular velocity ω must increase. If the skater reduces her moment of inertia by a factor of 2, she will then rotate with twice the angular velocity.

EXERCISE A When a spinning figure skater pulls in her arms, her moment of inertia decreases; to conserve angular momentum, her angular velocity increases. Does her rotational kinetic energy also increase? If so, where does the energy come from?

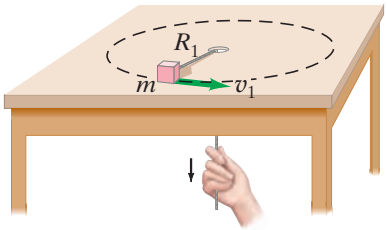
FIGURE 11–2 A diver rotates faster when arms and legs are tucked in than when they are outstretched. Angular momentum is conserved.



A similar example is the diver shown in Fig. 11–2. The push as she leaves the board gives her an initial angular momentum about her center of mass. When she curls herself into the tuck position, she rotates quickly one or more times. She then stretches out again, increasing her moment of inertia which reduces the angular velocity to a small value, and then she enters the water. The change in moment of inertia from the straight position to the tuck position can be a factor of as much as $3\frac{1}{2}$.

Note that for angular momentum to be conserved, the net torque must be zero, but the net force does not necessarily have to be zero. The net force on the diver in Fig. 11–2, for example, is not zero (gravity is acting), but the net torque about her CM is zero because the force of gravity can be considered to act at her center of mass.

FIGURE 11–3 Example 11–1.



EXAMPLE 11–1 Object rotating on a string of changing length. A small mass m attached to the end of a string revolves in a circle on a frictionless tabletop. The other end of the string passes through a hole in the table (Fig. 11–3). Initially, the mass revolves with a speed $v_1 = 2.4\,\text{m/s}$ in a circle of radius $R_1 = 0.80\,\text{m}$. The string is then pulled slowly through the hole so that the radius is reduced to $R_2 = 0.48\,\text{m}$. What is the speed, v_2, of the mass now?

APPROACH There is no net torque on the mass m because the force exerted by the string to keep it moving in a circle is exerted toward the axis; hence the lever arm is zero. We can thus apply conservation of angular momentum.

SOLUTION Conservation of angular momentum gives

$$I_1\omega_1 = I_2\omega_2.$$

Our small mass is essentially a particle whose moment of inertia about the hole is $I = mR^2$ (Eq. 10–11), so we have

$$mR_1^2\omega_1 = mR_2^2\omega_2,$$

or

$$\omega_2 = \omega_1\left(\frac{R_1^2}{R_2^2}\right).$$

Then, since $v = R\omega$, we can write

$$v_2 = R_2\omega_2 = R_2\omega_1\left(\frac{R_1^2}{R_2^2}\right) = R_2\frac{v_1}{R_1}\left(\frac{R_1^2}{R_2^2}\right) = v_1\frac{R_1}{R_2}$$
$$= (2.4\,\text{m/s})\left(\frac{0.80\,\text{m}}{0.48\,\text{m}}\right) = 4.0\,\text{m/s}.$$

The speed increases as the radius decreases.

EXAMPLE 11–2 **Clutch.** A simple clutch consists of two cylindrical plates that can be pressed together to connect two sections of an axle, as needed, in a piece of machinery. The two plates have masses $M_A = 6.0\,\text{kg}$ and $M_B = 9.0\,\text{kg}$, with equal radii $R_0 = 0.60\,\text{m}$. They are initially separated (Fig. 11–4). Plate M_A is accelerated from rest to an angular velocity $\omega_1 = 7.2\,\text{rad/s}$ in time $\Delta t = 2.0\,\text{s}$. Calculate (*a*) the angular momentum of M_A, and (*b*) the torque required to have accelerated M_A from rest to ω_1. (*c*) Next, plate M_B, initially at rest but free to rotate without friction, is placed in firm contact with freely rotating plate M_A, and the two plates then both rotate at a constant angular velocity ω_2, which is considerably less than ω_1. Why does this happen, and what is ω_2?

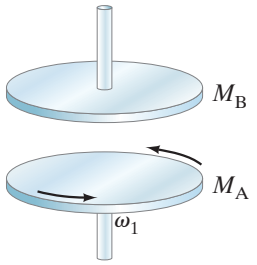


FIGURE 11–4 Example 11–2, two freely rotating plates. (In a clutch assembly, one plate is made to rotate by a motor. At a certain moment, the machinery is meant to have a second part also rotate; at that moment the two plates are forced together.) In this Example 11–2 we are dealing with two plates that are now not connected to a motor or anything else, and can spin freely.

APPROACH We use the definition of angular momentum $L = I\omega$ (Eq. 11–1) plus Newton's second law for rotation, Eq. 11–2.

SOLUTION (*a*) The angular momentum of M_A, a cylinder, is

$$L_A = I_A\omega_1 = \tfrac{1}{2}M_A R_0^2\omega_1 = \tfrac{1}{2}(6.0\,\text{kg})(0.60\,\text{m})^2(7.2\,\text{rad/s}) = 7.8\,\text{kg}\cdot\text{m}^2/\text{s}.$$

(*b*) The plate started from rest so the torque, assumed constant, was

$$\tau = \frac{\Delta L}{\Delta t} = \frac{7.8\,\text{kg}\cdot\text{m}^2/\text{s} - 0}{2.0\,\text{s}} = 3.9\,\text{m}\cdot\text{N}.$$

(*c*) Initially, before contact, M_A is rotating at constant ω_1 (we ignore friction). When plate B comes in contact, why is their joint rotation speed less? You might think in terms of the torque each exerts on the other upon contact. But quantitatively, it's easier to use conservation of angular momentum, since no external torques are assumed to act. Thus

$$\text{angular momentum before} = \text{angular momentum after}$$
$$I_A\omega_1 = (I_A + I_B)\omega_2.$$

Solving for ω_2 we find (after cancelling factors of R_0^2)

$$\omega_2 = \left(\frac{I_A}{I_A + I_B}\right)\omega_1 = \left(\frac{M_A}{M_A + M_B}\right)\omega_1 = \left(\frac{6.0\,\text{kg}}{15.0\,\text{kg}}\right)(7.2\,\text{rad/s}) = 2.9\,\text{rad/s}.$$

EXAMPLE 11–3 **ESTIMATE** **Neutron star.** Astronomers detect stars that are rotating extremely rapidly, known as neutron stars. A neutron star is believed to form from the inner core of a massive star that collapsed, under its own gravitation, to a star of very small radius and very high density. Before collapse, suppose the core of such a massive star is the size of our Sun $(r \approx 7 \times 10^5\,\text{km})$ with mass 2.0 times as great as the Sun, and is rotating at a frequency of 1.0 revolution every 100 days. If it were to undergo gravitational collapse to a neutron star of radius 10 km, what would its rotation frequency be? Assume the star is a uniform sphere at all times, and loses no mass.

PHYSICS APPLIED
Neutron star

APPROACH We assume the star is isolated (no external forces), so we can use conservation of angular momentum for this process. We use *r* for the radius of a sphere, as compared to *R* used for distance from an axis of rotation or cylindrical symmetry: see Fig. 10–2.

SOLUTION From conservation of angular momentum,

$$I_1\omega_1 = I_2\omega_2,$$

where the subscripts 1 and 2 refer to initial (normal star) and final (neutron star), respectively. Then, assuming no mass is lost in the process $(m_1 = m_2)$,

$$\omega_2 = \left(\frac{I_1}{I_2}\right)\omega_1 = \left(\frac{\frac{2}{5}m_1 r_1^2}{\frac{2}{5}m_2 r_2^2}\right)\omega_1 = \frac{r_1^2}{r_2^2}\omega_1.$$

The frequency $f = \omega/2\pi$, so

$$f_2 = \frac{\omega_2}{2\pi} = \frac{r_1^2}{r_2^2}f_1$$
$$= \left(\frac{7 \times 10^5\,\text{km}}{10\,\text{km}}\right)^2\left(\frac{1.0\,\text{rev}}{100\,\text{d}\,(24\,\text{h/d})\,(3600\,\text{s/h})}\right) \approx 6 \times 10^2\,\text{rev/s},$$

which is 600 Hz or $(600\,\text{rev/s})(60\,\text{s/min}) = 36{,}000\,\text{rpm}$.

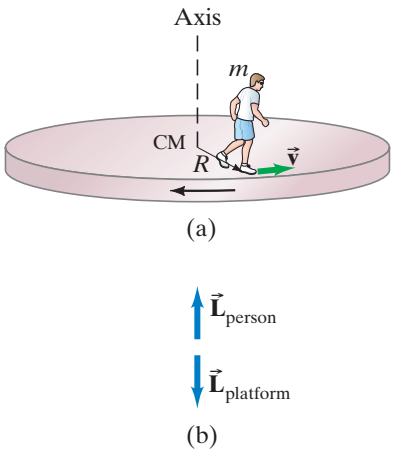


FIGURE 11–5 (a) A person standing on a circular platform, initially at rest, begins walking along the edge at speed v. The platform, mounted on nearly friction-free bearings, begins rotating in the opposite direction, so that (b) the total angular momentum remains zero ($\vec{\mathbf{L}}_{\text{platform}} = -\vec{\mathbf{L}}_{\text{person}}$).

Directional Nature of Angular Momentum

Angular momentum is a vector, as we shall discuss in detail later in this Chapter. For now we consider the simple case of an object rotating about a fixed axis, and the direction of $\vec{\mathbf{L}}$ is specified by a plus or minus sign, just as we did for one-dimensional linear motion in Chapter 2.

For a symmetrical object rotating about a symmetry axis (such as a cylinder or wheel), the direction of the angular momentum[†] can be taken as the direction of the angular velocity $\vec{\boldsymbol{\omega}}$. That is,

$$\vec{\mathbf{L}} = I\vec{\boldsymbol{\omega}}.$$

As a simple example, consider a person standing at rest on a circular platform capable of rotating friction-free about an axis through its center (that is, a simplified merry-go-round). If the person now starts to walk along the edge of the platform, Fig. 11–5a, the platform starts rotating in the opposite direction. Why? One explanation is that the person's foot exerts a force on the platform. Another explanation (and this is the most useful analysis here) is that this is an example of the conservation of angular momentum. If the person starts walking counterclockwise, the person's angular momentum points upward along the axis of rotation (remember how we defined the direction of $\vec{\boldsymbol{\omega}}$ using the right-hand rule in Section 10–2). The magnitude of the person's angular momentum will be $L = I\omega = (mR^2)(v/R)$, where v is the person's speed (relative to the Earth, not the platform), R is his distance from the rotation axis, m is his mass, and mR^2 is his moment of inertia if we consider him a particle (mass concentrated at one point, Eq. 10–11). The platform rotates in the opposite direction, so its angular momentum points downward. If the initial total angular momentum was zero (person and platform at rest), it will remain zero after the person starts walking. That is, the upward angular momentum of the person just balances the oppositely directed downward angular momentum of the platform (Fig. 11–5b), so the total vector angular momentum remains zero. Even though the person exerts a torque on the platform, the platform exerts an equal and opposite torque on the person. So the net torque on the *system* of person plus platform is zero (ignoring friction), and the total angular momentum remains constant.

EXAMPLE 11–4 **Running on a circular platform.** Suppose a 60-kg person stands at the edge of a 6.0-m-diameter circular platform, which is mounted on frictionless bearings and has a moment of inertia of $1800\ \text{kg}\cdot\text{m}^2$. The platform is at rest initially, but when the person begins running at a speed of 4.2 m/s (with respect to the Earth) around its edge, the platform begins to rotate in the opposite direction as in Fig. 11–5. Calculate the angular velocity of the platform.

APPROACH We use conservation of angular momentum. The total angular momentum is zero initially. Since there is no net torque, $\vec{\mathbf{L}}$ is conserved and will remain zero, as in Fig. 11–5. The person's angular momentum is $L_{\text{per}} = (mR^2)(v/R)$, and we take this as positive. The angular momentum of the platform is $L_{\text{plat}} = -I\omega$.

SOLUTION Conservation of angular momentum gives

$$L = L_{\text{per}} + L_{\text{plat}}$$

$$0 = mR^2\left(\frac{v}{R}\right) - I\omega.$$

So

$$\omega = \frac{mRv}{I} = \frac{(60\ \text{kg})(3.0\ \text{m})(4.2\ \text{m/s})}{1800\ \text{kg}\cdot\text{m}^2} = 0.42\ \text{rad/s}.$$

NOTE The frequency of rotation is $f = \omega/2\pi = 0.067\ \text{rev/s}$ and the period $T = 1/f = 15\ \text{s}$ per revolution.

[†]For more complicated situations of objects rotating about a fixed axis, there will be a component of $\vec{\mathbf{L}}$ along the direction of $\vec{\boldsymbol{\omega}}$ and its magnitude will be equal to $I\vec{\boldsymbol{\omega}}$, but there could be other components as well. If the total angular momentum is conserved, then the component $I\vec{\boldsymbol{\omega}}$ will also be conserved. So our results here can be applied to any rotation about a fixed axis.

EXERCISE B Suppose you are standing on the edge of a large freely rotating turntable. If you walk toward the center of the turntable, what will happen? (*a*) The turntable slows down. (*b*) The turntable speeds up. (*c*) Its rotation speed is unchanged. (*d*) The turntable stops. (*e*) You need to know the walking speed to answer.

CONCEPTUAL EXAMPLE 11–5 **Spinning bicycle wheel.** Your physics teacher is holding a spinning bicycle wheel while he stands on a stationary frictionless turntable (Fig. 11–6). What will happen if the teacher suddenly flips the bicycle wheel over so that it is spinning in the opposite direction?

RESPONSE We consider the system of turntable, teacher, and bicycle wheel. The total angular momentum initially is $\vec{\mathbf{L}}$ vertically upward (note black arrow showing direction of rotation). That is also what the system's angular momentum must be afterward, since $\vec{\mathbf{L}}$ is conserved when there is no net torque on the system as a whole. Thus, if the wheel's angular momentum after being flipped over is $-\vec{\mathbf{L}}$ downward, then the angular momentum of teacher plus turntable will have to be $+2\vec{\mathbf{L}}$ upward. We can safely predict that the turntable (and teacher) will begin revolving in the same direction the bike wheel was spinning originally (counterclockwise).



FIGURE 11–6 Example 11–5.

EXERCISE C Return to the Chapter-Opening Question, page 302, and answer it again now. Try to explain why you may have answered differently the first time.

EXERCISE D In Example 11–5, Fig. 11–6, what happens to the turntable if the teacher moves the spinning wheel's axis only 90° so it is horizontal? (*a*) It rotates in same direction and speed as in Example 11–5. (*b*) It rotates in that same direction, but slower. (*c*) It rotates in the opposite direction. (*d*) It doesn't start rotating.

11–2 Vector Cross Product; Torque as a Vector

Vector Cross Product

To deal with the vector nature of angular momentum and torque in general, we will need the concept of the *vector cross product* (often called simply the *vector product* or *cross product*). In general, the **vector** or **cross product** of two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ is defined as another vector $\vec{\mathbf{C}} = \vec{\mathbf{A}} \times \vec{\mathbf{B}}$ whose *magnitude is*

$$C = |\vec{\mathbf{A}} \times \vec{\mathbf{B}}| = AB \sin\theta, \qquad \textbf{(11–3a)}$$

where θ is the angle ($< 180°$) *between* $\vec{\mathbf{A}}$ *and* $\vec{\mathbf{B}}$, *and whose direction is perpendicular to both* $\vec{\mathbf{A}}$ *and* $\vec{\mathbf{B}}$ *in the sense of the right-hand rule*, Fig. 11–7. The angle θ is measured between $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ when their tails are together at the same point. According to the right-hand rule, as shown in Fig. 11–7, you orient your right hand so your fingers point along $\vec{\mathbf{A}}$, and when you bend your fingers they point along $\vec{\mathbf{B}}$. When your hand is correctly oriented in this way, your thumb will point along the direction of $\vec{\mathbf{C}} = \vec{\mathbf{A}} \times \vec{\mathbf{B}}$.

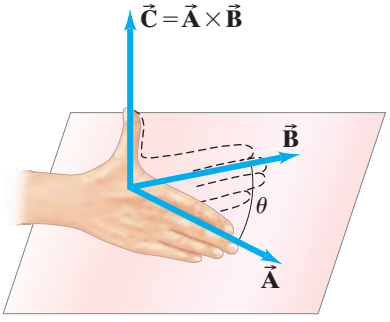
The cross product of two vectors, $\vec{\mathbf{A}} = A_x\hat{\mathbf{i}} + A_y\hat{\mathbf{j}} + A_z\hat{\mathbf{k}}$, and $\vec{\mathbf{B}} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}} + B_z\hat{\mathbf{k}}$, can be written in component form (see Problem 28) as

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \qquad \textbf{(11–3b)}$$

$$= (A_yB_z - A_zB_y)\hat{\mathbf{i}} + (A_zB_x - A_xB_z)\hat{\mathbf{j}} + (A_xB_y - A_yB_x)\hat{\mathbf{k}}. \qquad \textbf{(11–3c)}$$

Equation 11–3b is meant to be evaluated using the rules of determinants (and obtain Eq. 11–3c).[†] (Equation 11–3b is not literally a determinant, but a memory aid.)

FIGURE 11–7 The vector $\vec{\mathbf{C}} = \vec{\mathbf{A}} \times \vec{\mathbf{B}}$ is perpendicular to the plane containing $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$; its direction is given by the right-hand rule.



[†]A vector cross product is not a true vector, but a "pseudovector" or "axial vector" as discussed in Section 10–2. Imagine a mirror behind Fig. 11–7 (as in Fig. 10–10) and you will see that $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ reverse positions and $\vec{\mathbf{C}}$ does not, so $\vec{\mathbf{A}} \times \vec{\mathbf{B}} = -\vec{\mathbf{C}}$ in the mirror.

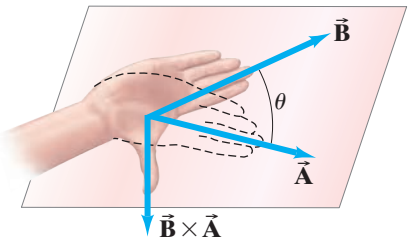


FIGURE 11–8 The vector $\vec{\mathbf{B}} \times \vec{\mathbf{A}}$ equals $-\vec{\mathbf{A}} \times \vec{\mathbf{B}}$; compare to Fig. 11–7.

FIGURE 11–9 Exercise E.

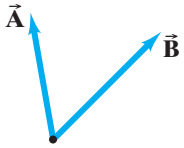


FIGURE 11–10 The torque due to the force $\vec{\mathbf{F}}$ (in the plane of the wheel) starts the wheel rotating counterclockwise so $\vec{\boldsymbol{\omega}}$ and $\vec{\boldsymbol{\alpha}}$ point out of the page.

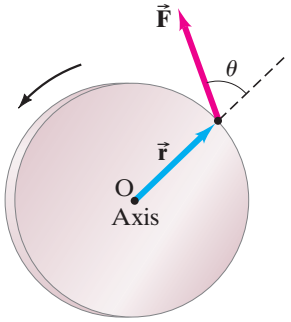
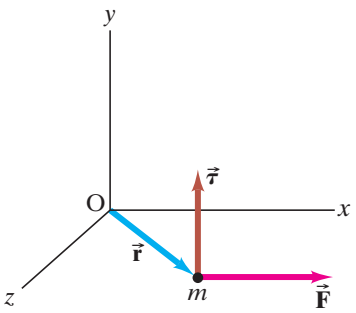


FIGURE 11–11 $\vec{\boldsymbol{\tau}} = \vec{\mathbf{r}} \times \vec{\mathbf{F}}$, where $\vec{\mathbf{r}}$ is the position vector.



Some properties of the cross product are the following:

$$\vec{\mathbf{A}} \times \vec{\mathbf{A}} = 0 \qquad \textbf{(11–4a)}$$

$$\vec{\mathbf{A}} \times \vec{\mathbf{B}} = -\vec{\mathbf{B}} \times \vec{\mathbf{A}} \qquad \textbf{(11–4b)}$$

$$\vec{\mathbf{A}} \times (\vec{\mathbf{B}} + \vec{\mathbf{C}}) = (\vec{\mathbf{A}} \times \vec{\mathbf{B}}) + (\vec{\mathbf{A}} \times \vec{\mathbf{C}}) \quad \text{[distributive law]} \qquad \textbf{(11–4c)}$$

$$\frac{d}{dt}(\vec{\mathbf{A}} \times \vec{\mathbf{B}}) = \frac{d\vec{\mathbf{A}}}{dt} \times \vec{\mathbf{B}} + \vec{\mathbf{A}} \times \frac{d\vec{\mathbf{B}}}{dt}. \qquad \textbf{(11–4d)}$$

Equation 11–4a follows from Eqs. 11–3 (since $\theta = 0$). So does Eq. 11–4b, since the magnitude of $\vec{\mathbf{B}} \times \vec{\mathbf{A}}$ is the same as that for $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$, but by the right-hand rule the direction is opposite (see Fig. 11–8). Thus the order of the two vectors is crucial. If you change the order, you change the result. That is, the commutative law does *not* hold for the cross product $(\vec{\mathbf{A}} \times \vec{\mathbf{B}} \neq \vec{\mathbf{B}} \times \vec{\mathbf{A}})$, although it does hold for the dot product of two vectors (Section 7–2) and for the product of scalars. Note in Eq. 11–4d that the order of quantities in the two products on the right must not be interchanged (because of Eq. 11–4b). Equation 11–4c follows from Eq. 11–3c (Problem 29).

EXERCISE E For the vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ in the plane of the page as shown in Fig. 11–9, in what direction is (i) $\vec{\mathbf{A}} \cdot \vec{\mathbf{B}}$, (ii) $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$, (iii) $\vec{\mathbf{B}} \times \vec{\mathbf{A}}$? (*a*) Into the page; (*b*) out of the page; (*c*) between $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$; (*d*) it is a scalar and has no direction; (*e*) it is zero and has no direction.

The Torque Vector

Torque is an example of a quantity that can be expressed as a cross product. To see this, let us take a simple example: the thin wheel shown in Fig. 11–10 which is free to rotate about an axle through its center at point O. A force $\vec{\mathbf{F}}$ acts at the edge of the wheel, at a point whose position relative to the center O is given by the position vector $\vec{\mathbf{r}}$ as shown. The force $\vec{\mathbf{F}}$ tends to rotate the wheel (assumed initially at rest) counterclockwise, so the angular velocity $\vec{\boldsymbol{\omega}}$ will point out of the page toward the viewer (remember the right-hand rule from Section 10–2). The torque due to $\vec{\mathbf{F}}$ will tend to increase $\vec{\boldsymbol{\omega}}$, so $\vec{\boldsymbol{\alpha}}$ also points outward along the rotation axis. The relation between angular acceleration and torque that we developed in Chapter 10 for an object rotating about a fixed axis is

$$\Sigma\tau = I\alpha,$$

(Eq. 10–14) where I is the moment of inertia. This scalar equation is the rotational equivalent of $\Sigma F = ma$, and we would like to make it a vector equation just as $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ is a vector equation. To do so in the case of Fig. 11–10 we must have the direction of $\vec{\boldsymbol{\tau}}$ point outward along the rotation axis, since $\vec{\boldsymbol{\alpha}}$ $(= d\vec{\boldsymbol{\omega}}/dt)$ has that direction; and the magnitude of the torque must be (see Eqs. 10–10 and Fig. 11–10) $\tau = rF_\perp = rF\sin\theta$. We can achieve this by defining the **torque vector** to be the cross product of $\vec{\mathbf{r}}$ and $\vec{\mathbf{F}}$:

$$\vec{\boldsymbol{\tau}} = \vec{\mathbf{r}} \times \vec{\mathbf{F}}. \qquad \textbf{(11–5)}$$

From the definition of the cross product above (Eq. 11–3a) the magnitude of $\vec{\boldsymbol{\tau}}$ will be $rF\sin\theta$ and the direction will be along the axis, as required for this special case.

We will see in Sections 11–3 through 11–5 that if we take Eq. 11–5 as the *general definition of torque*, then the vector relation $\Sigma\vec{\boldsymbol{\tau}} = I\vec{\boldsymbol{\alpha}}$ will hold in general. Thus we state now that Eq. 11–5 is the general definition of torque. It contains both magnitude and direction information. Note that this definition involves the position vector $\vec{\mathbf{r}}$ and thus the torque is being calculated about a point (and not R, the distance from the rotation axis, Fig. 10–2). We can choose that point O as we wish.

For a particle of mass m on which a force $\vec{\mathbf{F}}$ is applied, we define the torque about a point O as

$$\vec{\boldsymbol{\tau}} = \vec{\mathbf{r}} \times \vec{\mathbf{F}}$$

where $\vec{\mathbf{r}}$ is the position vector of the particle relative to O (Fig. 11–11). If we have a system of particles (which could be the particles making up a rigid object) the total torque $\vec{\boldsymbol{\tau}}$ on the system will be the sum of the torques on the individual particles:

$$\vec{\boldsymbol{\tau}}_i = \Sigma(\vec{\mathbf{r}}_i \times \vec{\mathbf{F}}_i),$$

where $\vec{\mathbf{r}}_i$ is the position vector of the i^{th} particle and $\vec{\mathbf{F}}_i$ is the net force on the i^{th} particle.

EXAMPLE 11–6 **Torque vector.** Suppose the vector $\vec{\mathbf{r}}$ is in the xz plane, as in Fig. 11–11, and is given by $\vec{\mathbf{r}} = (1.2\,\text{m})\,\hat{\mathbf{i}} + (1.2\,\text{m})\,\hat{\mathbf{k}}$. Calculate the torque vector $\vec{\tau}$ if $\vec{\mathbf{F}} = (150\,\text{N})\,\hat{\mathbf{i}}$.

APPROACH We use the determinant form, Eq. 11–3b.

SOLUTION
$$\vec{\tau} = \vec{\mathbf{r}} \times \vec{\mathbf{F}} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 1.2\,\text{m} & 0 & 1.2\,\text{m} \\ 150\,\text{N} & 0 & 0 \end{vmatrix} = 0\,\hat{\mathbf{i}} + (180\,\text{m}\cdot\text{N})\,\hat{\mathbf{j}} + 0\,\hat{\mathbf{k}}.$$

So τ has magnitude $180\,\text{m}\cdot\text{N}$ and points along the positive y axis.

EXERCISE F If $\vec{\mathbf{F}} = 5.0\,\text{N}\,\hat{\mathbf{i}}$ and $\vec{\mathbf{r}} = 2.0\,\text{m}\,\hat{\mathbf{j}}$, what is $\vec{\tau}$? (*a*) $10\,\text{m}\cdot\text{N}$, (*b*) $-10\,\text{m}\cdot\text{N}$, (*c*) $10\,\text{m}\cdot\text{N}\,\hat{\mathbf{k}}$, (*d*) $-10\,\text{m}\cdot\text{N}\,\hat{\mathbf{j}}$, (*e*) $-10\,\text{m}\cdot\text{N}\,\hat{\mathbf{k}}$.

11–3 Angular Momentum of a Particle

The most general way of writing Newton's second law for the translational motion of a particle (or system of particles) is in terms of the linear momentum $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$ as given by Eq. 9–2 (or 9–5):

$$\Sigma\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt}.$$

The rotational analog of linear momentum is *angular momentum*. Just as the rate of change of $\vec{\mathbf{p}}$ is related to the net force $\Sigma\vec{\mathbf{F}}$, so we might expect the rate of change of angular momentum to be related to the net torque. Indeed, we saw this was true in Section 11–1 for the special case of a rigid object rotating about a fixed axis. Now we will see it is true in general. We first treat a single particle.

Suppose a particle of mass m has momentum $\vec{\mathbf{p}}$ and position vector $\vec{\mathbf{r}}$ with respect to the origin O in some chosen inertial reference frame. Then the general definition of the **angular momentum**, $\vec{\mathbf{L}}$, of the particle about point O is the vector cross product of $\vec{\mathbf{r}}$ and $\vec{\mathbf{p}}$:

$$\vec{\mathbf{L}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}}. \qquad \text{[particle]} \quad \textbf{(11–6)}$$

Angular momentum is a vector whose direction is perpendicular to both $\vec{\mathbf{r}}$ and $\vec{\mathbf{p}}$ as given by the right-hand rule (Fig. 11–12). Its magnitude is given by

$$L = rp\sin\theta$$

or

$$L = rp_\perp = r_\perp p$$

where θ is the angle between $\vec{\mathbf{r}}$ and $\vec{\mathbf{p}}$ and $p_\perp\,(= p\sin\theta)$ and $r_\perp\,(= r\sin\theta)$ are the components of $\vec{\mathbf{p}}$ and $\vec{\mathbf{r}}$ perpendicular to $\vec{\mathbf{r}}$ and $\vec{\mathbf{p}}$, respectively.

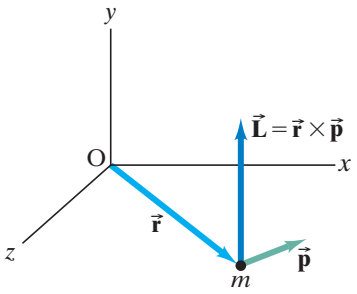


FIGURE 11–12 The angular momentum of a particle of mass m is given by $\vec{\mathbf{L}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}} = \vec{\mathbf{r}} \times m\vec{\mathbf{v}}$.

Now let us find the relation between angular momentum and torque for a particle. If we take the derivative of $\vec{\mathbf{L}}$ with respect to time we have

$$\frac{d\vec{\mathbf{L}}}{dt} = \frac{d}{dt}(\vec{\mathbf{r}} \times \vec{\mathbf{p}}) = \frac{d\vec{\mathbf{r}}}{dt} \times \vec{\mathbf{p}} + \vec{\mathbf{r}} \times \frac{d\vec{\mathbf{p}}}{dt}.$$

But

$$\frac{d\vec{\mathbf{r}}}{dt} \times \vec{\mathbf{p}} = \vec{\mathbf{v}} \times m\vec{\mathbf{v}} = m(\vec{\mathbf{v}} \times \vec{\mathbf{v}}) = 0,$$

since $\sin\theta = 0$ for this case. Thus

$$\frac{d\vec{\mathbf{L}}}{dt} = \vec{\mathbf{r}} \times \frac{d\vec{\mathbf{p}}}{dt}.$$

If we let $\Sigma\vec{\mathbf{F}}$ represent the resultant force on the particle, then in an inertial reference frame, $\Sigma\vec{\mathbf{F}} = d\vec{\mathbf{p}}/dt$ and

$$\vec{\mathbf{r}} \times \Sigma\vec{\mathbf{F}} = \vec{\mathbf{r}} \times \frac{d\vec{\mathbf{p}}}{dt} = \frac{d\vec{\mathbf{L}}}{dt}.$$

But $\vec{\mathbf{r}} \times \Sigma\vec{\mathbf{F}} = \Sigma\vec{\tau}$ is the net torque on our particle. Hence

$$\Sigma\vec{\tau} = \frac{d\vec{\mathbf{L}}}{dt}. \qquad \text{[particle, inertial frame]} \quad \textbf{(11–7)}$$

The rate of change of angular momentum of a particle equals the net torque applied to it. Equation 11–7 is the rotational equivalent of Newton's second law for a particle.

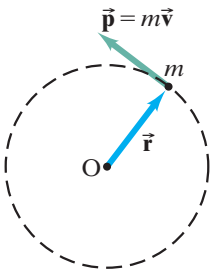


FIGURE 11–13 The angular momentum of a particle of mass m rotating in a circle of radius $\vec{\mathbf{r}}$ with velocity $\vec{\mathbf{v}}$ is $\vec{\mathbf{L}} = \vec{\mathbf{r}} \times m\vec{\mathbf{v}}$ (Example 11–7).

Equation 11–7 is valid only in an inertial frame since only then is it true that $\Sigma\vec{\mathbf{F}} = d\vec{\mathbf{p}}/dt$, which was used in the proof.

CONCEPTUAL EXAMPLE 11–7 **A particle's angular momentum.** What is the angular momentum of a particle of mass m moving with speed v in a circle of radius r in a counterclockwise direction?

RESPONSE The value of the angular momentum depends on the choice of the point O. Let us calculate $\vec{\mathbf{L}}$ with respect to the center of the circle, Fig. 11–13. Then $\vec{\mathbf{r}}$ is perpendicular to $\vec{\mathbf{p}}$ so $L = |\vec{\mathbf{r}} \times \vec{\mathbf{p}}| = rmv$. By the right-hand rule, the direction of $\vec{\mathbf{L}}$ is perpendicular to the plane of the circle, outward toward the viewer. Since $v = \omega r$ and $I = mr^2$ for a single particle rotating about an axis a distance r away, we can write

$$L = mvr = mr^2\omega = I\omega.$$

11–4 Angular Momentum and Torque for a System of Particles; General Motion

Relation Between Angular Momentum and Torque

Now we consider angular momentum in a more general way where the object is not necessarily symmetric and the rotation axis may not be fixed or through the CM.

Consider a system of n particles which have angular momenta $\vec{\mathbf{L}}_1, \vec{\mathbf{L}}_2, \ldots, \vec{\mathbf{L}}_n$. The system could be anything from a rigid object to a loose assembly of particles whose positions are not fixed relative to each other. The total angular momentum $\vec{\mathbf{L}}$ of the system is defined as the vector sum of the angular momenta of all the particles in the system:

$$\vec{\mathbf{L}} = \sum_{i=1}^{n} \vec{\mathbf{L}}_i. \qquad \textbf{(11–8)}$$

The resultant torque acting on the system is the sum of the net torques acting on all the particles:

$$\vec{\tau}_{\text{net}} = \sum \vec{\tau}_i.$$

This sum includes (1) internal torques due to internal forces that particles of the system exert on other particles of the system, and (2) external torques due to forces exerted by objects outside our system. By Newton's third law, the force each particle exerts on another is equal and opposite (and acts along the same line as) the force that the second particle exerts on the first. Hence the sum of all internal torques adds to zero, and

$$\vec{\tau}_{\text{net}} = \sum_i \vec{\tau}_i = \sum \vec{\tau}_{\text{ext}}.$$

Now we take the time derivative of Eq. 11–8 and use Eq. 11–7 for each particle to obtain

$$\frac{d\vec{\mathbf{L}}}{dt} = \sum_i \frac{d\vec{\mathbf{L}}_i}{dt} = \sum \vec{\tau}_{\text{ext}}$$

or

NEWTON'S SECOND LAW (rotation; system of particles)

$$\frac{d\vec{\mathbf{L}}}{dt} = \sum \vec{\tau}_{\text{ext}}. \qquad \text{[inertial reference frame]} \quad \textbf{(11–9a)}$$

This fundamental result states that the time rate of change of the total angular momentum of a system of particles (or a rigid object) equals the resultant external torque on the system. This is **Newton's second law for general rotational motion**. This law is the rotational equivalent of Eq. 9–5, $d\vec{\mathbf{P}}/dt = \Sigma\vec{\mathbf{F}}_{\text{ext}}$ for translational motion. Note that $\vec{\mathbf{L}}$ and $\Sigma\vec{\tau}$ must be calculated about the same origin O.

Equation 11–9a is valid when $\vec{\mathbf{L}}$ and $\vec{\tau}_{\text{ext}}$ are calculated with reference to a point fixed in an inertial reference frame. (In the derivation, we used Eq. 11–7 which is valid only in this case.) It is also valid when $\vec{\tau}_{\text{ext}}$ and $\vec{\mathbf{L}}$ are calculated about a point which is moving uniformly in an inertial reference frame since such

a point can be considered the origin of a second inertial reference frame. It is *not* valid in general when $\vec{\tau}_{\text{ext}}$ and $\vec{\mathbf{L}}$ are calculated about a point that is *accelerating*, except for one special (and very important) case—when that point is the center of mass (CM) of the system:

$$\frac{d\vec{\mathbf{L}}_{\text{CM}}}{dt} = \sum \vec{\tau}_{\text{CM}}. \qquad \text{[even if accelerating]} \quad \textbf{(11–9b)}$$

NEWTON'S SECOND LAW (for CM even if accelerating)

Equation 11–9b is valid no matter how the CM moves, and $\Sigma\boldsymbol{\tau}_{\text{CM}}$ is the net external torque calculated about the center of mass. The derivation is in the optional subsection below.

It is because of the validity of Eq. 11–9b that we are justified in describing the general motion of a system of particles, as we did in Chapter 10, as *translational* motion of the center of mass plus *rotation* about the center of mass. Equations 11–9b plus 9–5 $\left(d\vec{\mathbf{P}}_{\text{CM}}/dt = \Sigma\vec{\mathbf{F}}_{\text{ext}}\right)$ provide the more general statement of this principle. (See also Section 9–8.)

*Derivation of $d\vec{\mathbf{L}}_{\text{CM}}/dt = \Sigma\vec{\tau}_{\text{CM}}$

The proof of Eq. 11–9b is as follows. Let $\vec{\mathbf{r}}_i$ be the position vector of the i^{th} particle in an inertial reference frame, and $\vec{\mathbf{r}}_{\text{CM}}$ be the position vector of the center of mass of the system in this reference frame. The position of the i^{th} particle with respect to the CM is $\vec{\mathbf{r}}_i^*$ where (see Fig. 11–14)

$$\vec{\mathbf{r}}_i = \vec{\mathbf{r}}_{\text{CM}} + \vec{\mathbf{r}}_i^*.$$

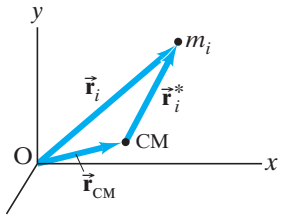


FIGURE 11–14 The position of m_i in the inertial frame is $\vec{\mathbf{r}}_i$; with regard to the CM (which could be accelerating) it is $\vec{\mathbf{r}}_i^*$, where $\vec{\mathbf{r}}_i = \vec{\mathbf{r}}_i^* + \vec{\mathbf{r}}_{\text{CM}}$ and $\vec{\mathbf{r}}_{\text{CM}}$ is the position of the CM in the inertial frame.

If we take the derivative of this equation and multiply each term by m_i, we can write the momentum, $p = mv$, of the i^{th} particle as

$$\vec{\mathbf{p}}_i = m_i\frac{d\vec{\mathbf{r}}_i}{dt} = m_i\frac{d}{dt}\left(\vec{\mathbf{r}}_i^* + \vec{\mathbf{r}}_{\text{CM}}\right) = m_i\vec{\mathbf{v}}_i^* + m_i\vec{\mathbf{v}}_{\text{CM}} = \vec{\mathbf{p}}_i^* + m_i\vec{\mathbf{v}}_{\text{CM}}.$$

Thus the CM momentum is $\vec{\mathbf{p}}_i^* = \vec{\mathbf{p}}_i - m_i\vec{\mathbf{v}}_{\text{CM}}$, and the angular momentum with respect to the CM is

$$\vec{\mathbf{L}}_{\text{CM}} = \sum_i\left(\vec{\mathbf{r}}_i^* \times \vec{\mathbf{p}}_i^*\right) = \sum_i\vec{\mathbf{r}}_i^* \times \left(\vec{\mathbf{p}}_i - m_i\vec{\mathbf{v}}_{\text{CM}}\right).$$

Then, taking the time derivative, we have

$$\frac{d\vec{\mathbf{L}}_{\text{CM}}}{dt} = \sum_i\left(\frac{d\vec{\mathbf{r}}_i^*}{dt} \times \vec{\mathbf{p}}_i^*\right) + \sum_i\left(\vec{\mathbf{r}}_i^* \times \frac{d}{dt}\left(\vec{\mathbf{p}}_i - m_i\vec{\mathbf{v}}_{\text{CM}}\right)\right).$$

The first term on the right is $\vec{\mathbf{v}}_i^* \times m\vec{\mathbf{v}}_i^*$ and equals zero because $\vec{\mathbf{v}}_i^*$ is parallel to itself $(\sin\theta = 0)$. Thus

$$\begin{aligned}\frac{d\vec{\mathbf{L}}_{\text{CM}}}{dt} &= \sum_i\vec{\mathbf{r}}_i^* \times \frac{d}{dt}\left(\vec{\mathbf{p}}_i - m_i\vec{\mathbf{v}}_{\text{CM}}\right)\\ &= \sum_i\vec{\mathbf{r}}_i^* \times \frac{d\vec{\mathbf{p}}_i}{dt} - \left(\sum_i m_i\vec{\mathbf{r}}_i^*\right) \times \frac{d\vec{\mathbf{v}}_{\text{CM}}}{dt}.\end{aligned}$$

The second term on the right is zero since, by Eq. 9–12, $\Sigma m_i\vec{\mathbf{r}}_i^* = M\vec{\mathbf{r}}_{\text{CM}}^*$, and $\vec{\mathbf{r}}_{\text{CM}}^* = 0$ by definition (the position of the CM is at the origin of the CM reference frame). Furthermore, by Newton's second law we have

$$\frac{d\vec{\mathbf{p}}_i}{dt} = \vec{\mathbf{F}}_i,$$

where $\vec{\mathbf{F}}_i$ is the net force on m_i. (Note that $d\vec{\mathbf{p}}_i^*/dt \neq \vec{\mathbf{F}}_i$ because the CM may be accelerating and Newton's second law does not hold in a noninertial reference frame.) Consequently

$$\frac{d\vec{\mathbf{L}}_{\text{CM}}}{dt} = \sum_i\vec{\mathbf{r}}_i^* \times \vec{\mathbf{F}}_i = \sum_i\left(\vec{\tau}_i\right)_{\text{CM}} = \sum\vec{\tau}_{\text{CM}},$$

where $\Sigma\vec{\tau}_{\text{CM}}$ is the resultant external torque on the entire system calculated about the CM. (By Newton's third law, the sum over all the $\vec{\tau}_i$ eliminates the net torque due to internal forces, as we saw on page 310.) This last equation is Eq. 11–9b, and this concludes its proof.

Limits on Eq. 11–9

To summarize, the relation

$$\sum \vec{\tau}_{\text{ext}} = \frac{d\vec{\mathbf{L}}}{dt}$$

is valid *only* when $\vec{\tau}_{\text{ext}}$ and $\vec{\mathbf{L}}$ are calculated with respect to either (1) the origin of an inertial reference frame or (2) the center of mass of a system of particles (or of a rigid object).

CAUTION

This equation is valid under certain circumstances

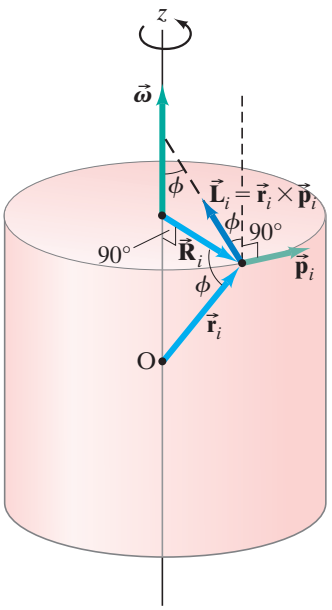


FIGURE 11–15
Calculating $L_\omega = L_z = \Sigma L_{iz}$. Note that $\vec{\mathbf{L}}_i$ is perpendicular to $\vec{\mathbf{r}}_i$, and $\vec{\mathbf{R}}_i$ is perpendicular to the z axis, so the three angles marked ϕ are equal.

11–5 Angular Momentum and Torque for a Rigid Object

Let us now consider the rotation of any rigid object about an axis that has a fixed direction in space (not necessarily through the CM), using the general principles just developed.

Let us calculate the component of angular momentum along the rotation axis (z) of the rotating object. We will call this component L_ω since the angular velocity $\vec{\boldsymbol{\omega}}$ points along the rotation axis. For each particle of the object,

$$\vec{\mathbf{L}}_i = \vec{\mathbf{r}}_i \times \vec{\mathbf{p}}_i.$$

Let ϕ be the angle between $\vec{\mathbf{L}}_i$ and the rotation axis. See Fig. 11–15, where we chose the rotation axis as the z axis (but ϕ is *not* the angle between $\vec{\mathbf{r}}_i$ and $\vec{\mathbf{p}}_i$, which is 90°). Then the component of $\vec{\mathbf{L}}_i$ along the rotation axis is

$$L_{i\omega} = r_i p_i \cos\phi = m_i v_i r_i \cos\phi,$$

where m_i is the mass and v_i the velocity of the i^{th} particle. Now $v_i = R_i\omega$ where ω is the angular velocity of the object and R_i is the perpendicular distance of m_i from the axis of rotation. Furthermore, $R_i = r_i \cos\phi$, as can be seen in Fig. 11–15, so

$$L_{i\omega} = m_i v_i (r_i \cos\phi) = m_i R_i^2 \omega.$$

We sum over all the particles to obtain

$$L_\omega = \sum_i L_{i\omega} = \left(\sum_i m_i R_i^2\right)\omega.$$

But $\Sigma m_i R_i^2$ is the moment of inertia I of the object about the axis of rotation. Therefore the component of the total angular momentum along the rotation axis is given by

$$L_\omega = I\omega. \qquad \textbf{(11–10)}$$

CAUTION

The component of $\vec{\mathbf{L}}$ along rotation axis is $L_\omega = I\omega$.

Note that we would obtain Eq. 11–10 no matter where we chose the point O (for measuring $\vec{\mathbf{r}}_i$) as long as it is on the axis of rotation. Equation 11–10 is the same as Eq. 11–1, which we have now proved from the general definition of angular momentum.

Although not true in general, L_ω is the only component of $\vec{\mathbf{L}}$ if the object rotates about a *symmetry* axis through the center of mass, as we now show. For each point on one side of the axis there will be a corresponding point on the opposite side. We can see from Fig. 11–15 that each $\vec{\mathbf{L}}_i$ has a component parallel to the axis $(L_{i\omega})$ and a component perpendicular to the axis. The components parallel to the axis add together for each pair of opposite points, but the components perpendicular to the axis for opposite points will have the same magnitude but opposite direction and so will cancel. Hence, for an object rotating about a symmetry axis, the angular momentum vector is parallel to the axis and we can write

$$\vec{\mathbf{L}} = I\vec{\boldsymbol{\omega}}, \qquad \text{[rotation axis = symmetry axis, through CM]} \quad \textbf{(11–11)}$$

where $\vec{\mathbf{L}}$ is measured relative to the center of mass.

The general relation between angular momentum and torque is Eq. 11–9:

$$\Sigma\vec{\tau} = \frac{d\vec{\mathbf{L}}}{dt}$$

where $\Sigma\vec{\tau}$ and $\vec{\mathbf{L}}$ are calculated either about (1) the origin of an inertial reference frame, or (2) the center of mass of the system. This is a vector relation, and must therefore be valid for each component. Hence, for a rigid object, the component along the rotation axis is

$$\Sigma\tau_{\text{axis}} = \frac{dL_\omega}{dt} = \frac{d}{dt}(I\omega) = I\frac{d\omega}{dt} = I\alpha,$$

which is valid for a rigid object rotating about an axis fixed relative to the object; also this axis must be either (1) fixed in an inertial system or (2) passing through the CM of the object. This is equivalent to Eqs. 10–14 and 10–15, which we now see are special cases of Eq. 11–9, $\Sigma\vec{\tau} = d\vec{\mathbf{L}}/dt$.

EXAMPLE 11–8 **Atwood's machine.** An *Atwood machine* consists of two masses, m_A and m_B, which are connected by an inelastic cord of negligible mass that passes over a pulley, Fig. 11–16. If the pulley has radius R_0 and moment of inertia I about its axle, determine the acceleration of the masses m_A and m_B, and compare to the situation where the moment of inertia of the pulley is ignored. Ignore friction at the axle O.

APPROACH We first determine the angular momentum of the system, and then apply Newton's second law, $\tau = dL/dt$.

SOLUTION We can calculate the angular momentum about an axis along the axle through the center O of the pulley. The pulley has angular momentum $I\omega$, where $\omega = v/R_0$ and v is the velocity of m_A and m_B at any instant. The angular momentum of m_A is $R_0 m_A v$ and that of m_B is $R_0 m_B v$. The total angular momentum is

$$L = (m_A + m_B)\,vR_0 + I\frac{v}{R_0}.$$

The external torque on the system, calculated about the axis O (taking clockwise as positive), is

$$\tau = m_B g R_0 - m_A g R_0.$$

(The normal force on the pulley exerted by the axle gives rise to no torque because the lever arm is zero.) We apply Eq. 11–9a:

$$\tau = \frac{dL}{dt}$$

$$(m_B - m_A)\,gR_0 = (m_A + m_B)\,R_0\frac{dv}{dt} + \frac{I}{R_0}\frac{dv}{dt}.$$

Solving for $a = dv/dt$, we get

$$a = \frac{dv}{dt} = \frac{(m_B - m_A)\,g}{(m_A + m_B) + I/R_0^2}.$$

If we were to ignore I, $a = (m_B - m_A)\,g/(m_B + m_A)$ and we see that the effect of the moment of inertia of the pulley is to slow down the system. This is just what we would expect.

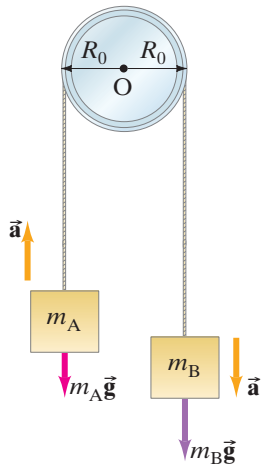


FIGURE 11–16 Atwood's machine, Example 11–8. We also discussed this in Example 4–13. We assume in this diagram that $m_B > m_A$.

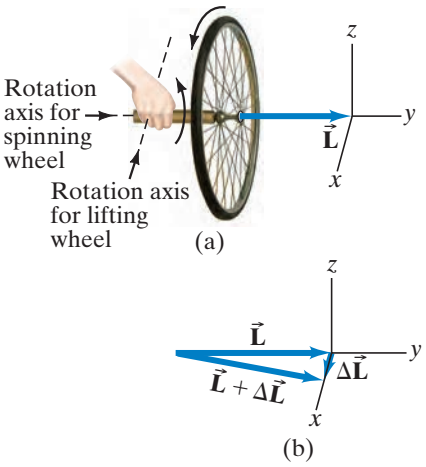


FIGURE 11–17 When you try to tilt a rotating bicycle wheel vertically upward, it swerves to the side instead.

CONCEPTUAL EXAMPLE 11–9 **Bicycle wheel.** Suppose you are holding a bicycle wheel by a handle connected to its axle as in Fig. 11–17a. The wheel is spinning so that its angular momentum $\vec{\mathbf{L}}$ points horizontally (along the y axis) as shown. Now you suddenly try to tilt the axle upward (along the z axis, so the CM moves vertically). You expect the wheel to go up, and it would if it weren't rotating. But it unexpectedly swerves to the right! Explain. Take the system as wheel, axle, and handle; your hand is external.

RESPONSE To explain this seemingly odd behavior—you may need to do it to believe it—we only need to use the relation $\vec{\boldsymbol{\tau}}_{\text{ext}} = d\vec{\mathbf{L}}/dt$. In the short time Δt, you exert an external torque (about an axis through your wrist) that points along the x axis perpendicular to $\vec{\mathbf{L}}$. Thus the change in $\vec{\mathbf{L}}$ is (using $\tau_{\text{ext}} \approx \Delta L/\Delta t$)

$$\Delta\vec{\mathbf{L}} \approx \vec{\boldsymbol{\tau}}_{\text{ext}}\,\Delta t;$$

so $\Delta\vec{\mathbf{L}}$ must point (approximately) along the x axis, because $\vec{\boldsymbol{\tau}}_{\text{ext}}$ does. Thus, if you are looking along the $+y$ axis, the new angular momentum, $\vec{\mathbf{L}} + \Delta\vec{\mathbf{L}}$, points to the right as shown in Fig. 11–17b. Since the angular momentum is directed along the axle of the wheel, we see that the axle, which now is along $\vec{\mathbf{L}} + \Delta\vec{\mathbf{L}}$, must move sideways to the right, which is what we observe.

CAUTION
$\vec{\mathbf{L}} = I\vec{\boldsymbol{\omega}}$ is not always valid

Although Eq. 11–11, $\vec{\mathbf{L}} = I\vec{\boldsymbol{\omega}}$, is not valid in general, it is valid in some useful cases, such as when the rotation axis is along a *symmetry* axis through the center of mass. Also, it can be shown that every rigid object, no matter what its shape, has three "principal axes" about which Eq. 11–11 is valid (we will not go into the details here). As an example of a case where Eq. 11–11 is not valid, consider the nonsymmetrical object shown in Fig. 11–18. It consists of two equal masses, m_{A} and m_{B}, attached to the ends of a rigid (massless) rod which makes an angle ϕ with the axis of rotation. We calculate the angular momentum about the CM at point O. At the instant shown, m_{A} is coming toward the viewer, and m_{B} is moving away, so $\vec{\mathbf{L}}_{\text{A}} = \vec{\mathbf{r}}_{\text{A}} \times \vec{\mathbf{p}}_{\text{A}}$ and $\vec{\mathbf{L}}_{\text{B}} = \vec{\mathbf{r}}_{\text{B}} \times \vec{\mathbf{p}}_{\text{B}}$ are as shown. The total angular momentum is $\vec{\mathbf{L}} = \vec{\mathbf{L}}_{\text{A}} + \vec{\mathbf{L}}_{\text{B}}$, which is clearly *not* along $\vec{\boldsymbol{\omega}}$ if $\phi \neq 90°$.

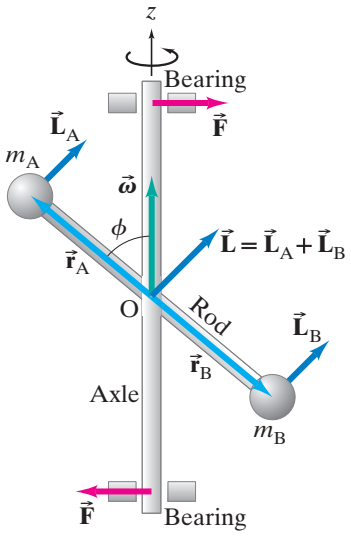


FIGURE 11–18 In this system $\vec{\mathbf{L}}$ and $\vec{\boldsymbol{\omega}}$ are not parallel. This is an example of rotational imbalance. The forces $\vec{\mathbf{F}}$ act on the axle, exerted by the bearing.

*Rotational Imbalance

Let us go one step further with the system shown in Fig. 11–18, since it is a fine illustration of $\Sigma\vec{\boldsymbol{\tau}}_{\text{ext}} = d\vec{\mathbf{L}}/dt$. If the system (axle, rod, and two masses) rotates with constant angular velocity, ω, the magnitude of $\vec{\mathbf{L}}$ will not change, but its direction will. As the rod and two masses rotate about the z axis, $\vec{\mathbf{L}}$ also rotates about the axis. At the moment shown in Fig. 11–18, $\vec{\mathbf{L}}$ is in the plane of the paper. A time dt later, when the rod has rotated through an angle $d\theta = \omega\,dt$, $\vec{\mathbf{L}}$ will also have rotated through an angle $d\theta$ ($\vec{\mathbf{L}}$ remains perpendicular to the rod). $\vec{\mathbf{L}}$ will then have a component pointing into the page. Thus $d\vec{\mathbf{L}}$ points into the page and so must $d\vec{\mathbf{L}}/dt$. From

$$\Sigma\vec{\boldsymbol{\tau}}_{\text{ext}} = \frac{d\vec{\mathbf{L}}}{dt},$$

we see that an external net torque, directed into the page at the moment shown, must be applied to the axle on which the rod is mounted. The torque is supplied by bearings (or other constraint) at the ends of the axle. The forces $\vec{\mathbf{F}}$ exerted by the bearings on the axle are shown in Fig. 11–18. The direction of each force $\vec{\mathbf{F}}$ rotates as the system does, always being in the plane of $\vec{\mathbf{L}}$ and $\vec{\boldsymbol{\omega}}$ for this system. If the torque due to these forces were not present, the system would not rotate about the fixed axis as desired.

PHYSICS APPLIED
Automobile wheel balancing

The axle tends to move in the direction of $\vec{\mathbf{F}}$ and thus tends to wobble as it rotates. This has many practical applications, such as wear on bearings and the vibrations felt in a car whose wheels are not balanced. Consider an automobile wheel that is symmetrical except for an extra mass m_{A} on one rim and an equal mass m_{B} opposite it on the other rim, as shown in Fig. 11–19. Because of the nonsymmetry of m_{A} and m_{B}, the wheel bearings would have to exert a force perpendicular to the axle at all times simply to keep the wheel rotating, just as in Fig. 11–18.

The bearings would wear excessively and the wobble of the wheel would be felt by occupants of the car. When the wheels are balanced, they rotate smoothly without wobble. This is why "dynamic balancing" of automobile wheels and tires is important. The wheel of Fig. 11–19 would balance *statically* just fine. If equal masses m_C and m_D are added symmetrically, below m_A and above m_B, the wheel will be balanced dynamically as well ($\vec{\mathbf{L}}$ will be parallel to $\vec{\boldsymbol{\omega}}$, and $\vec{\boldsymbol{\tau}}_{ext} = 0$).

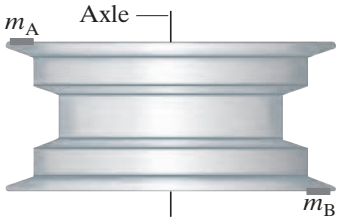
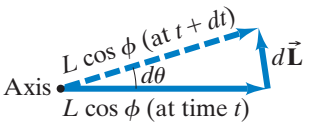


FIGURE 11–19 Unbalanced automobile wheel.

EXAMPLE 11–10 **Torque on unbalanced system.** Determine the magnitude of the external torque τ_{ext} needed to keep the system of axle, rod, and two masses turning in Fig. 11–18.

APPROACH Figure 11–20 is a view of the angular momentum vector, looking down along the rotation axis (z axis) of the system depicted in Fig. 11–18, as it rotates. (Think of your head being at the top of Fig. 11–18 and you are looking down along the z axis toward the bottom of the diagram.) $L\cos\phi$ is the component of $\vec{\mathbf{L}}$ perpendicular to the axle (points to the right in Fig. 11–18). We find dL from Fig. 11–20 and use $\tau_{ext} = dL/dt$.

FIGURE 11–20 Angular momentum vector looking down along the rotation axis (z) of the system of Fig. 11–18 as it rotates during a time dt.



SOLUTION In a time dt, $\vec{\mathbf{L}}$ changes by an amount (Fig. 11–20 and Eq. 10–2b, $\omega = d\theta/dt$)

$$dL = (L\cos\phi)\,d\theta = L\cos\phi\,\omega\,dt.$$

Hence

$$\tau_{ext} = \frac{dL}{dt} = \omega L\cos\phi.$$

But

$$L = L_A + L_B = r_A m_A v_A + r_B m_B v_B$$
$$= r_A m_A(\omega r_A \sin\phi) + r_B m_B(\omega r_B \sin\phi) = (m_A r_A^2 + m_B r_B^2)\,\omega\sin\phi.$$

The moment of inertia about the axis of rotation is $I = (m_A r_A^2 + m_B r_B^2)\sin^2\phi$, since the distance of each mass from the rotation axis is $R = r\sin\phi$. Thus $L = I\omega/\sin\phi$ and

$$\tau_{ext} = \omega L\cos\phi = (m_A r_A^2 + m_B r_B^2)\omega^2\sin\phi\cos\phi = I\omega^2/\tan\phi.$$

The situation of Fig. 11–18 illustrates the usefulness of the vector nature of torque and angular momentum. If we had considered only the components of angular momentum and torque along the rotation axis, we could not have calculated the external torque due to the bearings (since the forces $\vec{\mathbf{F}}$ act at the axle and hence produce no torque along that axis). By using the concept of vector angular momentum we have a far more powerful technique for understanding nature, and for attacking problems.

11–6 Conservation of Angular Momentum

In Chapter 9 we saw that the most general form of Newton's second law for the translational motion of a particle or system of particles is

$$\Sigma\vec{\mathbf{F}}_{ext} = \frac{d\vec{\mathbf{P}}}{dt},$$

where $\vec{\mathbf{P}}$ is the (linear) momentum, defined as $m\vec{\mathbf{v}}$ for a particle, or $M\vec{\mathbf{v}}_{CM}$ for a system of particles of total mass M whose CM moves with velocity $\vec{\mathbf{v}}_{CM}$, and $\Sigma\vec{\mathbf{F}}_{ext}$ is the net external force acting on the particle or system. This relation is valid only in an inertial reference frame.

In this Chapter, we have found a similar relation to describe the general rotation of a system of particles (including rigid objects):

$$\Sigma\vec{\boldsymbol{\tau}}_{ext} = \frac{d\vec{\mathbf{L}}}{dt},$$

where $\Sigma\vec{\boldsymbol{\tau}}_{ext}$ is the net external torque acting on the system, and $\vec{\mathbf{L}}$ is the total angular momentum. This relation is valid when $\Sigma\vec{\boldsymbol{\tau}}_{ext}$ and $\vec{\mathbf{L}}$ are calculated about a point fixed in an inertial reference frame, or about the CM of the system. It is a more general form of Eq. 11–2.

For translational motion, if the net force on the system is zero, $d\vec{\mathbf{P}}/dt = 0$, so the total linear momentum of the system remains constant. This is the law of conservation of linear momentum. For rotational motion, if the net external torque on the system is zero, then

$$\frac{d\vec{\mathbf{L}}}{dt} = 0 \quad \text{and} \quad \vec{\mathbf{L}} = \text{constant.} \qquad [\Sigma\vec{\tau}_{\text{ext}} = 0] \quad \textbf{(11–12)}$$

CONSERVATION OF ANGULAR MOMENTUM

In words:

The total angular momentum of a system remains constant if the net external torque acting on the system is zero.

This is the **law of conservation of angular momentum** in full vector form. It ranks with the laws of conservation of energy and linear momentum (and others to be discussed later) as one of the great laws of physics. In Section 11–1 we saw some Examples of this important law applied to the special case of a rigid object rotating about a fixed axis. Here we have it in general form. We use it now in interesting Examples.

EXAMPLE 11–11 Kepler's second law derived. Kepler's second law states that each planet moves so that a line from the Sun to the planet sweeps out equal areas in equal times (Section 6–5). Use conservation of angular momentum to show this.

APPROACH We determine the angular momentum of a planet in terms of the area swept out with the help of Fig. 11–21.

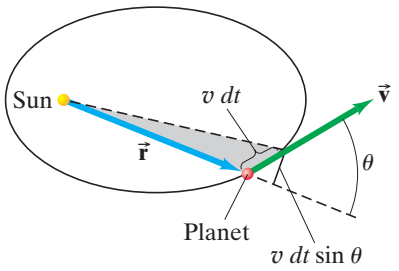


FIGURE 11–21 Kepler's second law of planetary motion.

SOLUTION The planet moves in an ellipse as shown in Fig. 11–21. In an infinitesimal time dt, the planet moves a distance $v\,dt$ and sweeps out an area dA equal to the area of a triangle of base r and height $v\,dt\sin\theta$ (shown exaggerated in Fig. 11–21). Hence

$$dA = \tfrac{1}{2}(r)(v\,dt\sin\theta)$$

and

$$\frac{dA}{dt} = \tfrac{1}{2}rv\sin\theta.$$

The magnitude of the angular momentum $\vec{\mathbf{L}}$ about the Sun is

$$L = |\vec{\mathbf{r}} \times m\vec{\mathbf{v}}| = mrv\sin\theta,$$

so

$$\frac{dA}{dt} = \frac{1}{2m}L.$$

But $L =$ constant, since the gravitational force $\vec{\mathbf{F}}$ is directed toward the Sun so the torque it produces is zero (we ignore the pull of the other planets). Hence $dA/dt =$ constant, which is what we set out to prove.

EXAMPLE 11–12 **Bullet strikes cylinder edge.** A bullet of mass m moving with velocity v strikes and becomes embedded at the edge of a cylinder of mass M and radius R_0, as shown in Fig. 11–22. The cylinder, initially at rest, begins to rotate about its symmetry axis, which remains fixed in position. Assuming no frictional torque, what is the angular velocity of the cylinder after this collision? Is kinetic energy conserved?

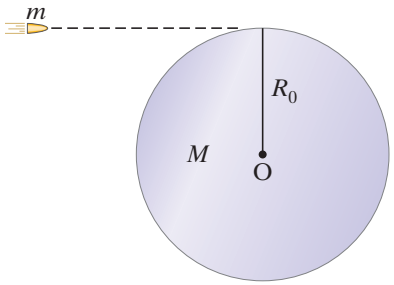


FIGURE 11–22 Bullet strikes and becomes embedded in cylinder at its edge (Example 11–12).

APPROACH We take as our system the bullet and cylinder, on which there is no net external torque. Thus we can use conservation of angular momentum, and we calculate all angular momenta about the center O of the cylinder.

SOLUTION Initially, because the cylinder is at rest, the total angular momentum about O is solely that of the bullet, of magnitude

$$L = |\vec{\mathbf{r}} \times \vec{\mathbf{p}}| = R_0 mv,$$

since R_0 is the perpendicular distance of $\vec{\mathbf{p}}$ from O. After the collision, the cylinder $\left(I_{\text{cyl}} = \frac{1}{2}MR_0^2\right)$ rotates with the bullet $\left(I_{\text{b}} = mR_0^2\right)$ embedded in it at angular velocity ω:

$$L = I\omega = \left(I_{\text{cyl}} + mR_0^2\right)\omega = \left(\tfrac{1}{2}M + m\right)R_0^2\omega.$$

Hence, because angular momentum is conserved, we find that ω is

$$\omega = \frac{L}{\left(\frac{1}{2}M + m\right)R_0^2} = \frac{mvR_0}{\left(\frac{1}{2}M + m\right)R_0^2} = \frac{mv}{\left(\frac{1}{2}M + m\right)R_0}.$$

Angular momentum is conserved in this collision. But kinetic energy, which initially is only that of the bullet, $K_{\text{i}} = \frac{1}{2}mv^2$ is not conserved:

$$\begin{aligned} K_{\text{f}} - K_{\text{i}} &= \tfrac{1}{2}I_{\text{cyl}}\omega^2 + \tfrac{1}{2}\left(mR_0^2\right)\omega^2 - \tfrac{1}{2}mv^2 \\ &= \tfrac{1}{2}\left(\tfrac{1}{2}MR_0^2\right)\omega^2 + \tfrac{1}{2}\left(mR_0^2\right)\omega^2 - \tfrac{1}{2}mv^2 \\ &= \tfrac{1}{2}\left(\tfrac{1}{2}M + m\right)\left(\frac{mv}{\frac{1}{2}M + m}\right)^2 - \tfrac{1}{2}mv^2 \\ &= -\frac{mM}{2M + 4m}v^2, \end{aligned}$$

which is less than zero. Hence $K_{\text{f}} < K_{\text{i}}$. This energy is transformed to thermal energy as a result of the inelastic collision.

*11–7 The Spinning Top and Gyroscope

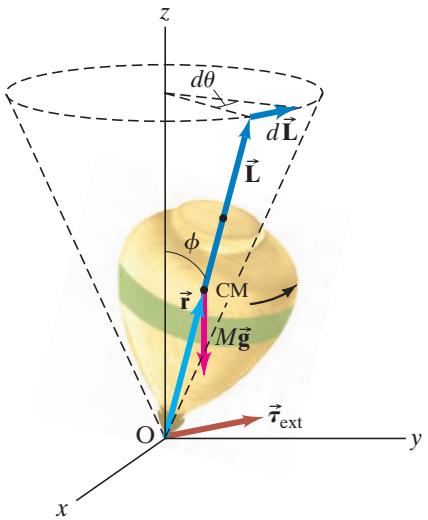
The motion of a spinning top, or a gyroscope, is another interesting example of rotational motion and of the use of the vector equation

$$\sum \vec{\boldsymbol{\tau}}_{\text{ext}} = \frac{d\vec{\mathbf{L}}}{dt}.$$

Consider a symmetrical top of mass M spinning rapidly about its symmetry axis, as in Fig. 11–23. The top is balanced on its tip at point O in an inertial reference frame. If the axis of the top makes an angle ϕ to the vertical (z axis), when the top is carefully released its axis will move, sweeping out a cone about the vertical as shown by the dashed lines in Fig. 11–23. This type of motion, in which a torque produces a change in the direction of the rotation axis, is called **precession**. The rate at which the rotation axis moves about the vertical (z) axis is called the angular velocity of precession, Ω (capital Greek omega). Let us now try to understand the reasons for this motion, and calculate Ω.

If the top was not spinning, it would immediately fall to the ground when released due to the pull of gravity. The apparent mystery of a top is that when it is spinning, it does not immediately fall to the ground but instead precesses—its rotation axis moves slowly around the vertical (z) axis. But this is not really so mysterious if we examine it from the point of view of angular momentum and torque, which we calculate about the point O.

FIGURE 11–23 Spinning top. It is difficult to release a spinning top so that it spins around a perfectly vertical axis. It usually makes some angle ϕ to the vertical.



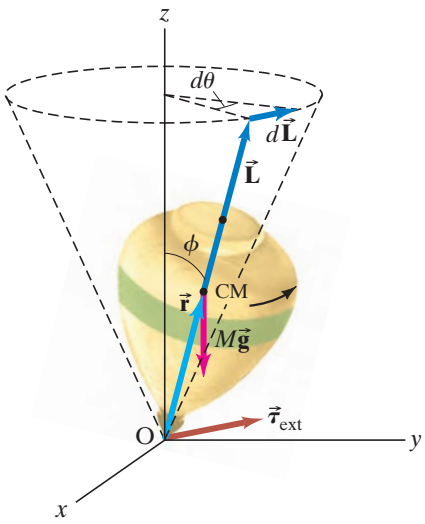


FIGURE 11–23 (Repeated.) A spinning top precesses.

When the top is spinning with angular velocity ω about its symmetry axis, it has an angular momentum $\vec{\mathbf{L}}$ directed along its axis, as shown in Fig. 11–23. (There is also angular momentum due to the precessional motion, so that the total $\vec{\mathbf{L}}$ is not exactly along the axis of the top; but if $\Omega \ll \omega$, which is usually the case, we can ignore this.) To change the angular momentum, a torque is required. If no torque were applied to the spinning top, $\vec{\mathbf{L}}$ would have to remain constant in magnitude and direction; the top would neither fall nor precess. But the slightest tip to the side (angle ϕ) means the spinning top experiences a net external torque about O, equal to $\vec{\boldsymbol{\tau}}_{\text{ext}} = \vec{\mathbf{r}} \times M\vec{\mathbf{g}}$, where $\vec{\mathbf{r}}$ is the position vector of the top's center of mass with respect to O, and M is the mass of the top. The direction of $\vec{\boldsymbol{\tau}}_{\text{ext}}$ is perpendicular to both $\vec{\mathbf{r}}$ and $M\vec{\mathbf{g}}$—and by the right-hand rule is in the horizontal (xy) plane, as shown in Fig. 11–23. The change in $\vec{\mathbf{L}}$ in a time dt is

$$d\vec{\mathbf{L}} = \vec{\boldsymbol{\tau}}_{\text{ext}}\, dt,$$

which is perpendicular to $\vec{\mathbf{L}}$ and horizontal (parallel to $\vec{\boldsymbol{\tau}}_{\text{ext}}$), as shown in Fig. 11–23. Since $d\vec{\mathbf{L}}$ is perpendicular to $\vec{\mathbf{L}}$, the magnitude of $\vec{\mathbf{L}}$ does not change. Only the direction of $\vec{\mathbf{L}}$ changes. Since $\vec{\mathbf{L}}$ points along the axis of the top, we see that this axis moves to the right in Fig. 11–23. That is, the upper end of the top's axis moves in a horizontal direction perpendicular to $\vec{\mathbf{L}}$. This explains why the top precesses rather than falls. The vector $\vec{\mathbf{L}}$ and the top's axis move together in a horizontal circle. As they do so, $\vec{\boldsymbol{\tau}}_{\text{ext}}$ and $d\vec{\mathbf{L}}$ rotate as well so as to be horizontal and perpendicular to $\vec{\mathbf{L}}$.

To determine Ω, we see from Fig. 11–23 that the angle $d\theta$ (which is in a horizontal plane) is related to dL by

$$dL = L \sin\phi\, d\theta,$$

since $\vec{\mathbf{L}}$ makes an angle ϕ to the z axis. The angular velocity of precession (using Ω, which is capital ω) is $\Omega = d\theta/dt$, which becomes (since $d\theta = dL/L\sin\phi$)

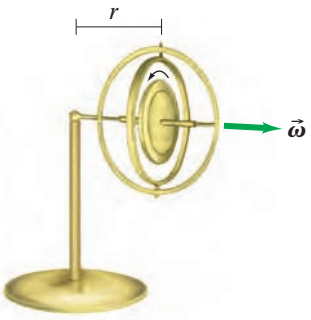
$$\Omega = \frac{1}{L\sin\phi}\frac{dL}{dt} = \frac{\tau}{L\sin\phi}. \qquad \text{[spinning top]} \quad \textbf{(11–13a)}$$

But $\tau_{\text{ext}} = |\vec{\mathbf{r}} \times M\vec{\mathbf{g}}| = rMg\sin\phi$ [because $\sin(\pi - \phi) = \sin\phi$] so we can also write

$$\Omega = \frac{Mgr}{L}. \qquad \text{[spinning top]} \quad \textbf{(11–13b)}$$

Thus the rate of precession does not depend on the angle ϕ; but it is inversely proportional to the top's angular momentum. The faster the top spins, the greater L is and the slower the top precesses.

FIGURE 11–24 A toy gyroscope.



From Eq. 11–1 (or Eq. 11–11) we can write $L = I\omega$, where I and ω are the moment of inertia and angular velocity of the spinning top about its spin axis. Then Eq. 11–13b for the top's precession angular velocity becomes

$$\Omega = \frac{Mgr}{I\omega}. \qquad \textbf{(11–13c)}$$

Equations 11–13 apply also to a toy gyroscope, which consists of a rapidly spinning wheel mounted on an axle (Fig. 11–24). One end of the axle rests on a support. The other end of the axle is free and will precess like a top if its "spin" angular velocity ω is large compared to the precession rate $(\omega \gg \Omega)$. As ω decreases due to friction and air resistance, the gyroscope will begin to fall, just as does a top.

11–8 Rotating Frames of Reference; Inertial Forces

Inertial and Noninertial Reference Frames

Up to now, we have examined the motion of objects, including circular and rotational motion, from the outside, as observers fixed on the Earth. Sometimes it is convenient to place ourselves (in theory, if not physically) into a reference frame that is rotating. Let us examine the motion of objects from the point of view, or frame of reference, of persons seated on a rotating platform such as a merry-go-round.

It looks to them as if the rest of the world is going around *them*. But let us focus attention on what they observe when they place a tennis ball on the floor of the rotating platform, which we assume is frictionless. If they put the ball down gently, without giving it any push, they will observe that it accelerates from rest and moves outward as shown in Fig. 11–25a. According to Newton's first law, an object initially at rest should stay at rest if no net force acts on it. But, according to the observers on the rotating platform, the ball starts moving even though there is no net force acting on it. To observers on the ground this is all very clear: the ball has an initial velocity when it is released (because the platform is moving), and it simply continues moving in a straight-line path as shown in Fig. 11–25b, in accordance with Newton's first law.

But what shall we do about the frame of reference of the observers on the rotating platform? Since the ball starts moving with no net force on it, Newton's first law, the law of inertia, does not hold in this rotating frame of reference. For this reason, such a frame is called a **noninertial reference frame**. An **inertial reference frame** (as discussed in Chapter 4) is one in which the law of inertia—Newton's first law—does hold, and so do Newton's second and third laws. In a noninertial reference frame, such as our rotating platform, Newton's second law also does not hold. For instance in the situation described above, there is no net force on the ball; yet, with respect to the rotating platform, the ball accelerates.

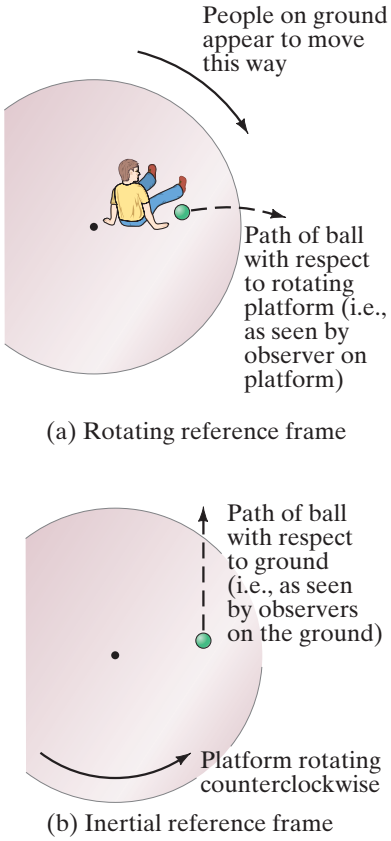


FIGURE 11–25 Path of a ball released on a rotating merry-go-round (a) in the reference frame of the merry-go-round, and (b) in a reference frame fixed on the ground.

Fictitious (Inertial) Forces

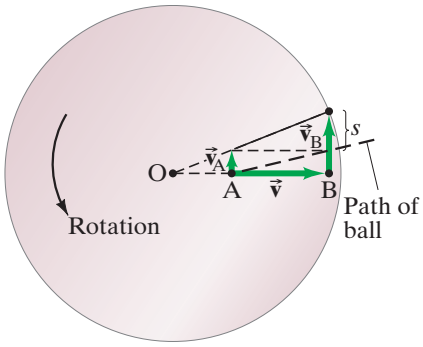
Because Newton's laws do not hold when observations are made with respect to a noninertial frame of reference, calculation of motion can be complicated. But there is a way around it for this rotating reference frame. The ball on the rotating platform of Fig. 11–25a flies outward when released (even though no force is actually acting on it). What we can do is write down the equation $\Sigma F = ma$ as if a force equal to mv^2/r (or $m\omega^2 r$) is acting radially outward on the object in addition to any other forces that may be acting. This extra force, which might be designated as "centrifugal force" since it *seems* to act outward, is called a **fictitious force** or **pseudoforce**. It is a pseudoforce ("pseudo" means "false") because there is no object that exerts this force. Furthermore, when viewed from an inertial reference frame, the effect doesn't exist at all (Fig. 11–25b). We have made up this pseudoforce so that we can make calculations in a noninertial frame using Newton's second law, $\Sigma F = ma$. Thus the observer in the noninertial frame of Fig. 11–25a uses Newton's second law for the ball's outward motion by assuming that a force equal to mv^2/r acts on it. Such pseudoforces are also called **inertial forces** because they "appear" only in a noninertial reference frame due to that frame's acceleration.

In Section 5–4 we discussed the forces on a person in a car going around a curve (Fig. 5–21) from the point of view of an inertial frame. The car, on the other hand, is not an inertial frame. Passengers in such a car could interpret this being pressed outward as the effect of a "centrifugal" force. But they need to recognize that it is a pseudoforce because it is an effect of being in a noninertial frame of reference, and there is no identifiable object exerting it.

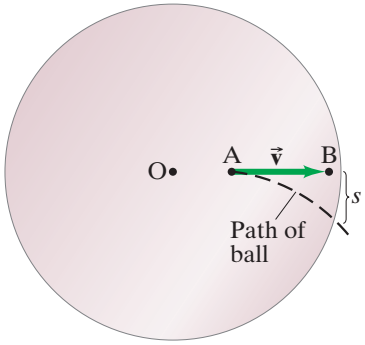
The Earth itself is rotating on its axis. Thus, strictly speaking, Newton's laws are not valid on the Earth. However, the effect of the Earth's rotation is usually so small that it can be ignored, although it does influence the movement of large air masses and ocean currents. Because of the Earth's rotation, the material of the Earth is concentrated slightly more at the equator. The Earth is thus not a perfect sphere but is slightly fatter at the equator than at the poles.

*11–9 The Coriolis Effect

In a reference frame that rotates at a constant angular speed ω (relative to an inertial frame), there exists another pseudoforce known as the *Coriolis force*. It appears to act on an object in a rotating reference frame only if the object is moving relative to that rotating reference frame, and it acts to deflect the object sideways. It, too, is an effect of the rotating reference frame being noninertial and hence is referred to as an *inertial force*. It also affects the weather.



(a) Inertial reference frame



(b) Rotating reference frame

FIGURE 11–26 The origin of the Coriolis effect. Looking down on a platform rotating counterclockwise, (a) as seen from a nonrotating inertial reference frame, and (b) as seen from the rotating platform as frame of reference.

To see how the Coriolis force arises, consider two people, A and B, at rest on a platform rotating with angular speed ω, as shown in Fig. 11–26a. They are situated at distances r_A and r_B from the axis of rotation (at O). The woman at A throws a ball with a horizontal velocity $\vec{\mathbf{v}}$ (in her reference frame) radially outward toward the man at B near the outer edge of the platform. In Fig. 11–26a, we view the situation from an inertial reference frame. The ball initially has not only the velocity $\vec{\mathbf{v}}$ radially outward, but also a tangential velocity $\vec{\mathbf{v}}_A$ due to the rotation of the platform. Now Eq. 10–4 tells us that $v_A = r_A\omega$, where r_A is the woman's radial distance from the axis of rotation at O. If the man at B had this same velocity v_A, the ball would reach him perfectly. But his speed is $v_B = r_B\omega$, which is greater than v_A because $r_B > r_A$. Thus, when the ball reaches the outer edge of the platform, it passes a point that the man at B has already gone by because his speed in that direction is greater than the ball's. So the ball passes behind him.

Figure 11–26b shows the situation as seen from the rotating platform as frame of reference. Both A and B are at rest, and the ball is thrown with velocity $\vec{\mathbf{v}}$ toward B, but the ball deflects to the right as shown and passes behind B as previously described. This is not a centrifugal-force effect, because that would act radially outward. Instead, this effect acts sideways, perpendicular to $\vec{\mathbf{v}}$, and is called a **Coriolis acceleration**; it is said to be due to the Coriolis force, which is a fictitious inertial force. Its explanation as seen from an inertial system was given above: it is an effect of being in a rotating system, for which a point farther from the rotation axis has a higher linear speed. On the other hand, when viewed from the rotating system, we can describe the motion using Newton's second law, $\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$, if we add a "pseudoforce" term corresponding to this Coriolis effect.

Let us determine the magnitude of the Coriolis acceleration for the simple case described above. (We assume v is large and distances short, so we can ignore gravity.) We do the calculation from the inertial reference frame (Fig. 11–26a). The ball moves radially outward a distance $r_B - r_A$ at speed v in a short time t given by

$$r_B - r_A = vt.$$

During this time, the ball moves to the side a distance s_A given by

$$s_A = v_A t.$$

The man at B, in this time t, moves a distance

$$s_B = v_B t.$$

The ball therefore passes behind him a distance s (Fig. 11–26a) given by

$$s = s_B - s_A = (v_B - v_A)t.$$

We saw earlier that $v_A = r_A\omega$ and $v_B = r_B\omega$, so

$$s = (r_B - r_A)\omega t.$$

We substitute $r_B - r_A = vt$ (see above) and get

$$s = \omega v t^2. \qquad \textbf{(11–14)}$$

This same s equals the sideways displacement as seen from the noninertial rotating system (Fig. 11–26b).

Equation 11–14 corresponds to motion at constant acceleration, because as we saw in Chapter 2 (Eq. 2–12b), $y = \frac{1}{2}at^2$ for a constant acceleration (with zero initial velocity in the y direction). Thus, if we write Eq. 11–14 in the form $s = \frac{1}{2}a_{Cor}t^2$, we see that the Coriolis acceleration a_{Cor} is

$$a_{Cor} = 2\omega v. \qquad \textbf{(11–15)}$$

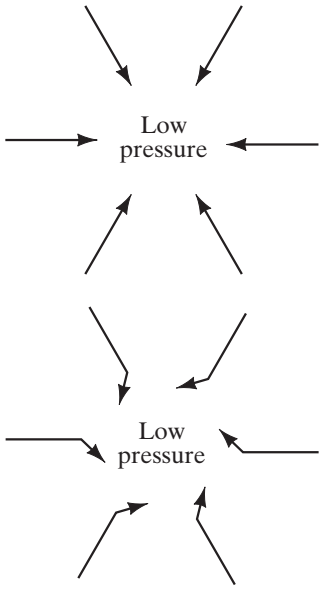
This relation is valid for any velocity in the plane of rotation perpendicular to the axis of rotation[†] (in Fig. 11–26, the axis through point O perpendicular to the page).

[†]The Coriolis acceleration can be written in general in terms of the vector cross product as $\vec{\mathbf{a}}_{Cor} = -2\vec{\boldsymbol{\omega}} \times \vec{\mathbf{v}}$ where $\vec{\boldsymbol{\omega}}$ has direction along the rotation axis; its magnitude is $a_{Cor} = 2\omega v_\perp$ where $v_\perp$ is the component of velocity perpendicular to the rotation axis.

Because the Earth rotates, the Coriolis effect has some interesting manifestations on the Earth. It affects the movement of air masses and thus has an influence on weather. In the absence of the Coriolis effect, air would rush directly into a region of low pressure, as shown in Fig. 11–27a. But because of the Coriolis effect, the winds are deflected to the right (Fig. 11–27b). This is true in the Northern Hemisphere, for which Fig. 11–26 applies as if looking down on Earth from above the North Pole and seeing the Earth rotate counterclockwise (the Earth rotates from west to east). So there tends to be a counterclockwise wind pattern around a low-pressure area. The reverse is true in the Southern Hemisphere (imagine looking down on Earth from above the South Pole and seeing the Earth rotate clockwise, the reverse of Fig. 11–26). Thus large storms, such as cyclones, hurricanes, and typhoons,[†] rotate counterclockwise in the Northern Hemisphere and clockwise in the Southern Hemisphere. The same effect explains the easterly trade winds near the equator: any winds heading south toward the equator will be deflected toward the west (that is, as if coming from the east).

In the vicinity of a *high*-pressure region, the air moves outward. So the winds, acted on by the Coriolis "force" to the right in the Northern Hemisphere, move in a clockwise direction. This is called an *anticyclone*, or anticyclonic behavior. In the Southern Hemisphere, winds circulate counterclockwise around a high-pressure, anticyclonic, region.

The Coriolis effect also acts on a falling object. An object released from the top of a high tower will not hit the ground directly below the release point, but will be deflected slightly to the east. Viewed from an inertial frame, this happens because the top of the tower revolves with the Earth at a slightly higher speed than the bottom of the tower.



(a)

(b)



(c)

FIGURE 11–27 (a) Winds (moving air masses) would flow directly toward a low-pressure area if the Earth did not rotate. (b) and (c): Because of the Earth's rotation, the winds are deflected to the right in the Northern Hemisphere (as in Fig. 11–26) as if a fictitious (Coriolis) force were acting.

[†]These three names are used in different places for the same phenomenon: a large cyclonic storm near Florida is a hurricane, if near Japan it is called a typhoon.

Summary

The **angular momentum** $\vec{\mathbf{L}}$ of a rigid object rotating about a fixed rotation axis is given by

$$L = I\omega. \qquad \textbf{(11–1)}$$

Newton's second law, in terms of angular momentum, is

$$\Sigma\tau = \frac{dL}{dt}. \qquad \textbf{(11–2)}$$

If the net torque on an object is zero, $dL/dt = 0$, so L = constant. This is the **law of conservation of angular momentum** for a rotating object.

The **vector product** or **cross product** of two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ is another vector $\vec{\mathbf{C}} = \vec{\mathbf{A}} \times \vec{\mathbf{B}}$ whose magnitude is $AB\sin\theta$ and whose direction is perpendicular to both $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ in the sense of the right-hand rule.

The **torque** $\vec{\boldsymbol{\tau}}$ due to a force $\vec{\mathbf{F}}$ is a vector quantity and is always calculated about some point O (the origin of a coordinate system) as follows:

$$\vec{\boldsymbol{\tau}} = \vec{\mathbf{r}} \times \vec{\mathbf{F}}, \qquad \textbf{(11–5)}$$

where $\vec{\mathbf{r}}$ is the position vector of the point at which the force $\vec{\mathbf{F}}$ acts.

Angular momentum is also a vector. For a particle having momentum $\vec{\mathbf{p}} = m\vec{\mathbf{v}}$, the angular momentum $\vec{\mathbf{L}}$ about some point O is

$$\vec{\mathbf{L}} = \vec{\mathbf{r}} \times \vec{\mathbf{p}}, \qquad \textbf{(11–6)}$$

where $\vec{\mathbf{r}}$ is the position vector of the particle relative to the point O at any instant. The net torque $\Sigma\vec{\boldsymbol{\tau}}$ on a particle is related to its angular momentum by

$$\Sigma\vec{\boldsymbol{\tau}} = \frac{d\vec{\mathbf{L}}}{dt}. \qquad \textbf{(11–7)}$$

For a system of particles, the total angular momentum $\vec{\mathbf{L}} = \Sigma\vec{\mathbf{L}}_i$. The total angular momentum of the system is related to the total net torque $\Sigma\vec{\boldsymbol{\tau}}$ on the system by

$$\Sigma\vec{\boldsymbol{\tau}}_{\text{ext}} = \frac{d\vec{\mathbf{L}}}{dt}. \qquad \textbf{(11–9a)}$$

This last relation is the vector rotational equivalent of Newton's second law. It is valid when $\vec{\mathbf{L}}$ and $\Sigma\vec{\boldsymbol{\tau}}_{\text{ext}}$ are calculated about an origin (1) fixed in an inertial reference system or (2) situated at the CM of the system. For a rigid object rotating about a fixed axis, the component of angular momentum about the rotation axis is given by $L_\omega = I\omega$. If an object rotates about an axis of symmetry, then the vector relation $\vec{\mathbf{L}} = I\vec{\boldsymbol{\omega}}$ holds, but this is not true in general.

If the total external torque on a system is zero, then the total vector angular momentum $\vec{\mathbf{L}}$ remains constant. This is the important **law of conservation of angular momentum**. It applies to the vector $\vec{\mathbf{L}}$ and therefore also to each of its components.

Questions

1. Can the diver of Fig. 11–2 do a somersault without having any initial rotation when she leaves the board? Explain.
2. When a motorcyclist leaves the ground on a jump and leaves the throttle on (so the rear wheel spins), why does the front of the cycle rise up?
3. Suppose you are standing on the edge of a large freely rotating turntable. What happens if you walk toward the center?
4. We claim that momentum and angular momentum are conserved. Yet most moving or rotating objects eventually slow down and stop. Explain.
5. A shortstop may leap into the air to catch a ball and throw it quickly. As he throws the ball, the upper part of his body rotates. If you look quickly you will notice that his hips and legs rotate in the opposite direction (Fig. 11–28). Explain.



FIGURE 11–28
Question 5. A shortstop in the air, throwing the ball.

6. If all the components of the vectors $\vec{\mathbf{V}}_1$ and $\vec{\mathbf{V}}_2$ were reversed in direction, how would this alter $\vec{\mathbf{V}}_1 \times \vec{\mathbf{V}}_2$?
7. A force $\vec{\mathbf{F}} = F\hat{\mathbf{j}}$ is applied to an object at a position $\vec{\mathbf{r}} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$ where the origin is at the CM. Does the torque about the CM depend on x? On y? On z?
8. A particle moves with constant speed along a straight line. How does its angular momentum, calculated about any point not on its path, change in time?
9. If the net force on a system is zero, is the net torque also zero? If the net torque on a system is zero, is the net force zero? Give examples.
10. Explain how a child "pumps" on a swing to make it go higher.
11. Describe the torque needed if the person in Fig. 11–17 is to tilt the axle of the rotating wheel directly upward without it swerving to the side.
12. An astronaut floats freely in a weightless environment. Describe how the astronaut can move her limbs so as to (*a*) turn her body upside down and (*b*) turn her body about-face.
13. On the basis of the law of conservation of angular momentum, discuss why a helicopter must have more than one rotor (or propeller). Discuss one or more ways the second propeller can operate in order to keep the helicopter stable.
14. A wheel is rotating freely about a vertical axis with constant angular velocity. Small parts of the wheel come loose and fly off. How does this affect the rotational speed of the wheel? Is angular momentum conserved? Is kinetic energy conserved? Explain.
15. Consider the following vector quantities: displacement, velocity, acceleration, momentum, angular momentum, torque. (*a*) Which of these are independent of the choice of origin of coordinates? (Consider different points as origin which are at rest with respect to each other.) (*b*) Which are independent of the (constant) velocity of the coordinate system?
16. How does a car make a right turn? Where does the torque come from that is needed to change the angular momentum?
17. In a rotating frame of reference, Newton's first and second laws remain useful if we assume that a pseudoforce equal to $m\omega^2 r$ is acting. What effect does this assumption have on the validity of Newton's third law?

*18. Why is it that at most locations on the Earth, a plumb bob does not hang precisely in the direction of the Earth's center?

*19. In the *battle* of the Falkland Islands in 1914, the shots of British gunners initially fell wide of their marks because their calculations were based on naval battles fought in the Northern Hemisphere. The Falklands are in the Southern Hemisphere. Explain the origin of their problem.

MisConceptual Questions

1. A small mass m on a string is rotating without friction in a circle. The string is shortened by pulling it through the axis of rotation without any external torque, Fig. 11–29. What happens to the angular velocity of the object?
(*a*) It increases.
(*b*) It decreases.
(*c*) It remains the same.

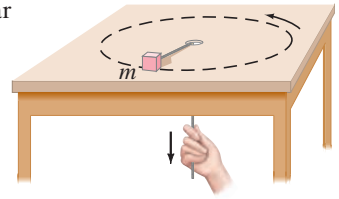


FIGURE 11–29
MisConceptual Questions 1 and 2.

2. A small mass m on a string is rotating without friction in a circle. The string is shortened by pulling it through the axis of rotation without any external torque, Fig. 11–29. What happens to the tangential velocity of the object?
(*a*) It increases.
(*b*) It decreases.
(*c*) It remains the same.

3. Suppose you are sitting on a rotating stool holding a 2-kg mass in each outstretched hand. If you suddenly drop the masses, your angular velocity will
(*a*) increase.
(*b*) decrease.
(*c*) stay the same.

4. Two spinning disks have the same angular momentum, but disk A has more kinetic energy than disk B. Which has the greater moment of inertia?
(*a*) Disk A.
(*b*) Disk B.
(*c*) Not enough information to tell.

5. The cross product of two vectors, $\vec{\mathbf{V}}_1 \times \vec{\mathbf{V}}_2$, would have to be zero if
(*a*) $V_1 = 0$.
(*b*) $V_2 = 0$.
(*c*) $\vec{\mathbf{V}}_1$ is parallel to $\vec{\mathbf{V}}_2$.
(*d*) $\vec{\mathbf{V}}_1$ is perpendicular to $\vec{\mathbf{V}}_2$.
(*e*) $\vec{\mathbf{V}}_1$ and $\vec{\mathbf{V}}_2$ are in the same plane.
(*f*) $\vec{\mathbf{V}}_1$ and $\vec{\mathbf{V}}_2$ make a 45° angle with each other.

6. If there were a great migration of people toward the Earth's equator, the length of the day would
(*a*) increase because of conservation of angular momentum.
(*b*) decrease because of conservation of angular momentum.
(*c*) decrease because of conservation of energy.
(*d*) increase because of conservation of energy.
(*e*) remain unaffected.

7. A particle moves with constant speed along a straight line. How does its angular momentum, calculated about any point O not on its path, change in time?
(*a*) Decreases as it approaches O, then increases.
(*b*) Increases as it approaches O, then decreases.
(*c*) Would remain the same.

8. The angular momentum vector can be written $\vec{\mathbf{L}} = I\vec{\boldsymbol{\omega}}$
(*a*) always.
(*b*) always for an object rotating about any axis.
(*c*) always for an object rotating about an axis fixed in space.
(*d*) for any rotation axis passing through the CM of the object.
(*e*) always for the component of $\vec{\mathbf{L}}$ that is along the (fixed) axis of rotation.

9. A person sits on a freely spinning chair (no friction), Fig. 11–30. When this person extends her arms,
(*a*) her moment of inertia increases and her angular velocity decreases.
(*b*) her moment of inertia decreases and her angular velocity decreases.
(*c*) her moment of inertia increases and her angular velocity increases.
(*d*) her moment of inertia increases and her angular velocity remains the same.
(*e*) her moment of inertia decreases and her angular velocity increases.

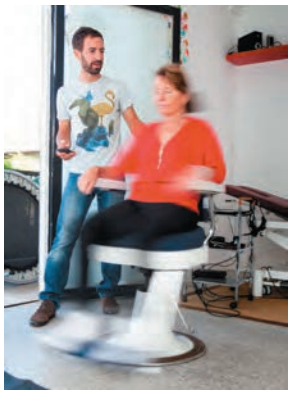


FIGURE 11–30 MisConceptual Question 9.

10. You are holding on to the ends of the horizontal axle of a spinning bicycle wheel with your hands. The top part of the wheel is rotating *toward* you. If you start to turn your whole body toward the left, you feel the right side of the axle
(*a*) pull on your right hand away from you.
(*b*) push on your right hand toward the left.
(*c*) push on your right hand vertically down.
(*d*) push on your right hand toward you.
(*e*) push on your right hand vertically up.

11. You are looking horizontally at a wheel which is rotating in a clockwise direction. Take the origin of coordinates at the center of the wheel. The x axis points toward you, the y axis points to the right, and the z axis is up. The direction of the angular momentum vector is
(*a*) clockwise.
(*b*) counterclockwise.
(*c*) $\hat{\mathbf{i}}$.
(*d*) $-\hat{\mathbf{i}}$.
(*e*) $\hat{\mathbf{j}}$.
(*f*) $-\hat{\mathbf{j}}$.
(*g*) $\hat{\mathbf{k}}$.
(*h*) $-\hat{\mathbf{k}}$.

Problems

11–1 Angular Momentum

1. (I) What is the angular momentum of a 0.210-kg ball revolving on the end of a thin string in a circle of radius 1.25 m at an angular speed of 10.4 rad/s?

2. (I) (*a*) What is the angular momentum of a 2.8-kg uniform cylindrical grinding wheel of radius 18 cm when rotating at 1500 rpm? (*b*) How much torque is required to stop it in 6.0 s?

3. (I) A nonrotating cylindrical disk of moment of inertia I is dropped onto an identical disk rotating at angular speed ω. Assuming no external torques, what is the final common angular speed of the two disks?

4. (II) A diver (such as the one shown in Fig. 11–2) can reduce her moment of inertia by a factor of about 3.5 when changing from the straight position to the tuck position. If she makes 2.0 rotations in 1.5 s when in the tuck position, what is her angular speed (rev/s) when in the straight position?

5. (II) A figure skater can increase her spin rotation rate from an initial rate of 1.0 rev every 1.5 s to a final rate of 2.5 rev/s. If her initial moment of inertia was $4.9\,\text{kg}\cdot\text{m}^2$, what is her final moment of inertia? How does she physically accomplish this change?

6. (II) A person stands, hands at his side, on a platform that is rotating at a rate of 0.80 rev/s. If he raises his arms to a horizontal position, Fig. 11–31, the speed of rotation decreases to 0.60 rev/s. (*a*) Why? (*b*) By what factor has his moment of inertia changed?



FIGURE 11–31
Problem 6.

7. (II) Determine the angular momentum of the Earth (*a*) about its rotation axis (assume the Earth is a uniform sphere), and (*b*) in its orbit around the Sun (treat the Earth as a particle orbiting the Sun).

8. (II) (*a*) What is the angular momentum of a figure skater spinning at 2.8 rev/s with arms in close to her body, assuming her to be a uniform cylinder with a height of 1.5 m, a radius of 15 cm, and a mass of 48 kg? (*b*) How much torque is required to slow her to a stop in 4.0 s, assuming she does *not* move her arms?

9. (II) A uniform disk turns at 4.1 rev/s around a frictionless central axis. A nonrotating rod, of the same mass as the disk and length equal to the disk's diameter, is dropped onto the freely spinning disk, Fig. 11–32. They then turn together around the spindle with their centers superposed. What is the angular frequency in rev/s of the combination?

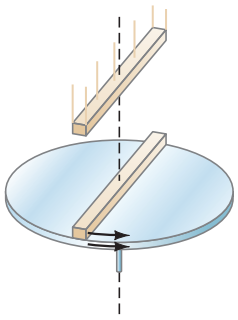


FIGURE 11–32
Problem 9.

10. (II) A person of mass 75 kg stands at the center of a rotating merry-go-round platform of radius 3.0 m and moment of inertia $920\ \text{kg}\cdot\text{m}^2$. The platform rotates without friction with angular velocity 0.95 rad/s. The person walks radially to the edge of the platform. (*a*) Calculate the angular velocity when the person reaches the edge. (*b*) Calculate the rotational kinetic energy of the system of platform plus person before and after the person's walk.

11. (II) A person stands on a platform, initially at rest, that can rotate freely without friction. The moment of inertia of the person plus the platform is I_P. The person holds a spinning bicycle wheel with its axis horizontal. The wheel has moment of inertia I_W and angular velocity ω_W. What will be the angular velocity ω_P of the platform if the person moves the axis of the spinning wheel so that it points (*a*) vertically upward, (*b*) at a 60° angle to the vertical, (*c*) vertically downward? (*d*) What will ω_P be if the person reaches up and stops the spinning wheel in part (*a*)?

12. (II) A potter's wheel is rotating around a vertical axis through its center at a frequency of 1.5 rev/s. The wheel can be considered a uniform disk of mass 5.0 kg and diameter 0.40 m. The potter then throws a 2.8-kg chunk of clay, approximately shaped as a flat disk of radius 8.0 cm, onto the center of the rotating wheel. What is the frequency of the wheel after the clay sticks to it?

13. (II) A 4.2-m-diameter merry-go-round is rotating freely with an angular velocity of 0.80 rad/s. Its total moment of inertia is $1630\ \text{kg}\cdot\text{m}^2$. Four people standing on the ground, each of mass 65 kg, suddenly step onto the edge of the merry-go-round. (*a*) What is the angular velocity of the merry-go-round now? (*b*) What if the people were on it initially and then jumped off in a radial direction (relative to the merry-go-round)?

14. (II) A woman of mass m stands at the edge of a solid cylindrical platform of mass M and radius R. At $t = 0$, the platform is rotating with negligible friction at angular velocity ω_0 about a vertical axis through its center, and the woman begins walking with speed v (relative to the platform) toward the center of the platform. (*a*) Determine the angular velocity of the system as a function of time. (*b*) What will be the angular velocity when the woman reaches the center?

15. (II) Model a figure skater's body (mass M) as a solid cylinder ($\approx 0.75M$ including head and thighs) and her arms as thin rods ($0.13M$ for both), making reasonable estimates for the dimensions. Then calculate the ratio of the angular speeds for a spinning skater with outstretched arms, and with arms held tightly against her body. (Ignore contributions from lower legs and feet.)

16. (II) A 3.0-kg cylindrical grindstone of radius 8.0 cm is attached to a motor. The angular momentum of the grindstone after the motor is turned on is given by

$$L = (1.2\ \text{kg}\cdot\text{m}^2/\text{s})\left(1 - e^{-0.050t}\right),$$

where t is in seconds. (*a*) What is the final rotation rate in rpm of the grindstone? (*b*) What is the net torque applied to the grindstone, as a function of time?

17. (II) A uniform horizontal rod of mass M and length ℓ rotates with angular velocity ω about a vertical axis through its center. Attached to each end of the rod is a small mass m. Determine the angular momentum of the system about the axis.

18. (II) Suppose our Sun eventually collapses into a *white dwarf*, losing about half its mass in the process, and ending up with a radius 1.0% of its existing radius. Assuming the lost mass carries away no angular momentum, (*a*) what would the Sun's new rotation rate be? Take the *Sun's* current *period* to be about 30 days. (*b*) What would be its final kinetic energy in terms of its initial kinetic energy of today?

19. (II) Hurricanes can involve winds in excess of 120 km/h at the outer edge. Make a rough estimate of (*a*) the energy, and (*b*) the angular momentum, of such a hurricane, approximating it as a rigidly rotating uniform cylinder of air (density $1.3\ \text{kg/m}^3$) of radius 85 km and height 4.5 km.

20. (III) An asteroid of mass 2.0×10^5 kg, traveling at a speed of 35 km/s relative to the Earth, hits the Earth at the equator tangentially, in the direction of Earth's rotation, and is embedded there. Use angular momentum to estimate the percent change in the angular speed of the Earth as a result of the collision.

21. (III) Suppose a 65-kg person stands at the edge of a 9.8-m-diameter merry-go-round turntable that is mounted on frictionless bearings and has a moment of inertia of $1850\ \text{kg}\cdot\text{m}^2$. The turntable is at rest initially, but when the person begins running at a speed of 3.8 m/s (with respect to the turntable) around its edge, the turntable begins to rotate in the opposite direction. Calculate the angular velocity of the turntable.

11–2 Vector Cross Product and Torque

22. (I) If vector $\vec{\mathbf{A}}$ points along the negative x axis and vector $\vec{\mathbf{B}}$ along the positive z axis, what is the direction of (*a*) $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ and (*b*) $\vec{\mathbf{B}} \times \vec{\mathbf{A}}$? (*c*) What is the magnitude of $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ and $\vec{\mathbf{B}} \times \vec{\mathbf{A}}$?

23. (I) Show that (*a*) $\hat{\mathbf{i}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}} \times \hat{\mathbf{k}} = 0$, (*b*) $\hat{\mathbf{i}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}}$, $\hat{\mathbf{i}} \times \hat{\mathbf{k}} = -\hat{\mathbf{j}}$, and $\hat{\mathbf{j}} \times \hat{\mathbf{k}} = \hat{\mathbf{i}}$.

24. (I) The directions of vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ are given below for several cases. For each case, state the direction of $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$. (*a*) $\vec{\mathbf{A}}$ points east, $\vec{\mathbf{B}}$ points south. (*b*) $\vec{\mathbf{A}}$ points east, $\vec{\mathbf{B}}$ points straight down. (*c*) $\vec{\mathbf{A}}$ points straight up, $\vec{\mathbf{B}}$ points north. (*d*) $\vec{\mathbf{A}}$ points straight up, $\vec{\mathbf{B}}$ points straight down.

25. (II) What is the angle θ between two vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$, if $|\vec{\mathbf{A}} \times \vec{\mathbf{B}}| = \vec{\mathbf{A}} \cdot \vec{\mathbf{B}}$?

26. (II) A particle is located at $\vec{\mathbf{r}} = (5.0\hat{\mathbf{i}} + 3.5\hat{\mathbf{j}} + 6.0\hat{\mathbf{k}})$ m. A force $\vec{\mathbf{F}} = (8.0\hat{\mathbf{j}} - 4.0\hat{\mathbf{k}})$ N acts on it. What is the torque, calculated about the origin?

27. (II) Consider a particle of a rigid object rotating about a fixed axis. Show that the tangential and radial vector components of the linear acceleration are:

$$\vec{\mathbf{a}}_{\tan} = \vec{\boldsymbol{\alpha}} \times \vec{\mathbf{r}} \quad \text{and} \quad \vec{\mathbf{a}}_{\mathrm{R}} = \vec{\boldsymbol{\omega}} \times \vec{\mathbf{v}}.$$

28. (II) Let us show that Eq. 11–3c is equivalent to the *definition* of *cross product*, Eq. 11–3a and the sentence below it including the right-hand rule and Fig. 11–7. We do this in three parts. (i) First show that the magnitude $\vec{\mathbf{C}}$ as given by Eq. 11–3c equals $AB\sin\theta$, Eq. 11–3a. Start with Eq. 11–3c and write $C^2 = |\vec{\mathbf{A}} \times \vec{\mathbf{B}}|^2$ in terms of the components in Eq. 11–3c. Use algebra to arrange the terms in the form $[A_x^2(B_y^2 + B_z^2) + \text{two similar terms}]$ plus a term $[-2(A_xB_xA_yB_y + 2 \text{ more terms})]$. To this long equation for C^2, add and subtract three terms of the form $[0 = A_x^2B_x^2 - A_x^2B_x^2, \ldots]$ and then gather terms together of the form $[(A_x^2 + \ldots)(B_x^2 + \ldots) - 2(A_xB_xA_yB_y + 2 \text{ more terms}) - (A_x^2B_x^2 + 2 \text{ more terms})]$. Simplify this long equation to the form $[A^2B^2 - (\vec{\mathbf{A}} \cdot \vec{\mathbf{B}})^2]$, from which show that $C^2 = A^2B^2\sin^2\theta$, which gives us Eq. 11–3a. (ii) Secondly, starting with Eq. 11–3c, show that $\vec{\mathbf{C}}$ is perpendicular to both $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ (as stated in the definition, the italic sentence below Eq. 11–3a) by taking the products $\vec{\mathbf{A}} \cdot \vec{\mathbf{C}}$ and $\vec{\mathbf{B}} \cdot \vec{\mathbf{C}}$ and showing they are zero (i.e. $\cos\theta = 0$ so $\theta = 90°$). (iii) Finally, choose a coordinate system so that $\vec{\mathbf{A}} = A_x\hat{\mathbf{i}}$ and $\vec{\mathbf{B}} = B_x\hat{\mathbf{i}} + B_y\hat{\mathbf{j}}$ and show that $\vec{\mathbf{C}} = \vec{\mathbf{A}} \times \vec{\mathbf{B}}$ follows the right-hand rule.

29. (II) Using Eq. 11–3c, show that the *distributive law* (Eq. 11–4c) is valid.

30. (II) An engineer estimates that under the most adverse expected weather conditions, the total force on the highway sign in Fig. 11–33 will be $\vec{\mathbf{F}} = (\pm 2.4\hat{\mathbf{i}} - 4.1\hat{\mathbf{j}})$ kN, acting at the CM. What torque does this force exert about the base O?

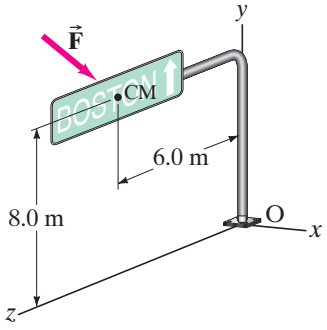


FIGURE 11–33 Problem 30.

31. (II) Determine (*a*) the vector product $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ and (*b*) the angle between $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ if $\vec{\mathbf{A}} = 7.4\hat{\mathbf{i}} - 3.5\hat{\mathbf{j}}$ and $\vec{\mathbf{B}} = -8.5\hat{\mathbf{i}} + 4.6\hat{\mathbf{j}} + 2.0\hat{\mathbf{k}}$. [*Hint*: To resolve an ambiguity, the dot product may prove useful.]

32. (II) The origin of a coordinate system is at the center of a wheel which rotates in the xy plane about its axle which is the z axis. A force $F = 215$ N acts in the xy plane, at a $+28.0°$ angle to the x axis, at the point $x = 28.0$ cm, $y = 33.5$ cm. Determine the magnitude and direction of the torque produced by this force about the axis.

33. (III) Show that the velocity $\vec{\mathbf{v}}$ of any point in an object rotating with *angular velocity* $\vec{\boldsymbol{\omega}}$ about a fixed axis can be written

$$\vec{\mathbf{v}} = \vec{\boldsymbol{\omega}} \times \vec{\mathbf{r}}$$

where $\vec{\mathbf{r}}$ is the position vector of the point relative to an origin O located on the axis of rotation. Can O be anywhere on the rotation axis? Will $\vec{\mathbf{v}} = \vec{\boldsymbol{\omega}} \times \vec{\mathbf{r}}$ if O is located at a point not on the axis of rotation?

11–3 Angular Momentum of a Particle

34. (I) What are the x, y, and z components of the angular momentum about the origin for a particle located at $\vec{\mathbf{r}} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$ which has momentum $\vec{\mathbf{p}} = p_x\hat{\mathbf{i}} + p_y\hat{\mathbf{j}} + p_z\hat{\mathbf{k}}$?

35. (I) Show that the kinetic energy K of a particle of mass m, moving in a circular path, is $K = L^2/2I$, where L is its angular momentum and I is its moment of inertia about the center of the circle.

36. (I) Calculate the angular momentum of a particle of mass m moving with constant velocity v for two cases (see Fig. 11–34): (*a*) about origin O, and (*b*) about O′.

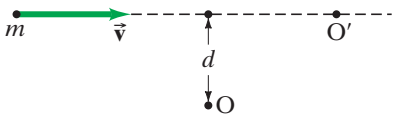


FIGURE 11–34 Problem 36.

37. (II) Two identical particles have equal but opposite momenta, $\vec{\mathbf{p}}$ and $-\vec{\mathbf{p}}$, but they are not traveling along the same line. Show that the total angular momentum of this system does not depend on the choice of origin.

38. (II) Determine the angular momentum of an 85-g particle about the origin of coordinates when the particle is at $x = 4.4$ m, $y = -6.0$ m, and it has velocity $v = (3.2\hat{\mathbf{i}} - 8.0\hat{\mathbf{k}})$ m/s.

39. (II) A particle is at the position $(x, y, z) = (1.0, 2.0, 3.0)$ m. It is traveling with a vector velocity $(-5.0, +2.8, -3.1)$ m/s. Its mass is 4.3 kg. What is its vector angular momentum about the origin?

11–4 and 11–5 Angular Momentum and Torque: General Motion; Rigid Objects

40. (II) An Atwood machine (Fig. 11–16) consists of two masses, $m_A = 7.0$ kg and $m_B = 8.2$ kg, connected by a cord that passes over a pulley free to rotate about a fixed axis. The pulley is a solid cylinder of radius $R_0 = 0.45$ m and mass 0.80 kg. (*a*) Determine the acceleration a of each mass. (*b*) What percentage of error in a would be made if the moment of inertia of the pulley were ignored? Ignore friction in the pulley bearings. Use angular momentum.

41. (II) Four identical particles of mass m are mounted at equal intervals on a thin rod of length ℓ and mass M, with one mass at each end of the rod. If the system is rotated with angular velocity ω about an axis perpendicular to the rod through one of the end masses, determine (*a*) the kinetic energy and (*b*) the angular momentum of the system.

42. (II) Two lightweight rods 24 cm in length are mounted perpendicular to an axle and at 180° to each other (Fig. 11–35). At the end of each rod is a 480-g mass. The rods are spaced 42 cm apart along the axle. The axle rotates at 4.5 rad/s. (*a*) What is the component of the total angular momentum along the axle? (*b*) What angle does the vector angular momentum make with the axle? [*Hint*: Remember that the vector angular momentum must be calculated about the *same point* for *both* masses, which could be the CM.]

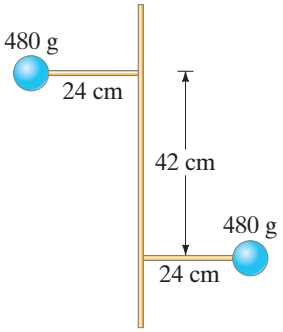


FIGURE 11–35 Problem 42.

43. (II) Figure 11–36 shows two masses connected by a cord passing over a pulley of radius R_0 and moment of inertia I. Mass M_A slides on a frictionless surface, and M_B hangs freely. Determine a formula for (*a*) the angular momentum of the system about the pulley axis, as a function of the speed v of mass M_A or M_B, and (*b*) the acceleration of the masses.

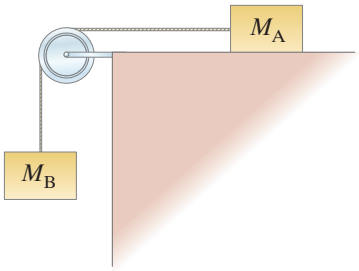


FIGURE 11–36 Problem 43.

***44.** (III) Show that the total angular momentum $\vec{\mathbf{L}} = \Sigma \vec{\mathbf{r}}_i \times \vec{\mathbf{p}}_i$ of a system of particles about the origin of an inertial reference frame can be written as the sum of the angular momentum about the CM, $\vec{\mathbf{L}}^*$ (spin angular momentum), plus the angular momentum of the CM about the origin (orbital angular momentum): $\vec{\mathbf{L}} = \vec{\mathbf{L}}^* + \vec{\mathbf{r}}_{CM} \times M\vec{\mathbf{v}}_{CM}$. [*Hint*: See the derivation of Eq. 11–9b.]

***45.** (III) (*a*) Determine the magnitude of the force $\vec{\mathbf{F}}$ exerted by each bearing in Fig. 11–18 (Example 11–10). The bearings are a distance d from point O. Ignore the effects of gravity. (*b*) Suppose in Fig. 11–18 that $m_A = m_B = 0.60$ kg, $r_A = r_B = 0.30$ m, and the distance between the bearings is 0.23 m. Evaluate the force that each bearing must exert on the axle if $\phi = 28.0°$, $\omega = 11.0$ rad/s.

***46.** (III) Suppose in Fig. 11–18 that $m_B = 0$; that is, only one mass, m_A, is actually present. If the bearings are each a distance d from O, determine the forces F_A and F_B at the upper and lower bearings respectively. [*Hint*: Choose an origin—different than O in Fig. 11–18—such that $\vec{\mathbf{L}}$ is parallel to $\vec{\boldsymbol{\omega}}$. Ignore effects of gravity.]

11–6 Angular Momentum Conservation

47. (II) A thin rod of mass M and length ℓ is suspended vertically from a frictionless pivot at its upper end. A mass m of putty traveling horizontally with a speed v strikes the rod at its CM and sticks there. How high does the bottom of the rod swing?

48. (II) A uniform stick 1.0 m long with a total mass of 270 g is pivoted at its center. A 3.5-g bullet is shot through the stick midway between the pivot and one end (Fig. 11–37). The bullet approaches at 250 m/s and leaves at 140 m/s. With what angular speed is the stick spinning after the collision?

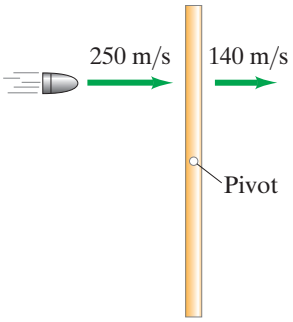


FIGURE 11–37 Problem 48.

49. (II) Two ice skaters, both of mass 68 kg, approach on parallel paths 1.6 m apart. Both are moving at 3.5 m/s with their arms outstretched. They join hands as they pass, still maintaining their 1.6-m separation, and begin rotating about one another. Treat the skaters as particles with regard to their rotational inertia. (*a*) What is their common angular speed after joining hands? (*b*) Calculate the change in kinetic energy for this process. (*c*) If they now pull on each other's hands, reducing their radius to half its original value, what is their common angular speed after reducing their radius? (*d*) Calculate the change in kinetic energy for this process. Why is it different than in part (*b*)?

50. (II) Suppose a 5.2×10^{10} kg meteorite struck the Earth at the equator with a speed $v = 2.2 \times 10^4$ m/s, as shown in Fig. 11–38 and remained stuck. By what factor would this affect the rotational frequency of the Earth (1 rev/day)?

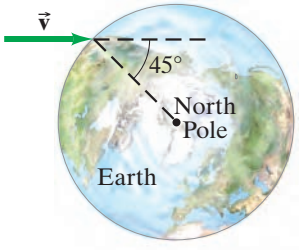


FIGURE 11–38 Problem 50.

51. (III) A 260-kg beam 2.7 m in length slides broadside down the ice with a speed of 18 m/s (Fig. 11–39). A 65-kg man at rest grabs one end as it goes past and hangs on as both he and the beam go spinning down the ice. Assume frictionless motion. (*a*) How fast does the center of mass of the system move after the collision? (*b*) With what angular velocity does the system rotate about its CM?

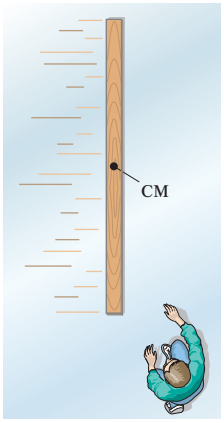


FIGURE 11–39
Problem 51.

52. (III) A thin uniform rod of mass M and length ℓ rests on a frictionless table and is struck at a point $\ell/4$ from its CM by a clay ball of mass m moving at speed v (Fig. 11–40). The ball sticks to the rod. Determine the translational and rotational motion of the rod after the collision.

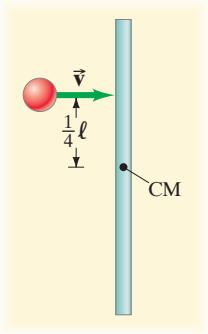


FIGURE 11–40
Problem 52.

53. (III) On a level billiards table a cue ball, initially at rest at point O on the table, is struck so that it leaves the cue stick with a center-of-mass speed v_0 and a "reverse" spin of angular speed ω_0 (see Fig. 11–41). A kinetic friction force acts on the ball as it initially skids across the table. (*a*) Explain why the ball's angular momentum is conserved about point O. (*b*) Using conservation of angular momentum, find the critical angular speed ω_C such that, if $\omega_0 = \omega_C$, kinetic friction will bring the ball to a complete (as opposed to momentary) stop. (*c*) If ω_0 is 10% smaller than ω_C, i.e., $\omega_0 = 0.90\,\omega_C$, determine the ball's CM velocity v_{CM} when it starts to roll without slipping. (*d*) If ω_0 is 10% larger than ω_C, i.e., $\omega_0 = 1.10\,\omega_C$, determine the ball's CM velocity v_{CM} when it starts to roll without slipping. [*Hint*: The ball possesses two types of angular momentum, the first due to the linear speed v_{CM} of its CM relative to point O, the second due to the spin at angular velocity ω about its own CM. The ball's total L about O is the sum of these two angular momenta.]

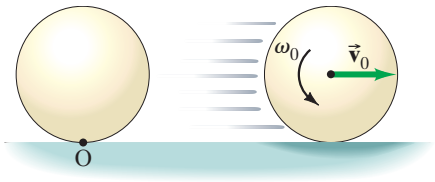


FIGURE 11–41 Problem 53.

*11–7 Spinning Top

***54.** (II) A 220-g top spinning at 15 rev/s makes an angle of 25° to the vertical and precesses at a rate of 1.00 rev per 5.5 s. If its CM is 3.5 cm from its tip along its symmetry axis, what is the moment of inertia of the top?

***55.** (II) A toy gyroscope consists of a 170-g disk with a radius of 5.5 cm mounted at the center of a thin axle 21 cm long (Fig. 11–42). The gyroscope spins at 45 rev/s. One end of its axle rests on a stand and the other end precesses horizontally about the stand. (*a*) How long does it take the gyroscope to precess once around? (*b*) If all the dimensions of the gyroscope were doubled (radius = 11 cm, axle = 42 cm), how long would it take to precess once?

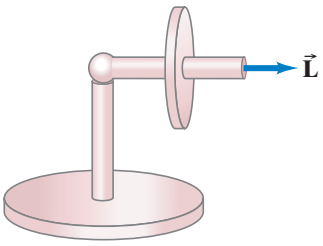


FIGURE 11–42
A wheel, rotating about a horizontal axle supported at one end, precesses. Problems 55, 56, and 57.

***56.** (II) Suppose the solid wheel of Fig. 11–42 has a mass of 260 g and rotates at 85 rad/s; it has radius 6.0 cm and is mounted at the center of a horizontal thin axle 25 cm long. At what rate does the axle precess?

***57.** (II) If a mass equal to half the mass of the wheel in Problem 56 is placed at the free end of the axle, what will be the precession rate now? Treat the extra mass as insignificant in size.

***58.** (II) A bicycle wheel of diameter 65 cm and mass m rotates on its axle; two 20-cm-long wooden handles, one on each side of the wheel, act as the axle. You tie a rope to a small hook on the end of one of the handles, and then spin the bicycle wheel with a flick of the hand. You release the spinning wheel, holding it by a short section of rope from your hand. The wheel then precesses about the vertical axis defined by the rope, instead of falling toward the ground (as it would if it were not spinning). Estimate the rate and direction of precession if the wheel rotates counterclockwise at 2.5 rev/s and its axle remains horizontal.

***59.** (III) The axis of the *Earth precesses* with a period of about 26,000 years (sometimes called the *precession of the equinoxes*). This is much like the precession of a top. Explain how the *Earth's equatorial bulge* gives rise to a torque exerted by the Sun and Moon on the Earth; see Fig. 11–43, which is drawn for the winter solstice (December 21), and has greatly exaggerated the Earth's equatorial bulge. About what axis would you expect the Earth's rotation axis to precess as a result of the torque due to the Sun? Does the torque exist three months later? Explain.

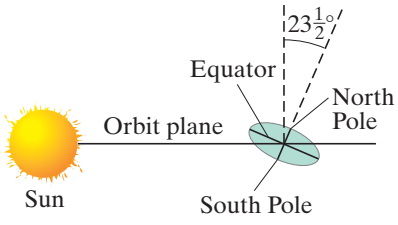


FIGURE 11–43 Problem 59. (Not to scale.)

11–8 Rotating Reference Frames; Inertial Forces

60. (II) If a plant is allowed to grow from seed on a rotating platform, it will grow at an angle, pointing inward. Calculate what this angle will be (put yourself in the rotating frame) in terms of g, r, and ω. Why does it grow inward rather than outward?

61. (III) Let $\vec{\mathbf{g}}'$ be the effective acceleration of gravity at a point on the rotating Earth, equal to the vector sum of the "true" value $\vec{\mathbf{g}}$ plus the effect of the rotating reference frame ($m\omega^2 r$ term). See Fig. 11–44. Determine the magnitude and direction of $\vec{\mathbf{g}}'$ relative to a radial line from the center of the Earth (*a*) at the North Pole, (*b*) at a latitude of 45.0° north, and (*c*) at the equator. Assume that g (if ω were zero) is a constant $9.80\,\text{m/s}^2$.

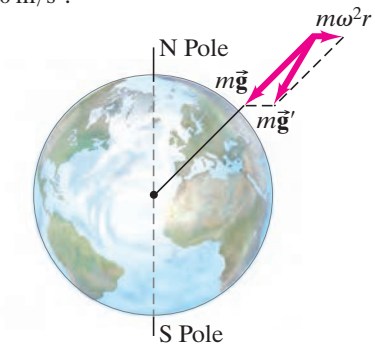


FIGURE 11–44 Problem 61.

*11–9 Coriolis Effect

***62.** (II) Suppose the man at B in Fig. 11–26 throws the ball toward the woman at A. (*a*) In what direction is the ball deflected as seen in the noninertial reference frame? (*b*) Determine a formula for the amount of deflection and for the (Coriolis) acceleration in this case.

***63.** (II) For what directions of velocity would the Coriolis effect on an object moving at the Earth's equator be zero?

***64.** (III) We can alter Eqs. 11–14 and 11–15 for use on Earth by considering only the component of $\vec{\mathbf{v}}$ perpendicular to the axis of rotation. From Fig. 11–45 we see that this is $v\cos\lambda$ for a vertically falling object, where λ is the latitude of the place on the Earth. If a lead ball is dropped vertically from a 110-m-high tower in Florence, Italy (latitude = 44°), how far from the base of the tower is it deflected by the Coriolis force?

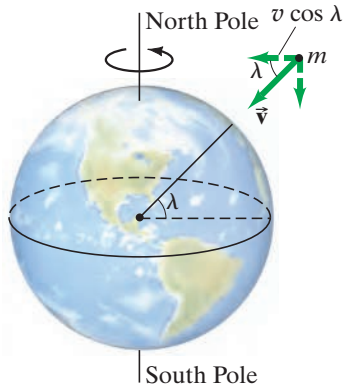


FIGURE 11–45 Problem 64. Object of mass m falling vertically to Earth at a latitude λ.

***65.** (III) An ant crawls with constant speed outward along a radial spoke of a wheel rotating at constant angular velocity ω about a vertical axis. Write a vector equation for all the forces (including inertial forces) acting on the ant. Take the x axis along the spoke, y perpendicular to the spoke pointing to the ant's left, and the z axis vertically upward. The wheel rotates counterclockwise as seen from above.

General Problems

66. A thin string is wrapped around a cylindrical hoop of radius R and mass M (Fig. 11–46). One end of the string is fixed, and the hoop is allowed to fall vertically, starting from rest, as the string unwinds. (*a*) Determine the angular momentum of the hoop about its CM as a function of time. (*b*) What is the tension in the string as a function of time?

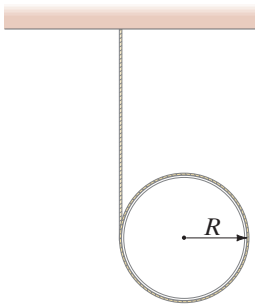


FIGURE 11–46 Problem 66.

67. A particle of mass 1.50 kg is moving with velocity $\vec{\mathbf{v}} = (7.0\,\hat{\mathbf{i}} + 6.0\,\hat{\mathbf{j}})\,\text{m/s}$. (*a*) Find the angular momentum $\vec{\mathbf{L}}$ relative to the origin when the particle is at $\vec{\mathbf{r}} = (2.0\,\hat{\mathbf{j}} + 4.0\,\hat{\mathbf{k}})\,\text{m}$. (*b*) At position $\vec{\mathbf{r}}$ a force of $\vec{\mathbf{F}} = 4.0\,\text{N}\,\hat{\mathbf{i}}$ is applied to the particle. Find the torque relative to the origin.

68. A merry-go-round with a moment of inertia equal to $860\,\text{kg}\cdot\text{m}^2$ and a radius of 3.0 m rotates with negligible friction at 1.70 rad/s. A child initially standing still next to the merry-go-round jumps onto the edge of the platform straight toward the axis of rotation causing the platform to slow to 1.25 rad/s. What is her mass?

69. Why might tall narrow SUVs and buses be prone to *rollover*? Consider a vehicle rounding a curve of radius R on a flat road. When just on the verge of rollover, its tires on the inside of the curve are about to leave the ground, so the friction and normal force on these two tires are zero. The total normal force on the outside tires is F_N and the total friction force is F_fr. Assume that the vehicle is not skidding. (*a*) Analysts define a *vehicle static stability factor* SSF = $w/2h$, where a vehicle's "track width" w is the distance between tires on the same axle, and h is the height of the CM above the ground. Show that the critical rollover speed is

$$v_\text{C} = \sqrt{Rg\left(\frac{w}{2h}\right)}.$$

[*Hint*: Take torques about an axis through the center of mass of the SUV, parallel to its direction of motion.] (*b*) Determine the ratio of highway curve radii (minimum possible) for a typical passenger car with SSF = 1.40 and an SUV with SSF = 1.05 at a speed of 90 km/h.

70. A projectile with mass m is launched from the ground and follows a trajectory given by

$$\vec{\mathbf{r}} = (v_{x0}t)\,\hat{\mathbf{i}} + \left(v_{y0}t - \frac{1}{2}gt^2\right)\hat{\mathbf{j}}$$

where v_{x0} and v_{y0} are the initial velocities in the x and y direction, respectively, and g is the acceleration due to gravity. The launch position is defined to be the origin. Determine the torque acting on the projectile about the origin using (*a*) $\vec{\boldsymbol{\tau}} = \vec{\mathbf{r}} \times \vec{\mathbf{F}}$, (*b*) $\vec{\boldsymbol{\tau}} = d\vec{\mathbf{L}}/dt$.

71. A small 250-g ball is moving in a 0.50-m-radius horizontal circle (the xy plane). The ball is moving clockwise when viewed from above and has kinetic energy 0.40 J. What is the ball's angular momentum?

72. A spherical asteroid with radius $r = 123\,\text{m}$ and mass $M = 2.65 \times 10^{10}\,\text{kg}$ rotates about an axis at 4.0 revolutions per day. A "tug" spaceship attaches itself to the asteroid's south pole (as defined by the axis of rotation) and fires its engine, applying a force F tangentially to the asteroid's surface as shown in Fig. 11–47. If $F = 265\,\text{N}$, how long will it take the tug to rotate the asteroid's axis of rotation through an angle of 6.0° by this method?

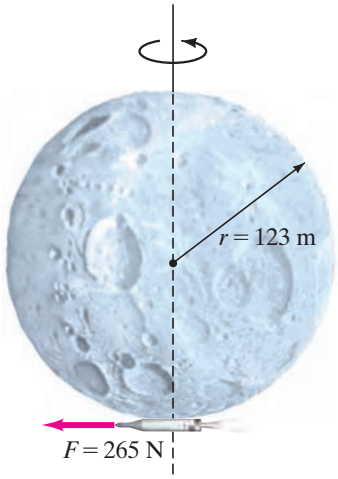


FIGURE 11–47 Problem 72.

73. The time-dependent position of a point object which moves counterclockwise along the circumference of a circle (radius R) in the xy plane with constant speed v is given by

$$\vec{\mathbf{r}} = \hat{\mathbf{i}} R\cos\omega t + \hat{\mathbf{j}} R\sin\omega t$$

where the constant $\omega = v/R$. Determine the velocity $\vec{\mathbf{v}}$ and angular velocity $\vec{\boldsymbol{\omega}}$ of this object and then show that these three vectors obey the relation $\vec{\mathbf{v}} = \vec{\boldsymbol{\omega}} \times \vec{\mathbf{r}}$.

74. The position of a particle with mass m traveling on a *helical path* (see Fig. 11–48) is given by

$$\vec{\mathbf{r}} = R\cos\left(\frac{2\pi z}{d}\right)\hat{\mathbf{i}} + R\sin\left(\frac{2\pi z}{d}\right)\hat{\mathbf{j}} + z\,\hat{\mathbf{k}}$$

where R and d are the radius and pitch of the helix, respectively, and z has time dependence $z = v_z t$ where v_z is the (constant) component of velocity in the z direction. Determine the time-dependent angular momentum $\vec{\mathbf{L}}$ of the particle about the origin.

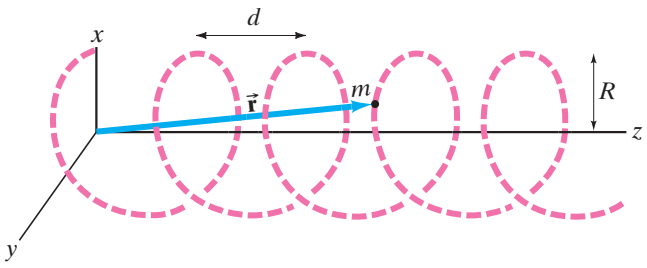


FIGURE 11–48 Problem 74.

75. A boy rolls a tire along a straight level street. The tire has mass 8.0 kg, radius 0.32 m and moment of inertia about its central axis of symmetry of $0.83\,\text{kg}\cdot\text{m}^2$. The boy pushes the tire forward away from him at a speed of 2.1 m/s and sees that the tire leans 12° to the right (Fig. 11–49). (*a*) How will the resultant torque due to gravity and the normal force $\vec{\mathbf{F}}_\text{N}$ affect the subsequent motion of the tire? (*b*) Compare the change in angular momentum caused by this torque in 0.20 s to the original magnitude of angular momentum.

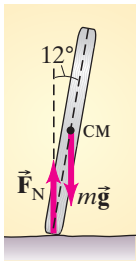


FIGURE 11–49 Problem 75.

76. A 70.0-kg person stands on a tiny rotating platform with arms outstretched. (*a*) Estimate the moment of inertia of the person using the following approximations: the body (including head and legs) is a 62.0-kg cylinder, 16 cm in radius and 1.70 m high; and each arm is a 4.0-kg thin rod, 60.0 cm long, attached to the cylinder. (*b*) Using the same approximations, estimate the moment of inertia when the arms are at the person's sides. (*c*) If one rotation takes 1.2 s when the person's arms are outstretched, what is the time for each rotation with arms at the sides? Ignore the moment of inertia of the lightweight platform. (*d*) Determine the change in kinetic energy when the arms are lifted from the sides to the horizontal position. (*e*) From your answer to part (*d*), would you expect it to be harder or easier to lift your arms when rotating or when at rest?

77. Water drives a waterwheel (or turbine) of radius $R = 3.0\,\text{m}$ as shown in Fig. 11–50. The water enters at a speed $v_1 = 7.0\,\text{m/s}$ and exits from the waterwheel at a speed $v_2 = 3.8\,\text{m/s}$. (*a*) If 85 kg of water passes through per second, what is the rate at which the water delivers angular momentum to the waterwheel? (*b*) What is the torque the water applies to the waterwheel? (*c*) If the water causes the waterwheel to make one revolution every 6.0 s, how much power is delivered to the wheel?

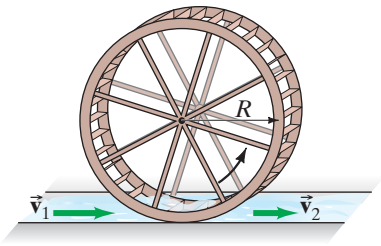


FIGURE 11–50 Problem 77.

78. In Example 11–1, the hand slowly pulls the string, causing the rotating mass $(m = 0.75\,\text{kg})$ to gain speed. (*a*) Calculate the work done by the hand. (*b*) Check that the work–energy theorem is satisfied.

79. A particle of mass m uniformly accelerates as it moves counterclockwise along the circumference of a circle of radius R:

$$\vec{\mathbf{r}} = \hat{\mathbf{i}} R\cos\theta + \hat{\mathbf{j}} R\sin\theta$$

with $\theta = \omega_0 t + \frac{1}{2}\alpha t^2$, where the constants ω_0 and α are the initial angular velocity and angular acceleration, respectively. Determine the object's tangential acceleration $\vec{\mathbf{a}}_\text{tan}$ and determine the torque acting on the object using (*a*) $\vec{\boldsymbol{\tau}} = \vec{\mathbf{r}} \times \vec{\mathbf{F}}$, (*b*) $\vec{\boldsymbol{\tau}} = I\vec{\boldsymbol{\alpha}}$.

80. The Moon orbits the Earth such that the same side always faces the Earth. Determine the ratio of the Moon's spin angular momentum (about its own axis) to its orbital angular momentum. (In the latter case, treat the Moon as a particle orbiting the Earth.)

81. Most of our Solar System's mass is contained in the Sun, and the planets possess almost all of the Solar System's angular momentum. This observation plays a key role in theories attempting to explain the formation of our Solar System. Estimate the fraction of the Solar System's total angular momentum from the planets using a simplified model which includes only the large outer planets with the most angular momentum. The central Sun (mass 1.99×10^{30} kg, radius 6.96×10^{8} m) spins about its axis once every 25 days and the planets Jupiter, Saturn, Uranus, and Neptune move in nearly circular orbits around the Sun with orbital data given in the Table below. Ignore each planet's spin about its own axis.

Planet	Mean Distance from Sun ($\times 10^6$ km)	Orbital Period (Earth Years)	Mass ($\times 10^{25}$ kg)
Jupiter	778	11.9	190
Saturn	1427	29.5	56.8
Uranus	2870	84.0	8.68
Neptune	4500	165	10.2

82. Competitive ice skaters commonly perform single, double, and triple axel jumps in which they rotate $1\frac{1}{2}$, $2\frac{1}{2}$, and $3\frac{1}{2}$ revolutions, respectively, about a vertical axis while airborne. For each of these jumps, a typical skater remains airborne for about 0.70 s. Suppose a skater leaves the ground in an "open" position (e.g., arms outstretched) with moment of inertia I_0 and rotational frequency $f_0 = 1.2$ rev/s, maintaining this position for 0.10 s. The skater then assumes a "closed" position (arms brought closer) with moment of inertia I, acquiring a rotational frequency f, which is maintained for 0.50 s. Finally, the skater immediately returns to the "open" position for 0.10 s until landing (see Fig. 11–51). (*a*) Why is angular momentum conserved during the skater's jump? Neglect air resistance. (*b*) Determine the minimum rotational frequency f during the flight's middle section for the skater to successfully complete a single and a triple axel. (*c*) Show that, according to this model, a skater must be able to reduce his or her moment of inertia in midflight by a factor of about 2 and 5 in order to complete a single and triple axel, respectively.

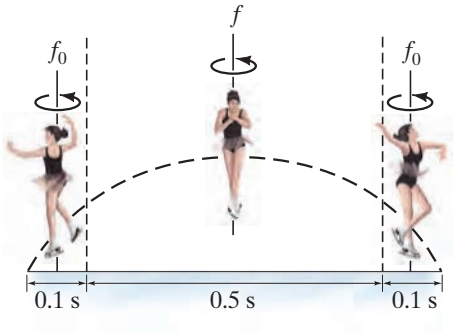


FIGURE 11–51 Problem 82.

83. Suppose a star the size of our Sun, but with mass 8.0 times as great, were rotating at a speed of 1.0 revolution every 9.0 days. If it were to undergo gravitational collapse to a *neutron star* of radius 12 km, losing 0.70 of its mass in the process, what would its rotation speed be? Assume the star is a uniform sphere at all times. Assume also that the thrown-off mass carries off either (*a*) no angular momentum, or (*b*) its proportional share (0.70) of the initial angular momentum.

84. A radio transmission tower has a mass of 76 kg and is 12 m high. The tower is anchored to the ground by a flexible joint at its base, but it is secured by three cables 120° apart (Fig. 11–52). In an analysis of a potential failure, a mechanical engineer needs to determine the behavior of the tower if one of the cables breaks. The tower would fall away from the broken cable, rotating about its base. Determine the speed of the top of the tower as a function of the rotation angle θ. Start your analysis with the rotational dynamics equation of motion $d\vec{\mathbf{L}}/dt = \vec{\tau}_{\text{ext}}$. Approximate the tower as a tall thin rod.

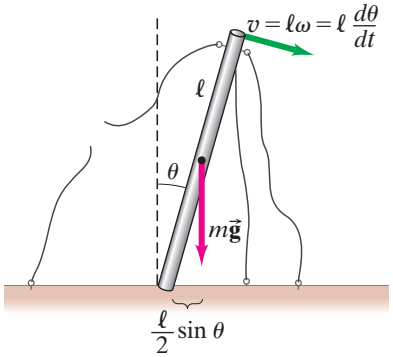


FIGURE 11–52 Problem 84.

85. A baseball bat has a *sweet spot* where a ball can be hit with almost effortless transmission of energy. A careful analysis of baseball dynamics shows that this special spot is located at the point where an applied force would result in pure rotation of the bat about the handle grip. Determine the location of the sweet spot, x_s, of the bat shown in Fig. 11–53. The linear mass density of the bat is given roughly by $(0.61 + 3.3x^2)$ kg/m, where x is in meters measured from the end of the handle. The entire bat is 0.84 m long. The desired rotation point should be 5.0 cm from the thin end where the bat is held. [*Hint*: Where is the CM of the bat?]

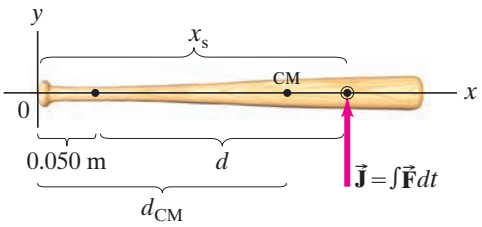


FIGURE 11–53 Problem 85.

86. Let $\vec{\mathbf{A}}, \vec{\mathbf{B}}$, and $\vec{\mathbf{C}}$ be three vectors, which for generality we assume do not all lie in the same plane. Show that $\vec{\mathbf{A}} \cdot (\vec{\mathbf{B}} \times \vec{\mathbf{C}}) = \vec{\mathbf{B}} \cdot (\vec{\mathbf{C}} \times \vec{\mathbf{A}}) = \vec{\mathbf{C}} \cdot (\vec{\mathbf{A}} \times \vec{\mathbf{B}})$.

ANSWERS TO EXERCISES

A: Yes. She does work to pull in her arms.
B: (*b*), it speeds up.
C: (*a*).
D: (*b*) The turntable revolves the same, counterclockwise but slower than in Example 11–5.
E: (i) (*d*); (ii) (*a*); (iii) (*b*).
F: (*e*).



Our whole built environment, from medieval cathedrals and modern bridges to skyscrapers, has required architects and engineers to determine the forces and stresses within these structures. The object is to keep these structures standing, or "static"—that is, not in motion, especially not falling down.

The study of statics applies equally well to the human body, including balance, the forces in muscles, joints, and bones, and ultimately the possibility of fracture.

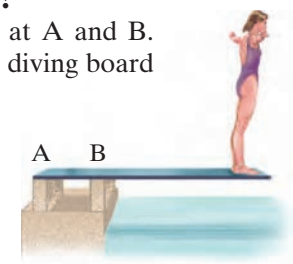
Six centuries ago the great Cathedral at the center of this photo in Florence, Italy, required great intuition and cleverness to construct the largest dome ever built (previous to prestressed concrete in the 20th century), as we will see in Section 12–8.

CHAPTER 12

Static Equilibrium; Elasticity and Fracture

CHAPTER-OPENING QUESTION—Guess now!
The diving board shown here is held by two supports at A and B. Which statement is true about the forces exerted *on* the diving board at A and B?

(a) $\vec{\mathbf{F}}_A$ is down, $\vec{\mathbf{F}}_B$ is up, and F_B is larger than F_A.
(b) Both forces are up and F_B is larger than F_A.
(c) $\vec{\mathbf{F}}_A$ is down, $\vec{\mathbf{F}}_B$ is up, and F_A is larger than F_B.
(d) Both forces are down and approximately equal.
(e) $\vec{\mathbf{F}}_B$ is down, $\vec{\mathbf{F}}_A$ is up, and they are equal.
(f) Both forces are up and are equal.



CONTENTS

12–1 The Conditions for Equilibrium
12–2 Solving Statics Problems
*12–3 Applications to Muscles and Joints
12–4 Stability and Balance
12–5 Elasticity; Stress and Strain
12–6 Fracture
*12–7 Trusses and Bridges
*12–8 Arches and Domes

In this Chapter, we will study a special case in mechanics—when the net force and the net torque on an object, or system of objects, are both zero. In this case both the linear acceleration and the angular acceleration of the object or system are zero. The object is either at rest, or its center of mass is moving at constant velocity. We will be concerned mainly with the first situation, in which the object or objects are all at rest, or *static* (= not moving).

The net force and the net torque can be zero, but this does not imply that no forces at all act on the objects. In fact it is virtually impossible to find an object on which no forces act. Just how and where these forces act can be very important, both for buildings and other structures, and in the human body.



FIGURE 12–1 Elevated walkway collapse in a Kansas City hotel in 1981. How a simple physics calculation could have prevented the tragic loss of over 100 lives is considered in Example 12–11.

Sometimes, as we shall see in this Chapter, the forces may be so great that the object is seriously *deformed*, or it may even *fracture* (break)—and avoiding such problems gives this field of *statics* even greater importance.

Statics is concerned with the calculation of the forces acting on and within structures that are in *equilibrium*. Determination of these forces, which occupies us in the first part of this Chapter, then allows a determination of whether the structures can sustain the forces without significant deformation or fracture, subjects we discuss later in this Chapter. These techniques can be applied in a wide range of fields. Architects and engineers must be able to calculate the forces on the structural components of buildings, bridges, machines, vehicles, and other structures, since any material will buckle or break if too much force is applied (Fig. 12–1). In the human body a knowledge of the forces in muscles, bones, and joints is of great value for doctors, physical therapists, and athletes.

12–1 The Conditions for Equilibrium

Objects in daily life have at least one force acting on them (gravity). If they are at rest, then there must be other forces acting on them as well so that the net force is zero. A book at rest on a table, for example, has two forces acting on it, the downward force of gravity and the normal force the table exerts upward on it (Fig. 12–2). Because the book is at rest, Newton's second law tells us that the net force on it is zero. Thus the upward force exerted by the table on the book must be equal in magnitude to the force of gravity acting downward on the book. Such an object is said to be in **equilibrium** (Latin for "equal forces" or "balance") under the action of these two forces.

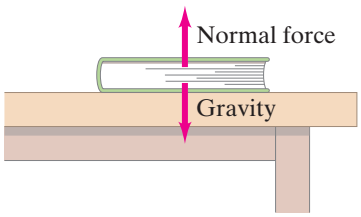


FIGURE 12–2 The book is in equilibrium; the net force on it is zero.

Do not confuse the two forces in Fig. 12–2 with the equal and opposite forces of Newton's third law, which act on different objects. In Fig. 12–2, both forces act on the same object; and they happen to add up to zero.

The First Condition for Equilibrium

For an object to be at rest, Newton's second law tells us that the sum of the forces acting on it must add up to zero. Since force is a vector, the components of the net force must each be zero. Hence, a condition for equilibrium is that

$$\Sigma F_x = 0, \qquad \Sigma F_y = 0, \qquad \Sigma F_z = 0. \tag{12–1}$$

We will mainly be dealing with forces that act in a plane, so we usually need only the x and y components. We must remember that if a particular force component points along the negative x or y axis, it must have a negative sign. Equations 12–1 represent the **first condition for equilibrium**.

We saw in Chapter 4 that to solve Problems involving forces, we need to draw a *free-body diagram*, indicating *all* the forces on a given object (see Section 4–7).

EXAMPLE 12–1 **Chandelier cord tension.** Calculate the tensions $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ in the two cords that are connected to the vertical cord supporting the 200-kg chandelier in Fig. 12–3. Ignore the mass of the cords.

APPROACH We need a free-body diagram, but for which object? If we choose the chandelier, the cord supporting it must exert a force equal to the chandelier's weight $mg = (200\,\text{kg})(9.80\,\text{m/s}^2) = 1960\,\text{N}$. But the forces $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ don't get involved. Instead, let us choose as our object the point where the three cords join (it could be a knot). The free-body diagram is then as shown in Fig. 12–3a. The three forces—$\vec{\mathbf{F}}_A$, $\vec{\mathbf{F}}_B$, and the tension in the vertical cord equal to the weight of the 200-kg chandelier—act at this point where the three cords join. For this junction point we write $\Sigma F_x = 0$ and $\Sigma F_y = 0$, since the problem is laid out in two dimensions. The directions of $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ are known, since tension in a cord can only be along the cord—any other direction would cause the cord to bend, as already pointed out in Chapter 4. Thus, our unknowns are the magnitudes F_A and F_B.

SOLUTION We first resolve $\vec{\mathbf{F}}_A$ into its horizontal (x) and vertical (y) components. Although we don't know the value of F_A, we can write (see Fig. 12–3b) $F_{Ax} = -F_A \cos 60°$ and $F_{Ay} = F_A \sin 60°$. $\vec{\mathbf{F}}_B$ has only an x component. In the vertical direction, we have the downward force exerted by the vertical cord equal to the weight of the chandelier $mg = (200\,\text{kg})(g)$, and the vertical component of $\vec{\mathbf{F}}_A$ upward:

$$\Sigma F_y = 0$$
$$F_A \sin 60° - (200\,\text{kg})(g) = 0$$

so

$$F_A = \frac{(200\,\text{kg})g}{\sin 60°} = (231\,\text{kg})g = (231\,\text{kg})(9.80\,\text{m/s}^2) = 2260\,\text{N}.$$

In the horizontal direction, with $\Sigma F_x = 0$,

$$\Sigma F_x = F_B - F_A \cos 60° = 0.$$

Thus

$$F_B = F_A \cos 60° = (231\,\text{kg})(g)(0.500) = (115\,\text{kg})g = 1130\,\text{N}.$$

The magnitudes of $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ determine the strength of cord or wire that must be used. In this case, the cord must be able to support a mass of more than 230 kg.

NOTE We didn't insert the value of g, the acceleration due to gravity, until the end. In this way we found the magnitude of the force in terms of g times the number of kilograms (which may be a more familiar quantity than newtons).

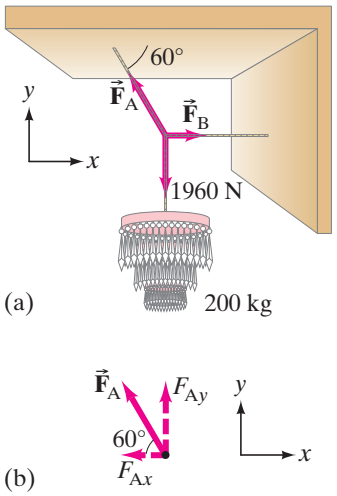


FIGURE 12–3 Example 12–1.

EXERCISE A In Example 12–1, F_A has to be greater than the chandelier's weight, mg. Why?

The Second Condition for Equilibrium

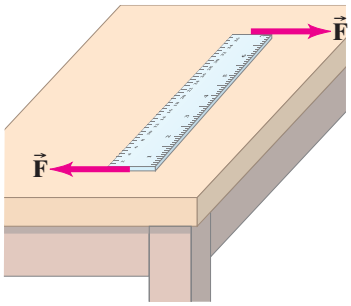
Although Eqs. 12–1 are a necessary condition for an object to be in equilibrium, they are not always a sufficient condition. Figure 12–4 shows an object on which the net force is zero. Although the two forces labeled $\vec{\mathbf{F}}$ add up to give zero net force on the object, they do give rise to a net torque that will rotate the object. Referring to Eq. 10–14, $\Sigma\tau = I\alpha$, we see that if an object is to remain at rest, the net torque applied to it (calculated about *any* axis) must be zero. Thus we have the **second condition for equilibrium**: that the sum of the torques acting on an object, as calculated about any axis, must be zero:

$$\Sigma\tau = 0. \qquad \textbf{(12–2)}$$

This condition will ensure that the angular acceleration, α, about any axis will be zero. If the object is not rotating initially ($\omega = 0$), it will not start rotating. Equations 12–1 and 12–2 are the only requirements for an object to be in equilibrium.

We will mainly consider cases in which the forces all act in a plane (we call it the xy plane). In such cases the torque is calculated about an axis that is perpendicular to the xy plane. *The choice of this axis is arbitrary*. If the object is at rest, then $\Sigma\tau = 0$ is valid about any axis. Therefore we can choose any axis that makes our calculation easier. Once the axis is chosen, all torques must be calculated about that axis.

FIGURE 12–4 Although the net force on the ruler is zero, it will move (rotate). A pair of equal forces acting in opposite directions but at different points on an object (as shown here) is referred to as a *couple*.



CAUTION

Axis choice for $\Sigma\tau = 0$ is arbitrary. All torques must be calculated about the same axis

PHYSICS APPLIED
The lever

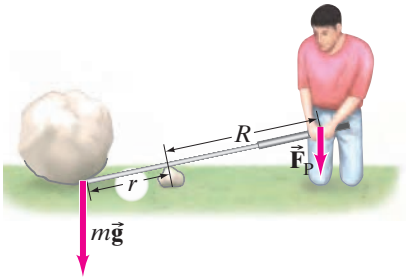


FIGURE 12–5 Example 12–2. A lever can "multiply" your force.

CONCEPTUAL EXAMPLE 12–2 **A lever.** The bar in Fig. 12–5 is being used as a lever to pry up a large rock. The small rock acts as a fulcrum (pivot point). The force F_P exerted at the long end of the bar can be quite a bit smaller than the rock's weight mg, since it is the *torques* that balance in the rotation about the fulcrum. If, however, the leverage isn't sufficient, and the large rock isn't budged, what are two ways to increase the lever arm (Section 10–4)?

RESPONSE One way is to increase the lever arm of the force F_P by slipping a pipe over the end of the bar and thereby pushing with a longer lever arm. A second way is to move the fulcrum closer to the large rock. This may change the long lever arm R only a little, but it changes the short lever arm r by a substantial fraction and therefore changes the ratio of R/r dramatically. In order to pry the rock, the torque due to F_P must at least balance the torque due to mg; that is, $mgr = F_P R$ and

$$\frac{r}{R} = \frac{F_P}{mg}.$$

With r smaller, the weight mg can be balanced with less force F_P. The ratio R/r is the **mechanical advantage** of the system. A lever is a "simple machine." We discussed another simple machine, the pulley, in Chapter 4, Example 4–14.

12–2 Solving Statics Problems

The subject of statics is important because it allows us to calculate certain forces on (or within) a structure when some of the forces on it are already known. We will mainly consider situations in which all the forces act in a plane, so we can have two force equations (x and y components) and one torque equation, for a total of three equations. Of course, you do not have to use all three equations if they are not needed. When using a torque equation, a torque that tends to rotate the object counterclockwise is usually considered positive, whereas a torque that tends to rotate it clockwise is considered negative. (But the opposite convention would be OK too.)

PROBLEM SOLVING
$\tau > 0$ *counterclockwise*
$\tau < 0$ *clockwise*

One of the forces that acts on objects is the force of gravity. As we discussed in Section 9–8, we can consider the force of gravity on an object as acting at its center of gravity (CG) or center of mass (CM), which for practical purposes are the same point. For uniform symmetrically shaped objects, the CG is at the geometric center. For more complicated objects, the CG can be determined as discussed in Section 9–8.

There is no single technique for attacking statics problems, but the following procedure may be helpful.

PROBLEM SOLVING

Statics

1. Choose one object at a time for consideration. Make a careful **free-body diagram** by showing all the forces acting on that object, including gravity, and the points at which these forces act. If you aren't sure of the direction of a force, choose a direction; if the actual direction of the force (or component of a force) is opposite, your eventual calculation will give a result with a minus sign.
2. Choose a convenient **coordinate system**, and resolve the forces into their components.
3. Using letters to represent unknowns, write down the **equilibrium equations** for the **forces**:
$$\Sigma F_x = 0$$
and
$$\Sigma F_y = 0,$$
assuming all the forces act in a plane.
4. For the **torque equation**,
$$\Sigma\tau = 0,$$
choose any axis perpendicular to the xy plane that might make the calculation easier. (For example, you can reduce the number of unknowns in the resulting equation by choosing the axis so that one of the unknown forces acts through that axis; then this force will have zero lever arm and produce zero torque, and so won't appear in the torque equation.) Pay careful attention to determining the lever arm for each force correctly. Give each torque a + or − sign to indicate torque direction. For example, if torques tending to rotate the object counterclockwise are positive, then those tending to rotate it clockwise are negative.
5. **Solve** these equations for the unknowns. Three equations allow a maximum of three unknowns to be solved for. They can be forces, distances, or even angles.

EXAMPLE 12–3 **Balancing a seesaw.** A board of mass $M = 4.0\,\text{kg}$ serves as a seesaw for two children, as shown in Fig. 12–6a. Child A has a mass of 30 kg and sits 2.5 m from the pivot point, P (his center of mass is 2.5 m from the pivot). At what distance x from the pivot must child B, of mass 25 kg, place herself to balance the seesaw? Assume the board is uniform and centered over the pivot.

PHYSICS APPLIED
Balancing a seesaw

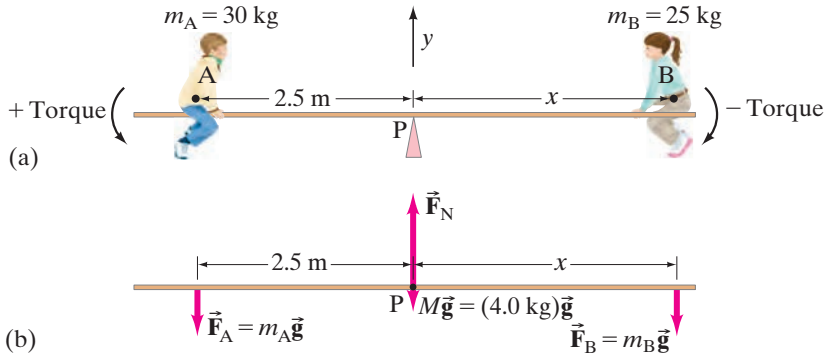


FIGURE 12–6 (a) Two children on a seesaw, Example 12–3. (b) Free-body diagram of the board.

APPROACH We follow the steps of the Problem Solving Strategy above.

SOLUTION

1. **Free-body diagram**. We choose the board as our object, and assume it is horizontal. Its free-body diagram is shown in Fig. 12–6b. The forces acting on the board are the forces exerted downward on it by each child, $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$, the upward force exerted by the pivot $\vec{\mathbf{F}}_N$, and the force of gravity on the board $(= M\vec{\mathbf{g}})$ which acts at the center of the uniform board.
2. **Coordinate system**. We choose y to be vertical, with positive upward, and x horizontal to the right, with origin at the pivot.
3. **Force equation**. All the forces are in the y (vertical) direction, so
$$\Sigma F_y = 0$$
$$F_N - m_A g - m_B g - Mg = 0,$$
where $F_A = m_A g$ and $F_B = m_B g$.
4. **Torque equation**. Let us calculate the torque about an axis through the board at the pivot point, P. Then the lever arms for F_N and for the weight of the board are zero, and they will contribute zero torque about point P. Thus the torque equation will involve only the forces $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$, which are equal to the weights of the children. The torque exerted by each child will be mg times the appropriate lever arm, which here is the distance of each child from the pivot point. $\vec{\mathbf{F}}_A$ tends to rotate the board counterclockwise $(+)$ and $\vec{\mathbf{F}}_B$ clockwise $(-)$, so the torque equation is
$$\Sigma\tau = 0$$
$$m_A g(2.5\,\text{m}) - m_B g x + Mg(0\,\text{m}) + F_N(0\,\text{m}) = 0$$
or
$$m_A g(2.5\,\text{m}) - m_B g x = 0,$$
where two terms were dropped because their lever arms were zero.
5. **Solve**. We solve the torque equation for x and find
$$x = \frac{m_A}{m_B}(2.5\,\text{m}) = \frac{30\,\text{kg}}{25\,\text{kg}}(2.5\,\text{m}) = 3.0\,\text{m}.$$
To balance the seesaw, child B must sit so that her CG is 3.0 m from the pivot point. This makes sense: since she is lighter, she must sit farther from the pivot than the heavier child in order to provide torques of equal magnitude.

EXERCISE B We did not need to use the force equation to solve Example 12–3 because of our choice of the axis. Use the force equation to find the force exerted by the pivot.

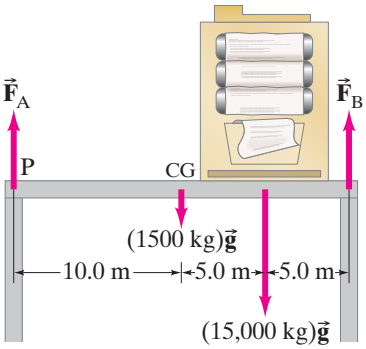


FIGURE 12–7 A 1500-kg beam supports a 15,000-kg machine. Example 12–4.

EXAMPLE 12–4 **Forces on a beam and supports.** A uniform 1500-kg beam, 20.0 m long, supports a 15,000-kg printing press 5.0 m from the right support column (Fig. 12–7). Calculate the force on each of the vertical support columns.

APPROACH We analyze the forces on the beam (the force the beam exerts on each column is equal and opposite to the force exerted by the column on the beam). We label the forces on the beam due to the supporting columns as $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ in Fig. 12–7. The force of gravity (= weight of the beam) acts at its center of gravity, 10.0 m from either end. We choose a convenient axis for writing the torque equation: the point of application of $\vec{\mathbf{F}}_A$ (labeled P), so $\vec{\mathbf{F}}_A$ will not enter the equation (its lever arm will be zero) and we will have an equation in only one unknown, F_B.

SOLUTION The torque equation, $\Sigma\tau = 0$, with the counterclockwise direction as positive, gives

$$\Sigma\tau = -(10.0\,\text{m})(1500\,\text{kg})g - (15.0\,\text{m})(15{,}000\,\text{kg})g + (20.0\,\text{m})F_B = 0.$$

Solving for F_B, we find $F_B = (12{,}000\,\text{kg})g = 118{,}000\,\text{N}$. To find F_A, we use $\Sigma F_y = 0$, with $+y$ upward:

$$\Sigma F_y = F_A - (1500\,\text{kg})g - (15{,}000\,\text{kg})g + F_B = 0.$$

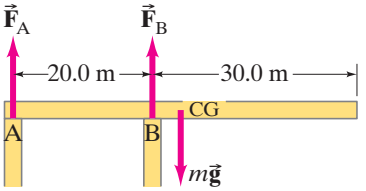
Putting in $F_B = (12{,}000\,\text{kg})g$, we find that $F_A = (4500\,\text{kg})g = 44{,}100\,\text{N}$.

PHYSICS APPLIED
Cantilever

PROBLEM SOLVING
If a force comes out negative

Figure 12–8 shows a uniform beam that extends beyond its support like a diving board. Such a beam is called a **cantilever**. The forces acting on the beam in Fig. 12–8 are those due to the supports, $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$, and the force of gravity which acts at the CG, 5.0 m to the right of the right-hand support. If you follow the procedure of the last Example and calculate F_A and F_B, assuming they point upward as shown in Fig. 12–8, you will find that F_A comes out negative. If the beam has a mass of 1200 kg and a weight $mg = 12{,}000\,\text{N}$, then $F_B = 15{,}000\,\text{N}$ and $F_A = -3000\,\text{N}$ (try it yourself: write ΣF_y and $\Sigma\tau$ equations). Whenever an unknown force comes out negative, it merely means that the force actually points in the opposite direction from what you assumed. Thus in Fig. 12–8, $\vec{\mathbf{F}}_A$ actually must pull downward (by means of bolts, screws, fasteners, and/or glue). To see why $\vec{\mathbf{F}}_A$ has to act downward, note that the board's weight acting at the CG would otherwise rotate the board clockwise about support B.

FIGURE 12–8 A cantilever.



EXERCISE C Return to the Chapter-Opening Question, page 331, and answer it again now. Try to explain why you may have answered differently the first time.

Our next Example involves a beam that is attached to a wall by a hinge and is supported by a cable or cord (Fig. 12–9). It is important to remember that a flexible cable can support a force only along its length. (If there were a component of force perpendicular to the cable, it would bend because it is flexible.) But for a rigid device, such as the hinge in Fig. 12–9, the force can be in any direction and we can know the direction only after solving the equations. (The hinge is assumed small and smooth, so it can exert no internal torque on the beam.)

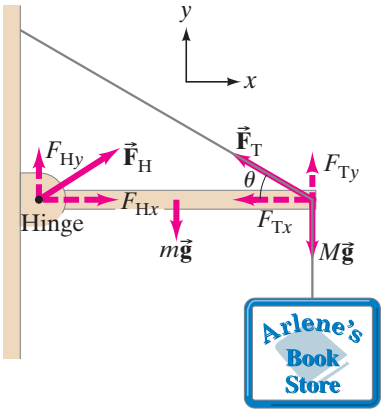


FIGURE 12–9 Hinged beam and cable, Example 12–5.

EXAMPLE 12–5 **Hinged beam and cable.** A uniform beam, 2.20 m long with mass $m = 25.0\,\text{kg}$, is mounted by a small hinge on a wall as shown in Fig. 12–9. The beam is held in a horizontal position by a cable that makes an angle $\theta = 30.0°$. The beam supports a sign of mass $M = 28.0\,\text{kg}$ suspended from its end. Determine the components of the force $\vec{\mathbf{F}}_\text{H}$ that the (smooth) hinge exerts on the beam, and the tension F_T in the supporting cable.

APPROACH Figure 12–9 is the free-body diagram for the beam, showing all the forces acting on the beam. It also shows the components of $\vec{\mathbf{F}}_\text{T}$ and a guess for the direction of $\vec{\mathbf{F}}_\text{H}$. We have three unknowns, $F_{\text{H}x}$, $F_{\text{H}y}$, and F_T (we are given θ), so we will need all three equations, $\Sigma F_x = 0$, $\Sigma F_y = 0$, $\Sigma\tau = 0$.

SOLUTION The sum of the forces in the vertical (y) direction is

$$\begin{aligned} \Sigma F_y &= 0 \\ F_{\text{H}y} + F_{\text{T}y} - mg - Mg &= 0. \end{aligned} \qquad \textbf{(i)}$$

In the horizontal (x) direction, the sum of the forces is

$$\begin{aligned} \Sigma F_x &= 0 \\ F_{\text{H}x} - F_{\text{T}x} &= 0. \end{aligned} \qquad \textbf{(ii)}$$

For the torque equation, we choose the axis at the point where $\vec{\mathbf{F}}_\text{T}$ and $M\vec{\mathbf{g}}$ act. Then our torque equation will contain only one unknown, $F_{\text{H}y}$, because the lever arms for $\vec{\mathbf{F}}_\text{T}$, $M\vec{\mathbf{g}}$, and $F_{\text{H}x}$ are zero. We choose torques that tend to rotate the beam counterclockwise as positive. The weight mg of the (uniform) beam acts at its center, so we have

$$\begin{aligned} \Sigma\tau &= 0 \\ -(F_{\text{H}y})(2.20\,\text{m}) + mg(1.10\,\text{m}) &= 0. \end{aligned}$$

We solve for $F_{\text{H}y}$:

$$F_{\text{H}y} = \left(\frac{1.10\,\text{m}}{2.20\,\text{m}}\right)mg = (0.500)(25.0\,\text{kg})\left(9.80\,\text{m/s}^2\right) = 123\,\text{N}. \qquad \textbf{(iii)}$$

Next, since the tension $\vec{\mathbf{F}}_\text{T}$ in the cable acts along the cable ($\theta = 30.0°$), we see from Fig. 12–9 that $\tan\theta = F_{\text{T}y}/F_{\text{T}x}$, or

$$F_{\text{T}y} = F_{\text{T}x}\tan\theta = F_{\text{T}x}(\tan 30.0°). \qquad \textbf{(iv)}$$

Equation (i) above gives

$$F_{\text{T}y} = (m + M)g - F_{\text{H}y} = (53.0\,\text{kg})\left(9.80\,\text{m/s}^2\right) - 123\,\text{N} = 396\,\text{N}.$$

Equations (iv) and (ii) give

$$\begin{aligned} F_{\text{T}x} &= F_{\text{T}y}/\tan 30.0° = 396\,\text{N}/\tan 30.0° = 686\,\text{N}; \\ F_{\text{H}x} &= F_{\text{T}x} = 686\,\text{N}. \end{aligned}$$

The components of $\vec{\mathbf{F}}_\text{H}$ are $F_{\text{H}y} = 123\,\text{N}$ and $F_{\text{H}x} = 686\,\text{N}$. The tension in the wire is $F_\text{T} = \sqrt{F_{\text{T}x}^2 + F_{\text{T}y}^2} = \sqrt{(686\,\text{N})^2 + (396\,\text{N})^2} = 792\,\text{N}$.[†]

Alternate Solution Let us see the effect of choosing a different axis for calculating torques, such as an axis through the hinge. Then the lever arm for F_H is zero, and the torque equation ($\Sigma\tau = 0$) becomes

$$-mg(1.10\,\text{m}) - Mg(2.20\,\text{m}) + F_{\text{T}y}(2.20\,\text{m}) = 0.$$

We solve this for $F_{\text{T}y}$ and find

$$F_{\text{T}y} = \frac{m}{2}g + Mg = (12.5\,\text{kg} + 28.0\,\text{kg})\left(9.80\,\text{m/s}^2\right) = 397\,\text{N}.$$

We get the same result, within the precision of our significant figures.

NOTE It doesn't matter which axis we choose for $\Sigma\tau = 0$. Using a second axis can serve as a check.

[†]Our calculation used numbers rounded off to 3 significant figures. If you keep an extra digit, or leave the numbers in your calculator, you get $F_{\text{T}y} = 396.5\,\text{N}$, $F_{\text{T}x} = F_{\text{H}x} = 686.8\,\text{N}$, $F_\text{T} = 793\,\text{N}$, and $F_{\text{H}y} = 122.5\,\text{N}$, all within the expected precision of 3 significant figures (Section 1–3).

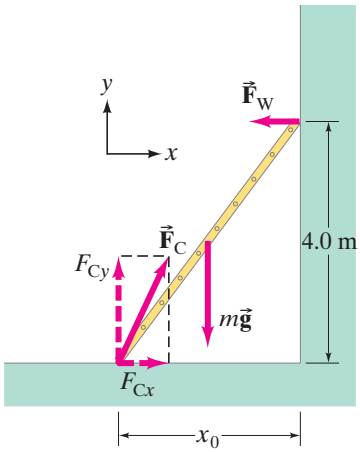


FIGURE 12–10 A ladder leaning against a wall. Example 12–6. The force $\vec{\mathbf{F}}_C$ that the cement floor exerts on the ladder need not be along the ladder which (unlike a cord) is rigid.

EXAMPLE 12–6 Ladder. A 5.0-m-long ladder leans against a smooth wall at a point 4.0 m above a cement floor as shown in Fig. 12–10. The ladder is uniform and has mass $m = 12.0\,\text{kg}$. Assuming the wall is frictionless, but the floor is not, determine the forces exerted on the ladder by the floor and by the wall.

APPROACH Figure 12–10 is the free-body diagram for the ladder, showing all the forces acting on the ladder. The wall, since it is frictionless, can exert a force only perpendicular to the wall, and we label that force $\vec{\mathbf{F}}_W$. The cement floor exerts a force $\vec{\mathbf{F}}_C$ which has both horizontal and vertical force components: F_{Cx} is frictional and F_{Cy} is the normal force. Finally, gravity exerts a force $mg = (12.0\,\text{kg})\,(9.80\,\text{m/s}^2) = 118\,\text{N}$ on the ladder at its midpoint, since the ladder is uniform.

SOLUTION Again we use the equilibrium conditions, $\Sigma F_x = 0$, $\Sigma F_y = 0$, $\Sigma\tau = 0$. We will need all three since there are three unknowns: F_W, F_{Cx}, and F_{Cy}. The y component of the force equation is

$$\Sigma F_y = F_{Cy} - mg = 0,$$

so immediately we have

$$F_{Cy} = mg = 118\,\text{N}.$$

The x component of the force equation is

$$\Sigma F_x = F_{Cx} - F_W = 0.$$

To determine both F_{Cx} and F_W, we need a torque equation. If we choose to calculate torques about an axis through the point where the ladder touches the cement floor, then $\vec{\mathbf{F}}_C$, which acts at this point, will have a lever arm of zero and so won't enter the equation. The ladder touches the floor a distance $x_0 = \sqrt{(5.0\,\text{m})^2 - (4.0\,\text{m})^2} = 3.0\,\text{m}$ from the wall (right triangle, $c^2 = a^2 + b^2$). The lever arm for mg is half this, or 1.5 m, and the lever arm for F_W is 4.0 m, Fig. 12–10. The torque equation about the ladder's contact point on the cement is

$$\Sigma\tau = (4.0\,\text{m})F_W - (1.5\,\text{m})mg = 0.$$

Thus

$$F_W = \frac{(1.5\,\text{m})\,(12.0\,\text{kg})\,(9.8\,\text{m/s}^2)}{4.0\,\text{m}} = 44\,\text{N}.$$

Then, from the x component of the force equation,

$$F_{Cx} = F_W = 44\,\text{N}.$$

Since the components of $\vec{\mathbf{F}}_C$ are $F_{Cx} = 44\,\text{N}$ and $F_{Cy} = 118\,\text{N}$, then

$$F_C = \sqrt{(44\,\text{N})^2 + (118\,\text{N})^2} = 126\,\text{N} \approx 130\,\text{N}$$

(rounded off to two significant figures), and it acts at an angle to the floor of

$$\theta = \tan^{-1}(118\,\text{N}/44\,\text{N}) = 70°.$$

NOTE The force $\vec{\mathbf{F}}_C$ does *not* have to act along the ladder's direction because the ladder is rigid and not flexible like a cord or cable (see page 99).

EXERCISE D Why is it reasonable to ignore friction along the wall, but not reasonable to ignore it along the floor?

*12–3 Applications to Muscles and Joints

The techniques we have been discussing for calculating forces on objects in equilibrium can be applied to the human body to study forces on muscles, bones, and joints. Generally a muscle is attached, via tendons, to two different bones, as in Fig. 12–11. The points of attachment are called *insertions*. Two bones are flexibly connected at a *joint*, such as at the elbow, knee, shoulder, and hip. A muscle exerts a pull when its fibers contract under stimulation by a nerve, but a muscle can *not* exert a push. Muscles that tend to bring two limbs closer together, such as the biceps muscle in the upper arm (Fig. 12–11), are called *flexors*; those that act to extend a limb outward, such as the triceps muscle in Fig. 12–11, are *extensors*. You use the flexor muscle in the upper arm when lifting an object in your hand; you use the extensor muscle when throwing a ball.

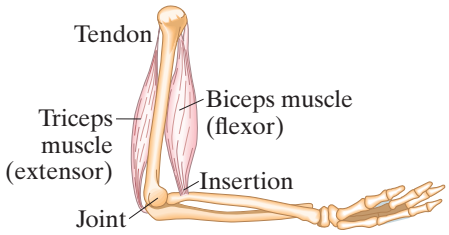


FIGURE 12–11 The biceps (flexor) and triceps (extensor) muscles in the human arm.

PHYSICS APPLIED
Forces in muscles and joints

EXAMPLE 12–7 **Force exerted by biceps muscle.** How much force must the biceps muscle exert when a 5.0-kg ball is held in the hand (*a*) with the arm horizontal as in Fig. 12–12a, and (*b*) when the arm is at a 45° angle as in Fig. 12–12b? The biceps muscle is connected to the forearm by a tendon attached 5.0 cm from the elbow joint. Assume that the mass of forearm and hand together is 2.0 kg and their CG is as shown.

APPROACH The free-body diagram for the forearm is shown in Fig. 12–12; the forces are the weights of the forearm and ball, the upward force $\vec{\mathbf{F}}_M$ exerted by the muscle, and a force $\vec{\mathbf{F}}_J$ exerted at the joint by the bone in the upper arm (all assumed to act vertically). We wish to find the magnitude of $\vec{\mathbf{F}}_M$, which can be done using the torque equation and by choosing our axis through the joint so that $\vec{\mathbf{F}}_J$ contributes zero torque.

SOLUTION (*a*) We calculate torques about the point where $\vec{\mathbf{F}}_J$ acts in Fig. 12–12a. The $\Sigma\tau = 0$ equation gives

$$(0.050\,\text{m})F_M - (0.15\,\text{m})(2.0\,\text{kg})g - (0.35\,\text{m})(5.0\,\text{kg})g = 0.$$

We solve for F_M:

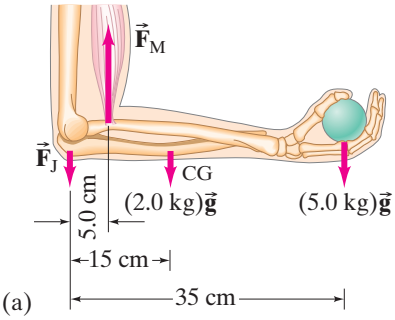
$$F_M = \frac{(0.15\,\text{m})(2.0\,\text{kg})g + (0.35\,\text{m})(5.0\,\text{kg})g}{0.050\,\text{m}} = (41\,\text{kg})g = 400\,\text{N}.$$

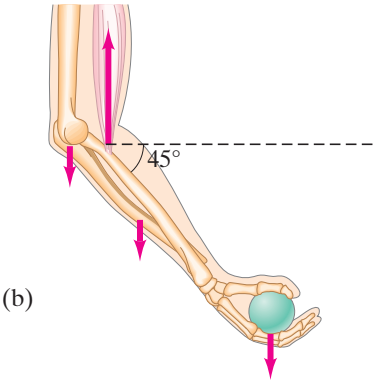
(*b*) The lever arm, as calculated about the joint, is reduced by the factor cos 45° for all three forces. Our torque equation will look like the one just above, except that each term will have its lever arm reduced by the same factor, which will cancel out. The same result is obtained, $F_M = 400\,\text{N}$.

NOTE The force required of the muscle (400 N) is quite large compared to the weight of the object lifted ($= mg = 49\,\text{N}$). Indeed, the muscles and joints of the body are generally subjected to quite large forces.

NOTE Forces exerted on joints can be large and even painful or injurious. Using $\Sigma F_y = 0$ we calculate for this case $F_J = F_M - (2.0\,\text{kg})g - (5.0\,\text{kg})g = 330\,\text{N}$.

FIGURE 12–12 Example 12–7, forces on forearm.





EXERCISE E How much mass could the person in Example 12–7 hold in the hand with a biceps force of 450 N if the tendon was attached 6.0 cm from the elbow instead of 5.0 cm?

Note that an increase in the distance of the joint to the point of insertion of the biceps muscle can be a considerable advantage for lifting and throwing. Champion athletes are often found to have muscle insertions farther from the joint than the average person, and if this applies to one muscle, it usually applies to all.

PHYSICS APPLIED
What can make an athlete

PHYSICS APPLIED
Forces on the spine, and back pain

As another example of the large forces acting within the human body, we consider the muscles used to support the body when you bend forward (Fig. 12–13a). The lowest vertebra on the spinal column (fifth lumbar vertebra) acts as a fulcrum for this bending position. Figure 12–13b is a simplified schematic drawing showing the forces on the upper body. The force exerted by the back muscles is $\vec{\mathbf{F}}_M$, the force on the base of the spine at the lowest vertebra is $\vec{\mathbf{F}}_V$, and $\vec{\mathbf{w}}_H$, $\vec{\mathbf{w}}_A$, and $\vec{\mathbf{w}}_T$ represent the weights of the **h**ead, freely hanging **a**rms, and **t**orso, respectively. Values are approximations. The distances (in cm) refer to a person 180 cm tall, but are approximately in the same ratio of 3 : 1 : 2 for an average person of any height; the result in the following Example is then independent of the height of the person.

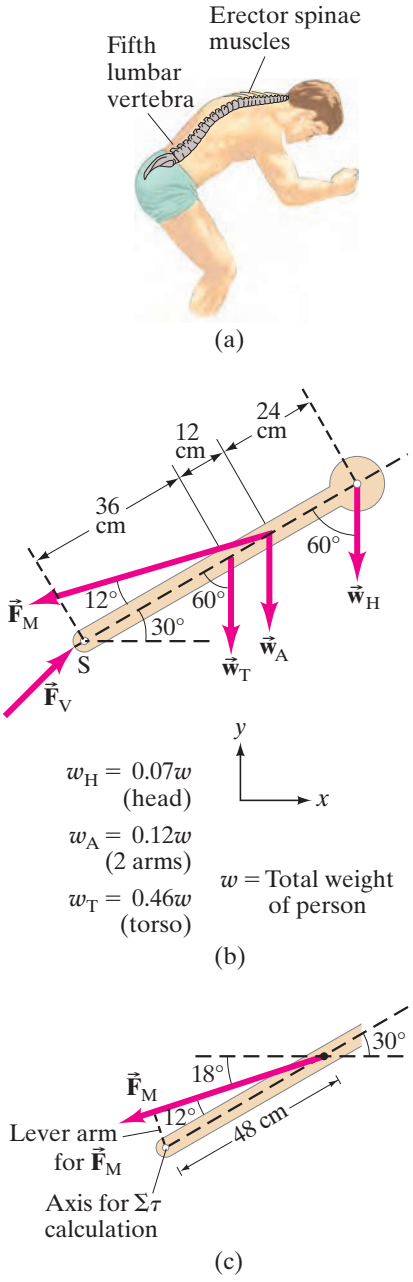


FIGURE 12–13 (a) A person bending over. (b) Forces on the back exerted by the back muscles ($\vec{\mathbf{F}}_M$) and by the vertebrae ($\vec{\mathbf{F}}_V$) when a person bends over. (c) Finding the lever arm for $\vec{\mathbf{F}}_M$.

EXAMPLE 12–8 **Forces on your back.** (*a*) Estimate the magnitude of the force $\vec{\mathbf{F}}_M$ the muscles exert on the back to support the upper body of a person bending forward, Fig. 12–13b. (*b*) Estimate the magnitude and direction of the force $\vec{\mathbf{F}}_V$ acting on the fifth lumbar vertebra (exerted by the spine below).

APPROACH We can calculate F_M using the torque equation if we take the axis at the base of the spine (point S); with this choice, the other unknown, F_V, doesn't appear in the equation because its lever arm is zero. To determine the lever arms, we need to use trigonometric functions.

SOLUTION (*a*) For $\vec{\mathbf{F}}_M$, the lever arm (perpendicular distance from axis to line of action of the force) will be the real distance to where the force acts (48 cm) multiplied by sin 12°, as shown in Fig. 12–13c. The lever arms for $\vec{\mathbf{w}}_H$, $\vec{\mathbf{w}}_A$, and $\vec{\mathbf{w}}_T$ can be seen from Fig. 12–13b to be their respective distances from S times sin 60°. F_M tends to rotate the torso counterclockwise, which we take to be positive. Then $\vec{\mathbf{w}}_H$, $\vec{\mathbf{w}}_A$, $\vec{\mathbf{w}}_T$ will contribute negative torques. Thus $\Sigma\tau = 0$ gives

$$(0.48\,\text{m})(\sin 12^\circ)(F_M) - (0.72\,\text{m})(\sin 60^\circ)(w_H) - (0.48\,\text{m})(\sin 60^\circ)(w_A) - (0.36\,\text{m})(\sin 60^\circ)(w_T) = 0.$$

Solving for F_M and putting in the values for w_H, w_A, w_T given in Fig. 12–13b, we find

$$F_M = \frac{(0.72\,\text{m})(0.07w) + (0.48\,\text{m})(0.12w) + (0.36\,\text{m})(0.46w)}{(0.48\,\text{m})(\sin 12^\circ)}(\sin 60^\circ)$$
$$= 2.37w \approx 2.4w,$$

where w is the total weight of the person.

(*b*) To get the components of $\vec{\mathbf{F}}_V$ we use the x and y components of the force equation (noting that $30^\circ - 12^\circ = 18^\circ$):

$$\Sigma F_y = F_{Vy} - F_M \sin 18^\circ - w_H - w_A - w_T = 0$$

so

$$F_{Vy} = (2.37 \sin 18^\circ + 0.07 + 0.12 + 0.46)w = 1.38w \approx 1.4w;$$

the x-component $\Sigma F_x = 0$ gives

$$\Sigma F_x = F_{Vx} - F_M \cos 18^\circ = 0$$

so

$$F_{Vx} = 2.37 \cos 18^\circ w = 2.25w \approx 2.3w,$$

where we keep 3 significant figures for calculating, but round off to 2 for giving the answer. Then

$$F_V = \sqrt{F_{Vx}^2 + F_{Vy}^2} = 2.6w.$$

The angle θ that F_V makes with the horizontal is given by $\tan\theta = F_{Vy}/F_{Vx} = 1.38/2.25 = 0.613$, so $\theta = 32^\circ$.

NOTE The force on the lowest vertebra is over $2\frac{1}{2}$ times the total body weight! The disks at the base of the spine are clearly being compressed under very large forces. [If the body was less bent over (say, the 30° angle in Fig. 12–13b becomes 40° or 50°), then the stress on the lower back will be less (see Problem 32).]

If the person in Fig. 12–13 has a mass of 90 kg and is holding 20 kg in his hands (this increases w_A to $0.34w$), then F_V is increased to almost four times the person's weight ($3.7w$). For this 200-lb person, the force on the disk would be over 700 lb! It is little wonder so many people suffer from low back pain at one time or another.

12–4 Stability and Balance

An object in static equilibrium, if left undisturbed, will undergo no translational or rotational acceleration since the sum of all the forces and the sum of all the torques acting on it are zero. However, if the object is displaced slightly, three outcomes are possible: (1) the object returns to its original position, in which case it is said to be in **stable equilibrium**; (2) the object moves even farther from its original position, and it is said to be in **unstable equilibrium**; or (3) the object remains in its new position, and it is said to be in **neutral equilibrium**.

Consider the following examples. A ball suspended freely from a string is in stable equilibrium, for if it is displaced to one side, it will return to its original position (Fig. 12–14a) due to the net force and torque exerted on it. On the other hand, a pencil standing on its point is in unstable equilibrium. If its center of gravity is directly over its tip (Fig. 12–14b), the net force and net torque on it will be zero. But if it is displaced ever so slightly as shown—say, by a slight vibration or tiny air current—there will be a torque on it, and this torque acts to make the pencil continue to fall in the direction of the original displacement. Finally, an example of an object in neutral equilibrium is a sphere resting on a horizontal tabletop. If it is moved slightly to one side, it will remain in its new position—no net torque acts on it.

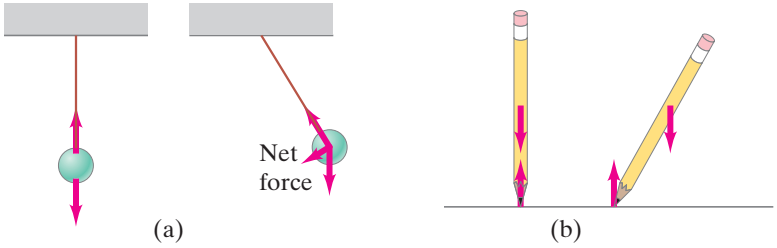


FIGURE 12–14 (a) Stable equilibrium, and (b) unstable equilibrium.

In most situations, such as in the design of structures and in working with the human body, we are interested in maintaining stable equilibrium, or *balance*, as we sometimes say. In general, an object whose center of gravity (CG) is below its point of support, such as a ball on a string, will be in stable equilibrium. If the CG is above the base of support, we have a more complicated situation. Consider a standing refrigerator (Fig. 12–15a). If it is tipped slightly, it will return to its original position due to the torque on it as shown in Fig. 12–15b. But if it is tipped too far, Fig. 12–15c, it will fall over. The critical point is reached when the CG shifts from one side of the pivot point to the other. When the CG is on one side, the torque pulls the object back onto its original base of support, Fig. 12–15b. If the object is tipped further, the CG goes past the pivot point and the torque causes the object to topple, Fig. 12–15c. In general, *an object whose center of gravity is above its base of support will be stable if a vertical line projected downward from the CG falls within the base of support*. This is because the normal force upward on the object (which balances out gravity) can be exerted only within the area of contact, so if the force of gravity acts beyond this area, a net torque will act to topple the object.

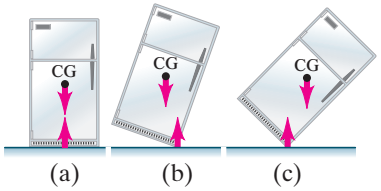


FIGURE 12–15 Equilibrium of a refrigerator resting on a flat floor.

Stability, then, can be relative. A brick lying on its widest face is more stable than a brick standing on its end, for it will take more of an effort to tip it over. In the extreme case of the pencil in Fig. 12–14b, the base is practically a point and the slightest disturbance will topple it. In general, the larger the base and the lower the CG, the more stable the object.

PHYSICS APPLIED
Humans and balance

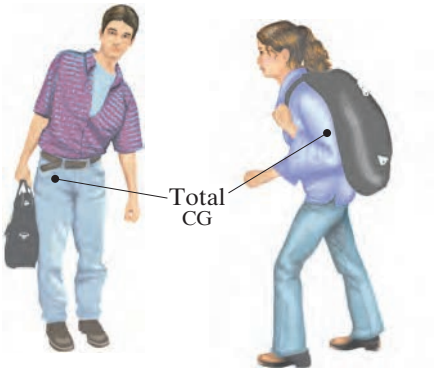


FIGURE 12–16 Humans adjust their posture to achieve stability when carrying loads.

FIGURE 12–17 Hooke's law: $\Delta\ell \propto$ applied force.

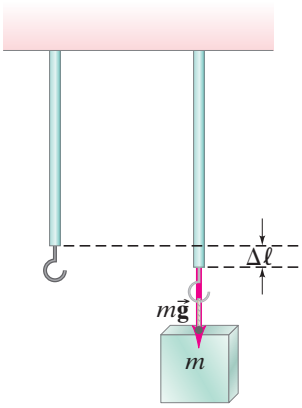
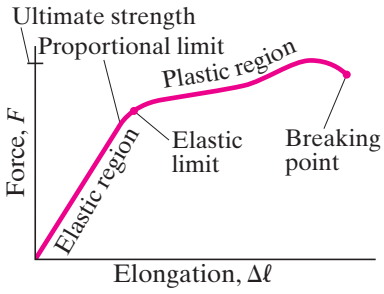


FIGURE 12–18 Applied force vs. elongation for a typical metal under tension.



In this sense, humans are less stable than four-legged mammals, which have a larger base of support because of their four legs, and most also have a lower center of gravity. When walking and performing other kinds of movement, a person continually shifts the body so that its CG is over the feet, although in the normal adult this requires no conscious thought. Even as simple a movement as bending over requires moving the hips backward so that the CG remains over the feet, and you do this repositioning without thinking about it. To see this, position yourself with your heels and back to a wall and try to touch your toes. You won't be able to do it without falling. Persons carrying heavy loads automatically adjust their posture so that the CG of the total mass is over their feet, Fig. 12–16.

12–5 Elasticity; Stress and Strain

In the first part of this Chapter we studied how to calculate the forces on objects in equilibrium. In this Section we study the effects of these forces: any object changes shape under the action of applied forces. If the forces are great enough, the object will break, or *fracture*, as we will discuss in the next Section.

Elasticity and Hooke's Law

If a force is exerted on an object, such as the vertically suspended metal rod shown in Fig. 12–17, the length of the object changes. If the amount of elongation, $\Delta\ell$, is small compared to the length of the object, experiment shows that $\Delta\ell$ is proportional to the force exerted on the object. This proportionality can be written as an equation:

$$F = k\,\Delta\ell. \qquad \textbf{(12–3)}$$

Here F represents the force pulling on the object, $\Delta\ell$ is the change in length, and k is a proportionality constant. Equation 12–3, which is sometimes called **Hooke's law**† after Robert Hooke (1635–1703), who first noted it, is found to be valid for almost any solid material from iron to bone—but it is valid only up to a point. If the force is too great, the object stretches excessively and eventually breaks.

Figure 12–18 shows a typical graph of applied force versus elongation. Up to a point called the **proportional limit**, Eq. 12–3 is a good approximation for many common materials, and the curve is a straight line. Beyond this point, the graph deviates from a straight line, and no simple relationship exists between F and $\Delta\ell$. Nonetheless, up to a point farther along the curve called the **elastic limit**, the object will return to its original length if the applied force is removed. The region from the origin to the elastic limit is called the *elastic region*. If the object is stretched beyond the elastic limit, it enters the *plastic region*: it does not return to the original length upon removal of the external force, but remains permanently deformed (such as a bent paper clip). The maximum elongation is reached at the *breaking point*. The maximum force that can be applied without breaking is called the **ultimate strength** of the material (actually, force per unit area, as we discuss in the next Section).

Young's Modulus

The amount of elongation of an object, such as the rod shown in Fig. 12–17, depends not only on the force applied to it, but also on the material of which it is made and on its dimensions. That is, the constant k in Eq. 12–3 can be written in terms of these factors.

If we compare rods made of the same material but of different lengths and cross-sectional areas, it is found that for the same applied force, the amount of stretch (again assumed small compared to the total length) is proportional to the original length and inversely proportional to the cross-sectional area.

†The term "law" applied to this relation is historical, but today it is not really appropriate. First of all, it is only an approximation, and second, it refers only to a limited set of phenomena. Most physicists today prefer to reserve the word "law" for those relations that are deeper and more encompassing and precise, such as Newton's laws of motion or the law of conservation of energy.

That is, the longer the object, the more it elongates for a given force; and the thicker it is, the less it elongates. These findings can be combined with Eq. 12–3 to yield

$$\Delta\ell = \frac{1}{E}\frac{F}{A}\ell_0, \qquad \textbf{(12–4)}$$

where ℓ_0 is the original length of the object, A is the cross-sectional area, and $\Delta\ell$ is the change in length due to the applied force F. E is a constant of proportionality[‡] known as the **elastic modulus**, or **Young's modulus**; its value depends only on the material. The value of Young's modulus for various materials is given in Table 12–1 (the shear modulus and bulk modulus in this Table are discussed later in this Section). Because E is a property only of the material and is independent of the object's size or shape, Eq. 12–4 is far more useful for practical calculation than Eq. 12–3.

TABLE 12–1 Elastic Moduli

Material	Young's Modulus, E (N/m^2)	Shear Modulus, G (N/m^2)	Bulk Modulus, B (N/m^2)
Solids			
Iron, cast	100×10^9	40×10^9	90×10^9
Steel	200×10^9	80×10^9	140×10^9
Brass	100×10^9	35×10^9	80×10^9
Aluminum	70×10^9	25×10^9	70×10^9
Concrete	20×10^9		
Brick	14×10^9		
Marble	50×10^9		70×10^9
Granite	45×10^9		45×10^9
Wood (pine) (parallel to grain)	10×10^9		
(perpendicular to grain)	1×10^9		
Nylon	$\approx 3 \times 10^9$		
Bone (limb)	15×10^9	80×10^9	
Liquids			
Water			2.0×10^9
Alcohol (ethyl)			1.0×10^9
Mercury			25×10^9
Gases[†]			
Air, H_2, He, CO_2			1.01×10^5

[†]At normal atmospheric pressure; no variation in temperature during process.

EXAMPLE 12–9 Tension in piano wire. A 1.60-m-long steel piano wire has a diameter of 0.20 cm. How great is the tension in the wire if it stretches 0.25 cm when tightened?

APPROACH We assume Hooke's law holds, and use it in the form of Eq. 12–4, finding E for steel in Table 12–1.

SOLUTION We solve for F in Eq. 12–4 and note that the area of the wire is $A = \pi r^2 = (3.14)(0.0010\,\text{m})^2 = 3.14 \times 10^{-6}\,\text{m}^2$, so F is

$$F = E\frac{\Delta\ell}{\ell_0}A = (2.0 \times 10^{11}\,\text{N/m}^2)\left(\frac{0.0025\,\text{m}}{1.60\,\text{m}}\right)(3.14 \times 10^{-6}\,\text{m}^2) = 980\,\text{N}.$$

NOTE The large tension in all the wires in a piano must be supported by a strong frame.

EXERCISE F Two steel wires have the same length and are under the same tension. But wire A has twice the diameter of wire B. Which of the following is true? (*a*) Wire B stretches twice as much as wire A. (*b*) Wire B stretches four times as much as wire A. (*c*) Wire A stretches twice as much as wire B. (*d*) Wire A stretches four times as much as wire B. (*e*) Both wires stretch the same amount.

[‡]The fact that E is in the denominator, so $1/E$ is the actual proportionality constant, is merely a convention. When we rewrite Eq. 12–4 to get Eq. 12–5, E is found in the numerator.

Stress and Strain

From Eq. 12–4, we see that the change in length of an object is directly proportional to the product of the object's length ℓ_0 and the force per unit area F/A applied to it. It is general practice to define the force per unit area as the **stress**:

$$\text{stress} = \frac{\text{force}}{\text{area}} = \frac{F}{A},$$

which has SI units of N/m^2. The **strain** is defined to be the ratio of the change in length to the original length:

$$\text{strain} = \frac{\text{change in length}}{\text{original length}} = \frac{\Delta\ell}{\ell_0},$$

and is dimensionless (no units). Strain is thus the fractional change in length of the object, and is a measure of how much the object has been deformed. Stress is applied to the material by external agents, whereas strain is the material's response to the stress. Equation 12–4 can be rewritten as

$$\frac{F}{A} = E\frac{\Delta\ell}{\ell_0} \qquad \textbf{(12–5)}$$

or

$$E = \frac{F/A}{\Delta\ell/\ell_0} = \frac{\text{stress}}{\text{strain}}.$$

Thus we see that the strain is directly proportional to the stress, in the linear (elastic) region of Fig. 12–18.

Tension, Compression, and Shear Stress

The rod shown in Fig. 12–19a is said to be under *tension* or **tensile stress**. Not only is there a force pulling down on the rod at its lower end, but since the rod is in equilibrium we know that the support at the top is exerting an equal[†] upward force on the rod at its upper end, Fig. 12–19a. In fact, this tensile stress exists throughout the material. Consider, for example, the lower half of a suspended rod as shown in Fig. 12–19b. This lower half is in equilibrium, so there must be an upward force on it to balance the downward force at its lower end. What exerts this upward force? It must be the upper part of the rod. Thus we see that external forces applied to an object give rise to internal forces, or stress, within the material itself.

Strain or deformation due to tensile stress is only one type of stress to which materials can be subjected. There are two other common types of stress: compressive and shear. **Compressive stress** is the exact opposite of tensile stress. Instead of being stretched, the material is compressed: the forces act inwardly on the object. Columns that support a weight, such as the columns of a Greek temple (Fig. 12–20), are subjected to compressive stress. Equations 12–4 and 12–5 apply equally well to compression and tension, and the values for the modulus E are usually the same.

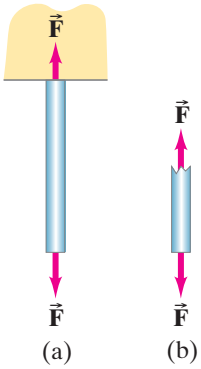


FIGURE 12–19 Stress exists *within* the material.

[†]Or a greater force if the weight of the rod cannot be ignored compared to F.



FIGURE 12–20 This Greek temple, in Agrigento, Sicily, built 2500 years ago, shows the post-and-beam construction. The columns are under compression.

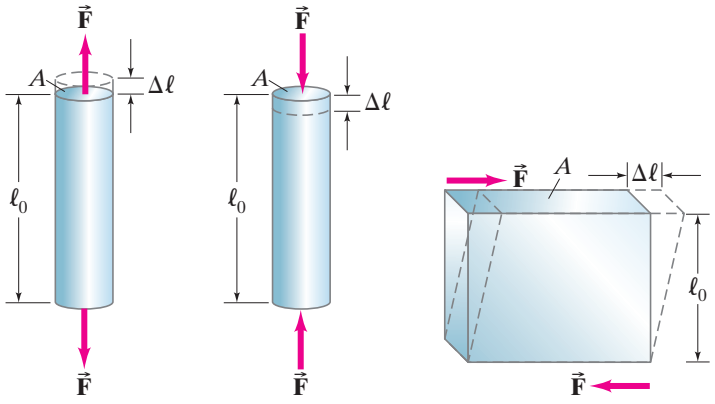


FIGURE 12–21 The three types of stress for rigid objects.

Figure 12–21 compares tensile and compressive stresses as well as the third type, shear stress. An object under **shear stress** has equal and opposite forces applied *across* its opposite faces. A simple example is a book or brick firmly attached to a tabletop, on which a force is exerted parallel to the top surface. The table exerts an equal and opposite force along the bottom surface. Although the dimensions of the object do not change significantly, the shape of the object does change, Fig. 12–21c. An equation similar to Eq. 12–4 can be applied to calculate shear strain:

$$\Delta\ell = \frac{1}{G}\frac{F}{A}\ell_0 \qquad \textbf{(12–6)}$$

but $\Delta\ell$, ℓ_0, and A must be reinterpreted as indicated in Fig. 12–21c. Note that A is the area of the surface *parallel* to the applied force (and not perpendicular as for tension and compression), and $\Delta\ell$ is *perpendicular* to ℓ_0. The constant of proportionality G is called the **shear modulus** and is generally one-half to one-third the value of Young's modulus E (see Table 12–1). Figure 12–22 suggests why $\Delta\ell \propto \ell_0$: the fatter book shifts more for the same shearing force.

FIGURE 12–22 The fatter book (a) shifts more than the thinner book (b) with the same applied shear force.



(a)

(b)

Volume Change—Bulk Modulus

If an object is subjected to inward forces from all sides, its volume will decrease. A common situation is an object submerged in a fluid. In this case, the fluid exerts a pressure on the object in all directions, as we shall discuss in Chapter 13. *Pressure* is defined as force per unit area, and thus is the equivalent of stress. For this situation the change in volume, ΔV, is proportional to the original volume, V_0, and to the change in the pressure, ΔP. We thus obtain a relation of the same form as Eq. 12–4 but with a proportionality constant called the **bulk modulus** B:

$$\frac{\Delta V}{V_0} = -\frac{1}{B}\Delta P \qquad \textbf{(12–7)}$$

or

$$B = -\frac{\Delta P}{\Delta V/V_0}.$$

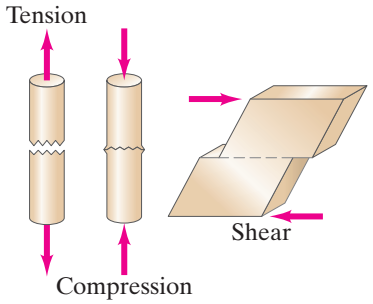
The minus sign means the volume *decreases* with an increase in pressure.

Values for the bulk modulus are given in Table 12–1. Since liquids and gases do not have fixed shapes, only the bulk modulus (not the Young's or shear moduli) applies to them.

12–6 Fracture

If the stress on a solid object is too great, the object fractures, or breaks (Fig. 12–23). Table 12–2 lists the **ultimate strengths** for tension, compression, and shear for a variety of materials. These values give the maximum force per unit area, or stress, that an object can withstand under each of these three types of stress for various types of material. They are, however, representative values only, and the actual value for a given specimen can differ considerably. It is therefore necessary to maintain a **safety factor** of from 3 to perhaps 10 or more: so the actual stresses should not exceed one-tenth to one-third of the values given in the Table.

FIGURE 12–23 Fracture as a result of the three types of stress.



You may encounter tables of "allowable stresses" in which appropriate safety factors have already been included.

TABLE 12–2 Ultimate Strengths of Materials (force/area)

Material	Tensile Strength (N/m^2)	Compressive Strength (N/m^2)	Shear Strength (N/m^2)
Iron, cast	170×10^6	550×10^6	170×10^6
Steel, typical	500×10^6	500×10^6	250×10^6
Steel, high-carbon	$\approx 2500 \times 10^6$	$\approx 2000 \times 10^6$	
Brass	250×10^6	250×10^6	200×10^6
Aluminum	200×10^6	200×10^6	200×10^6
Concrete	2×10^6	20×10^6	2×10^6
Brick		35×10^6	
Marble		80×10^6	
Granite		170×10^6	
Wood (pine) (parallel to grain)	40×10^6	35×10^6	5×10^6
(perpendicular to grain)		10×10^6	
Nylon	500×10^6		
Bone (limb)	130×10^6	170×10^6	

EXAMPLE 12–10 ESTIMATE Breaking the piano wire. The steel piano wire we discussed in Example 12–9 was 1.60 m long with a diameter of 0.20 cm. Approximately what tension force would break it?

APPROACH We set the tensile stress F/A equal to the ultimate tensile strength of typical steel given in Table 12–2.

SOLUTION The wire's area is $A = \pi r^2$, where $r = 0.10\,\text{cm} = 1.0 \times 10^{-3}\,\text{m}$. Table 12–2 tells us

$$\frac{F}{A} = 500 \times 10^6\,\text{N/m}^2,$$

so the wire would likely break if the force exceeded

$$F = (500 \times 10^6\,\text{N/m}^2)(\pi)(1.0 \times 10^{-3}\,\text{m})^2 = 1600\,\text{N}.$$

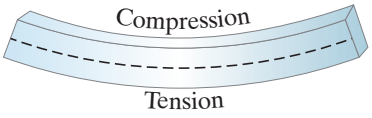


FIGURE 12–24 A beam sags, at least a little (but is exaggerated here), even under its own weight. The beam thus changes shape: the upper edge is compressed, and the lower edge is under tension (elongated). Shearing stress also occurs within the beam.

As can be seen in Table 12–2, concrete (like stone and brick) is reasonably strong under compression but extremely weak under tension. Thus concrete can be used as vertical columns placed under compression, but is of little value as a beam because it cannot withstand the tensile forces that result from the inevitable sagging of the lower edge of a beam (see Fig. 12–24). *Reinforced concrete*, in which iron rods are embedded in the concrete (Fig. 12–25), is much stronger than plain concrete. But the concrete on the lower edge of a loaded beam still tends to crack because it is weak under tension. This problem is solved with *prestressed concrete*, which also contains iron rods or a wire mesh, but during the pouring of the concrete, the rods or wire are held under tension. After the concrete dries, the tension on the iron is released, putting the concrete under compression. The amount of compressive stress is carefully predetermined so that when loads are applied to the beam, the compression on the lower edge is never allowed to be reduced so far as to put the concrete into tension.

FIGURE 12–25 Steel rods around which concrete is poured for strength.



CONCEPTUAL EXAMPLE 12–11 A tragic substitution. Two walkways, one above the other, are suspended from vertical rods attached to the ceiling of a high hotel lobby, Fig. 12–26a. The original design called for single rods 14 m long, but when such long rods proved to be unwieldy to install, it was decided to replace each long rod with two shorter ones as shown schematically in Fig. 12–26b. Determine the net force exerted by the rods on the supporting pin A (assumed to be the same size) for each design. Assume each vertical rod supports a mass m of each bridge.

RESPONSE The single long vertical rod in Fig. 12–26a exerts an upward force equal to mg on pin A to support the mass m of the upper bridge (Fig. 12–26c). Why? Because the pin is in equilibrium, and the other force that balances this is the downward force mg exerted on pin A by the upper bridge. There is thus a shear stress on the pin because the rod pulls up on one side of the pin, and the bridge pulls down on the other side. The situation when two shorter rods support the bridges (Fig. 12–26b) is shown in Fig. 12–26d, in which only the connections at the upper bridge are shown. The lower rod exerts a force mg downward on the lower of the two pins because it supports the lower bridge. The upper rod exerts a force $2mg$ on the upper pin (labelled A) because the upper rod supports both bridges. Thus we see that when the builders substituted two shorter rods for each single long one, the stress in the supporting pin A was *doubled*. What perhaps seemed like a simple substitution did, in fact, lead to a tragic collapse in 1981 with a loss of life of over 100 people (see Fig. 12–1). Having a feel for physics, and being able to make calculations based on physics, can have a great effect, literally, on people's lives.

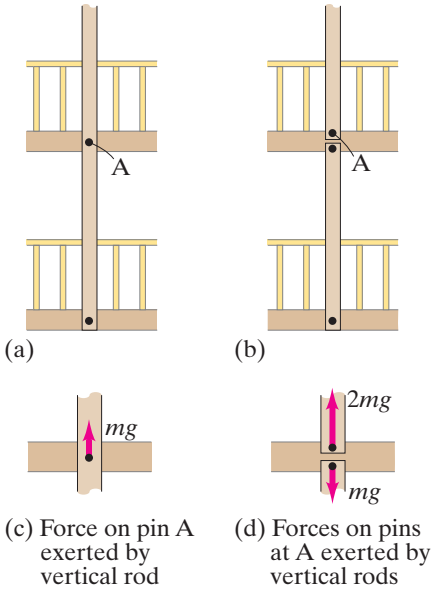
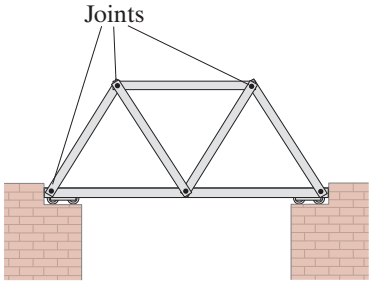


FIGURE 12–26 Example 12–11.

*12–7 Trusses and Bridges

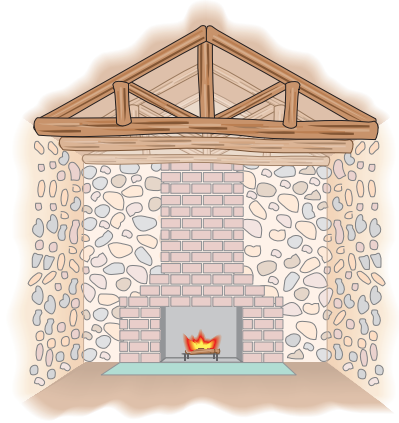
A beam used to span a wide space, as for a bridge, is subject to strong stresses of all three types as we saw in Fig. 12–24: compression, tension, and shear. A basic engineering device to support large spans is the *truss*, an example of which is shown in Fig. 12–27. Wooden truss bridges were first designed by the great architect Andrea Palladio (1518–1580), famous for his design of public buildings and villas. With the introduction of steel in the nineteenth century, much stronger steel trusses came into use, although wood trusses are still used to support the roofs of houses and mountain lodges (Fig. 12–28).

FIGURE 12–27 A truss bridge.



Basically, a **truss** is a framework of rods or struts joined together at their ends by **pins** (bolts) or rivets, always arranged as triangles. (Triangles are relatively stable, as compared to a rectangle, which easily becomes a parallelogram under sideways forces and then collapses.) The place where the struts are joined by a pin is called a **joint**.

FIGURE 12–28 (Below.) A roof truss.



It is commonly assumed that the struts of a truss are under pure compression or pure tension—that is, the forces act along the length of each strut, Fig. 12–29a. This is an ideal, valid only if a strut has no mass and supports no weight along its length, in which case a strut has only two forces on it, at the ends, as shown in Fig. 12–29a. If the strut is in equilibrium, these two forces must be equal and opposite in direction $(\Sigma\vec{\mathbf{F}} = 0)$. But couldn't they be at an angle, as in Fig. 12–29b? No, because then $\Sigma\vec{\boldsymbol{\tau}}$ would not be zero. The two forces *must* act along the strut if the strut is in equilibrium. But in a real case of a strut with mass, there are three forces on the strut, as shown in Fig. 12–29c, and $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ do not act along the strut; the vector diagram in Fig. 12–29d shows $\Sigma\vec{\mathbf{F}} = \vec{\mathbf{F}}_1 + \vec{\mathbf{F}}_2 + m\vec{\mathbf{g}} = 0$. Can you see why $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ both point *above* the strut? (Do $\Sigma\tau$ about each end.)

In Example 12–5, Fig. 12–9, the force $\vec{\mathbf{F}}_H$ at the hinge is *not* along the beam, but acts at an upward angle. If that beam were massless, we see from Eq. (iii) in Example 12–5 with $m = 0$, that $F_{Hy} = 0$, and $\vec{\mathbf{F}}_H$ would be along the beam.

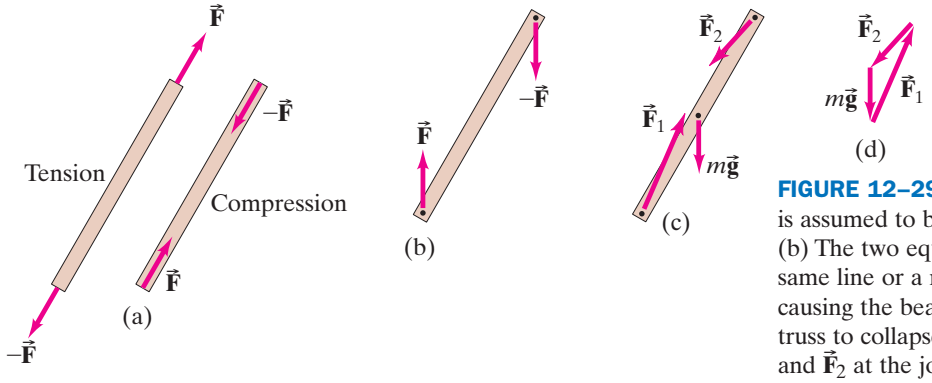


FIGURE 12–29 (a) If you ignore its mass, a strut (or rod) is assumed to be under pure tension or compression. (b) The two equal and opposite forces must be along the same line or a net torque would exist, as shown here, causing the beam to rotate (no longer be static) and the truss to collapse. (c) Real struts have mass, so the forces $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ at the joints or pins do not act precisely along the strut. (d) Vector diagram of part (c).

The assumption that the forces in each strut of a truss are purely along the strut is still very useful whenever the loads act only at the joints and are much greater than the weight of the struts themselves.

EXAMPLE 12–12 **A truss bridge.** Determine the tension or compression in each of the struts of the truss bridge shown in Fig. 12–30a. The bridge is 64 m long and supports a uniform level concrete roadway whose total mass is 1.40×10^6 kg. Use the **method of joints**, which involves (1) drawing a free-body diagram of the truss as a whole, and (2) drawing a free-body diagram for each of the pins (joints), one by one, and setting $\Sigma\vec{\mathbf{F}} = 0$ for each pin. Ignore the mass of the struts. Assume all triangles are equilateral.

APPROACH Any truss bridge has two trusses, one on each side of the roadway. Consider only one truss, Fig. 12–30a, and it will support half the weight of the roadway. That is, our truss supports a total mass $M = 7.0 \times 10^5$ kg. First we draw a free-body diagram for the entire truss as a single unit, which we assume rests on supports at either end that exert upward forces $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$, Fig. 12–30b. We assume the mass of the roadway acts entirely at the center, on pin C, as shown. From *symmetry* we can see that each of the end supports carries half the weight [or do a torque equation about, say, point A: $(F_2)(\ell) - Mg(\ell/2) = 0$], so

$$F_1 = F_2 = \tfrac{1}{2}Mg.$$

SOLUTION We look at pin A and apply $\Sigma\vec{\mathbf{F}} = 0$ to it. We label the forces on pin A due to each strut with two subscripts: $\vec{\mathbf{F}}_{AB}$ means the force exerted by the strut AB on pin A, and $\vec{\mathbf{F}}_{AC}$ is the force exerted by strut AC on A. $\vec{\mathbf{F}}_{AB}$ and $\vec{\mathbf{F}}_{AC}$ act along their respective struts; but not knowing whether each is compressive or tensile, we can draw four different free-body diagrams that satisfy this requirement, as shown in Fig. 12–30c. Only the one on the left could provide $\Sigma\vec{\mathbf{F}} = 0$, so we immediately know the directions of $\vec{\mathbf{F}}_{AB}$ and $\vec{\mathbf{F}}_{AC}$.† These forces act on pin A. The force that pin A exerts on strut AB is opposite in direction to $\vec{\mathbf{F}}_{AB}$ (Newton's third law), so strut AB is under compression, and strut AC, in turn, is under tension. Now let's calculate the magnitudes of $\vec{\mathbf{F}}_{AB}$ and $\vec{\mathbf{F}}_{AC}$. At pin A:

$$\Sigma F_x = F_{AC} - F_{AB}\cos 60^\circ = 0$$
$$\Sigma F_y = F_1 - F_{AB}\sin 60^\circ = 0.$$

Thus (recall $\cos 60^\circ = \sin 30^\circ = \frac{1}{2}$, and $\sin 60^\circ = \sqrt{1-\cos^2 60^\circ} = \sqrt{1-(\frac{1}{2})^2} = \sqrt{3}/2$):

$$F_{AB} = \frac{F_1}{\sin 60^\circ} = \frac{\frac{1}{2}Mg}{\frac{1}{2}\sqrt{3}} = \frac{1}{\sqrt{3}}Mg,$$

which equals $(7.0 \times 10^5\ \text{kg})(9.8\ \text{m/s}^2)/\sqrt{3} = 4.0 \times 10^6\ \text{N}$.
Then

$$F_{AC} = F_{AB}\cos 60^\circ = \frac{1}{2\sqrt{3}}Mg.$$

Next we look at pin B, and Fig. 12–30d is the free-body diagram. [Convince yourself that if $\vec{\mathbf{F}}_{BD}$ or $\vec{\mathbf{F}}_{BC}$ were in the opposite direction, $\Sigma\vec{\mathbf{F}}$ could not be zero; note that $\vec{\mathbf{F}}_{BA} = -\vec{\mathbf{F}}_{AB}$ (and $F_{BA} = F_{AB}$) because now we are at the opposite end of strut AB.] We see that BC is under tension and BD compression. (Recall that the forces on the struts are opposite to the forces shown which are on the pin.) We set $\Sigma\vec{\mathbf{F}} = 0$ on pin B (Fig. 12–30d):

$$\Sigma F_x = F_{BA}\cos 60^\circ + F_{BC}\cos 60^\circ - F_{BD} = 0$$
$$\Sigma F_y = F_{BA}\sin 60^\circ - F_{BC}\sin 60^\circ = 0.$$

Then, because $F_{BA} = F_{AB}$, we have from the ΣF_y equation

$$F_{BC} = F_{AB} = \frac{1}{\sqrt{3}}Mg.$$

From the ΣF_x equation, Fig. 12–30d,

$$F_{BD} = F_{BA}\cos 60^\circ + F_{BC}\cos 60^\circ = \frac{1}{\sqrt{3}}Mg(\tfrac{1}{2}) + \frac{1}{\sqrt{3}}Mg(\tfrac{1}{2}) = \frac{1}{\sqrt{3}}Mg.$$

The solution is complete. By symmetry, $F_{DE} = F_{AB}$, $F_{CE} = F_{AC}$, and $F_{CD} = F_{BC}$.

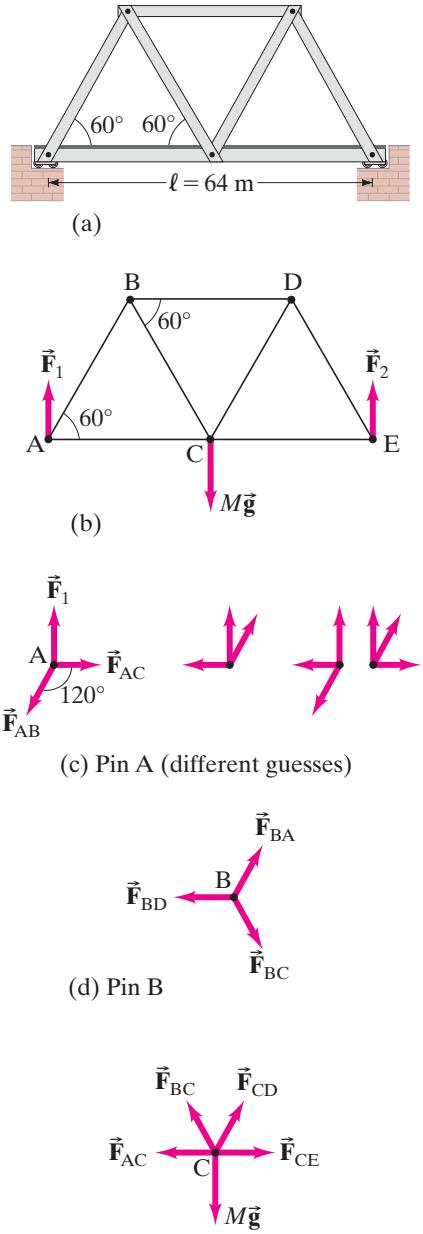


FIGURE 12–30 Example 12–12. (a) A truss bridge. Then free-body diagrams: (b) for the entire truss, (c) for pin A (different guesses), (d) for pin B and (e) for pin C.

†If we were to choose the direction of a force on a diagram opposite to what it really is, we would get a minus sign.

NOTE As a check, calculate ΣF_x and ΣF_y for pin C and see if they equal zero. Figure 12–30e shows the free-body diagram.

Example 12–12 put the roadway load at the center, C. Now consider a heavy load, such as a heavy truck, supported by strut AC at its middle, as shown in Fig. 12–31a. The strut AC sags under this load, telling us there is shear stress in strut AC. Figure 12–31b shows the forces exerted on strut AC: the weight of the truck $m\vec{\mathbf{g}}$, and the forces $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_C$ that pins A and C exert on the strut. [Note that $\vec{\mathbf{F}}_1$ does not appear because it is a force (exerted by external supports) that acts on pin A, not on strut AC.] The forces that pins A and C exert on strut AC will act not only along the strut, but will have vertical components too, perpendicular to the strut to balance the weight of the truck, $m\vec{\mathbf{g}}$, creating shear stress. The other struts, not bearing weight, remain under pure tension or compression. Problem 57 deals with this situation, and an early step in its solution is to calculate the forces $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_C$ by using torque equations for the strut.

For very large bridges, truss structures are too heavy. One solution is to build suspension bridges, with the load being carried by relatively light suspension cables under tension, supporting the roadway by means of closely spaced vertical wires, as shown in Fig. 12–32a.

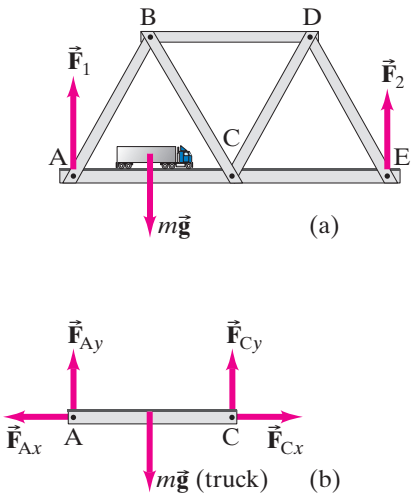


FIGURE 12–31 (a) Truss with truck of mass m at center of strut AC. (b) Forces on strut AC.

EXAMPLE 12–13 **Suspension bridge.** Determine the shape of the cable between the two towers of a suspension bridge (as in Fig. 12–32a), assuming the weight of the roadway is supported uniformly along its length. Ignore the weight of the cable.

APPROACH We take $x = 0$, $y = 0$ at the center of the span, as shown in Fig. 12–32b. Let $\vec{\mathbf{F}}_{T0}$ be the tension in the cable at $x = 0$; it acts horizontally as shown. Let F_T be the tension in the cable at some other place where the horizontal coordinate is x, as shown. This section of cable supports a portion of the roadway whose weight w is proportional to the distance x, since the roadway is assumed uniform; that is,

$$w = \lambda x$$

where λ is the weight per unit length.

SOLUTION We set $\Sigma\vec{\mathbf{F}} = 0$:

$$\Sigma F_x = F_T \cos\theta - F_{T0} = 0$$
$$\Sigma F_y = F_T \sin\theta - w = 0$$

We divide these two equations and we use $w = \lambda x$,

$$\tan\theta = \frac{w}{F_{T0}} = \frac{\lambda x}{F_{T0}}.$$

The slope of our curve (the cable) at any point is

$$\frac{dy}{dx} = \tan\theta$$

or, since we just saw $\tan\theta = \lambda x/F_{T0}$,

$$\frac{dy}{dx} = \frac{\lambda}{F_{T0}} x.$$

We integrate this:

$$\int dy = \frac{\lambda}{F_{T0}} \int x\, dx$$
$$y = Ax^2 + B$$

where we set $A = \lambda/F_{T0}$ and B is a constant of integration. This is just the equation of a parabola.

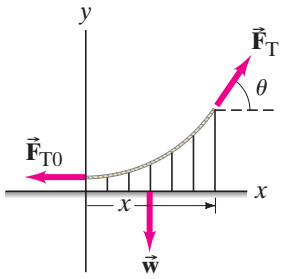
NOTE Real bridges have cables that do have mass, so the cables hang only approximately as a parabola, although often it is quite close.

PHYSICS APPLIED
Suspension bridge

FIGURE 12–32 (a) Suspension bridge (Golden Gate bridge, California). (b) Free-body diagram for Example 12–13.



(a)



(b)

*12–8 Arches and Domes

There are various ways that engineers and architects can span an interior or exterior space. In this Section we discuss arches and domes.



(a)



(b)

FIGURE 12–33 (a) Round arches in the Roman Forum, over 2000 years old. The one in the background is the Arch of Titus. (b) A modern arch is used here to good effect for a bridge over a chasm on the California coast.

PHYSICS APPLIED
Architecture: Beams, arches and domes

The semicircular **arch** (Fig. 12–33a) was introduced by the ancient Romans 2000 years ago. Aside from its aesthetic appeal, it was a tremendous technological innovation. The advantage of the "true" or semicircular arch is that, if well designed, its wedge-shaped stones experience stress which is mainly compressive even when supporting a large load such as the wall and roof of a cathedral. Because the wedge-shaped stones are forced to squeeze against one another, they are mainly under compression (see Fig. 12–34). A round arch consisting of many well-shaped stones could span a very wide space. When supporting a vertical "load" above it, the arch would tend to split apart if there were not horizontal forces exerted on it by some sort of buttressing, as we discuss shortly.

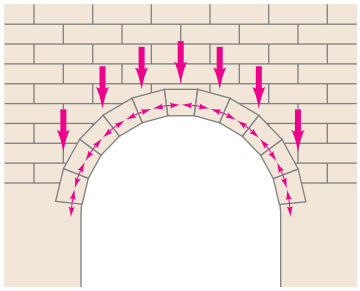


FIGURE 12–34 Stones in a round arch (see Fig. 12–33a) are mainly under compression.

FIGURE 12–35 Flying buttresses (on the cathedral of Notre Dame, in Paris).



The pointed arch came into use about A.D. 1100 and became the hallmark of the great Gothic cathedrals. It too was an important technical innovation, and was first used to support heavy loads such as the tower and central arch of a cathedral. Apparently the builders realized that, because of the steepness of the pointed arch, the forces due to the weight above would act more nearly vertically, so less horizontal buttressing would be needed. The pointed arch reduced the load on the walls, so there could be more window openings and light. The smaller buttressing needed was provided on the outside by graceful flying buttresses (Fig. 12–35).

The technical innovation of the pointed arch was achieved not through calculation but through experience and intuition; it was not until much later that detailed calculations, such as those presented earlier in this Chapter, came into use.

To make an accurate analysis of a stone arch is quite difficult in practice. But if we make some simplifying assumptions, we can show why the horizontal component of the external force needed at the base is less for a pointed arch than for a round one. Figure 12–36 shows a round arch and a pointed arch, each with an 8.0-m span. The height of the round arch is thus 4.0 m, whereas that of the pointed arch is larger and has been chosen to be 8.0 m. Each arch

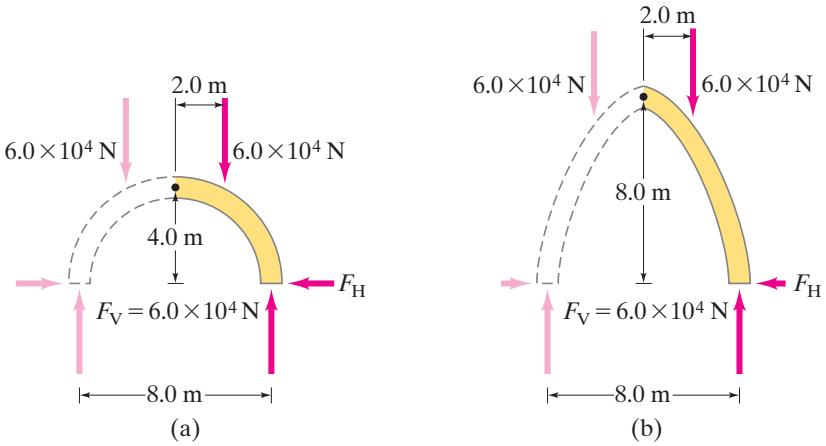


FIGURE 12–36 (a) Forces in a round arch, compared (b) with those in a pointed arch.

supports a weight of $12.0 \times 10^4\,\text{N}$ $(= 12{,}000\,\text{kg} \times g)$ which, for simplicity, we have divided into two parts (each $6.0 \times 10^4\,\text{N}$) acting on the two halves of each arch as shown. To be in equilibrium, each of the supports below the arch must exert an upward force of $6.0 \times 10^4\,\text{N}$. For rotational equilibrium, each support also exerts a horizontal force, F_H, at the base of the arch, and it is this we want to calculate. Due to the *symmetry*, we can focus only on the right half of each arch. We set equal to zero the total torque calculated about the apex of the arch due to the three forces exerted on that half arch. The torque equation $(\Sigma\tau = 0)$ contains three terms: the weight above the arch, the support F_V below, and the horizontal force F_H, which for the round arch (see Fig. 12–36a) is

$$-(2.0\,\text{m})(6.0 \times 10^4\,\text{N}) + (4.0\,\text{m})(6.0 \times 10^4\,\text{N}) - (4.0\,\text{m})(F_\text{H}) = 0.$$

Thus $F_\text{H} = 3.0 \times 10^4\,\text{N}$ for the round arch. For the pointed arch, the torque equation is (see Fig. 12–36b)

$$-(2.0\,\text{m})(6.0 \times 10^4\,\text{N}) + (4.0\,\text{m})(6.0 \times 10^4\,\text{N}) - (8.0\,\text{m})(F_\text{H}) = 0.$$

Solving, we find that $F_\text{H} = 1.5 \times 10^4\,\text{N}$—only half as much as for the round arch! From this calculation we can see that the horizontal buttressing force required for a pointed arch is less because the arch is higher, and there is therefore a longer lever arm for this force. Indeed, the steeper the arch, the less the horizontal component of the force needs to be, and hence the more nearly vertical is the force exerted at the base of the arch.

Whereas an arch spans a two-dimensional space, a **dome**—which is basically an arch rotated about a vertical axis—spans a three-dimensional space. The Romans built the first large domes. Their shape was hemispherical and some still stand, such as that of the Pantheon in Rome (Fig. 12–37), built 2000 years ago.



(a) (b)

FIGURE 12–37 Interior of the Pantheon in Rome, Italy, built almost 2000 years ago. (a) This view, showing the great dome and its central opening for light, was painted about 1740 by Panini. Photographs (b) do not capture it as well as this painting does.



FIGURE 12–38 Brunelleschi's dome on the cathedral of Florence, Italy.

Fourteen centuries later, a new cathedral being built in Florence was to have a dome 43 m in diameter to rival that of the Pantheon. This new dome was to rest on a "drum" with no external abutments. Filippo Brunelleschi (1377–1446) designed a pointed dome (Fig. 12–38), since a pointed dome, like a pointed arch, exerts a smaller side thrust against its base. A dome, like an arch, is not stable until all the stones are in place. To support smaller domes during construction, wooden frameworks were used. But no trees big enough or strong enough could be found to span the 43-m space required. Brunelleschi decided to try to build the dome in horizontal layers, each bonded to the previous one, holding it in place until the last stone of the circle was placed. Each closed ring was then strong enough to support the next layer. It was an amazing feat, an engineering marvel, and so its picture is also on the cover. Only in the twentieth century were larger domes built. One of the largest is the Superdome in New Orleans, completed in 1975, whose 200-m-diameter dome is made of steel trusses and concrete.

EXAMPLE 12–14 A modern dome. The 1.2×10^6 kg dome of the Small Sports Palace in Rome (Fig. 12–39a) is supported by 36 buttresses positioned at a 38° angle so that they connect smoothly with the dome. Calculate the components of the force, F_H and F_V, that each buttress exerts on the dome so that the force acts purely in compression—that is, at a 38° angle (Fig. 12–39b).

APPROACH We can find the vertical component F_V exerted upward by each buttress because each supports $\frac{1}{36}$ of the dome's weight. We find F_H knowing that the buttress needs to be under compression so $\vec{\mathbf{F}} = \vec{\mathbf{F}}_V + \vec{\mathbf{F}}_H$ acts at a 38° angle.

SOLUTION The vertical load on *each* buttress is $\frac{1}{36}$ of the total weight. Thus

$$F_V = \frac{mg}{36} = \frac{(1.2 \times 10^6\,\text{kg})(9.8\,\text{m/s}^2)}{36} = 330{,}000\,\text{N}.$$

The force must act at a 38° angle at the base of the dome in order to be purely compressive. Thus

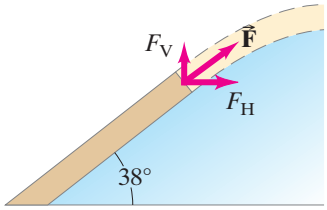
$$\tan 38° = \frac{F_V}{F_H};$$

$$F_H = \frac{F_V}{\tan 38°} = \frac{330{,}000\,\text{N}}{\tan 38°} = 420{,}000\,\text{N}.$$

NOTE For each buttress to exert this 420,000-N horizontal force, a prestressed-concrete tension ring surrounds the base of the buttresses beneath the ground (see Problem 58 and Fig. 12–83).



(a)



(b)

FIGURE 12–39 Example 12–14. (a) The dome of the Small Sports Palace in Rome, built by Pier Luigi Nervi for the 1960 Olympics. (b) The force components that each buttress exerts on the dome.

Summary

An object at rest is said to be in **equilibrium**. The subject concerned with the determination of the forces within a structure at rest is called **statics**.

The two necessary conditions for an object to be in equilibrium are (1) the vector sum of all the forces on it must be zero, and (2) the sum of all the torques (calculated about any arbitrary axis) must also be zero. For a two-dimensional problem we can write

$$\Sigma F_x = 0, \quad \Sigma F_y = 0, \quad \Sigma\tau = 0. \qquad \textbf{(12–1, 12–2)}$$

Apply these equilibrium equations to one object at a time.

An object in static equilibrium is said to be in (*a*) **stable**, (*b*) **unstable**, or (*c*) **neutral equilibrium**, depending on whether a slight displacement leads to (*a*) a return to the original position, (*b*) further movement away from the original position, or (*c*) rest in the new position. An object in stable equilibrium is also said to be in *balance*.

Hooke's law applies to many elastic solids, and states that the change in length of an object is proportional to the applied force:

$$F = k\,\Delta\ell. \qquad \textbf{(12–3)}$$

If the force is too great, the object will exceed its **elastic limit**, which means it will no longer return to its original shape when the distorting force is removed. If the force is even greater, the **ultimate strength** of the material can be exceeded, and the object will **fracture**. The force per unit area acting on an object is the **stress**, and the resulting fractional change in length is the **strain**.

The stress on an object is present within the object and can be of three types: **compression**, **tension**, or **shear**. The ratio of stress to strain is called the **elastic modulus** of the material. **Young's modulus** applies for compression and tension, and the **shear modulus** for shear. **Bulk modulus** applies to an object (or a fluid) whose volume changes as a result of pressure on all sides. All three moduli are constants for a given material when distorted within the elastic region.

[*Arches and domes are special ways to span a space that allow the stresses to be managed well. **Truss** structures, as on a bridge, require detailed statics calculations.]

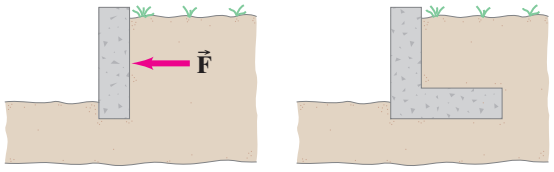
Questions

1. Describe several situations in which an object is not in equilibrium, even though the net force on it is zero.
2. A bungee jumper momentarily comes to rest at the bottom of the dive before he springs back upward. At that moment, is the bungee jumper in equilibrium? Explain.
3. You can find the center of gravity of a meter stick by resting it horizontally on your two index fingers, and then slowly drawing your fingers together. First the meter stick will slip on one finger, and then on the other, but eventually the fingers meet at the CG. Why does this work?
4. Your doctor's scale has arms on which weights slide to counter your weight, Fig. 12–40. These weights are much lighter than you are. How does this work?



FIGURE 12–40 Question 4.

5. A ground retaining wall is shown in Fig. 12–41a. The ground, particularly when wet, can exert a significant force F on the wall. (*a*) What force produces the torque to keep the wall upright? (*b*) Explain why the retaining wall in Fig. 12–41b would be much less likely to overturn than that in Fig. 12–41a.



(a) (b)

FIGURE 12–41 Question 5.

6. Can the sum of the torques on an object be zero while the net force on the object is nonzero? Explain.
7. A ladder, leaning against a wall, makes a 60° angle with the ground. When is it more likely to slip: when a person stands on the ladder near the top or near the bottom? Explain.
8. Figure 12–42 shows a cone. Explain how to lay it on a flat table so that it is in (*a*) stable equilibrium, (*b*) unstable equilibrium, (*c*) neutral equilibrium.

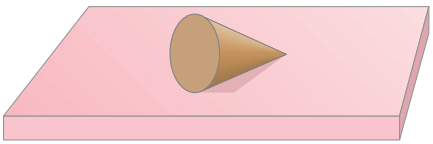
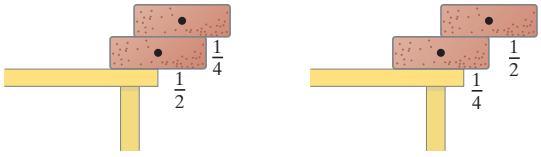


FIGURE 12–42 Question 8.

9. Stand facing a wall with your toes tight against the wall, and go up on your tiptoes. Try it again after turning around and push your back against the wall with your heels tight against the wall. Explain why you can or cannot perform these motions.
10. Place yourself facing the edge of an open door. Position your feet astride the door with your nose and abdomen touching the door's edge. Try to rise on your tiptoes. Why can't this be done?
11. Why is it not possible to sit upright in a chair and rise to your feet without first leaning forward?
12. Why is it more difficult to do sit-ups when your knees are bent than when your legs are stretched out?
13. Explain why touching your toes while you are seated on the floor with outstretched legs produces less stress on the lower spinal column than when touching your toes from a standing position. Use a diagram.
14. Which configurations of brick, Fig. 12–43a or Fig. 12–43b, is more likely to be stable? Why?



(a) (b)

FIGURE 12–43 Question 14. The dots indicate the CG of each brick (assumed uniform). The fractions $\frac{1}{4}$ and $\frac{1}{2}$ indicate what portion of each brick is hanging beyond its support.

15. Is the Young's modulus for an elastic bungee cord smaller or larger than that for an ordinary rope?
16. Examine how a pair of scissors or shears cuts through a piece of cardboard. Is the name "shears" justified? Explain.
17. Materials such as ordinary concrete and stone are very weak under tension or shear. Would it be wise to use such a material for either of the supports of the cantilever shown in Fig. 12–8? If so, which one(s)? Explain.

MisConceptual Questions

1. A uniform meter stick supported at the 25-cm mark is in equilibrium when a 1-kg rock is suspended at the 0-cm end (as shown in Fig. 12–44). The mass of the meter stick is
(*a*) greater than the mass of the rock.
(*b*) equal to the mass of the rock.
(*c*) less than the mass of the rock.
(*d*) none of these.

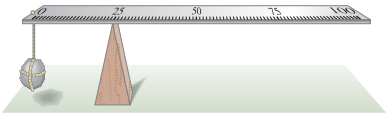


FIGURE 12–44 MisConceptual Question 1.

2. When you apply the torque equation $\Sigma\tau = 0$ to an object in equilibrium, the axis about which torques are calculated
(*a*) must be located at a pivot.
(*b*) must be located at the object's center of gravity.
(*c*) should be located at the edge of the object.
(*d*) can be located anywhere.

3. A uniform beam is hinged at one end and held in a horizontal position by a cable, as shown in Fig. 12–45. The tension in the cable
(*a*) must be at least half the weight of the beam, no matter what the angle of the cable.
(*b*) could be less than half the beam's weight for some angles.
(*c*) will be half the beam's weight for all angles.
(*d*) will equal the beam's weight for all angles.

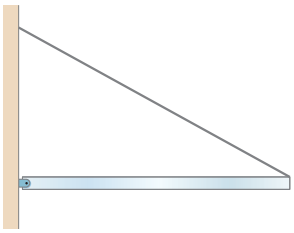


FIGURE 12–45
MisConceptual Question 3: beam and cable.

4. A heavy ball suspended by a cable is pulled to the side by a horizontal force $\vec{\mathbf{F}}$ as shown in Fig. 12–46. If angle θ is small, the magnitude of the force F can be less than the weight of the ball because:
(*a*) the force holds up only part of the ball's weight.
(*b*) even though the ball is stationary, it is not really in equilibrium.
(*c*) $\vec{\mathbf{F}}$ is equal to only the x component of the tension in the cable.
(*d*) the original statement is not true. To move the ball, $\vec{\mathbf{F}}$ must be at least equal to the ball's weight.

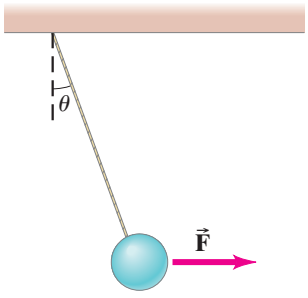


FIGURE 12–46
MisConceptual Question 4.

5. Two children are balanced on opposite sides of a seesaw. If one child leans inward toward the pivot point, her side will
(*a*) rise.
(*b*) fall.
(*c*) neither rise nor fall.

6. A 10.0-N weight is suspended by two cords as shown in Fig. 12–47. What can you say about the tension in the two cords?
(*a*) The tension in both cords is 5.0 N.
(*b*) The tension in both cords is equal but not 5.0 N.
(*c*) The tension in cord A is greater than that in cord B.
(*d*) The tension in cord B is greater than that in cord A.

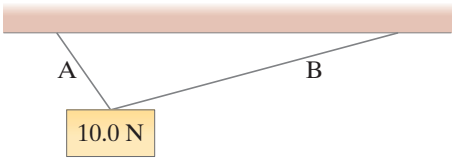


FIGURE 12–47 MisConceptual Question 6.

7. As you increase the force that you apply while pulling on a rope, which of the following is NOT affected?
(*a*) The stress on the rope.
(*b*) The strain on the rope.
(*c*) The Young's modulus of the rope.
(*d*) All of the above.
(*e*) None of the above.

8. A woman is balancing on a high wire which is tightly strung, dipping only slightly from horizontal, as shown in Fig. 12–48. The tension in the wire is
(*a*) about half the woman's weight.
(*b*) about twice the woman's weight.
(*c*) about equal to the woman's weight.
(*d*) much less than the woman's weight.
(*e*) much more than the woman's weight.

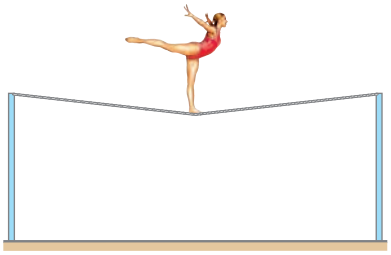


FIGURE 12–48 MisConceptual Question 8.

9. A parking garage is designed for two levels of cars. To make more money, the owner decides to double the height of the garage. For the support columns to hold up four floors of cars instead of two, how should he change the columns' diameter?
(*a*) Double the area of the columns by increasing their diameter by a factor of 2.
(*b*) Double the area of the columns by increasing their diameter by a factor of $\sqrt{2}$.
(*c*) Quadruple the area of the columns by increasing their diameter by a factor of 2.
(*d*) Increase the area of the columns by a factor of 8 by increasing their diameter by a factor of $2\sqrt{2}$.
(*e*) He doesn't need to increase the diameter of the columns.

10. A perfectly elastic rubber band is stretched by 1.0 cm when a force of 0.35 N is applied to each end. If instead a force of 0.70 N is applied to each end, estimate how far the rubber band will stretch from its unstretched length:
(*a*) 0.25 cm.
(*b*) 0.5 cm.
(*c*) 1.0 cm.
(*d*) 2.0 cm.
(*e*) 4.0 cm.

11. A larger boy and a smaller girl are balanced on a massless seesaw board. If they both move forward so that they are one-half their original distances from the pivot point, what will happen to the seesaw?
(*a*) It is impossible to say without knowing the masses.
(*b*) Nothing, the seesaw will still be balanced.
(c) The side the girl is sitting on will tilt downward.
(*d*) The side the boy is sitting on will tilt downward.
(*e*) It is impossible to say without knowing the distances.

Problems

12–1 and 12–2 Equilibrium

1. (I) Three forces are applied to a tree sapling, as shown in Fig. 12–49, to stabilize it. If $\vec{\mathbf{F}}_A = 385\text{ N}$ and $\vec{\mathbf{F}}_B = 515\text{ N}$, find $\vec{\mathbf{F}}_C$ in magnitude and direction.

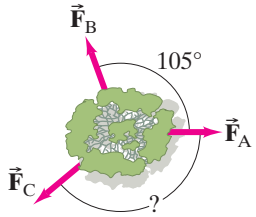


FIGURE 12–49 Problem 1.

2. (I) Calculate the mass m needed in order to suspend the leg shown in Fig. 12–50. Assume the leg (with cast) has a mass of 15.0 kg, and its CG is 35.0 cm from the hip joint; the cord holding the sling is 78.0 cm from the hip joint.

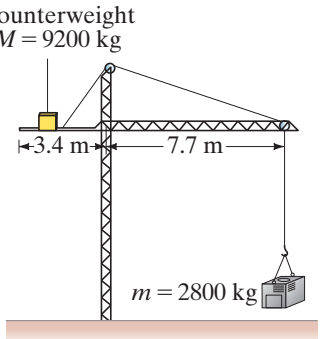


FIGURE 12–50 Problem 2.

3. (I) A tower crane (Fig. 12–51a) must always be carefully balanced so that there is no net torque tending to tip it. A particular crane at a building site is about to lift a 2800-kg air-conditioning unit. The crane's dimensions are shown in Fig. 12–51b. (*a*) Where must the crane's 9200-kg counterweight be placed when the load is lifted from the ground? (The counterweight is usually moved automatically via sensors and motors to precisely compensate for the load.) (*b*) Determine the maximum load that can be lifted with this counterweight when it is placed at its full extent. Ignore the mass of the beam.



(a)



(b)

FIGURE 12–51 Problem 3.

4. (I) What is the mass of the diver in Fig. 12–52 if she exerts a torque of 1800 m · N on the board, relative to the left (A) support post?

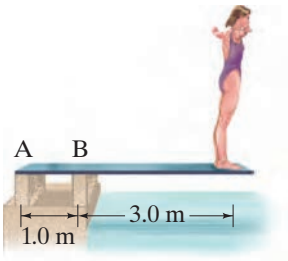


FIGURE 12–52 Problems 4 and 5.

5. (II) Calculate the forces F_A and F_B that the supports exert on the diving board of Fig. 12–52 when a 58-kg person stands at its tip. (*a*) Ignore the weight of the board. (*b*) Take into account the board's mass of 28 kg. Assume the board's CG is at its center.

6. (II) Figure 12–53 shows a pair of forceps used to hold a thin plastic rod firmly. If the thumb and finger each squeeze with a force $F_T = F_F = 11.0\text{ N}$, what force do the forceps jaws exert on the plastic rod?

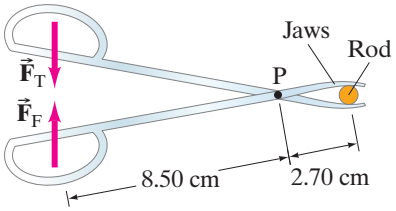


FIGURE 12–53 Problem 6.

7. (II) A 110-kg horizontal beam is supported at each end. A 320-kg piano rests a quarter of the way from one end. What is the vertical force on each of the supports?

8. (II) The two trees in Fig. 12–54 are 6.2 m apart. A backpacker is trying to lift his pack out of the reach of bears. Calculate the magnitude of the force $\vec{\mathbf{F}}$ that he must exert downward to hold a 19-kg backpack so that the rope sags at its midpoint by (*a*) 1.5 m, (*b*) 0.15 m.

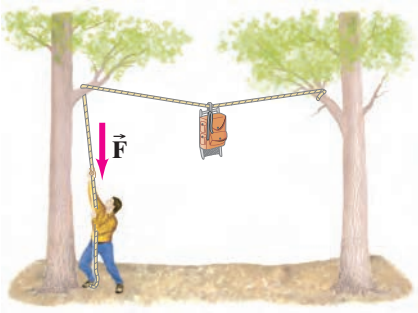


FIGURE 12–54 Problem 8.

9. (II) A 75-kg adult sits at one end of a 9.0-m-long board. His 25-kg child sits on the other end. (*a*) Where should the pivot be placed so that the board is balanced, ignoring the board's mass? (*b*) Find the pivot point if the board is uniform and has a mass of 18 kg.

10. (II) Find the tension in the two cords shown in Fig. 12–55. Neglect the mass of the cords, and assume that the angle θ is 33° and the mass m is 170 kg.

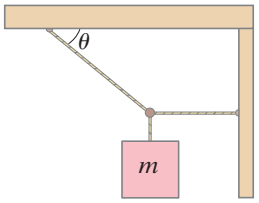


FIGURE 12–55 Problem 10.

11. (II) Find the tension in the two wires supporting the traffic light shown in Fig. 12–56.

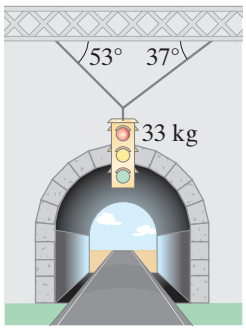


FIGURE 12–56
Problem 11.

12. (II) How close to the edge of the 27.0-kg table shown in Fig. 12–57 can a 66.0-kg person sit without tipping it over?

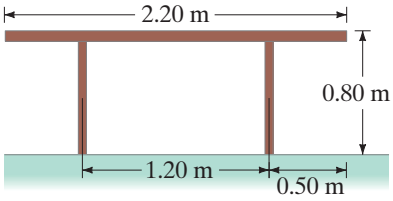


FIGURE 12–57 Problem 12.

13. (II) The force required to pull the cork out of the top of a wine bottle is in the range of 200 to 400 N. What range of forces F is required to open a wine bottle with the bottle opener shown in Fig. 12–58?

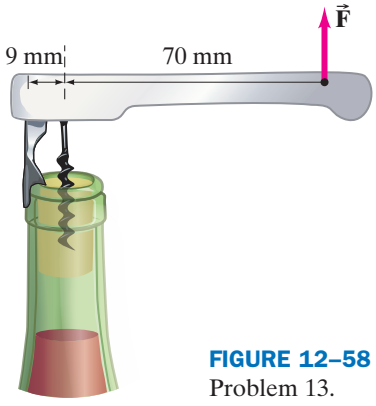


FIGURE 12–58
Problem 13.

14. (II) A shop sign weighing 215 N hangs from the end of a uniform 135-N beam as shown in Fig. 12–59. Find the tension in the supporting wire (at 35.0°), and the horizontal and vertical forces exerted by the hinge on the beam at the wall.

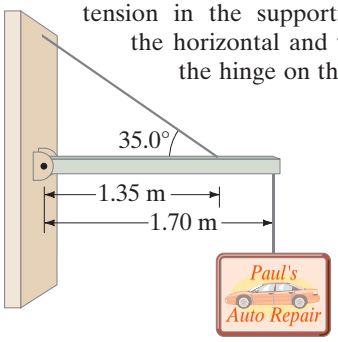


FIGURE 12–59
Problem 14.

15. (II) Three children are trying to balance on a seesaw, which includes a fulcrum rock acting as a pivot at the center, and a very light board 3.2 m long (Fig. 12–60). Two playmates are already on either end. Boy A has a mass of 45 kg, and boy B a mass of 35 kg. Where should girl C, whose mass is 25 kg, place herself so as to balance the seesaw?

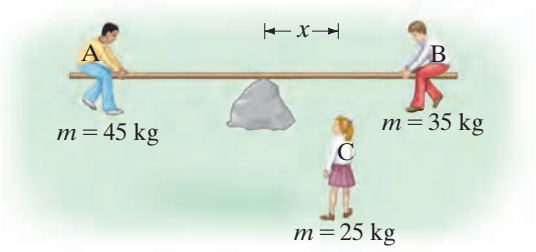


FIGURE 12–60 Problem 15.

16. (II) A uniform steel beam has a mass of 980 kg. On it is resting half of an identical beam, as shown in Fig. 12–61. What is the vertical support force at each end?

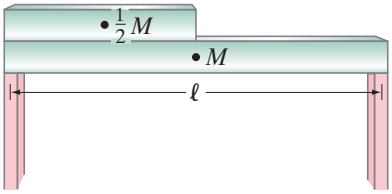


FIGURE 12–61 Problem 16.

17. (II) A traffic light hangs from a pole as shown in Fig. 12–62. The uniform aluminum pole AB is 7.20 m long and has a mass of 12.0 kg. The mass of the traffic light is 21.5 kg. Determine (*a*) the tension in the horizontal massless cable CD, and (*b*) the vertical and horizontal components of the force exerted by the pivot A on the aluminum pole.

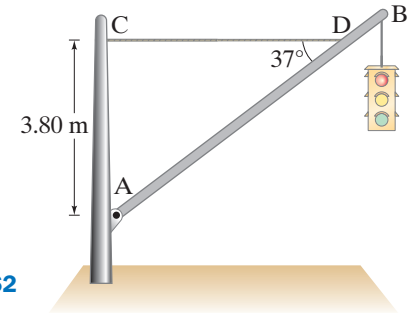


FIGURE 12–62
Problem 17.

18. (II) Figure 12–63 shows how a scale with a capacity of only 250 N can be used to weigh an object that is heavier than that. The board is 3.0 m long, has a mass of 3.4 kg, and is of uniform density. What is the weight of an object that is 1.2 m from the pivot end, if the scale reads 210 N? Assume that the beam remains nearly horizontal.

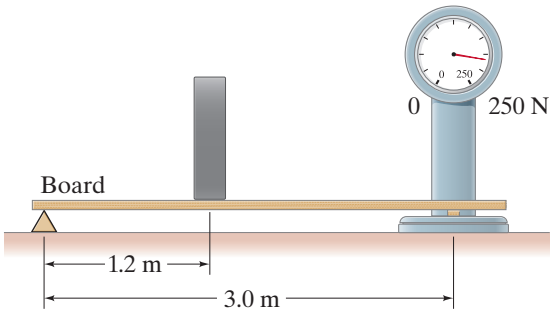


FIGURE 12–63 Problem 18.

19. (II) A 2300-kg trailer is attached to a stationary truck at point B, Fig. 12–64. Determine the normal force exerted by the road on the rear tires at A, and the vertical force exerted on the trailer by the support B.

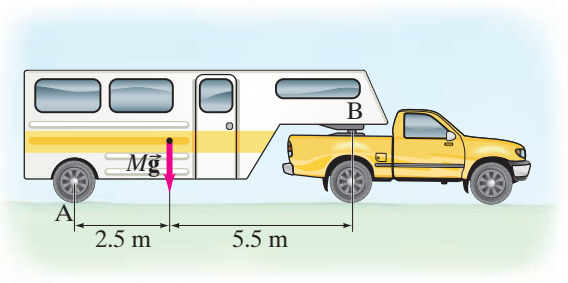


FIGURE 12–64 Problem 19.

20. (II) A 172-cm-tall person lies on a light (massless) board which is supported by two scales, one under the top of her head and one beneath the bottom of her feet (Fig. 12–65). The two scales read, respectively, 35.1 and 31.6 kg. What distance is the center of gravity of this person from the top of her head?



FIGURE 12–65 Problem 20.

21. (II) A man doing push-ups pauses in the position shown in Fig. 12–66. His mass $m = 68$ kg. Determine the normal force exerted by the floor (*a*) on each hand; (*b*) on each foot.

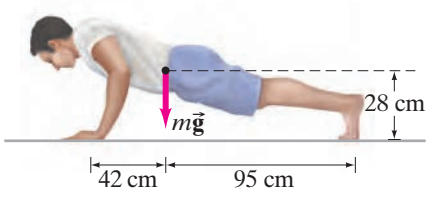


FIGURE 12–66 Problem 21.

22. (II) You are on a pirate ship and being forced to walk the plank (Fig. 12–67). You are standing at the point marked C. The plank is nailed onto the deck at point A, and rests on the support 0.75 m away from A. The center of mass of the uniform plank is located at point B. Your mass is 65 kg and the mass of the plank is 45 kg. What is the minimum downward force the nails must exert on the plank to hold it in place?

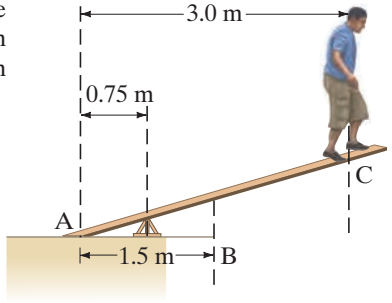


FIGURE 12–67 Problem 22.

23. (II) (*a*) A large 66.0-kg board is propped at a 45.0° angle against the edge of a barn door that is 2.6 m wide. How great a horizontal force must a person behind the door exert (at the edge) in order to open it? Assume that there is negligible friction between the door and the board but that the board is firmly set against the ground. (*b*) Repeat, assuming the coefficient of static friction between the board and the door is 0.35.

24. (II) A door 2.30 m high and 1.30 m wide has a mass of 13.0 kg. A hinge 0.40 m from the top and another hinge 0.40 m from the bottom each support half the door's weight (Fig. 12–68). Assume that the center of gravity is at the geometrical center of the door, and determine the horizontal and vertical force components exerted by each hinge on the door.

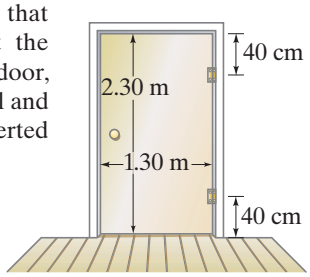


FIGURE 12–68 Problem 24.

25. (II) A uniform rod AB of length 4.5 m and mass $M = 3.8$ kg is hinged at A and held in equilibrium by a light cord, as shown in Fig. 12–69. A load $W = 22$ N hangs from the rod at a distance d so that the tension in the cord is 85 N. (*a*) Draw a free-body diagram for the rod. (*b*) Determine the vertical and horizontal forces on the rod exerted by the hinge. (*c*) Determine d from the appropriate torque equation.

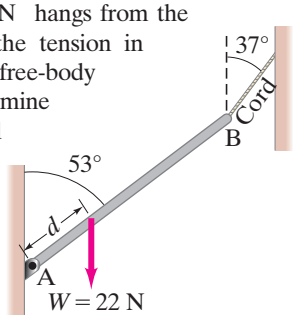


FIGURE 12–69 Problem 25.

26. (III) Two wires run from the top of a pole 2.6 m tall that supports a volleyball net. The two wires are anchored to the ground 2.0 m apart, and each is 2.0 m from the pole (Fig. 12–70). The tension in each wire is 125 N. What is the tension in the net, assumed horizontal and attached at the top of the pole?

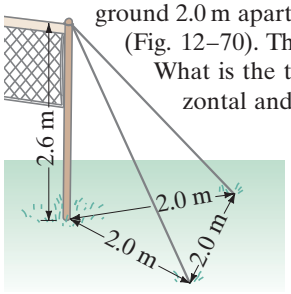


FIGURE 12–70 Problem 26.

27. (III) Consider a ladder with a painter climbing up it (Fig. 12–71). The mass of the uniform ladder is 12.0 kg, and the mass of the painter is 55.0 kg. If the ladder begins to slip at its base when the painter's feet are 70% of the way up the length of the ladder, what is the coefficient of static friction between the ladder and the floor? Assume the wall is frictionless.

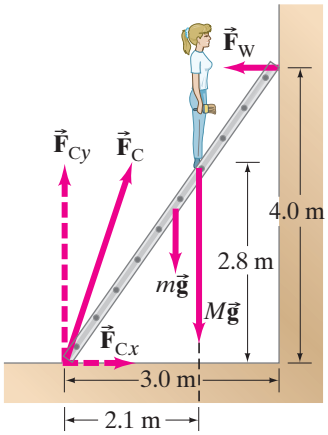


FIGURE 12–71
Problem 27.

28. (III) A uniform ladder of mass m and length ℓ leans at an angle θ against a frictionless wall, Fig. 12–72. If the coefficient of static friction between the ladder and the ground is μ_s, determine a formula for the minimum angle at which the ladder will not slip.

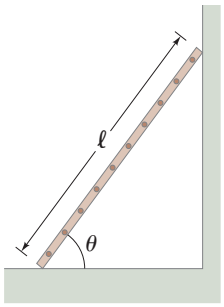


FIGURE 12–72
Problem 28.

29. (III) A refrigerator is approximately a uniform rectangular solid 1.9 m tall, 1.0 m wide, and 0.75 m deep. If it sits upright on a truck with its 1.0-m dimension in the direction of travel, and if the refrigerator cannot slide on the truck, how rapidly can the truck accelerate without tipping the refrigerator over? [*Hint*: The normal force would act at one corner.]

30. (III) A 56.0-kg person stands 2.0 m from the bottom of the stepladder shown in Fig. 12–73. Determine (*a*) the tension in the horizontal tie rod, which is halfway up the ladder, (*b*) the normal force the ground exerts on each side of the ladder, and (*c*) the force (magnitude and direction) that the left side of the ladder exerts on the right side at the hinge on the top. Ignore the mass of the ladder and assume the ground is frictionless. [*Hint*: Consider free-body diagrams for both sections of the ladder.]

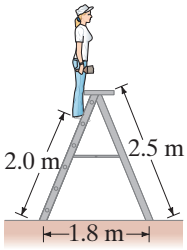


FIGURE 12–73
Problem 30.

*12–3 Muscles and Joints

***31.** (I) Approximately what magnitude force, F_M, must the extensor muscle in the upper arm exert on the lower arm to hold a 7.3-kg shot put (Fig. 12–74)? Assume the lower arm has a mass of 2.3 kg and its CG is 12.0 cm from the elbow-joint pivot.

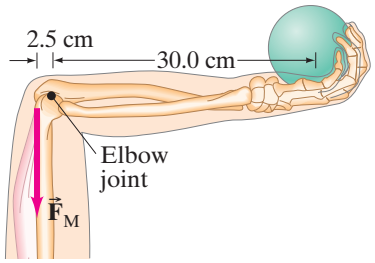


FIGURE 12–74
Problem 31.

***32.** (II) Redo Example 12–8 assuming now that the person is less bent over so that the 30° in Fig. 12–13b is instead 45°. What will be the magnitude of F_V on the vertebra?

***33.** (II) (*a*) Calculate the magnitude of the force, F_M, required of the "deltoid" muscle to hold up the outstretched arm shown in Fig. 12–75. The total mass of the arm is 3.6 kg. (*b*) Calculate the magnitude of the force F_J exerted by the shoulder joint on the upper arm and the angle (to the horizontal) at which it acts.

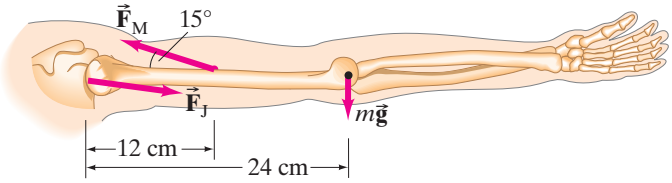


FIGURE 12–75 Problems 33 and 34.

***34.** (II) Suppose the hand in Problem 33 holds an 8.5-kg mass. What force, F_M, is required of the deltoid muscle, assuming the mass is 52 cm from the shoulder joint?

***35.** (II) The Achilles tendon is attached to the rear of the foot as shown in Fig. 12–76. When a person stands on one foot and lifts the heel to stand on the "ball of one foot," estimate the tension F_T in the Achilles tendon (pulling upward), and the (downward) force F_B exerted by the lower leg bone on the foot. Assume the person has a mass of 72 kg and D is twice as long as d.

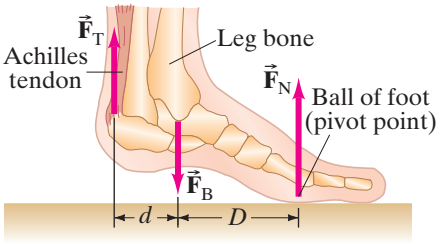


FIGURE 12–76 Problem 35.

12–4 Stability and Balance

36. (II) The Leaning Tower of Pisa is 55 m tall and about 7.7 m in radius. The top is 4.5 m off center. Is the tower in stable equilibrium? If so, how much farther can it lean before it becomes unstable? Assume the tower is of uniform composition.

37. (II) The center of gravity of a loaded truck depends on how the truck is packed. If it is 4.0 m high and 2.4 m wide, and its CG is 2.2 m above the ground, on how steep a slope can the truck be parked without tipping over (Fig. 12–77)?

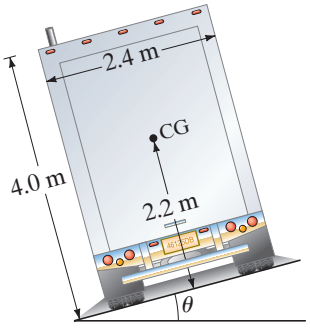


FIGURE 12–77 Problem 37.

38. (III) Four bricks are to be stacked at the edge of a table, each brick overhanging the one below it, so that the top brick extends as far as possible beyond the edge of the table. (*a*) To achieve this, show that successive bricks must extend no more than (starting at the top) $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{6}$, and $\frac{1}{8}$ of their length beyond the one below (Fig. 12–78a). (*b*) Is the top brick completely beyond the base? (*c*) Determine a general formula for the maximum total distance spanned by *n* bricks if they are to remain stable. (*d*) A builder wants to construct a *corbeled arch* (Fig. 12–78b) based on the principle of stability discussed in (*a*) and (*c*) above. What minimum number of bricks, each 0.30 m long and uniform, is needed if the arch is to span 1.0 m?

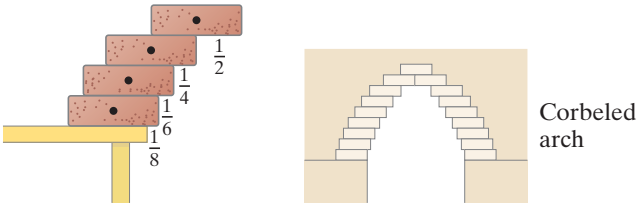


FIGURE 12–78 Problem 38.

12–5 Elasticity; Stress and Strain

39. (I) A nylon string on a tennis racket is under a tension of 315 N. If its diameter is 1.00 mm, by how much is it lengthened from its untensioned length of 30.0 cm?

40. (I) A marble column of cross-sectional area 1.4 m^2 supports a mass of 22,000 kg. (*a*) What is the stress within the column? (*b*) What is the strain? (*c*) By how much is the column shortened if it is 8.6 m high?

41. (I) A sign (mass 1700 kg) hangs from the bottom end of a vertical steel girder with a cross-sectional area of 0.012 m^2. (*a*) What is the stress within the girder? (*b*) What is the strain on the girder? (*c*) If the girder is 9.50 m long, how much is it lengthened? (Ignore the mass of the girder itself.)

42. (II) A 15-cm-long tendon was found to stretch 3.7 mm by a force of 13.4 N. The tendon was approximately round with an average diameter of 8.5 mm. Calculate Young's modulus of this tendon.

43. (II) How much pressure is needed to compress the volume of an iron block by 0.10%? Express your answer in N/m^2, and compare it to atmospheric pressure $(1.0 \times 10^5\ N/m^2)$.

44. (II) At depths of 2000 m in the sea, the pressure is about 200 times atmospheric pressure $(1\ atm = 1.0 \times 10^5\ N/m^2)$. By what percentage does the interior space of an iron bathysphere's volume change at this depth?

45. (II) A steel wire 2.6 mm in diameter stretches by 0.030% when a mass is suspended from it. How large is the mass?

12–6 Fracture

46. (I) The femur bone in the human leg has a minimum effective cross section of about 3.0 cm^2 $(= 3.0 \times 10^{-4}\ m^2)$. How much compressive force can it withstand before breaking?

47. (I) What diameter aluminum wire would have the same tensile breaking force as a 2.0-mm-diameter steel wire?

48. (II) (*a*) What is the maximum tension possible in a 1.00-mm-diameter nylon tennis racket string? (*b*) If you want tighter strings, what do you do to prevent breakage: use thinner or thicker strings? Why? What causes strings to break when they are hit by the ball?

49. (II) If a compressive force of 3.3×10^4 N is exerted on the end of a 22-cm-long bone of cross-sectional area 3.6 cm^2, (*a*) will the bone break, and (*b*) if not, by how much does it shorten?

50. (II) (*a*) What is the minimum cross-sectional area required of a vertical steel cable from which is suspended a 240-kg chandelier? Assume a safety factor of 7.0. (*b*) If the cable is 7.5 m long, how much does it elongate?

51. (II) Assume the supports of the uniform cantilever shown in Fig. 12–79 $(m = 2900\ kg)$ are made of wood. Calculate the minimum cross-sectional area required of each, assuming a safety factor of 9.0.

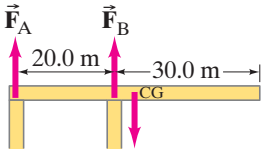


FIGURE 12–79
Problem 51.

52. (II) An iron bolt is used to connect two iron plates together. The bolt must withstand shear forces up to about 3600 N. Calculate the minimum diameter for the bolt, based on a safety factor of 7.0.

53. (III) A steel cable is to support an elevator whose total (loaded) mass is not to exceed 3100 kg. If the maximum acceleration of the elevator is 1.8 m/s^2, calculate the diameter of cable required. Assume a safety factor of 8.0.

*12–7 Trusses and Bridges

***54.** (II) Figure 12–80 shows a simple truss that carries a load at the center (C) of 1.65×10^4 N. (*a*) Calculate the force on each strut at the pins, A, B, C, and D. (*b*) Determine which struts (ignore their masses) are under tension and which are under compression.

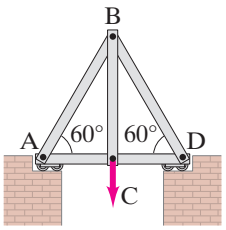


FIGURE 12–80
Problem 54.

*55. (II) A heavy load $Mg = 62.0\,\text{kN}$ hangs at point E of the single cantilever truss shown in Fig. 12–81. (*a*) Use a torque equation for the truss as a whole to determine the tension F_T in the support cable, and then determine the force $\vec{\mathbf{F}}_A$ on the truss at pin A. (*b*) Determine the force in each member of the truss. Neglect the weight of the trusses, which is small compared to the load.

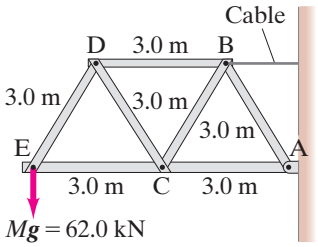


FIGURE 12–81 Problem 55.

*56. (III) (*a*) Consider again Example 12–12 but this time assume the roadway is supported also at the center C of Fig. 12–30b, as well as at the ends. That is, $\frac{1}{2}$ the mass M $(= 7.0 \times 10^5\,\text{kg})$ acts at the center and $\frac{1}{4}M$ at each end support (think of the bridge as two spans, AC and CE, so the center pin supports two span ends). Calculate the magnitude of the force in each truss member, and whether compression or tension. Compare to Example 12–12. (*b*) This time determine the force in each truss assuming $\frac{1}{2}M$ acts downward at the center of span AC, and $\frac{1}{2}M$ at the center of span CE. (*c*) Which of these approaches is the most valid? [*Hint*: The forces on the main span ACE may not be purely horizontal; see paragraph after Example 12–12].

*57. (III) The truss shown in Fig. 12–82 supports a railway bridge. Determine the compressive or tension force in each strut if a 53-ton $(1\,\text{ton} = 10^3\,\text{kg})$ train locomotive is stopped at the midpoint between the center and one end. Ignore the masses of the rails and truss, and use only $\frac{1}{2}$ the mass of train because there are two trusses (one on each side of the train). Assume all triangles are equilateral. [*Hint*: See Fig. 12–31.]

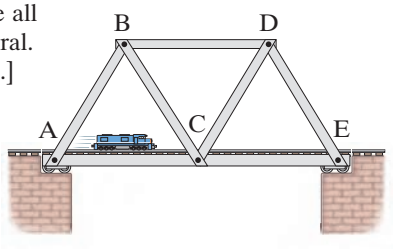


FIGURE 12–82 Problem 57.

*12–8 Arches and Domes

*58. (II) The subterranean tension ring that surrounds the dome in Fig. 12–39 exerts the balancing horizontal force on the abutments for the dome and is 36-sided, so each segment makes a 10° angle with the adjacent one (Fig. 12–83). Calculate the tension F that must exist in each segment so that the required force of $4.2 \times 10^5\,\text{N}$ can be exerted at each corner (Example 12–14).

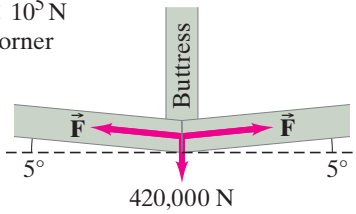


FIGURE 12–83 Problem 58.

*59. (II) How high must a pointed arch be if it is to span a space 8.0 m wide and exert one-third the horizontal force at its base that a round arch would?

General Problems

60. A woman holds a 1.8-m-long uniform 10.0-kg pole as shown in Fig. 12–84. (*a*) Determine the forces she must exert with each hand (magnitude and direction). To what position should she move her left hand so that neither hand has to exert a force greater than (*b*) 150 N? (*c*) 85 N?

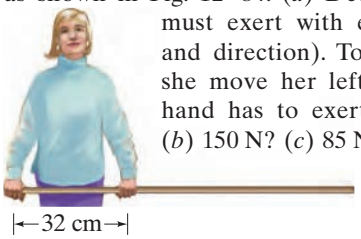


FIGURE 12–84 Problem 60.

61. A uniform cube of side ℓ rests on a rough floor. It is subjected to a steady horizontal pull F, exerted a distance h above the floor as shown in Fig. 12–85. As F is increased, the block will either begin to slide, or begin to tip over. Determine the coefficient of static friction μ_S so that: (*a*) the block begins to slide rather than tip; (*b*) the block begins to tip. [*Hint*: Where will the normal force on the block act if it tips?]

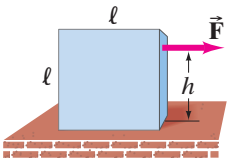


FIGURE 12–85 Problem 61.

62. A 50-story building is being planned. It is to be 180.0 m high with a base 46.0 m by 76.0 m. Its total mass will be about $1.8 \times 10^7\,\text{kg}$, and its weight therefore about $1.8 \times 10^8\,\text{N}$. Suppose a 200-km/h wind exerts a force of $950\,\text{N/m}^2$ over the 76.0-m-wide face (Fig. 12–86). Calculate the torque about the potential pivot point, the rear edge of the building (where $\vec{\mathbf{F}}_E$ acts in Fig. 12–86), and determine whether the building will topple. Assume the total force of the wind acts at the midpoint of the building's face, and that the building is not anchored in bedrock. [*Hint*: $\vec{\mathbf{F}}_E$ in Fig. 12–86 represents the force that the Earth would exert on the building in the case where the building would just begin to tip.]

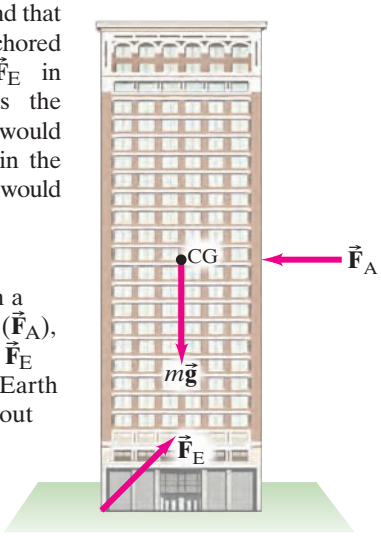


FIGURE 12–86 Forces on a building subjected to wind ($\vec{\mathbf{F}}_A$), gravity ($m\vec{\mathbf{g}}$), and the force $\vec{\mathbf{F}}_E$ on the building due to the Earth if the building were just about to tip. Problem 62.

63. When a wood shelf of mass 6.6 kg is fastened inside a slot in a vertical support as shown in Fig. 12–87, the support exerts a torque on the shelf. (*a*) Draw a free-body diagram for the shelf, assuming three vertical forces (two exerted by the support slot—explain why). Then calculate (*b*) the magnitudes of the three forces and (*c*) the torque exerted by the support (about the left end of the shelf).

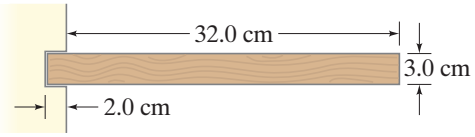


FIGURE 12–87 Problem 63.

64. A pole projects horizontally from the front wall of a shop. A 6.1-kg sign hangs from the pole at a point 2.2 m from the wall (Fig. 12–88). (*a*) What is the torque due to this sign calculated about the point where the pole meets the wall? (*b*) If the pole is not to fall off, there must be another torque exerted to balance it. What exerts this torque? Use a diagram to show how this torque must act. (*c*) Discuss whether compression, tension, and/or shear play a role in part (*b*).

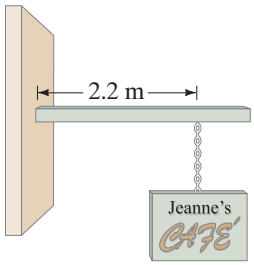


FIGURE 12–88
Problem 64.

65. A uniform meter stick with a mass of 180 g is supported horizontally by two vertical strings, one at the 0-cm mark and the other at the 90-cm mark (Fig. 12–89). What is the tension in the string (*a*) at 0 cm? (*b*) at 90 cm?

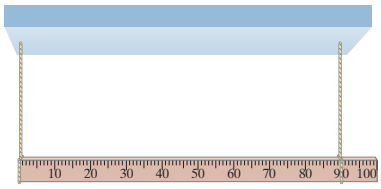


FIGURE 12–89 Problem 65.

66. There is a maximum height of a uniform vertical column made of any material that can support itself without buckling, and it is independent of the cross-sectional area (why?). Calculate this height for (*a*) ordinary steel (density 7.8×10^3 kg/m^3), and (*b*) granite (density 2.7×10^3 kg/m^3).

67. When a mass of 25 kg is hung from the middle of a fixed straight aluminum wire, the wire sags to make an angle of 12° with the horizontal as shown in Fig. 12–90. Determine the radius of the wire.

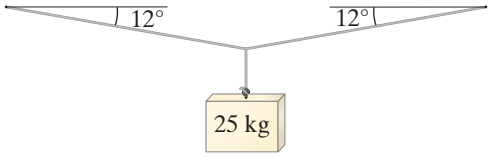


FIGURE 12–90 Problem 67.

68. The *mobile* in Fig. 12–91 is in equilibrium. Object B has mass of 0.748 kg. Determine the masses of objects A, C, and D. (Neglect the weights of the crossbars.)

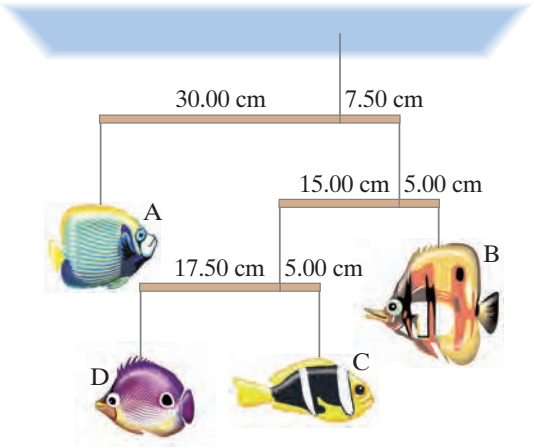


FIGURE 12–91 Problem 68.

69. A 65.0-kg painter is on a uniform 25-kg scaffold supported from above by ropes (Fig. 12–92). There is a 4.0-kg pail of paint to one side, as shown. Can the painter walk safely to both ends of the scaffold? If not, which end(s) is dangerous, and how close to the end can he approach safely?

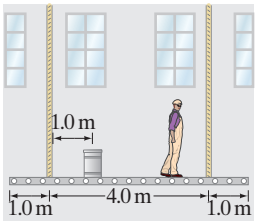


FIGURE 12–92
Problem 69.

70. The roof over a 9.0-m × 10.0-m room in a school has a total mass of 12,400 kg. The roof is to be supported by vertical wooden "2 × 4s" (2 × 4 in inches, but actually about 4.0 cm × 9.0 cm) equally spaced along the 10.0-m sides. How many supports are required on each side, and how far apart must they be? Consider only compression, and assume a safety factor of 12.

71. A 25-kg object is being lifted by two people pulling on the ends of a 1.15-mm-diameter nylon cord that goes over two 3.00-m-high poles that are 4.5 m apart, as shown in Fig. 12–93. How high above the floor will the object be when the cord breaks?

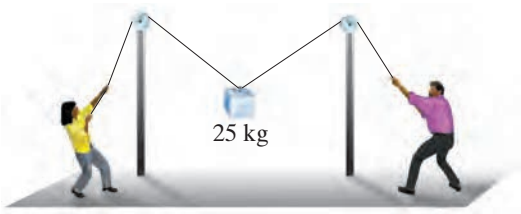


FIGURE 12–93 Problem 71.

72. A cubic crate of side $s = 2.00$ m is top-heavy: its CG is 18 cm above its true center. How steep an incline can the crate rest on without tipping over? [*Hint*: The normal force would act at the lowest corner.]

73. A uniform 95-kg flagpole of length 8.4 m is being erected by pulling on a rope attached 2/3 of the way to the top (Fig. 12–94). When the pole is inclined at 35° and the rope makes an angle with the ground of 18°, what is the tension in the rope?

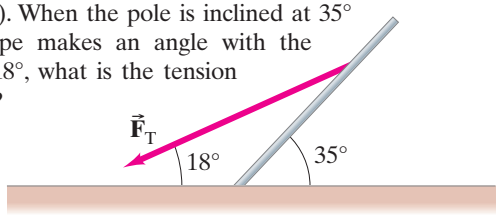


FIGURE 12–94
Problem 73.

74. Two identical, uniform beams are symmetrically set up against each other (Fig. 12–95) on a floor with which they have a coefficient of friction $\mu_s = 0.45$. What is the minimum angle the beams can make with the floor and still not fall?

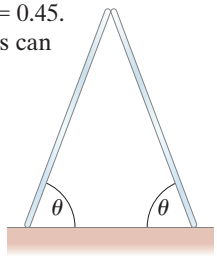


FIGURE 12–95
Problem 74.

75. A steel rod of radius $R = 15$ cm and length ℓ_0 stands upright on a firm surface. A 78-kg man climbs atop the rod. (*a*) Determine the percent decrease in the rod's length. (*b*) When a metal is compressed, each atom throughout its bulk moves closer to its neighboring atom by exactly the same fractional amount. If iron atoms in steel are normally 2.0×10^{-10} m apart, by what distance did this interatomic spacing have to change in order to produce the normal force required to support the man? [*Note*: Neighboring atoms repel each other, and this repulsion accounts for the observed normal force.]

76. A home mechanic wants to raise the 310-kg engine out of a car. The plan is to stretch a rope vertically from the engine to a branch of a tree 6.0 m above, and back to the bumper (Fig. 12–96). When the mechanic climbs up a stepladder and pulls horizontally on the rope at its midpoint, the engine rises out of the car. (*a*) How much force must the mechanic exert to hold the engine 0.50 m above its normal position? (*b*) What is the system's mechanical advantage?

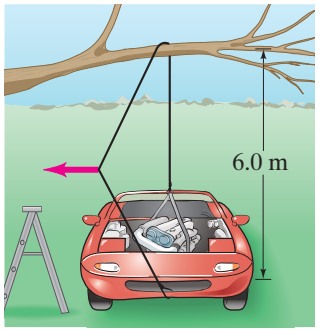


FIGURE 12–96
Problem 76.

77. A 2.0-m-high box with a 1.0-m-square base is moved across a rough floor as in Fig. 12–97. The uniform box weighs 250 N and has a coefficient of static friction with the floor of 0.60. What minimum force must be exerted on the box to make it slide? What is the maximum height h above the floor that this force can be applied without tipping the box over? Note that if the box tips, the normal force and the friction force will act at the lowest corner.

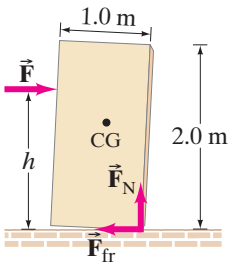


FIGURE 12–97
Problem 77.

78. Two springs, both having stiffness constant 225 N/m, are attached to a table and to a 0.500-kg uniform thin wooden board (Fig. 12–98). The board is exactly horizontal. What are the natural lengths of each spring? [*Hint*: One of the springs is stretched, the other compressed, from their natural equilibrium lengths.]

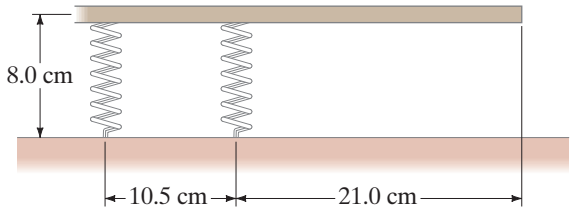


FIGURE 12–98 Problem 78.

79. A tightly stretched horizontal "high wire" is 36 m long. It sags vertically 0.90 m when a 56.0-kg *tightrope walker* stands at its center. What is the tension in the wire? Is it possible to increase the tension in the wire so that there is no sag?

80. *Parachutists* whose chutes have failed to open have been known to survive if they land in deep snow. Assume that a 75-kg parachutist hits the snow with an area of impact of 0.30 m^2 at a velocity of 55 m/s, and that the ultimate strength of body tissue is 5×10^5 N/m^2. Assume that the person is brought to rest in 1.0 m of snow. Show that the person may escape serious injury.

81. A uniform flexible steel cable of weight mg is suspended between two points at the same elevation as shown in Fig. 12–99, where $\theta = 62°$. Determine the tension in the cable (*a*) at its lowest point, and (*b*) at the points of attachment. (*c*) What is the direction of the tension force in each case?

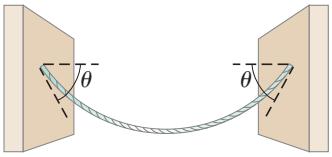


FIGURE 12–99 Problem 81.

82. One rod of the square frame shown in Fig. 12–100 contains a turnbuckle which, when turned, can put the rod under tension or compression. If the turnbuckle puts rod AB under a compressive force F, determine the forces produced in the other rods. Ignore the mass of the rods and assume the diagonal rods cross each other freely at the center without friction. [*Hint*: Use the symmetry of the situation.]

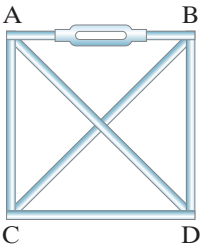


FIGURE 12–100
Problem 82.

83. A uniform beam of mass M and length ℓ is mounted on a hinge at a wall as shown in Fig. 12–101. It is held in a horizontal position by a wire making an angle θ as shown. A mass m is placed on the beam a distance x from the wall, and this distance can be varied. Determine, as a function of x, (*a*) the tension in the wire and (*b*) the components of the force exerted by the beam on the hinge.

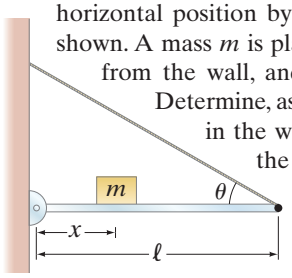


FIGURE 12–101
Problem 83.

84. If 25 kg is the maximum mass m that a person can hold in a hand when the arm is positioned with a 105° angle at the elbow as shown in Fig. 12–102, what is the maximum force F_{max} that the biceps muscle exerts on the forearm? Assume the forearm and hand have a total mass of 2.0 kg with a CG that is 15 cm from the elbow, and that the biceps muscle attaches 5.0 cm from the elbow.

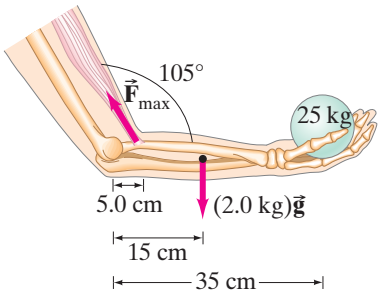


FIGURE 12–102 Problem 84.

85. A uniform 6.0-m-long ladder of mass 16.0 kg leans against a smooth wall (so the force exerted by the wall, $\vec{\mathbf{F}}_W$, is perpendicular to the wall). The ladder makes an angle of 20.0° with the vertical wall, and the ground is rough. Determine the coefficient of static friction at the base of the ladder if the ladder is not to slip when a 76.0-kg person stands three-fourths of the way up the ladder.

86. In Fig. 12–103, consider the right-hand (northernmost) section of the Golden Gate Bridge, which has a length $d_1 = 343$ m. Assume the CG of this span is halfway between the tower and anchor. Determine F_{T1} and F_{T2} (which act on the northernmost cable) in terms of mg, the weight of the northernmost span, and calculate the tower height h needed for equilibrium. Assume the roadway is supported only by the suspension cables, and neglect the mass of the cables and vertical wires. [*Hint*: F_{T3} does not act on this section.]

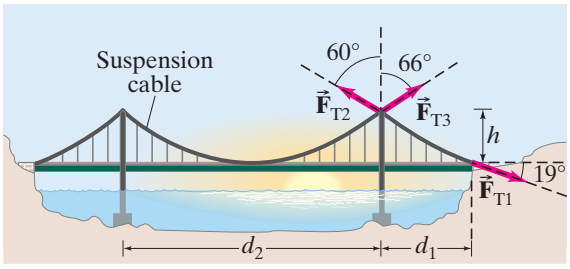


FIGURE 12–103 Problems 86 and 87.

87. Assume that a single-span suspension bridge such as the Golden Gate Bridge has the symmetrical configuration indicated in Fig. 12–103. Assume that the roadway is uniform over the length of the bridge and that each segment of the suspension cable provides the sole support for the roadway directly below it. The ends of the cable are anchored to the ground only, not to the roadway. What must the ratio of d_2 to d_1 be so that the suspension cable exerts no net horizontal force on the towers? Neglect the mass of the cables and the fact that the roadway isn't precisely horizontal.

88. A uniform sphere of weight mg and radius r_0 is tethered to a wall by a rope of length ℓ. The rope is tied to the wall a distance h above the contact point of the sphere, as shown in Fig. 12–104. The rope makes an angle θ with respect to the wall and is not in line with the ball's center. The coefficient of static friction between the wall and sphere is μ. (*a*) Determine the value of the frictional force on the sphere due to the wall. [*Hint*: A wise choice of axis will make this calculation easy.] (*b*) Suppose the sphere is just on the verge of slipping. Derive an expression for μ in terms of h and θ.

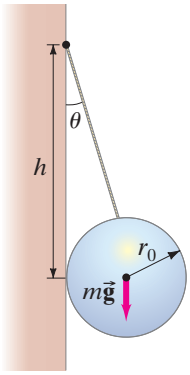


FIGURE 12–104
Problem 88.

89. A uniform ladder of mass m and length ℓ leans at an angle θ against a wall, Fig. 12–105. The coefficients of static friction between ladder–ground and ladder–wall are μ_G and μ_W, respectively. The ladder will be on the verge of slipping when both the static friction forces due to the ground and due to the wall take on their maximum values. (*a*) Show that the ladder will be stable if $\theta \geq \theta_{min}$, where the minimum angle θ_{min} is given by

$$\tan\theta_{min} = \frac{1}{2\mu_G}(1 - \mu_G\mu_W).$$

(*b*) "Leaning ladder problems" are often analyzed under the seemingly unrealistic assumption that the wall is frictionless (see Example 12–6). You wish to investigate the magnitude of error introduced by modeling the wall as frictionless, if in reality it is frictional. Using the relation found in part (*a*), calculate the true value of θ_{min} for a frictional wall, taking $\mu_G = \mu_W = 0.40$. Then, determine the approximate value of θ_{min} for the "frictionless wall" model by taking $\mu_G = 0.40$ and $\mu_W = 0$. Finally, determine the percent deviation of the approximate value of θ_{min} from its true value.

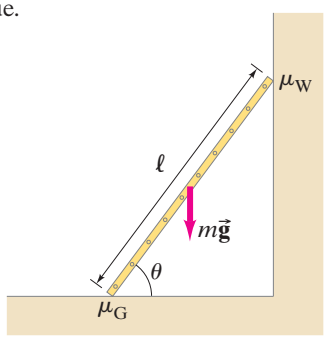


FIGURE 12–105 Problem 89.

90. A 28-kg uniform round table is supported by three vertical legs equal distances apart on the edge. What minimum mass, placed on the table's edge, will cause the table to overturn?

91. The CM of a 95,000-kg train locomotive starts across a 260-m-long bridge at time $t = 0$. The bridge is a uniform beam of mass 23,000 kg and the train travels at a constant 80.0 km/h. What are the magnitudes of the vertical forces, $F_A(t)$ and $F_B(t)$, on the two end supports, written as a function of time during the train's passage?

92. A 23-kg sphere rests between two smooth planes as shown in Fig. 12–106. Determine the magnitude of the force acting on the sphere exerted by each plane.

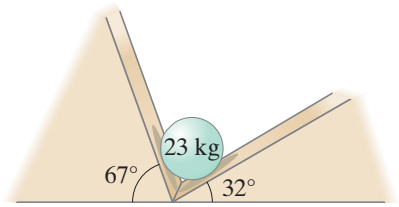


FIGURE 12–106 Problem 92.

93. A uniform 55-kg bar is 2.0 m long and is attached to a wall by a frictionless hinge. The center of mass of the bar is connected to the wall, Fig. 12–107, by a horizontal steel wire (cross-sectional area $0.0050\,\text{m}^2$). When a force $\vec{\mathbf{F}}$ is applied to the end of the bar, and perpendicular to it, the wire stretches 0.10 mm. (*a*) What is the magnitude of $\vec{\mathbf{F}}$? (*b*) Determine the x and y components of the force $\vec{\mathbf{F}}_\text{H}$ exerted on the bar at the hinge.

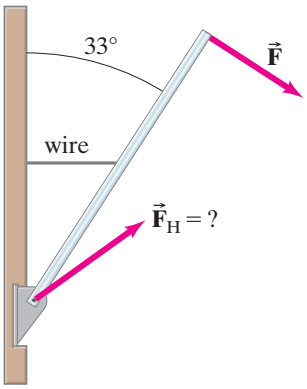


FIGURE 12–107 Problem 93.

94. The forces acting on an 85,000-kg aircraft flying at constant velocity are shown in Fig. 12–108. The engine thrust, $F_\text{T} = 5.0 \times 10^5\,\text{N}$, acts on a line 1.6 m below the CM. Determine the drag force F_D and the distance above the CM that it acts. Assume $\vec{\mathbf{F}}_\text{D}$ and $\vec{\mathbf{F}}_\text{T}$ are horizontal. ($\vec{\mathbf{F}}_\text{L}$ is the "lift" force on the wing.)

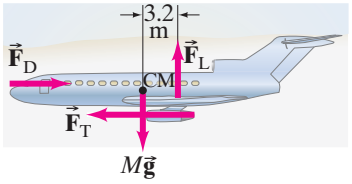


FIGURE 12–108
Problem 94.

95. Suppose a 65-kg person jumps from a height of 3.0 m down to the ground. (*a*) What is the speed of the person just before landing (Chapter 2)? (*b*) Estimate the average force on the person's feet exerted by the ground to bring the person to rest, if the knees are bent so the person's CG moves a distance $d = 50\,\text{cm}$ during the deceleration period (Fig. 12–109). [*Hint*: This force exerted by the ground $\neq$ net force. You may want to consult Chapters 2, 4, and 7, and be sure to draw a careful free-body diagram of the person.] (*c*) Estimate the decelerating force if the person lands stiff-legged so $d \approx 1.0\,\text{cm}$. (*d*) Estimate the stress in the tibia (a lower leg bone of area $= 3.0 \times 10^{-4}\,\text{m}^2$), and determine whether or not the bone will *fracture* if the landing is made with bent legs ($d = 50\,\text{cm}$). (*e*) Estimate the stress and determine if the tibia will break in a stiff-legged landing ($d = 1.0\,\text{cm}$).

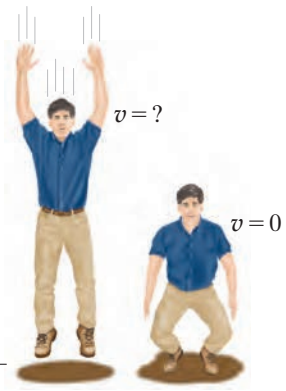


FIGURE 12–109
Problem 95.

ANSWERS TO EXERCISES

A: F_A also has a component to balance the sideways force F_B. So the resultant force will be greater than $F_{\text{A}y} = mg$.

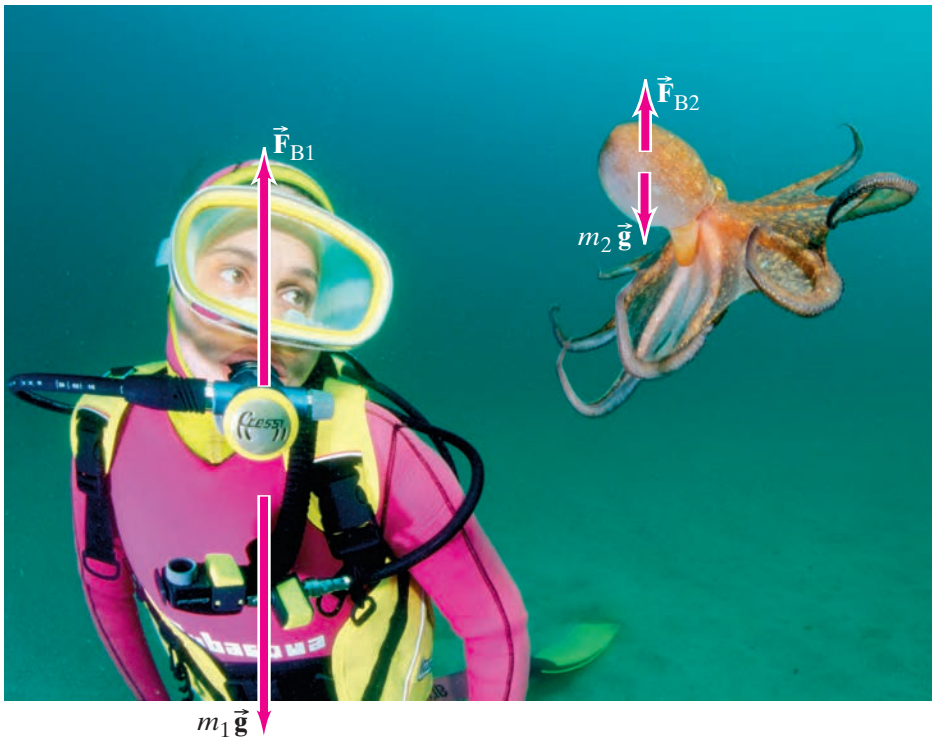
B: $F_\text{N} = m_\text{A}\,g + m_\text{B}\,g + Mg$
$= (30\,\text{kg} + 25\,\text{kg} + 4.0\,\text{kg})\,g = 580\,\text{N}$.

C: (*a*).

D: Static friction at the cement floor ($= F_{Cx}$) is crucial, or else the ladder would slip. At the top, the ladder can move and adjust, so we wouldn't need or expect a strong static friction force there.

E: 7.0 kg.

F: (*b*).



Underwater divers and sea creatures experience a buoyant force ($\vec{\mathbf{F}}_B$) that closely balances their weight $m\vec{\mathbf{g}}$. The buoyant force is equal to the weight of the volume of fluid displaced (Archimedes' principle) and arises because the pressure increases with depth in the fluid. Sea creatures have a density very close to that of water, so their weight very nearly equals the buoyant force. Humans have a density slightly less than water, so they can float if not weighed down.

When fluids flow, interesting effects occur because the pressure in the fluid is lower where the fluid velocity is higher (Bernoulli's principle).

CHAPTER 13

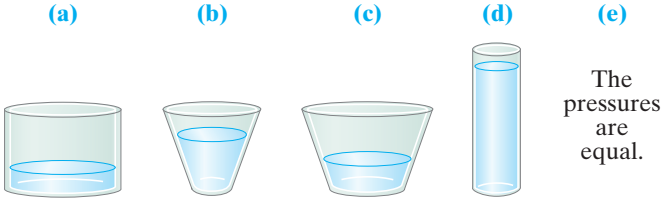
Fluids

CONTENTS

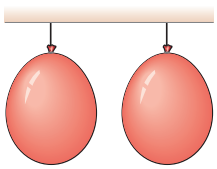
13–1 Phases of Matter
13–2 Density and Specific Gravity
13–3 Pressure in Fluids
13–4 Atmospheric Pressure and Gauge Pressure
13–5 Pascal's Principle
13–6 Measurement of Pressure; Gauges and the Barometer
13–7 Buoyancy and Archimedes' Principle
13–8 Fluids in Motion; Flow Rate and the Equation of Continuity
13–9 Bernoulli's Equation
13–10 Applications of Bernoulli's Principle: Torricelli, Airplanes, Baseballs, Blood Flow
13–11 Viscosity
*13–12 Flow in Tubes: Poiseuille's Equation, Blood Flow
*13–13 Surface Tension and Capillarity
13–14 Pumps, and the Heart

CHAPTER-OPENING QUESTIONS—Guess now!

1. Which container below has the largest pressure at the bottom? Assume each container holds the same volume of water.



2. Two balloons are tied and hang with their nearest edges about 3 cm apart. If you blow between the balloons (not *at* the balloons, but at the opening between them), what will happen?



(a) Nothing.
(b) The balloons will move closer together.
(c) The balloons will move farther apart.

In previous Chapters we considered objects that were solid and assumed they maintain their shape except for a small amount of elastic deformation. We sometimes treated objects as point particles. Now we are going to shift our attention to materials that are very deformable and can flow. Such "fluids" include liquids and gases. We will examine fluids both at rest (fluid statics) and in motion (fluid dynamics).

13–1 Phases of Matter

The three common **phases**, or **states**, of matter are solid, liquid, and gas. A simple way to distinguish these three phases is as follows. A **solid** maintains a generally fixed size and shape; usually it requires a large force to change the volume or shape of a solid[†] (although a thin object might bend). A **liquid** does not maintain a fixed shape—it takes on the shape of its container, and it can flow; but like a solid it is not readily compressible, and its volume can be changed significantly only by a very large force. A **gas** has neither a fixed shape nor a fixed volume—it will expand to fill its container. For example, when air is pumped into an automobile tire, the air does not all run to the bottom of the tire as a liquid would; it spreads out to fill the whole volume of the tire.

Because liquids and gases do not maintain a fixed shape, they both have the ability to flow. They are thus referred to collectively as **fluids**.

The division of matter into three phases is not always simple. How, for example, should butter be classified? Furthermore, a fourth phase of matter can be distinguished, the **plasma** phase, which occurs only at very high temperatures and consists of ionized atoms (electrons separated from the nuclei). Some scientists believe that *colloids* (suspensions of tiny particles in a liquid) should also be considered a separate phase of matter. **Liquid crystals**, used in TV, cell phone, and computer screens, can be considered a phase of matter in between solids and liquids. For now, we will be interested in the three ordinary phases of matter.

13–2 Density and Specific Gravity

It is sometimes said that iron is "heavier" than wood. This cannot really be true since a large log clearly weighs more than an iron nail. What we should say is that iron is more *dense* than wood.

The **density**, ρ, of a substance (ρ is the lowercase Greek letter rho) is defined as its mass per unit volume:

$$\rho = \frac{m}{V}, \tag{13–1}$$

where m is the mass of a sample of the substance and V its volume. Density is a characteristic property of any pure substance. Objects made of a particular pure substance, such as pure gold, can have any size or mass, but the density will be the same for each.

We can use density, Eq. 13–1, to write the mass of an object as

$$m = \rho V,$$

and the weight of an object is thus

$$mg = \rho V g.$$

The SI unit for density is kg/m^3. Sometimes densities are given in g/cm^3. Note that a density given in g/cm^3 must be multiplied by 1000 to give the result in kg/m^3 $\left[1\,\text{kg/m}^3 = 1000\,\text{g}/(100\,\text{cm})^3 = 10^3\,\text{g}/10^6\,\text{cm}^3 = 10^{-3}\,\text{g/cm}^3\right]$. For example, the density of aluminum is $\rho = 2.70\,\text{g/cm}^3$, which equals $2700\,\text{kg/m}^3$. The densities of various substances are given in Table 13–1. The Table specifies temperature and atmospheric pressure because they affect density (the effect is slight for liquids and solids). Note that air is about 1000 times less dense than water.

TABLE 13–1
Densities of Substances[‡]

Substance	Density, ρ (kg/m^3)
Solids	
Aluminum	2.70×10^3
Iron and steel	7.8×10^3
Copper	8.9×10^3
Lead	11.3×10^3
Gold	19.3×10^3
Concrete	2.3×10^3
Granite	2.7×10^3
Wood (typical)	$0.3–0.9 \times 10^3$
Glass, common	$2.4–2.8 \times 10^3$
Ice (H_2O)	0.917×10^3
Bone	$1.7–2.0 \times 10^3$
Liquids	
Water (4°C)	1.000×10^3
Sea water	1.025×10^3
Blood, plasma	1.03×10^3
Blood, whole	1.05×10^3
Mercury	13.6×10^3
Alcohol, ethyl	0.79×10^3
Gasoline	$0.7–0.8 \times 10^3$
Gases	
Air	1.29
Helium	0.179
Carbon dioxide	1.98
Water (steam) (100°C)	0.598

[‡]Densities are given at 0°C and 1 atm pressure unless otherwise specified.

EXAMPLE 13–1 **Mass, given volume and density.** What is the mass of a solid iron wrecking ball of radius 18 cm?

APPROACH First we use the standard formula $V = \frac{4}{3}\pi r^3$ (see inside rear cover) to obtain the sphere's volume. Then Eq. 13–1 and Table 13–1 give us the mass m.

SOLUTION The volume of the sphere is

$$V = \tfrac{4}{3}\pi r^3 = \tfrac{4}{3}(3.14)(0.18\,\text{m})^3 = 0.024\,\text{m}^3.$$

From Table 13–1, the density of iron is $\rho = 7800\,\text{kg/m}^3$, so Eq. 13–1 gives

$$m = \rho V = (7800\,\text{kg/m}^3)(0.024\,\text{m}^3) = 190\,\text{kg}.$$

[†]Section 12–5.

The **specific gravity** of a substance is defined as the ratio of the density of that substance to the density of water at 4.0°C. Because specific gravity (abbreviated SG) is a ratio, it is a simple number without dimensions or units. For example (see Table 13–1), the specific gravity of lead is 11.3 $\left[(11.3 \times 10^3\,\text{kg/m}^3)/(1.00 \times 10^3\,\text{kg/m}^3)\right]$. The SG of alcohol is 0.79.

The concepts of density and specific gravity are especially helpful in the study of fluids because we are not always dealing with a fixed volume or mass.

13–3 Pressure in Fluids

Pressure and force are related, but they are not the same thing. **Pressure** is defined as force per unit area, where the force F is understood to be the magnitude of the force acting perpendicular to the surface area A:

$$\text{pressure} = P = \frac{F}{A}. \qquad \textbf{(13–2)}$$

Although force is a vector, pressure is a scalar. Pressure has magnitude only. The SI unit of pressure is N/m^2. This unit has the official name **pascal** (Pa), in honor of Blaise Pascal (see Section 13–5); that is, $1\ \text{Pa} = 1\ \text{N/m}^2$. However, for simplicity, we will often use N/m^2. Other units sometimes used are dynes/cm^2 (cgs),[†] and in the British system lb/in^2 (pounds per square inch, abbreviated "psi"). Several other units for pressure are discussed in Sections 13–4 and 13–6, along with conversions between them (see also the Table inside the front cover).

CAUTION
Pressure is a scalar, not a vector

EXAMPLE 13–2 Calculating pressure. A 60-kg person's two feet cover an area of 500 cm^2. (*a*) Determine the pressure exerted by the two feet on the ground. (*b*) If the person stands on one foot, what will be the pressure under that foot?

APPROACH Assume the person is at rest. Then the ground pushes up on her with a force equal to her weight mg, and she exerts a force mg on the ground where her feet (or foot) contact it. Because $1\ \text{cm}^2 = (10^{-2}\,\text{m})^2 = 10^{-4}\,\text{m}^2$, then $500\ \text{cm}^2 = 0.050\ \text{m}^2$.

SOLUTION (*a*) The pressure on the ground exerted by the two feet is

$$P = \frac{F}{A} = \frac{mg}{A} = \frac{(60\,\text{kg})(9.8\,\text{m/s}^2)}{(0.050\,\text{m}^2)} = 12 \times 10^3\,\text{N/m}^2.$$

(*b*) If the person stands on one foot, the force is still equal to the person's weight, but the area will be half as much, so the pressure (F/A) will be twice as much: $24 \times 10^3\,\text{N/m}^2$.

Pressure is particularly useful for dealing with fluids. It is an experimental observation that *a fluid exerts pressure in every direction*. This is well known to swimmers and divers who feel the water pressure on all parts of their bodies. At any depth in a fluid at rest, the pressure is the same in all directions at that given depth. To see why, consider a tiny cube of the fluid (Fig. 13–1) which is so small that we can consider it a point and can ignore the force of gravity on it. The pressure on one side of it must equal the pressure on the opposite side. If this weren't true, there would be a net force on the cube and it would start moving. If the fluid is not flowing, then the pressures must be equal.

For a fluid at rest, the force due to fluid pressure always acts *perpendicular* to any solid surface it touches. If there were a component of the force parallel to the surface, as shown in Fig. 13–2, then according to Newton's third law the solid surface would exert a force back on the fluid, which would cause the fluid to flow—in contradiction to our assumption that the fluid is at rest. Thus the force due to the pressure in a fluid at rest is always perpendicular to the surface in contact with the fluid.

[†]Sections 1–4 and 4–4.

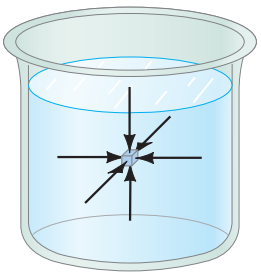
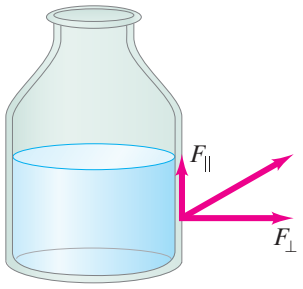


FIGURE 13–1 Pressure is the same in every direction in a nonmoving fluid at a given depth. If this weren't true, the fluid would be moving (or flowing).

FIGURE 13–2 If there were a component of force parallel to the solid surface of the container, the liquid would move in response to it. For a liquid at rest, $F_{\parallel} = 0$.



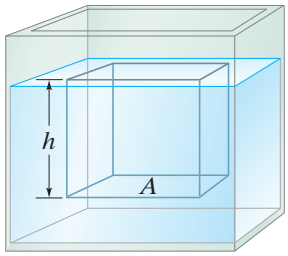


FIGURE 13–3 Calculating the pressure at a depth h in a liquid, due to the weight of the liquid above.

We now calculate quantitatively how the pressure in a liquid of uniform density varies with *depth*. Let us look at a depth h below the surface of the liquid, as shown in Fig. 13–3 (that is, the liquid's top surface is a height h above this level). The pressure due to the liquid at this depth h is due to the weight of the column of liquid above it. Thus the force due to the weight of liquid acting on the area A is $F = mg = (\rho V)g = \rho Ahg$, where Ah is the volume of the liquid above area A, ρ is the density of the liquid (assumed to be constant), and g is the acceleration of gravity. The pressure P due to the weight of liquid is then

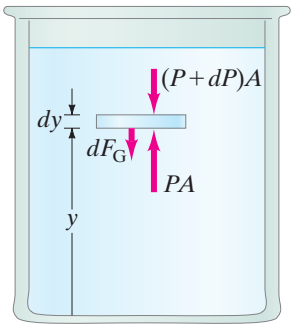
$$P = \frac{F}{A} = \frac{\rho Ahg}{A}$$

$$P = \rho gh. \qquad \text{[liquid]} \quad \textbf{(13–3)}$$

Note that the area A doesn't affect the pressure at a given depth. The fluid pressure is directly proportional to the density of the liquid and to the depth within the liquid. In general, *the pressure at equal depths within a uniform liquid is the same*.

EXERCISE A Return to Chapter-Opening Question 1, page 365, and answer it again now. Try to explain why you may have answered differently the first time.

FIGURE 13–4 Forces on a flat, slablike volume of fluid for determining the pressure P as a function of height y in the fluid.



Equation 13–3 tells us what the pressure is at a depth h in the liquid, due to the liquid itself. But what if there is additional pressure exerted at the surface of the liquid, such as the pressure of the atmosphere or a piston pushing down? And what if the density of the fluid is not constant? Gases are quite compressible and hence their density can vary significantly with depth. Liquids, too, can be compressed, although we can often ignore the variation in density. (One exception is in the depths of the ocean where the great weight of water above significantly compresses the water and increases its density.) To cover these, and other cases, we now treat the general case of determining how the pressure in a fluid varies with depth.

As shown in Fig. 13–4, let us determine the pressure at any height y above some reference point[†] (such as the ocean floor or the bottom of a tank or swimming pool). Within this fluid, at the height y, we consider a tiny, flat, slablike volume of the fluid whose area is A and whose (infinitesimal) thickness is dy, as shown. Let the pressure acting upward on its lower surface (at height y) be P. The pressure acting downward on the top surface of our tiny slab (at height $y + dy$) is designated $P + dP$. The fluid pressure acting on our slab thus exerts a force equal to PA upward on our slab and a force equal to $(P + dP)A$ downward on it. The only other force acting vertically on the slab is the (infinitesimal) force of gravity dF_G, which on our slab of mass dm is

$$dF_G = (dm)g = \rho g\, dV = \rho gA\, dy,$$

where ρ is the density of the fluid at the height y. Since the fluid is assumed to be at rest, our slab is in equilibrium so the net force on it must be zero. Therefore we have

$$PA - (P + dP)A - \rho gA\, dy = 0,$$

which when simplified becomes

$$\frac{dP}{dy} = -\rho g. \qquad \textbf{(13–4)}$$

This relation tells us how the pressure within the fluid varies with height above any reference point. The minus sign indicates that the pressure decreases with an increase in height; or that the pressure increases with depth (reduced height).

If the pressure at a height y_1 in the fluid is P_1, and at height y_2 it is P_2, then we can integrate Eq. 13–4 to obtain

$$\int_{P_1}^{P_2} dP = -\int_{y_1}^{y_2} \rho g\, dy$$

$$P_2 - P_1 = -\int_{y_1}^{y_2} \rho g\, dy, \qquad \textbf{(13–5)}$$

where we assume ρ is a function of height y: $\rho = \rho(y)$. This is a general relation,

[†]Now we are measuring y positive upwards, the reverse of what we did to get Eq. 13–3 where we measured the depth (i.e. downward as positive).

and we apply it now to two special cases: (1) pressure in liquids of uniform density and (2) pressure variations in the Earth's atmosphere.

For liquids in which any variation in density can be ignored, ρ = constant and Eq. 13–5 is readily integrated:

$$P_2 - P_1 = -\rho g(y_2 - y_1). \qquad \textbf{(13–6a)}$$

For the everyday situation of a liquid in an open container—such as water in a glass, a swimming pool, a lake, or the ocean—there is a free surface at the top exposed to the atmosphere. It is convenient to measure distances from this top surface. That is, we let h be the *depth* in the liquid where $h = y_2 - y_1$ as shown in Fig. 13–5. If we let y_2 be the position of the top surface, then P_2 represents the atmospheric pressure, P_0, at the top surface. Then, from Eq. 13–6a, the pressure $P\ (= P_1)$ at a depth h in the fluid is

$$P = P_0 + \rho gh. \qquad [h \text{ is depth in liquid}] \quad \textbf{(13–6b)}$$

Note that Eq. 13–6b is simply the liquid pressure (Eq. 13–3) plus the pressure P_0 due to the atmosphere above.

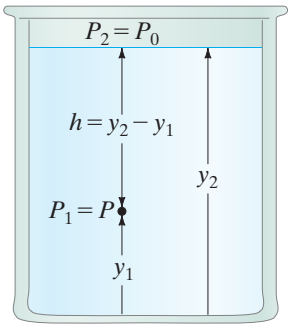


FIGURE 13–5 Pressure at a depth $h = (y_2 - y_1)$ in a liquid of density ρ is $P = P_0 + \rho gh$, where P_0 is the external pressure at the liquid's top surface.

EXAMPLE 13–3 **Pressure at a faucet.** The surface of the water in a storage tank is 30 m above a water faucet in the kitchen of a house, Fig. 13–6. Calculate the difference in water pressure between the faucet and the surface of the water in the tank.

APPROACH Water is practically incompressible, so ρ is constant even for $h = 30$ m when used in Eq. 13–6b. Only h matters; we can ignore the "route" of the pipe and its bends.

SOLUTION We assume the atmospheric pressure at the surface of the water in the storage tank is the same as at the faucet. So, the water pressure difference between the faucet and the surface of the water in the tank is

$$\Delta P = \rho gh = (1.0 \times 10^3\,\text{kg/m}^3)(9.8\,\text{m/s}^2)(30\,\text{m}) = 2.9 \times 10^5\,\text{N/m}^2.$$

NOTE The height h is sometimes called the **pressure head**. In this Example, the head of water is 30 m at the faucet. The very different diameters of the tank and faucet don't affect the result—only height does.

PHYSICS APPLIED
Water supply

CAUTION
Pressure depends only on h; *ignore pipe bends*

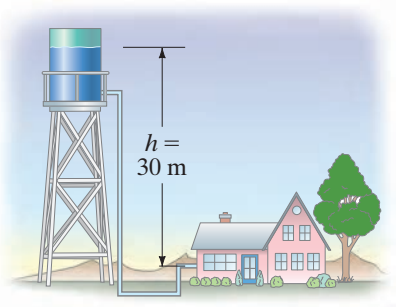


FIGURE 13–6 Example 13–3.

EXAMPLE 13–4 **Force on aquarium window.** Calculate the force due to water pressure exerted on a $1.0\,\text{m} \times 3.0\,\text{m}$ aquarium viewing window whose top edge is 1.0 m below the water surface, Fig. 13–7.

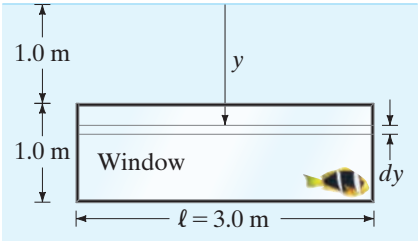
APPROACH At a depth h, the pressure due to the water is given by Eq. 13–6b. Divide the window up into thin horizontal strips of length $\ell = 3.0$ m and thickness dy, as shown in Fig. 13–7. We choose a coordinate system with $y = 0$ at the surface of the water and y is positive downward. (With this choice, the minus sign in Eq. 13–6a becomes plus, or we use Eq. 13–6b with $y = h$.) The force due to water pressure on each strip is $dF = PdA = \rho g y \ell\, dy$.

SOLUTION The total force on the window is given by the integral:

$$\int_{y_1 = 1.0\,\text{m}}^{y_2 = 2.0\,\text{m}} \rho g y \ell\, dy = \tfrac{1}{2}\rho g \ell (y_2^2 - y_1^2)$$
$$= \tfrac{1}{2}(1000\,\text{kg/m}^3)(9.8\,\text{m/s}^2)(3.0\,\text{m})[(2.0\,\text{m})^2 - (1.0\,\text{m})^2] = 44{,}000\,\text{N}.$$

NOTE To check our answer, we can do an estimate: multiply the area of the window $(3.0\,\text{m}^2)$ times the pressure at the middle of the window $(h = 1.5\,\text{m})$ using Eq. 13–3, $P = \rho gh = (1000\,\text{kg/m}^3)(9.8\,\text{m/s}^2)(1.5\,\text{m}) \approx 1.5 \times 10^4\,\text{N/m}^2$. So $F = PA \approx (1.5 \times 10^4\,\text{N/m}^2)(3.0\,\text{m})(1.0\,\text{m}) \approx 4.5 \times 10^4\,\text{N}$. Good!

FIGURE 13–7 Example 13–4.



EXERCISE B A dam holds back a lake that is 85 m deep at the dam. If the lake is 20 km long, how much thicker should the dam be than if the lake were smaller, only 1.0 km long?

Now let us apply Eq. 13–4 or 13–5 to gases. The density of gases is normally quite small, so the difference in pressure at different heights can usually be ignored if $y_2 - y_1$ is not large (we used this in Example 13–3). Indeed, for most ordinary containers of gas, we can assume that the pressure is the same throughout. However, if $y_2 - y_1$ is very large, we cannot make this assumption. An interesting example is the air of Earth's atmosphere, whose pressure at sea level is about $1.013 \times 10^5\ \text{N/m}^2$ and decreases gradually with altitude.

PHYSICS APPLIED
How pressure in the atmosphere decreases with altitude

Elevation effect on atmospheric pressure

We can determine the variation in pressure in the Earth's atmosphere as a function of height y above sea level, assuming g is constant and that the density of the air is proportional to the pressure. (This last assumption is not terribly accurate, in part because temperature and other weather effects are important.)

We start with Eq. 13–4 and integrate it from the surface of the Earth where $y = 0$ and $P = P_0$, up to height y at pressure P. We are assuming that ρ is proportional to P, so we can write

$$\frac{\rho}{\rho_0} = \frac{P}{P_0},$$

where $P_0 = 1.013 \times 10^5\ \text{N/m}^2$ is atmospheric pressure at sea level and $\rho_0 = 1.29\ \text{kg/m}^3$ is the density of air at sea level at 0°C (Table 13–1). (These are "average" values for P_0 and ρ_0 because they in fact vary with the weather.) From the differential change in pressure with height, Eq. 13–4, we have

$$\frac{dP}{dy} = -\rho g = -P\left(\frac{\rho_0}{P_0}\right)g,$$

so

$$\frac{dP}{P} = -\frac{\rho_0}{P_0}\,g\,dy.$$

We integrate this from $y = 0$ (Earth's surface) where $P = P_0$, to the height y where the pressure is P:

$$\int_{P_0}^{P} \frac{dP}{P} = -\frac{\rho_0}{P_0}\,g \int_0^y dy$$

$$\ln\frac{P}{P_0} = -\frac{\rho_0}{P_0}\,gy,$$

since $\ln P - \ln P_0 = \ln(P/P_0)$: see Appendix A–7. Then

$$P = P_0 e^{-(\rho_0 g/P_0)y}. \qquad \textbf{(13–6c)}$$

So, based on our assumptions, we find that the air pressure in our atmosphere decreases approximately exponentially with height above Earth's surface.

The atmosphere does not have a distinct top surface, so there is no natural point from which to measure *depth* in the atmosphere, as we can do for a liquid.

PHYSICS APPLIED
Altitude where air pressure drops to half

EXAMPLE 13–5 **Elevation where $P = \frac{1}{2}$ atm.** At what altitude above sea level is the atmospheric pressure equal to half the pressure at sea level?

APPROACH We use Eq. 13–6c setting $P = \frac{1}{2}P_0$ and solve for y.

SOLUTION The constant $(\rho_0 g/P_0)$ has the value

$$\frac{\rho_0 g}{P_0} = \frac{(1.29\ \text{kg/m}^3)(9.80\ \text{m/s}^2)}{(1.013 \times 10^5\ \text{N/m}^2)} = 1.25 \times 10^{-4}\ \text{m}^{-1}.$$

Then, when we set $P = \frac{1}{2}P_0$ in Eq. 13–6c, we obtain

$$\tfrac{1}{2} = e^{-(1.25\times10^{-4}\ \text{m}^{-1})y}.$$

Now we take natural logarithms of both sides:

$$\ln\tfrac{1}{2} = (-1.25 \times 10^{-4}\ \text{m}^{-1})y.$$

So (recall $\ln\frac{1}{2} = -\ln 2$, Appendix A–7, Eq. iii)

$$y = (\ln 2.00)/(1.25 \times 10^{-4}\ \text{m}^{-1}) = 5550\ \text{m}.$$

At an elevation of about 5500 m (about 18,000 ft), atmospheric pressure is half what it is at sea level. Mountain climbers often use oxygen tanks at very high altitudes.

13–4 Atmospheric Pressure and Gauge Pressure

Atmospheric Pressure

The pressure of the air at a given place on Earth varies slightly according to the weather and your elevation above sea level. At sea level, the pressure of the atmosphere on average is $1.013 \times 10^5\,\text{N/m}^2$ $(\text{or } 14.7\,\text{lb/in.}^2)$. This value lets us define a commonly used unit of pressure, the **atmosphere** (abbreviated atm):

$$1\,\text{atm} = 1.013 \times 10^5\,\text{N/m}^2 = 101.3\,\text{kPa}.$$

Another unit of pressure sometimes used (in meteorology and on weather maps) is the **bar**, which is defined as

$$1\,\text{bar} = 1.000 \times 10^5\,\text{N/m}^2.$$

Thus **standard atmospheric pressure**, at sea level, is slightly more than 1 bar.

The pressure due to the weight of the atmosphere is exerted on all objects immersed in this great sea of air, including our bodies. How does a human body withstand the enormous pressure on its surface? The answer is that living cells maintain an internal pressure that closely equals the external pressure, just as the pressure inside a balloon closely matches the outside pressure of the atmosphere. An automobile tire, because of its rigidity, can maintain internal pressures much greater than the external pressure.

PHYSICS APPLIED
Pressure on living cell

CONCEPTUAL EXAMPLE 13–6 **Finger holds water in a straw.** You insert a straw of length ℓ into a tall glass of water. You place your finger over the top of the straw, capturing some air above the water but preventing any additional air from getting in or out, and then you lift the straw from the water. You find that the straw retains most of the water (Fig. 13–8a). Does the air in the space between your finger and the top of the water have a pressure P that is greater than, equal to, or less than, the atmospheric pressure P_0 outside the straw?

RESPONSE Consider the forces on the column of water (Fig. 13–8b). Atmospheric pressure outside the straw pushes upward on the water at the bottom of the straw, gravity pulls the water downward, and the air pressure inside the top of the straw pushes downward on the water. Since the water is in equilibrium, the upward force due to atmospheric pressure P_0 must balance the two downward forces. The only way this is possible is for the air pressure P inside the straw at the top to be *less than* the atmosphere pressure outside the straw. (When you initially remove the straw from the water glass, a little water may leave the bottom of the straw, thus increasing the volume of trapped air and reducing its density and pressure.)

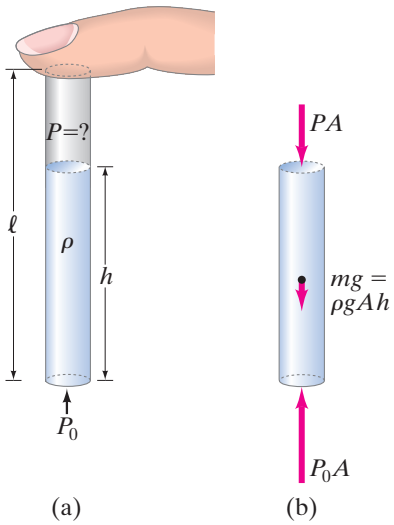


FIGURE 13–8 Example 13–6.

Gauge Pressure

It is important to note that tire gauges, and most other pressure gauges, register the pressure above atmospheric pressure. This is called **gauge pressure**. Thus, to get the **absolute pressure**, P, we must add the atmospheric pressure, P_0, to the gauge pressure, P_G:

$$P = P_G + P_0.$$

If a tire gauge registers 220 kPa, the absolute pressure within the tire is 220 kPa + 101 kPa = 321 kPa, equivalent to about 3.2 atm (2.2 atm gauge pressure).

13–5 Pascal's Principle

The Earth's atmosphere exerts a pressure on all objects with which it is in contact, including other fluids. External pressure acting on a fluid is transmitted throughout that fluid. For instance, according to Eq. 13–3, the pressure due to the water at a depth of 100 m below the surface of a lake is $P = \rho g\,\Delta h = (1000\,\text{kg/m}^3)(9.8\,\text{m/s}^2)(100\,\text{m}) = 9.8 \times 10^5\,\text{N/m}^2$, or 9.7 atm.

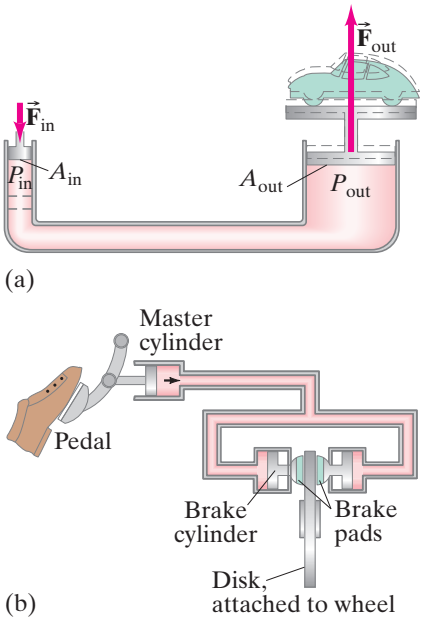


FIGURE 13–9 Applications of Pascal's principle: (a) hydraulic lift; (b) hydraulic brakes in a car.

However, the total pressure 100 m down in a lake is due to the pressure of water plus the pressure of the air above it. Hence the total pressure (if the lake is near sea level) is 9.7 atm + 1.0 atm = 10.7 atm. This is just one example of a general principle attributed to the French philosopher and scientist Blaise Pascal (1623–1662). **Pascal's principle** states that *if an external pressure is applied to a confined fluid, the pressure at every point within the fluid increases by that amount.*

A number of practical devices make use of Pascal's principle. One example is the hydraulic lift, illustrated in Fig. 13–9a, in which a small input force is used to exert a large output force by making the area of the output piston larger than the area of the input piston. To see how this works, we assume the input and output pistons are at the same height (at least approximately). Then the external input force F_{in}, by Pascal's principle, increases the pressure equally throughout the fluid. Therefore, at the same level (see Fig. 13–9a),

$$P_{out} = P_{in}$$

where the input quantities are represented by the subscript "in" and the output by "out." Since $P = F/A$, we write the above equality as

$$\frac{F_{out}}{A_{out}} = \frac{F_{in}}{A_{in}},$$

or

$$\frac{F_{out}}{F_{in}} = \frac{A_{out}}{A_{in}}.$$

The quantity F_{out}/F_{in} is called the **mechanical advantage** of the hydraulic lift, and it is equal to the ratio of the areas. For example, if the area of the output piston is 20 times that of the input cylinder, the force is multiplied by a factor of 20. Thus a force of 200 lb could lift a 4000-lb car.

PHYSICS APPLIED
Hydraulic lift

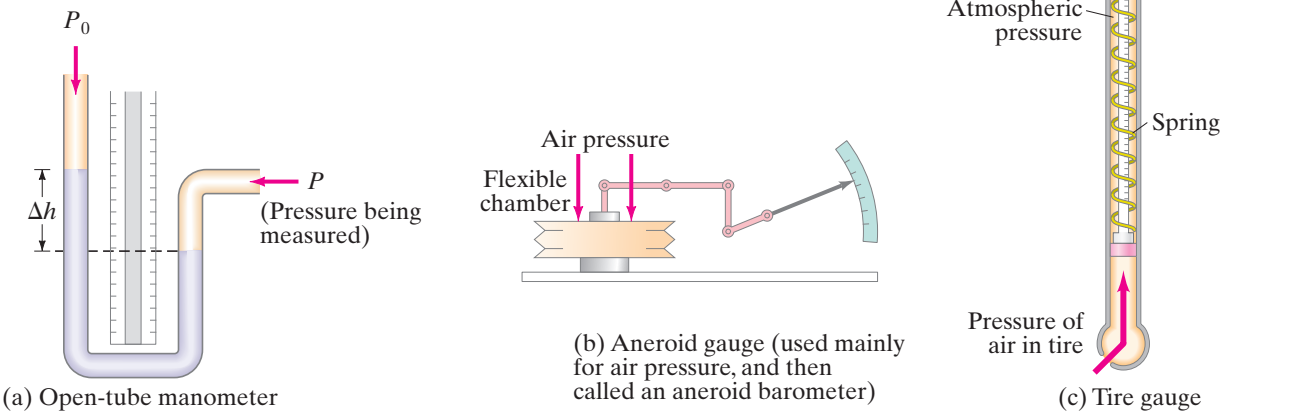
PHYSICS APPLIED
Hydraulic brakes

Figure 13–9b illustrates the brake system of a car. When the driver presses the brake pedal, the pressure in the master cylinder increases. This pressure increase occurs throughout the brake fluid, thus pushing the brake pads against the disk attached to the car's wheel. A leak, though, can be disastrous.

13–6 Measurement of Pressure; Gauges and the Barometer

Many devices have been invented to measure pressure, some of which are shown in Fig. 13–10. The simplest is the *open-tube* **manometer** (Fig. 13–10a) which is a U-shaped tube partially filled with a liquid, usually mercury or water. The pressure P

FIGURE 13–10 Pressure gauges: (a) open-tube manometer, (b) aneroid gauge, and (c) common tire pressure gauge.



being measured is related to the difference in height Δh of the two levels of the liquid by the relation

$$P = P_0 + \rho g\,\Delta h, \qquad \text{[manometer]}$$

where P_0 is atmospheric pressure (acting on the top of the liquid in the left-hand tube), and ρ is the density of the liquid in the U-shaped tube. Note that the quantity $\rho g\,\Delta h$ is the gauge pressure—the amount by which P exceeds atmospheric pressure P_0. If the liquid in the left-hand column is lower than that in the right-hand column, P is less than atmospheric pressure (and Δh is negative).

Instead of calculating the product $\rho g\,\Delta h$, sometimes only the change in height Δh is specified. In fact, pressures are sometimes specified as so many "millimeters of mercury" (mm-Hg), if ρ is for mercury, or "mm of water" (mm-H_2O). The unit mm-Hg is equivalent to a pressure of $133\,\text{N/m}^2$, because $\rho g\,\Delta h$ for 1 mm $(= 1.0 \times 10^{-3}\,\text{m})$ of mercury gives

$$\begin{aligned}\rho g\,\Delta h &= (13.6 \times 10^3\,\text{kg/m}^3)(9.80\,\text{m/s}^2)(1.00 \times 10^{-3}\,\text{m}) \\ &= 1.33 \times 10^2\,\text{N/m}^2.\end{aligned}$$

The unit mm-Hg is also called the **torr** in honor of the Italian scientist Evangelista Torricelli (1608–1647), a student of Galileo's who invented the barometer (see Fig. 13–11).

Conversion factors among the various units of pressure (an incredible nuisance!) are given in Table 13–2. It is important that only $\text{N/m}^2 = \text{Pa}$, the proper SI unit, be used in calculations involving other quantities specified in SI units.

PROBLEM SOLVING
Use SI unit in calculations:
1 Pa = 1 N/m^2

TABLE 13–2 Conversion Factors Between Different Units of Pressure

In Terms of 1 Pa = 1 N/m^2	1 atm in Different Units
1 atm = $1.013 \times 10^5\,\text{N/m}^2$ = $1.013 \times 10^5\,\text{Pa} = 101.3\,\text{kPa}$	1 atm = $1.013 \times 10^5\,\text{N/m}^2$
1 bar = $1.000 \times 10^5\,\text{N/m}^2$	1 atm = 1.013 bar
1 dyne/cm^2 = $0.1\,\text{N/m}^2$	1 atm = 1.013×10^6 dyne/cm^2
1 lb/in.^2 = $6.90 \times 10^3\,\text{N/m}^2$	1 atm = 14.7 lb/in.^2
1 lb/ft^2 = $47.9\,\text{N/m}^2$	1 atm = 2.12×10^3 lb/ft^2
1 cm-Hg = $1.33 \times 10^3\,\text{N/m}^2$	1 atm = 76.0 cm-Hg
1 mm-Hg = $133\,\text{N/m}^2$	1 atm = 760 mm-Hg
1 torr = $133\,\text{N/m}^2$	1 atm = 760 torr
1 mm-H_2O (4°C) = $9.80\,\text{N/m}^2$	1 atm = 1.03×10^4 mm-H_2O (4°C) ≈ 10 m of water

Another type of pressure gauge is the **aneroid gauge** (Fig. 13–10b) in which the pointer is linked to the flexible ends of an evacuated thin metal chamber. In electronic gauges, the pressure may be applied to a thin metal diaphragm whose resulting deformation is translated into an electrical signal by a transducer. A common tire gauge uses a spring, as shown in Fig. 13–10c.

Atmospheric pressure can be measured by a modified kind of mercury manometer with one end closed, called a mercury **barometer** (Fig. 13–11). The glass tube is completely filled with mercury and then inverted into the bowl of mercury. If the tube is long enough, the level of the mercury will drop, leaving a vacuum at the top of the tube, since atmospheric pressure can support a column of mercury only about 76 cm high (exactly 76.0 cm at standard atmospheric pressure). That is, a column of mercury 76 cm high exerts the same pressure as the atmosphere:†

$$\begin{aligned}P &= \rho g\,\Delta h \\ &= (13.6 \times 10^3\,\text{kg/m}^3)(9.80\,\text{m/s}^2)(0.760\,\text{m}) = 1.013 \times 10^5\,\text{N/m}^2 = 1.00\,\text{atm}.\end{aligned}$$

Household barometers are usually of the aneroid type (Fig. 13–10b), either mechanical (with dial) or electronic.

†This calculation confirms the entry in Table 13–2, 1 atm = 76.0 cm-Hg.

FIGURE 13–11 A mercury barometer, invented by Torricelli, is shown here when the air pressure is standard atmospheric, 76.0 cm-Hg.

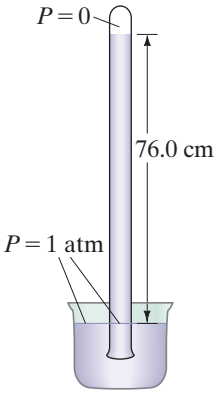




FIGURE 13–12 A water barometer: a full tube of water (longer than 10 m), closed at the top, is inserted into a tub of water. When the submerged bottom end of the tube is unplugged, some water flows out of the tube into the tub, leaving a vacuum at the top of the tube above the water's upper surface. Why? Because air pressure can support a column of water only 10 m high.

A calculation similar to that just done for Hg will show that atmospheric pressure can maintain a column of water 10.3 m high in a tube whose top is under vacuum (Fig. 13–12). No matter how good a vacuum pump is, water cannot be made to rise more than 10.3 m under normal atmospheric pressure. To pump water out of deep mine shafts with a vacuum pump requires multiple stages for depths greater than 10 m. Galileo studied this problem, and his student Torricelli was the first to explain it. The point is that a pump does not really suck water up a tube—it merely reduces the pressure at the top of the tube. Atmospheric air pressure *pushes* the water up the tube if the top end is at lower pressure (under a vacuum), just as it is air pressure that pushes (or maintains) the mercury 76 cm high in a barometer. [Force pumps, Section 13–14, can push higher.]

CONCEPTUAL EXAMPLE 13–7 **Suction.** A novice engineer proposes suction cup shoes for space shuttle astronauts working on the exterior of a spacecraft. Having just studied this Chapter, you gently remind him of the fallacy of this plan. What is it?

RESPONSE Suction cups work by pushing out the air underneath the cup. What holds the suction cup in place is the air pressure outside it. (This can be a substantial force when on Earth. For example, a 10-cm-diameter suction cup has an area of $7.9 \times 10^{-3}\,\text{m}^2$. The force of the atmosphere on it is $(7.9 \times 10^{-3}\,\text{m}^2)(1.0 \times 10^5\,\text{N/m}^2) \approx 800\,\text{N}$, about 180 lbs!) But in outer space, there is no air pressure to push the suction cup onto the spacecraft.

We sometimes mistakenly think of suction as something we actively do. For example, we intuitively think that we pull the soda up through a straw. Instead, what we do is use our mouth to lower the pressure at the top of the straw, and the atmosphere *pushes* the soda up the straw.

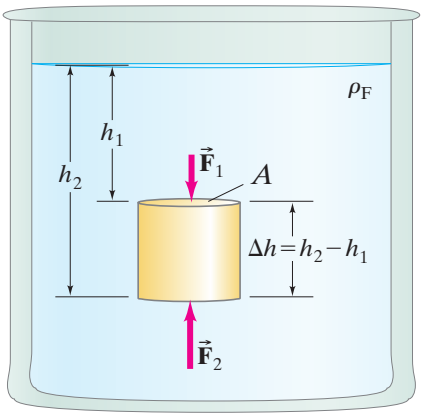
Vacuum

A "vacuum" ideally means a space devoid of matter. In real life it refers to a reduced pressure, usually much less than atmospheric pressure. (The top of the barometer in Fig. 13–11 would seem a perfect vacuum; but some molecules of mercury evaporate into the gas state, just like water evaporates.)

13–7 Buoyancy; Archimedes' Principle

Objects submerged in a fluid appear to weigh less than they do when outside the fluid. For example, a large rock that you would have difficulty lifting off the ground can often be easily lifted from the bottom of a stream. When you lift the rock through the surface of the water, it suddenly seems to be much heavier. Many objects, such as wood, float on the surface of water. These two effects are examples of **buoyancy**. In each example, the force of gravity is acting downward. But in addition, an upward *buoyant force* is exerted by the liquid. The buoyant force on fish and underwater divers (as in the Chapter-Opening photo) almost exactly balances the force of gravity downward, and allows them to "hover" in equilibrium.

FIGURE 13–13 Determination of the buoyant force.



The buoyant force occurs because the pressure in a fluid increases with depth. Thus the upward pressure on the bottom surface of a submerged object is greater than the downward pressure on its top surface. To see this effect, consider a cylinder of height Δh whose top and bottom ends have an area A and which is completely submerged in a fluid of density ρ_F, as shown in Fig. 13–13. The fluid exerts a pressure $P_1 = \rho_F g h_1$ at the top surface of the cylinder (Eq. 13–3). The force due to this pressure on top of the cylinder is $F_1 = P_1 A = \rho_F g h_1 A$, and it is directed downward. Similarly, the fluid exerts an upward force on the bottom of the cylinder equal to $F_2 = P_2 A = \rho_F g h_2 A$. The net force on the cylinder exerted by the fluid pressure, which is the **buoyant force**, $\vec{\mathbf{F}}_B$, acts upward and has the magnitude

$$\begin{aligned} F_B &= F_2 - F_1 = \rho_F g A (h_2 - h_1) \\ &= \rho_F g A\, \Delta h \\ &= \rho_F V g \\ &= m_F g, \end{aligned}$$

where $V = A\,\Delta h$ is the volume of the cylinder; the product $\rho_F V$ is the mass of the fluid displaced, and $\rho_F V g = m_F g$ is the weight of fluid which takes up

a volume equal to the volume of the cylinder. Thus the buoyant force on the cylinder is equal to the weight of fluid displaced by the cylinder.

This result is valid no matter what the shape of the object. Its discovery is credited to Archimedes (287?–212 B.C.), and it is called **Archimedes' principle**:

> **the buoyant force on an object immersed in a fluid is equal to the weight of the fluid displaced by that object.**

By "fluid displaced," we mean a volume of fluid equal to the submerged volume of the object (or that part of the object that is submerged). If the object is placed in a glass or tub initially filled to the brim with water, the water that flows over the top represents the water displaced by the object.

We can derive Archimedes' principle in general by the following simple but elegant argument. The irregularly shaped object D shown in Fig. 13–14a is acted on by the force of gravity (its weight, $m\vec{\mathbf{g}}$, downward) and the buoyant force, $\vec{\mathbf{F}}_B$, upward. We wish to determine F_B. To do so, we next consider a body (D′ in Fig. 13–14b), this time made of the fluid itself, with the same shape and size as the original object, and located at the same depth. You might think of this body of fluid as being separated from the rest of the fluid by an imaginary membrane. The buoyant force F_B on this body of fluid will be exactly the same as that on the original object since the surrounding fluid, which exerts F_B, is in exactly the same configuration. This body of fluid D′ is in equilibrium (the fluid as a whole is at rest). Therefore, $F_B = m'g$, where $m'g$ is the weight of the body of fluid D′. Hence the buoyant force F_B is equal to the weight of the body of fluid whose volume equals the volume of the original submerged object, which is Archimedes' principle.

Archimedes' discovery was made by experiment. What we have done is show that Archimedes' principle can be derived from Newton's laws.

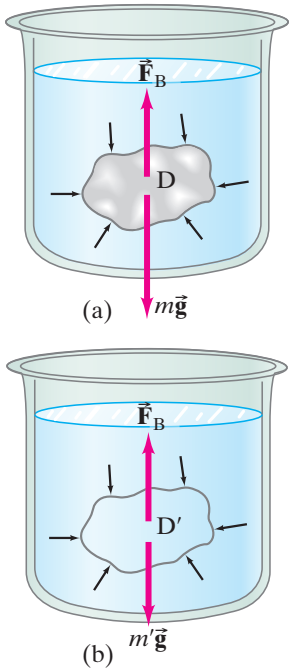


FIGURE 13–14 Archimedes' principle.

CONCEPTUAL EXAMPLE 13–8 **Two pails of water.** Consider two identical pails of water filled to the brim. One pail contains only water, the other has a piece of wood floating in it. Which pail has the greater weight?

RESPONSE Both pails weigh the same. Recall Archimedes' principle: the wood displaces a volume of water with weight equal to the weight of the wood. Some water will have overflowed from the pail, but Archimedes' principle tells us the spilled water has weight equal to the weight of the wood; so that pail will have the same weight as the other pail.

EXAMPLE 13–9 **Recovering a submerged statue.** A 70-kg ancient statue lies at the bottom of the sea. Its volume is $3.0 \times 10^4\,\text{cm}^3$. How much force is needed to lift it (without acceleration)?

APPROACH The force F needed to lift the statue is equal to the statue's weight mg minus the buoyant force F_B. Figure 13–15 is the free-body diagram.

SOLUTION We apply Newton's second law, $\Sigma F = ma = 0$, which gives $F + F_B - mg = 0$ or

$$F = mg - F_B.$$

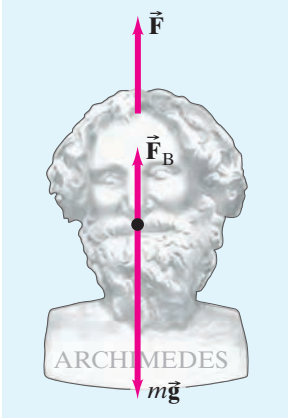
The buoyant force on the statue due to the water is equal to the weight of $3.0 \times 10^4\,\text{cm}^3 = 3.0 \times 10^{-2}\,\text{m}^3$ of water (for seawater, $\rho = 1.025 \times 10^3\,\text{kg/m}^3$):

$$F_B = m_{H_2O}g = \rho_{H_2O}Vg = (1.025 \times 10^3\,\text{kg/m}^3)(3.0 \times 10^{-2}\,\text{m}^3)(9.8\,\text{m/s}^2)$$
$$= 3.0 \times 10^2\,\text{N},$$

where we use the chemical symbol for water, H_2O, as a subscript. The weight of the statue is $mg = (70\,\text{kg})(9.8\,\text{m/s}^2) = 6.9 \times 10^2\,\text{N}$. Hence the force F needed to lift it is $690\,\text{N} - 300\,\text{N} = 390\,\text{N}$. It feels like the statue has a mass of only $(390\,\text{N})/(9.8\,\text{m/s}^2) = 40\,\text{kg}$.

NOTE Here $F = 390\,\text{N}$ is the force needed to lift the statue without acceleration when it is under water. As the statue comes *out* of the water, the force F increases, reaching 690 N when the statue is fully out of the water.

FIGURE 13–15 Example 13–9. The force needed to lift the statue is $\vec{\mathbf{F}}$.



Archimedes is said to have discovered his principle in his bath while thinking how he might determine whether the king's new crown was pure gold or a fake. Gold has a specific gravity of 19.3, somewhat higher than that of most metals, but a determination of specific gravity or density is not readily done directly because, even if the mass is known, the volume of an irregularly shaped object is not easily calculated. However, if the object is weighed in air $(= w)$ and also "weighed" while it is under water $(= w')$, the density can be determined using Archimedes' principle, as the following Example shows. The quantity w' is called the *apparent weight* in water, and is what a spring scale reads when the object is submerged in water (see Fig. 13–16); w' equals the true weight $(w = mg)$ minus the buoyant force.

EXAMPLE 13–10 **Archimedes: Is the crown gold?** When a crown of mass 14.7 kg is submerged in water, an accurate scale reads only 13.4 kg. Is the crown made of gold?

APPROACH If the crown is gold, its density and specific gravity must be very high, SG = 19.3 (see Table 13–1). We determine the specific gravity of the crown using Archimedes' principle and the two free-body diagrams shown in Fig. 13–16.

SOLUTION The apparent weight of the submerged object (the crown) is w' (what the scale reads in Fig. 13–16b), and is the force pulling down on the scale hook. By Newton's third law, w' equals the force F'_T that the scale exerts on the crown in Fig. 13–16b. The sum of the forces on the crown at rest is zero, so w' equals the actual weight $w\ (= mg)$ minus the buoyant force F_B:

$$w' = F'_T = w - F_B$$

so

$$w - w' = F_B.$$

Let V be the volume of the completely submerged object (the crown) and ρ_O the object's density (so $\rho_O V$ is its mass), and let ρ_F be the density of the fluid (water). Then $(\rho_F V)g$ is the weight of fluid displaced $(= F_B)$. Now we can write

$$w = mg = \rho_O Vg$$
$$w - w' = F_B = \rho_F Vg. \qquad \text{(Archimedes' principle)}$$

We divide these two equations and obtain

$$\frac{w}{w - w'} = \frac{\rho_O Vg}{\rho_F Vg} = \frac{\rho_O}{\rho_F}.$$

We see that $w/(w - w')$ is equal to the specific gravity of the object (the crown) if the fluid in which it is submerged is water $(\rho_F = 1.00 \times 10^3\ \text{kg/m}^3)$. Thus

$$\frac{\rho_O}{\rho_{H_2O}} = \frac{w}{w - w'} = \frac{(14.7\ \text{kg})g}{(14.7\ \text{kg} - 13.4\ \text{kg})g} = \frac{14.7\ \text{kg}}{1.3\ \text{kg}} = 11.3.$$

This corresponds to a density of $11{,}300\ \text{kg/m}^3$. The crown is not gold, but seems to be made of lead (see Table 13–1).

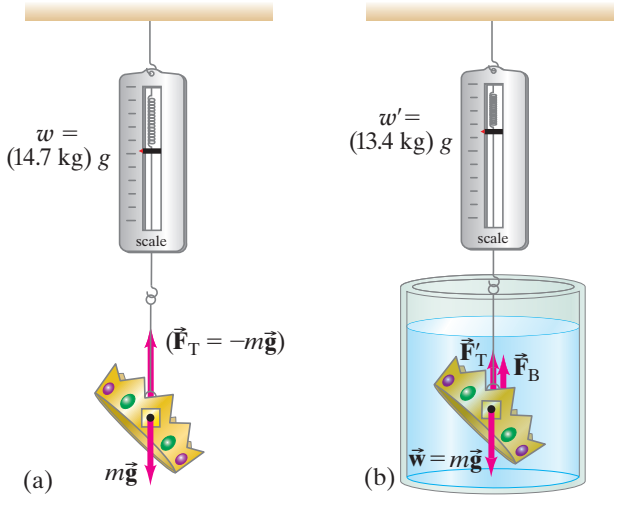


FIGURE 13–16 (a) A scale reads the mass of an object in air—in this case the crown of Example 13–10. All objects are at rest, so the tension F_T in the connecting cord equals the weight w of the object: $F_T = mg$. We show the free-body diagram of the crown, and F_T is what causes the scale reading (it is equal to the net downward force on the scale, by Newton's third law). (b) Submerged, the crown has an additional force on it, the buoyant force F_B. The net force is zero, so $F'_T + F_B = mg\ (= w)$. The scale now reads $m' = 13.4$ kg, where m' is related to the effective weight by $w' = m'g$. Thus $F'_T = w' = w - F_B$.

Archimedes' principle applies equally well to objects that float, such as wood. In general, *an object floats on a fluid if its density* (ρ_O) *is less than that of the fluid* (ρ_F). This is readily seen from Fig. 13–17a, where a submerged log of mass m_O will experience a net upward force and float to the surface if $F_B > m_O g$; that is, if $\rho_F V g > \rho_O V g$ or $\rho_F > \rho_O$. At equilibrium—that is, when floating—the buoyant force on an object has magnitude equal to the weight of the object. For example, a log whose specific gravity is 0.60 and whose volume is 2.0 m^3 has a mass

$$m_O = \rho_O V = (0.60 \times 10^3\ \text{kg/m}^3)(2.0\ \text{m}^3) = 1200\ \text{kg}.$$

If the log is fully submerged, it will displace a mass of water

$$m_F = \rho_F V = (1000\ \text{kg/m}^3)(2.0\ \text{m}^3) = 2000\ \text{kg}.$$

Hence the buoyant force on the submerged log $(= m_F g)$ will be greater than its weight, and it will be pushed upward to float on the surface (Fig. 13–17). The log will come to equilibrium when it displaces 1200 kg of water, which means that 1.2 m^3 of its volume will be submerged. This 1.2 m^3 corresponds to 60% of the volume of the log $(= 1.2/2.0 = 0.60)$, so 60% of the log is submerged.

In general when an object floats, we have $F_B = m_O g$, which we can write as (see Fig. 13–18)

$$F_B = m_O g$$
$$\rho_F V_{\text{displ}} g = \rho_O V_O g,$$

where V_O is the full volume of the object and V_{displ} is the volume of fluid it displaces (= volume submerged). Thus

$$\frac{V_{\text{displ}}}{V_O} = \frac{\rho_O}{\rho_F}.$$

That is, the fraction of the object submerged is given by the ratio of the object's density to that of the fluid. If the fluid is water, this fraction equals the specific gravity of the object.

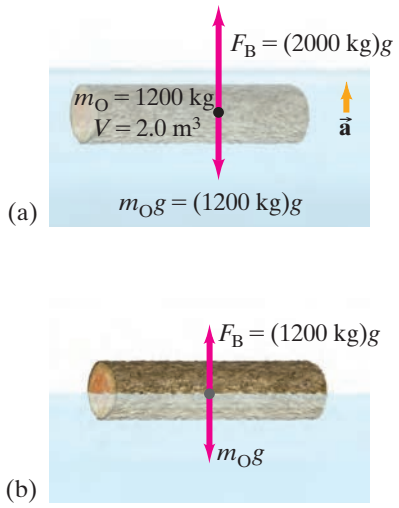
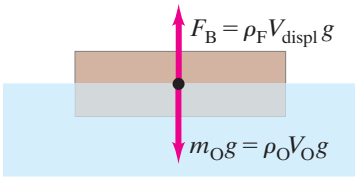


FIGURE 13–17 (a) The fully submerged log accelerates upward because $F_B > m_O g$. It comes to equilibrium (b) when $\Sigma F = 0$, so $F_B = m_O g = (1200\ \text{kg})g$. Then 1200 kg, or 1.2 m^3, of water is displaced.

FIGURE 13–18 An object floating in equilibrium: $F_B = m_O g$.



EXAMPLE 13–11 **Hydrometer calibration.** A **hydrometer** is a simple instrument used to measure the specific gravity of a liquid by indicating how deeply the instrument sinks in the liquid. A particular hydrometer (Fig. 13–19) consists of a glass tube, weighted at the bottom, which is 25.0 cm long and 2.00 cm^2 in cross-sectional area, and has a mass of 45.0 g. How far from the weighted end should the 1.000 mark be placed?

APPROACH The hydrometer will float in water if its density ρ is less than $\rho_{H_2O} = 1.000\ \text{g/cm}^3$, the density of water. The fraction of the hydrometer submerged $(V_{\text{displaced}}/V_{\text{total}})$ is equal to the density ratio ρ/ρ_{H_2O}.

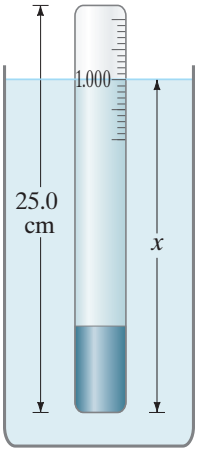
SOLUTION The hydrometer has an overall density

$$\rho = \frac{m}{V} = \frac{45.0\ \text{g}}{(2.00\ \text{cm}^2)(25.0\ \text{cm})} = 0.900\ \text{g/cm}^3.$$

Thus, when placed in water, it will come to equilibrium when 0.900 of its volume is submerged. Since it is of uniform cross section, $(0.900)(25.0\ \text{cm}) = 22.5\ \text{cm}$ of its length will be submerged. The specific gravity of water is defined to be 1.000, so the mark should be placed 22.5 cm from the weighted end.

NOTE Hydrometers can be used to measure the density of liquids like car antifreeze coolant, car battery acid (a measure of its charge), wine and beer fermenting, and many others.

FIGURE 13–19 A hydrometer. Example 13–11.



EXERCISE C On the hydrometer of Example 13–11, will the marks above the 1.000 mark represent higher or lower values of density of the liquid in which it is submerged?

EXERCISE D Which of the following objects, submerged in water, experiences the largest magnitude of the buoyant force? (*a*) A 1-m^3 helium balloon; (*b*) 1 m^3 of wood; (*c*) 1 m^3 of ice; (*d*) 1 m^3 of iron; (*e*) all the same.

PHYSICS APPLIED
Continental drift—plate tectonics

Archimedes' principle is also useful in geology. According to the theories of plate tectonics and continental drift, the continents float on a fluid "sea" of slightly deformable rock (mantle rock). Some interesting calculations can be done using very simple models, which we consider in the Problems at the end of the Chapter.

Air is a fluid, and it too exerts a buoyant force. Ordinary objects weigh less in air than they do in a vacuum. Because the density of air is so small, the effect for ordinary solids is slight. There are objects, however, that *float* in air—helium-filled balloons, for example, because the density of helium is less than the density of air.

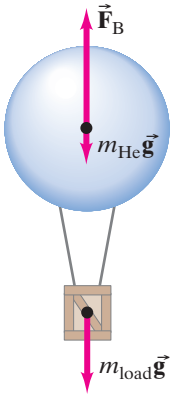


FIGURE 13–20 Example 13–12.

EXAMPLE 13–12 Helium balloon. What volume V of helium is needed if a balloon is to lift a load of 180 kg (including the weight of the empty balloon)?

APPROACH The buoyant force on the helium balloon, F_B, which is equal to the weight of displaced air, must be at least equal to the weight of the helium plus the weight of the balloon and load (Fig. 13–20), if they are to float upward. Table 13–1 gives the density of helium as $0.179\ \text{kg/m}^3$.

SOLUTION The buoyant force must have a minimum value of

$$F_B = (m_{He} + 180\ \text{kg})g.$$

This equation can be written in terms of density using Archimedes' principle:

$$\rho_{air} Vg = (\rho_{He} V + 180\ \text{kg})g.$$

Solving now for V, we find

$$V = \frac{180\ \text{kg}}{\rho_{air} - \rho_{He}} = \frac{180\ \text{kg}}{(1.29\ \text{kg/m}^3 - 0.179\ \text{kg/m}^3)} = 160\ \text{m}^3.$$

NOTE This is the minimum volume needed near the Earth's surface, where $\rho_{air} = 1.29\ \text{kg/m}^3$. To reach a high altitude, a greater volume would be needed since the density of air decreases with altitude.

CONCEPTUAL EXAMPLE 13–13 Throwing a rock overboard. A rowboat carrying a large granite rock floats in a small lake. If the rock (SG $\approx$ 3, Table 13–1) is thrown overboard and sinks, does the lake level drop, rise, or stay the same?

RESPONSE Together the boat and rock float, so the buoyant force on them equals their total weight. The boat and rock displace a mass of water whose weight is equal to the weight of boat plus rock. When the rock is thrown into the lake, it displaces only its own volume, which is smaller than the volume of water the rock displaced when in the boat ($\approx \frac{1}{3}$ as much because the rock's density is $\approx$ 3 times greater than water). So less lake water is displaced and the water level of the lake *drops* when the rock is in the lake.

Maybe numbers can help. Suppose the boat and the rock each has a mass of 60 kg. Then the boat carrying the rock displaces 120 kg of water, which is a volume of $0.12\ \text{m}^3$ ($\rho = 1000\ \text{kg/m}^3$ for water, Table 13–1). When the rock is thrown into the lake, the boat alone now displaces $0.06\ \text{m}^3$. The rock displaces only its own volume of $0.02\ \text{m}^3$ ($\rho = m/V \approx 3$ so $V \approx 0.06\ \text{m}^3/3$). Thus a total of $0.08\ \text{m}^3$ of water is displaced. Less water is displaced so the water level of the lake drops.

EXERCISE E If you throw a flat 60-kg aluminum plate into water, the plate sinks. But if that aluminum is shaped into a rowboat, it floats. Explain.

13–8 Fluids in Motion; Flow Rate and the Equation of Continuity

We now turn to the subject of fluids in motion, which is called **fluid dynamics**, or (especially if the fluid is water) **hydrodynamics**.

We can distinguish two main types of fluid flow. If the flow is smooth, such that neighboring layers of the fluid slide by each other smoothly, the flow is said to be **streamline** or **laminar flow**.† In streamline flow, each particle of the fluid follows a smooth path, called a **streamline**, and these paths do not cross

†The word laminar means "in layers."

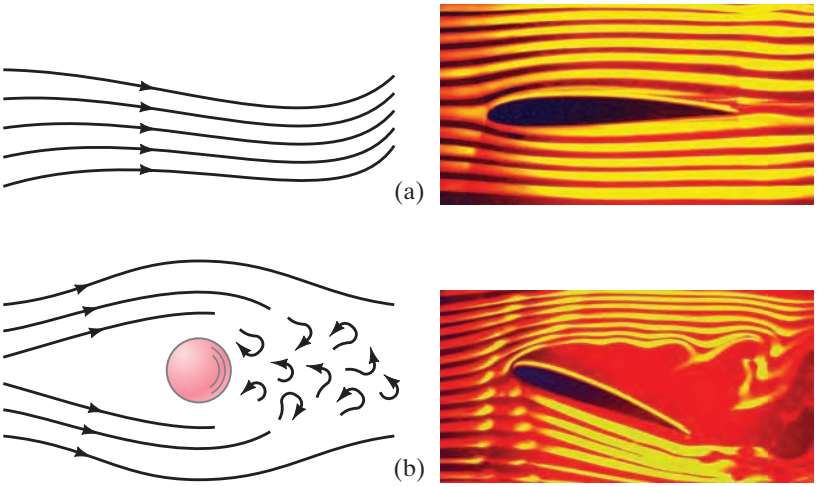


FIGURE 13–21 (a) Streamline, or laminar, flow; (b) turbulent flow. The photos show airflow around an airfoil or airplane wing (more in Section 13–10).

one another (Fig. 13–21a). Above a certain speed (and depending on the shape of an object immersed in the fluid) the flow becomes turbulent. **Turbulent flow** is characterized by erratic, small, whirlpool-like circles called *eddy currents* or *eddies* (Fig. 13–21b). Eddies absorb a great deal of energy, and although a certain amount of internal friction called **viscosity** is present even during streamline flow, it is much greater when the flow is turbulent. A few tiny drops of ink or food coloring dropped into a moving liquid can quickly reveal whether the flow is streamline or turbulent.

Let us consider the steady laminar flow of a fluid through an enclosed tube or pipe as shown in Fig. 13–22. First we determine how the speed of the fluid changes when the diameter of the tube changes. The mass **flow rate** is defined as the mass Δm of fluid that passes a given point per unit time Δt:

$$\text{mass flow rate} = \frac{\Delta m}{\Delta t}.$$

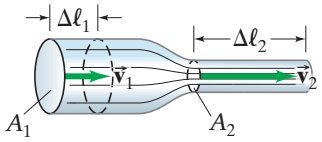


FIGURE 13–22 Fluid flow through a pipe of varying diameter.

In Fig. 13–22, the volume of fluid passing point 1 (through area A_1) in a time Δt is $A_1\,\Delta\ell_1$, where $\Delta\ell_1$ is the distance the fluid moves in time Δt. The velocity[†] of fluid (density ρ_1) passing point 1 is $v_1 = \Delta\ell_1/\Delta t$. Then the mass flow rate $\Delta m_1/\Delta t$ through area A_1 is

$$\frac{\Delta m_1}{\Delta t} = \frac{\rho_1\,\Delta V_1}{\Delta t} = \frac{\rho_1 A_1\,\Delta\ell_1}{\Delta t} = \rho_1 A_1 v_1,$$

where $\Delta V_1 = A_1\,\Delta\ell_1$ is the volume of mass Δm_1. Similarly, at point 2 (through area A_2), the flow rate is $\rho_2 A_2 v_2$. Since no fluid flows in or out the sides of the tube, the flow rates through A_1 and A_2 must be equal. Thus

$$\frac{\Delta m_1}{\Delta t} = \frac{\Delta m_2}{\Delta t},$$

and

$$\rho_1 A_1 v_1 = \rho_2 A_2 v_2. \qquad \textbf{(13–7a)}$$

This is called the **equation of continuity**.

If the fluid is incompressible (ρ doesn't change with pressure), which is an excellent approximation for liquids under most circumstances (and sometimes for gases as well), then $\rho_1 = \rho_2$, and the equation of continuity becomes

$$A_1 v_1 = A_2 v_2. \qquad [\rho = \text{constant}] \quad \textbf{(13–7b)}$$

The product Av represents the *volume rate of flow* (volume of fluid passing a given point per second), since $\Delta V/\Delta t = A\,\Delta\ell/\Delta t = Av$, which in SI units is m^3/s. Equation 13–7b tells us that where the cross-sectional area is large, the velocity is small, and where the area is small, the velocity is large.

[†]If there were no viscosity, the velocity would be the same across a cross section of the tube. Real fluids have viscosity, and this internal friction causes different layers of the fluid to flow at different speeds. In this case v_1 and v_2 represent the average speeds at each cross section.

That this is reasonable can be seen by looking at a river. A river flows slowly through a meadow where it is broad, but speeds up to torrential speed when passing through a narrow gorge.

PHYSICS APPLIED
Blood flow

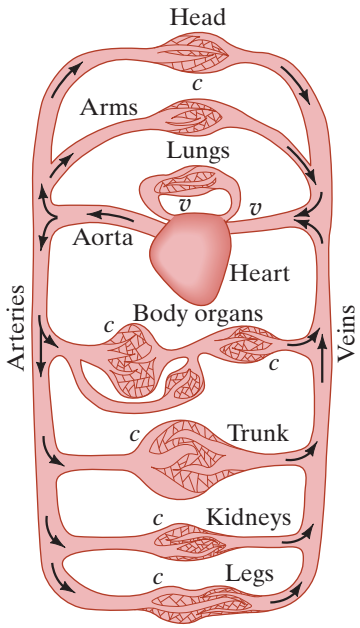


FIGURE 13–23 Human circulatory system.

EXAMPLE 13–14 **ESTIMATE** **Blood flow.** In humans, blood flows from the heart into the aorta, from which it passes into the major arteries, Fig. 13–23. These branch into the small arteries (arterioles), which in turn branch into myriads of tiny capillaries. The blood returns to the heart via the veins. The radius of the aorta is about 1.2 cm, and the blood passing through it has a speed of about 40 cm/s. A typical capillary has a radius of about 4×10^{-4} cm, and blood flows through it at a speed of about 5×10^{-4} m/s. Estimate the number of capillaries that are in the body.

APPROACH We assume the density of blood doesn't vary significantly from the aorta to the capillaries. By the equation of continuity, the volume flow rate in the aorta must equal the volume flow rate through *all* the capillaries. The total area of all the capillaries is given by the area of a typical capillary multiplied by the total number N of capillaries.

SOLUTION Let A_1 be the area of the aorta and A_2 be the area of *all* the capillaries through which blood flows. Then $A_2 = N\pi r_{\text{cap}}^2$, where $r_{\text{cap}} \approx 4 \times 10^{-4}$ cm is the estimated average radius of one capillary. From the equation of continuity (Eq. 13–7b), we have

$$\begin{aligned} v_2 A_2 &= v_1 A_1 \\ v_2 N\pi r_{\text{cap}}^2 &= v_1 \pi r_{\text{aorta}}^2 \end{aligned}$$

so

$$N = \frac{v_1}{v_2}\frac{r_{\text{aorta}}^2}{r_{\text{cap}}^2} = \left(\frac{0.40\ \text{m/s}}{5 \times 10^{-4}\ \text{m/s}}\right)\left(\frac{1.2 \times 10^{-2}\ \text{m}}{4 \times 10^{-6}\ \text{m}}\right)^2 \approx 7 \times 10^9,$$

or on the order of 10 billion capillaries.

EXAMPLE 13–15 **Heating duct to a room.** What area must a heating duct have if air moving 3.0 m/s along it can replenish the air every 15 minutes in a room of volume 300 m^3? Assume the air's density remains constant.

APPROACH We apply the equation of continuity at constant density, Eq. 13–7b, to the air that flows through the duct (point 1 in Fig. 13–24) and then into the room (point 2). The volume flow rate in the room equals the volume of the room divided by the 15-min replenishing time.

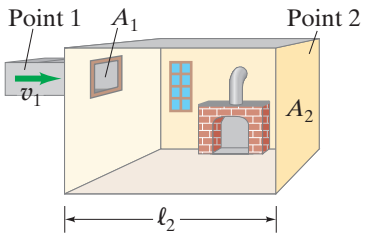
SOLUTION Consider the room as a large section of the duct, Fig. 13–24, and think of air equal to the volume of the room as passing by point 2 in $t = 15\ \text{min} = 900\ \text{s}$. Reasoning in the same way we did to obtain Eq. 13–7a (changing Δt to t), we write $v_2 = \ell_2/t$ so $A_2 v_2 = A_2 \ell_2/t = V_2/t$, where V_2 is the volume of the room. Then the equation of continuity becomes $A_1 v_1 = A_2 v_2 = V_2/t$ and

$$A_1 = \frac{V_2}{v_1 t} = \frac{300\ \text{m}^3}{(3.0\ \text{m/s})\,(900\ \text{s})} = 0.11\ \text{m}^2.$$

NOTE If the duct is square, then each side has length $\ell = \sqrt{A} = 0.33$ m, or 33 cm. A rectangular duct 20 cm × 55 cm will also do.

PHYSICS APPLIED
Heating duct

FIGURE 13–24 Example 13–15.



13–9 Bernoulli's Equation

Have you ever wondered why an airplane can fly, or how a sailboat can move against the wind? These are examples of a principle worked out by Daniel Bernoulli (1700–1782) concerning fluids in motion. In essence, **Bernoulli's principle** states that *where the velocity of a fluid is high, the pressure is low, and where the velocity is low, the pressure is high*. For example, if the pressure in the fluid is measured at points 1

and 2 of Fig. 13–22, it will be found that the pressure is lower at point 2, where the velocity is greater, than it is at point 1, where the velocity is smaller. At first glance, this might seem strange; you might expect that the greater speed at point 2 would imply a higher pressure. But this cannot be the case: if the pressure in the fluid at point 2 were higher than at point 1, this higher pressure would slow the fluid down, whereas in fact it has sped up in going from point 1 to point 2. Thus the pressure at point 2 must be less than at point 1, to be consistent with the fact that the fluid accelerates.

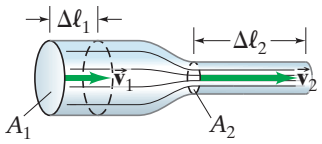


FIGURE 13–22 (Repeated.) Fluid flow through a pipe of varying diameter.

To help clarify any misconceptions, a faster fluid might indeed exert a greater force bouncing off an obstacle placed in its path. But that is not what we mean by the pressure in a fluid. We are examining smooth streamline flow, with no obstacles that interrupt the flow. The fluid pressure is exerted on the walls of a tube or pipe, or on the surface of a material the fluid passes over.

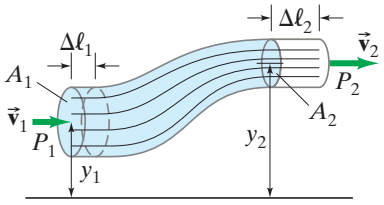
Bernoulli developed an equation that expresses this principle quantitatively. To derive Bernoulli's equation, we assume the flow is steady and laminar, the fluid is incompressible, and the viscosity is small enough to be ignored. To be general, we assume the fluid is flowing in a tube of nonuniform cross section that varies in height above some reference level, Fig. 13–25. We will consider the volume of fluid shown in color and calculate the work done to move it from the position shown in Fig. 13–25a to that shown in Fig. 13–25b. In this process, fluid entering area A_1 flows a distance $\Delta\ell_1$ and forces the fluid at area A_2 to move a distance $\Delta\ell_2$. The fluid to the left of area A_1 exerts a pressure P_1 on our section of fluid and does an amount of work

$$W_1 = F_1\,\Delta\ell_1 = P_1 A_1\,\Delta\ell_1,$$

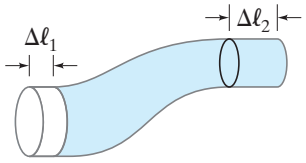
(since $P = F/A$). At point 2 (area A_2), the work done on our section of fluid is

$$W_2 = -P_2 A_2\,\Delta\ell_2.$$

FIGURE 13–25 Fluid flow: for derivation of Bernoulli's equation.



(a)



(b)

The negative sign is present because the force exerted on the fluid is opposite to the displacement. Work is also done on the fluid by the force of gravity. The net effect of the process shown in Fig. 13–25 is to move a mass m of volume $A_1\,\Delta\ell_1$ $(= A_2\,\Delta\ell_2,$ since the fluid is incompressible) from point 1 to point 2, so the work done by gravity is

$$W_3 = -mg(y_2 - y_1),$$

where y_1 and y_2 are heights of the center of the tube above some (arbitrary) reference level. In the case shown in Fig. 13–25, this term is negative since the motion is uphill against the force of gravity. The net work W done on the fluid is thus

$$W = W_1 + W_2 + W_3$$
$$W = P_1 A_1\,\Delta\ell_1 - P_2 A_2\,\Delta\ell_2 - mgy_2 + mgy_1.$$

According to the work-energy principle (Section 7–4), the net work done on a system is equal to its change in kinetic energy. Hence

$$\tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2 = P_1 A_1\,\Delta\ell_1 - P_2 A_2\,\Delta\ell_2 - mgy_2 + mgy_1.$$

The mass m has volume $A_1\,\Delta\ell_1 = A_2\,\Delta\ell_2$ for an incompressible fluid. Thus we can substitute $m = \rho A_1\,\Delta\ell_1 = \rho A_2\,\Delta\ell_2$, and then divide through by $A_1\,\Delta\ell_1 = A_2\,\Delta\ell_2$, to obtain

$$\tfrac{1}{2}\rho v_2^2 - \tfrac{1}{2}\rho v_1^2 = P_1 - P_2 - \rho gy_2 + \rho gy_1,$$

which we rearrange to get

$$P_2 + \tfrac{1}{2}\rho v_2^2 + \rho gy_2 = P_1 + \tfrac{1}{2}\rho v_1^2 + \rho gy_1. \qquad \textbf{(13–8)}$$

Bernoulli's equation

This is **Bernoulli's equation**. Since points 1 and 2 can be any two points along a tube of flow, Bernoulli's equation can be written as

$$P + \tfrac{1}{2}\rho v^2 + \rho gy = \text{constant}$$

at every point in the fluid, where y is the height of the center of the tube above a fixed reference level. [Note that if there is no flow $(v_1 = v_2 = 0)$, then Eq. 13–8 reduces to the hydrostatic equation, Eq. 13–6a: $P_1 - P_2 = \rho g(y_2 - y_1)$.]

Bernoulli's equation is an expression of the law of energy conservation, since we derived it from the work-energy principle.

EXERCISE F As water in a level pipe passes from a narrow cross section of pipe to a wider cross section, how does the pressure against the walls change?

PHYSICS APPLIED
Hot-water heating system

EXAMPLE 13–16 Flow and pressure in a hot-water heating system. Water circulates throughout a house in a hot-water heating system. If the water is pumped at a speed of 0.50 m/s through a 4.0-cm-diameter pipe in the basement under a pressure of 3.0 atm, what will be the flow speed and pressure in a 2.6-cm-diameter pipe on the second floor 5.0 m above? Assume the pipes do not divide into branches.

APPROACH We use the equation of continuity at constant density to determine the flow speed on the second floor, and then Bernoulli's equation to find the pressure.

SOLUTION We take v_2 in the equation of continuity, Eq. 13–7, as the flow speed on the second floor, and v_1 as the flow speed in the basement. Noting that the areas are proportional to the radii squared $(A = \pi r^2)$, we obtain

$$v_2 = \frac{v_1 A_1}{A_2} = \frac{v_1 \pi r_1^2}{\pi r_2^2} = (0.50\,\text{m/s})\,\frac{(0.020\,\text{m})^2}{(0.013\,\text{m})^2} = 1.2\,\text{m/s}.$$

To find the pressure on the second floor, we use Bernoulli's equation (Eq. 13–8):

$$\begin{aligned} P_2 &= P_1 + \rho g(y_1 - y_2) + \tfrac{1}{2}\rho(v_1^2 - v_2^2) \\ &= (3.0\times10^5\,\text{N/m}^2) + (1.0\times10^3\,\text{kg/m}^3)(9.8\,\text{m/s}^2)(-5.0\,\text{m}) \\ &\quad + \tfrac{1}{2}(1.0\times10^3\,\text{kg/m}^3)[(0.50\,\text{m/s})^2 - (1.2\,\text{m/s})^2] \\ &= (3.0\times10^5\,\text{N/m}^2) - (4.9\times10^4\,\text{N/m}^2) - (6.0\times10^2\,\text{N/m}^2) \\ &= 2.5\times10^5\,\text{N/m}^2 = 2.5\,\text{atm}. \end{aligned}$$

NOTE The velocity term contributes very little in this case.

13–10 Applications of Bernoulli's Principle: Torricelli, Airplanes, Baseballs, Blood Flow

Bernoulli's equation can be applied to many situations. One example is to calculate the velocity, v_1, of a liquid flowing out of a spigot at the bottom of a reservoir, Fig. 13–26. We choose point 2 in Eq. 13–8 to be the top surface of the liquid. Assuming the diameter of the reservoir is large compared to that of the spigot, v_2 will be almost zero. Points 1 (the spigot) and 2 (top surface) are open to the atmosphere, so the pressure at both points is equal to atmospheric pressure: $P_1 = P_2$. Then Bernoulli's equation becomes

$$\tfrac{1}{2}\rho v_1^2 + \rho g y_1 = \rho g y_2$$

or

$$v_1 = \sqrt{2g(y_2 - y_1)}. \qquad \textbf{(13–9)}$$

This result is called **Torricelli's theorem**. Although it is seen to be a special case of Bernoulli's equation, it was discovered a century earlier by Evangelista Torricelli (see page 373). Equation 13–9 tells us that the liquid leaves the spigot with the same speed that a freely falling object would attain if falling from the same height. This should not be too surprising since the derivation of Bernoulli's equation relies on the conservation of energy.

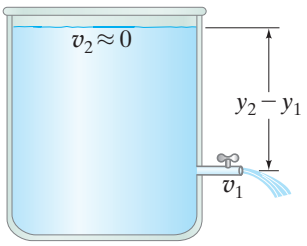


FIGURE 13–26 Torricelli's theorem: $v_1 = \sqrt{2g(y_2 - y_1)}$.

Another special case of Bernoulli's equation arises when a fluid is flowing horizontally with no appreciable change in height; that is, $y_1 = y_2$. Then Eq. 13–8 becomes

$$P_1 + \tfrac{1}{2}\rho v_1^2 = P_2 + \tfrac{1}{2}\rho v_2^2, \qquad \textbf{(13–10)}$$

which tells us quantitatively that the speed is high where the pressure is low, and vice versa. It explains many common phenomena, some of which are illustrated in Figs. 13–27 to 13–33. The pressure in the air blown at high speed across the top of the vertical tube of a perfume atomizer (Fig. 13–27a) is less than the normal air pressure acting on the surface of the liquid in the bowl. Thus atmospheric pressure in the bowl pushes the perfume up the tube because of the lower pressure at the top. A Ping-Pong ball can be made to float above a blowing jet of air (a hair dryer or a vacuum cleaner that can also blow air), Fig. 13–27b; if the ball begins to leave the jet of air, the higher pressure in the still air outside the jet pushes the ball back in.

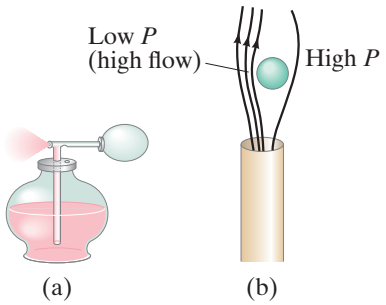


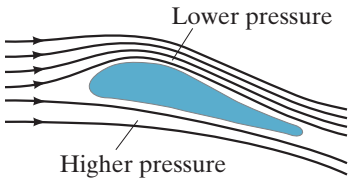
FIGURE 13–27 Examples of Bernoulli's principle: (a) atomizer, (b) Ping-Pong ball in jet of air.

EXERCISE G Return to Chapter-Opening Question 2, page 365, and answer it again now. Try to explain why you may have answered differently the first time. Try it and see.

Airplane Wings and Dynamic Lift

Airplanes experience a "lift" force on their wings, keeping them up in the air, if they are moving at a sufficiently high speed relative to the air and the wing is tilted upward at a small angle (the *attack angle*). See Fig. 13–28, where streamlines of air are shown rushing by the wing (we are in the reference frame of the wing, as if sitting on the wing.) The upward tilt, as well as the rounded upper surface of the wing, causes the streamlines to be forced upward and to be crowded together above the wing. The area for air flowing between any two streamlines is smaller as the streamlines get closer together, so from the equation of continuity $(A_1 v_1 = A_2 v_2)$, the air speed increases above the wing where the streamlines are squished together. (Recall also how the crowded streamlines in a pipe constriction, Fig. 13–22, indicate the velocity is higher in the constriction.) Thus the air speed is greater above the wing than below it, so the pressure above the wing is less than the pressure below the wing (Bernoulli's principle). Hence there is a net upward force on the wing called **dynamic lift**. Experiments show that the speed of air above the wing can even be double the speed of the air below it. (Friction between the air and wing exerts a *drag force*, toward the rear, which must be overcome by the plane's engines.)

FIGURE 13–28 Lift on an airplane wing. We are in the reference frame of the wing, seeing the air flow by.



PHYSICS APPLIED
Airplanes and dynamic lift

A flat wing, or one with symmetric cross section, will experience lift as long as the front of the wing is tilted upward (attack angle). The wing shown in Fig. 13–28 can experience lift even if the attack angle is zero, because the rounded upper surface deflects air up, squeezing the streamlines together. Being both rounded and tilted is better. Airplanes can fly upside down, experiencing lift, if the attack angle is sufficient to deflect streamlines up and closer together.

Our picture considers streamlines; but if the attack angle is larger than about 15°, turbulence sets in (Fig. 13–21b) leading to greater drag and less lift, causing the plane to "stall" and then to drop dangerously.

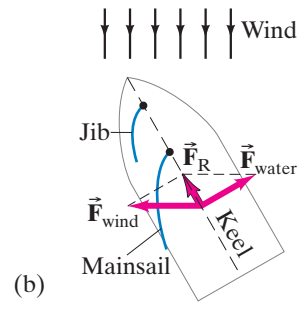
From another point of view, the upward tilt of a wing means the air moving horizontally in front of the wing is deflected downward by the lower surface of the wing; the change in momentum of the rebounding air molecules results in an upward force on the wing (Newton's third law).

Sailboats

A sailboat can move "against" the wind, with the aid of the Bernoulli effect, by setting the sails at an angle, as shown in Fig. 13–29. The air traveling rapidly over the bulging front surface of the mainsail exerts a smaller pressure than the relatively still air behind the sail. The result is a net force on the sail, $\vec{\mathbf{F}}_{\text{wind}}$, as shown in Fig. 13–29b. This force would tend to make the boat move sideways if it weren't for the *keel* that extends vertically downward beneath the water: the water exerts a force $(\vec{\mathbf{F}}_{\text{water}})$ on the keel nearly perpendicular to the keel. The resultant of these two forces $(\vec{\mathbf{F}}_{\text{R}})$ is almost directly forward as shown.

FIGURE 13–29 Sailboat (a) sailing against the wind with (b) analysis.


(a)



(b)

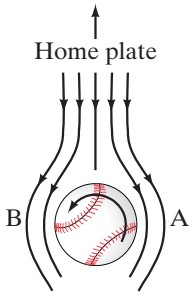
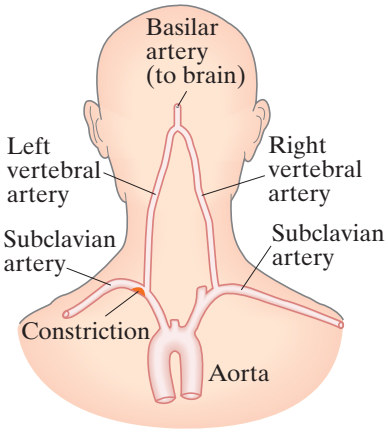


FIGURE 13–30 Looking down on a pitched baseball heading toward home plate. We are in the reference frame of the baseball, with the air flowing by.

FIGURE 13–31 Rear of the head and shoulders showing arteries leading to the brain and to the arms. High blood velocity past the constriction in the left subclavian artery causes low pressure in the left vertebral artery, in which a reverse (downward) blood flow can then occur, resulting in a TIA, a loss of blood to the brain.



Baseball Curve

Why a spinning pitched baseball (soccer ball, tennis ball) curves can also be explained using Bernoulli's principle. It is simplest if we put ourselves in the reference frame of the ball, with the air rushing by, just as we did for the airplane wing. Suppose the ball is rotating counterclockwise as seen from above, Fig. 13–30. A thin layer of air (*boundary layer*) is being dragged around by the ball. We are looking down on the ball, and at point A in Fig. 13–30, this boundary layer tends to slow down the oncoming air. At point B, the air rotating with the ball adds its speed to that of the oncoming air, so the air speed is higher at B than at A. The higher speed at B means the pressure is lower at B than at A, resulting in a net force toward B. The ball's path curves toward the left (as seen by the pitcher).

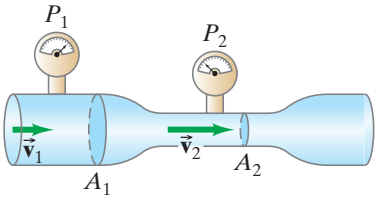
Lack of Blood to the Brain—TIA

In medicine, one of many applications of Bernoulli's principle is to explain a TIA, a *transient ischemic attack* (meaning a temporary lack of blood supply to the brain). A person suffering a TIA may experience symptoms such as dizziness, double vision, headache, and weakness of the limbs. A TIA can occur as follows. Blood normally flows up to the brain at the back of the head via the two vertebral arteries—one going up each side of the neck—which meet to form the basilar artery just below the brain, as shown in Fig. 13–31. Each vertebral artery connects to the subclavian artery, as shown, before the blood passes to the arms. When an arm is exercised vigorously, blood flow increases to meet the needs of the arm's muscles. If the subclavian artery on one side of the body is partially blocked, however, as in arteriosclerosis (hardening of the arteries), the blood velocity will have to be higher on that side to supply the needed blood. (Recall the equation of continuity: smaller area means larger velocity for the same flow rate, Eqs. 13–7.) The increased blood velocity past the opening to the vertebral artery results in lower pressure there (Bernoulli's principle). Thus blood rising in the vertebral artery on the "good" side at normal pressure can be *diverted down* into the other vertebral artery because of the low pressure on that side, instead of passing upward to the brain. Hence the blood supply to the brain is reduced, causing dizziness, or worse.

Other Applications

A **venturi tube** is essentially a pipe with a narrow constriction (the throat). The flowing fluid speeds up as it passes through this constriction, so the pressure is lower in the throat. A **venturi meter**, Fig. 13–32, is used to measure the flow speed of gases and liquids, including blood velocity in arteries. The velocity v_1 can be determined by measuring the pressure P_1 and P_2, the areas A_1 and A_2, as well as the density of the fluid. (The formula is given in Problem 60.)

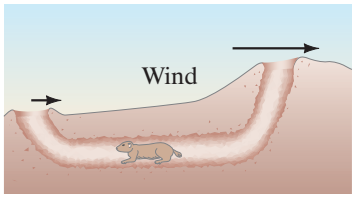
FIGURE 13–32 Venturi meter.



PHYSICS APPLIED
Smoke up a chimney, and underground air circulation

Why does smoke go up a chimney? It's partly because hot air rises (it's less dense and therefore buoyant). But Bernoulli's principle also plays a role. When wind blows across the top of a chimney, the pressure is less there than inside the house. Hence, air and smoke are pushed up the chimney by the higher indoor pressure. Even on an apparently still night there is usually enough ambient air flow at the top of a chimney to assist upward flow of smoke.

FIGURE 13–33 Bernoulli's principle explains air flow in underground burrows.



If gophers, prairie dogs, rabbits, and other animals that live underground are to avoid suffocation, the air must circulate in their burrows. The burrows always have at least two entrances (Fig. 13–33). The speed of air flow across different holes will usually be slightly different. This results in a slight pressure difference, which forces a flow of air through the burrow via Bernoulli's principle. The flow of air is enhanced if one hole is higher than the other (animals often build mounds) since wind speed tends to increase with height.

Bernoulli's equation ignores the effects of friction (viscosity) and the compressibility of the fluid. The energy that is transformed to internal (or potential) energy due to compression and to thermal energy by friction can be taken into account by adding terms to Eq. 13–8. These terms are difficult to calculate theoretically and are normally determined empirically for given situations. They do not significantly alter the explanations for the phenomena described above.

13–11 Viscosity

Real fluids have a certain amount of internal friction called **viscosity**, as mentioned in Section 13–8. Viscosity exists in both liquids and gases, and is essentially a frictional force between adjacent layers of fluid as the layers move past one another. In liquids, viscosity is due to the electrical cohesive forces between the molecules. In gases, it arises from collisions between the molecules.

The viscosity of different fluids can be expressed quantitatively by a *coefficient of viscosity*, η (the Greek lowercase letter eta), which is defined in the following way. A thin layer of fluid is placed between two flat plates. One plate is stationary and the other is made to move, Fig. 13–34. The fluid directly in contact with each plate is held to the surface by the adhesive force between the molecules of the liquid and those of the plate. Thus the upper surface of the fluid moves with the same speed v as the upper plate, whereas the fluid in contact with the stationary plate remains stationary. The stationary layer of fluid retards the flow of the layer just above it, which in turn retards the flow of the next layer, and so on. Thus the velocity varies continuously from 0 to v, as shown. The *increase* in velocity divided by the distance over which this change is made—equal to v/ℓ—is called the **velocity gradient** $(= \Delta v/\ell = (v - 0)/\ell = v/\ell)$. To move the upper plate requires a force, which you can verify by moving a flat plate across a puddle of syrup on a table. For a given fluid, it is found that the force required, F, is proportional to the area of fluid in contact with each plate, A, and to the speed, v, and is inversely proportional to the separation, ℓ, of the plates: $F \propto vA/\ell$. For different fluids, the more viscous the fluid, the greater is the required force. The proportionality constant for this equation is defined as the coefficient of viscosity, η:

$$F = \eta A \frac{v}{\ell}. \qquad \textbf{(13–11)}$$

Solving for η, we find $\eta = F\ell/vA$. The SI unit for η is $\mathrm{N \cdot s/m^2 = Pa \cdot s}$ (pascal · second). In the cgs system, the unit is $\mathrm{dyne \cdot s/cm^2}$, which is called a *poise* (P). Viscosities are often given in centipoise $(1\,\mathrm{cP} = 10^{-2}\,\mathrm{P} = 10^{-3}\,\mathrm{Pa \cdot s})$. Table 13–3 lists the coefficient of viscosity for various fluids. The temperature is also specified, since it has a strong effect; the viscosity of liquids such as motor oil, for example, decreases rapidly as temperature increases.‡

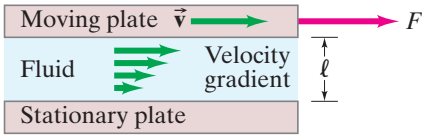


FIGURE 13–34 Determination of viscosity.

TABLE 13–3
Coefficients of Viscosity

Fluid (temperature in °C)	Coefficient of Viscosity, η (Pa · s)†
Water (0°)	1.8×10^{-3}
(20°)	1.0×10^{-3}
(100°)	0.3×10^{-3}
Whole blood (37°)	$\approx 4 \times 10^{-3}$
Blood plasma (37°)	$\approx 1.5 \times 10^{-3}$
Ethyl alcohol (20°)	1.2×10^{-3}
Engine oil (30°) (SAE 10)	200×10^{-3}
Glycerine (20°)	1500×10^{-3}
Air (20°)	0.018×10^{-3}
Hydrogen (0°)	0.009×10^{-3}
Water vapor (100°)	0.013×10^{-3}

†1 Pa · s = 10 poise (P) = 1000 cP.

*13–12 Flow in Tubes: Poiseuille's Equation, Blood Flow

If a fluid had no viscosity, it could flow through a level tube or pipe without a force being applied. Viscosity acts like a sort of friction (between fluid layers moving at slightly different speeds), so a pressure difference between the ends of a level tube is necessary for the steady flow of any real fluid, be it water or oil in a pipe, or blood in the circulatory system of a human.

‡The Society of Automotive Engineers assigns numbers to represent the viscosity of oils: 30-weight (SAE 30) is more viscous than 10-weight. Multigrade oils, such as 20–50, are designed to maintain viscosity as temperature increases; 20–50 means the oil acts like 20-weight when cool and is like 50-weight when it is hot (engine running temperature). In other words, the viscosity does not drop precipitously as the oil warms up, as a simple 20-weight oil would.

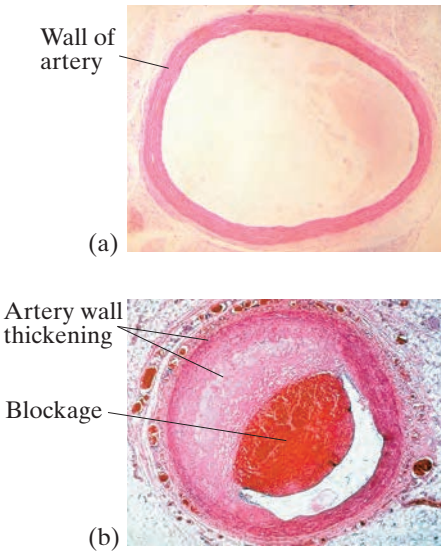


FIGURE 13–35 A cross section of a human artery that (a) is healthy, (b) is partly blocked as a result of arteriosclerosis.

PHYSICS APPLIED

Blood flow and heart disease

The French scientist J. L. Poiseuille (1799–1869), who was interested in the physics of blood circulation (and after whom the "poise" is named), determined how the variables affect the flow rate of an incompressible fluid undergoing laminar flow in a cylindrical tube. His result, known as **Poiseuille's equation**, is:

$$Q = \frac{\pi R^4 (P_1 - P_2)}{8\eta\ell}, \qquad \textbf{(13–12)}$$

where R is the inside radius of the tube, ℓ is the tube length, $P_1 - P_2$ is the pressure difference between the ends, η is the coefficient of viscosity, and Q is the volume rate of flow (volume of fluid flowing past a given point per unit time which in SI has units of m^3/s). Equation 13–12 applies only to laminar (streamline) flow.

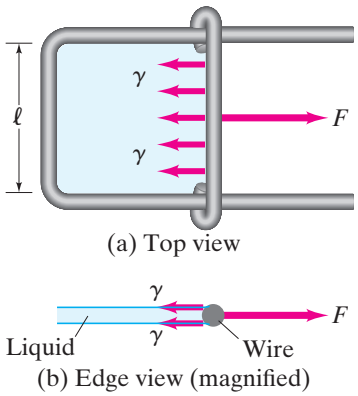
Poiseuille's equation tells us that the flow rate Q is directly proportional to the **pressure gradient**, $(P_1 - P_2)/\ell$, and it is inversely proportional to the viscosity of the fluid. This is just what we might expect. It may be surprising, however, that Q also depends on the *fourth* power of the tube's radius. This means that for the same pressure gradient, if the tube radius is halved, the flow rate is decreased by a factor of 16! Thus the rate of flow, or alternately the pressure required to maintain a given flow rate, is greatly affected by only a small change in tube radius.

An interesting example of this R^4 dependence is *blood flow* in the human body. Poiseuille's equation is valid only for the streamline flow of an incompressible fluid. So it cannot be precisely accurate for blood whose flow is not without turbulence and that contains blood cells (whose diameter is almost equal to that of a capillary). Nonetheless, Poiseuille's equation does give a reasonable first approximation. Because the radius of arteries is reduced as a result of arteriosclerosis (thickening and hardening of artery walls, Fig. 13–35) and by cholesterol buildup, the pressure gradient must be increased to maintain the same flow rate. If the radius is reduced by half, the heart would have to increase the pressure by a factor of about $2^4 = 16$ in order to maintain the same blood-flow rate. The heart must work much harder under these conditions, but usually cannot maintain the original flow rate. Thus, high blood pressure is an indication both that the heart is working harder and that the blood-flow rate is reduced.



FIGURE 13–36 Spherical water droplets, dew on a blade of grass.

FIGURE 13–37 U-shaped wire apparatus holding a film of liquid to measure surface tension $(\gamma = F/2\ell)$.



*13–13 Surface Tension and Capillarity

*Surface Tension

The *surface* of a liquid at rest behaves in an interesting way, almost as if it were a stretched membrane under tension. For example, a drop of water on the end of a dripping faucet, or hanging from a thin branch in the early morning dew (Fig. 13–36), forms into a nearly spherical shape as if it were a tiny balloon filled with water. A steel needle can be made to float on the surface of water even though it is denser than the water. The surface of a liquid acts like it is under tension, and this tension, acting along the surface, arises from the attractive forces between the molecules. This effect is called **surface tension**. More specifically, a quantity called the *surface tension*, γ (the Greek letter gamma), is defined as the force F per unit length ℓ that acts perpendicular to any line or cut in a liquid surface, tending to pull the surface closed:

$$\gamma = \frac{F}{\ell}. \qquad \textbf{(13–13)}$$

To understand this, consider the U-shaped apparatus shown in Fig. 13–37 which encloses a thin film of liquid (such as a liquid soap film). Because of surface tension, a force F is required to pull the movable wire and thus increase the surface area of the liquid. The liquid contained by the wire apparatus is a thin film having both a top and a bottom surface. Hence the total length of the surface being increased is 2ℓ, and the surface tension is $\gamma = F/2\ell$. A delicate apparatus of this type can be used to measure the surface tension of various liquids. The surface tension of water is 0.072 N/m at 20°C. Table 13–4 gives the values for several substances. Note that temperature has a considerable effect on the surface tension.

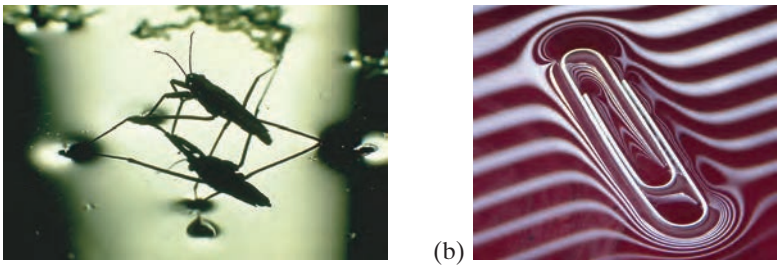


FIGURE 13–38 (a) Water strider. (b) Paper clip (light coming through window blinds).

Because of surface tension, some insects (Fig. 13–38a) can walk on water, and objects more dense than water, such as a paper clip (Fig. 13–38b), can float on the surface. Figure 13–39a shows how the surface tension can support the weight w of an object. Actually, the object sinks slightly into the fluid, so w is the "effective weight" of that object—its true weight less the buoyant force.

EXAMPLE 13–17 ESTIMATE **Insect walks on water.** The base of an insect's leg is approximately spherical in shape, with a radius of about 2.0×10^{-5} m. The 0.0030-g mass of the insect is supported equally by its six legs. Estimate the angle θ at which the surface tension force acts (see Fig. 13–39) for an insect on the surface of water. Assume the water temperature is 20°C.

APPROACH Since the insect is in equilibrium, the upward surface tension force is equal to the pull of gravity downward on each leg. We ignore the buoyant force for this estimate.

SOLUTION For each leg, we assume the surface tension force acts all around a circle of radius r, at an angle θ, as shown in Fig. 13–39a. Only the vertical component, $\gamma \cos\theta$, acts to balance the weight mg. We approximate the length ℓ in Eq. 13–13 as the circumference of the circle, $\ell \approx 2\pi r$. Then the net upward force due to surface tension is $F_y \approx (\gamma \cos\theta)\ell \approx 2\pi r\gamma \cos\theta$. We set this surface tension force equal to one-sixth the weight of the insect since it has six legs:

$$\begin{aligned} 2\pi r\gamma \cos\theta &\approx \tfrac{1}{6}mg \\ (6.28)\,(2.0 \times 10^{-5}\,\text{m})\,(0.072\,\text{N/m})\cos\theta &\approx \tfrac{1}{6}(3.0 \times 10^{-6}\,\text{kg})\,(9.8\,\text{m/s}^2) \\ \cos\theta &\approx 0.54. \end{aligned}$$

So $\theta \approx 57°$.

NOTE If $\cos\theta$ had come out greater than 1, the surface tension would not have been great enough to support the insect's weight. If the insect is very light, it will sink less into the water and θ (Fig. 13–39a) will be larger than calculated above.

NOTE Our estimate ignored the buoyant force and ignored any difference between the radius of the insect's "foot" and the radius of the surface depression.

Soaps and detergents lower the surface tension of water. This is desirable for washing and cleaning since the high surface tension of pure water prevents it from penetrating easily between the fibers of material and into tiny crevices. Substances that reduce the surface tension of a liquid are called *surfactants*.

*Capillarity

Surface tension plays a role in another interesting phenomenon, *capillarity*. It is a common observation that water in a glass container rises up slightly where it touches the glass, Fig. 13–40a. The water is said to "wet" the glass. Mercury, on the other hand, is depressed when it touches the glass, Fig. 13–40b; the mercury does not wet the glass. Whether a liquid wets a solid surface is determined by the relative strength of the cohesive forces between the molecules of the liquid compared to the adhesive forces between the molecules of the liquid and those of the container. **Cohesion** refers to the force between molecules of the same type, whereas **adhesion** refers to the force between molecules of different types.

TABLE 13–4 Surface Tension of Some Substances

Substance	Surface Tension (N/m)
Mercury (20°C)	0.44
Blood, whole (37°C)	0.058
Blood, plasma (37°C)	0.073
Alcohol, ethyl (20°C)	0.023
Water (0°C)	0.076
(20°C)	0.072
(100°C)	0.059
Benzene (20°C)	0.029
Soap solution (20°C)	≈0.025
Oxygen (−193°C)	0.016

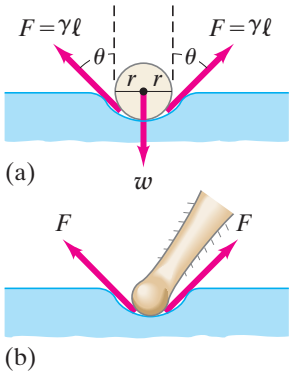
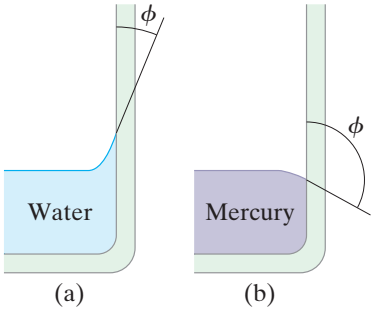


FIGURE 13–39 Surface tension acting on (a) a sphere, and (b) an insect leg. Example 13–17.

PHYSICS APPLIED
Soaps and detergents

FIGURE 13–40 (a) Water "wets" the surface of glass, whereas (b) mercury does not "wet" the glass.



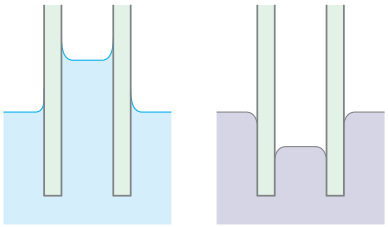


FIGURE 13–41 Capillarity.

Water wets glass because the water molecules are more strongly attracted to the glass molecules than they are to other water molecules. The opposite is true for mercury: the cohesive forces are stronger than the adhesive forces.

In tubes having very small diameters, liquids are observed to rise or fall relative to the level of the surrounding liquid. This phenomenon is called **capillarity**, or **capillary action**, and such thin tubes are called **capillaries**. Whether the liquid rises or falls (Fig. 13–41) depends on the relative strengths of the adhesive and cohesive forces. Thus water rises in a glass tube, whereas mercury falls. The actual amount of rise (or fall) depends on the surface tension—which is what keeps the liquid surface from breaking apart.

13–14 Pumps, and the Heart

We conclude this Chapter with a brief discussion of pumps, including the heart. Pumps can be classified into categories according to their function. A *vacuum pump* is designed to reduce the pressure (usually of air) in a given vessel. The best vacuum attainable in the laboratory today is on the order of 10^{-14} atm.

A *force pump*, on the other hand, is a pump that is intended to increase the pressure—for example, to lift a liquid (such as water from a well) or to push a fluid through a pipe.

Figure 13–42 illustrates the principle behind a simple reciprocating pump. It could be a vacuum pump, in which case the intake is connected to the vessel to be evacuated. A similar mechanism is used in some force pumps, and in this case the fluid is forced under increased pressure through the outlet.

Another type of pump is the centrifugal pump, shown in Fig. 13–43. It, or any force pump, can be used as a *circulating pump*—that is, to circulate a fluid around a closed path, such as the cooling water or lubricating oil in an automobile.

FIGURE 13–42 One kind of pump (reciprocating type): the intake valve opens and air (or fluid that is being pumped) fills the empty space when the piston moves to the left. When the piston moves to the right (not shown), the outlet valve opens and fluid is forced out.

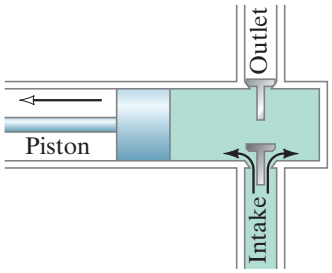
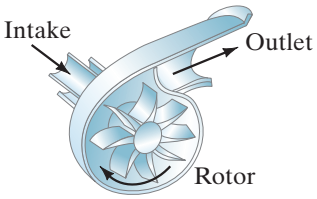


FIGURE 13–43 Centrifugal pump: the rotating blades force fluid through the outlet pipe; this kind of pump is used in vacuum cleaners and as a water pump in automobiles.



PHYSICS APPLIED
Heart as a pump

The heart of a human (and of other animals as well) is essentially a circulating pump. The action of a human heart is shown in Fig. 13–44. There are actually two separate paths for blood flow. The longer path takes blood to the parts of the body, via the arteries, bringing oxygen to body tissues and picking up carbon dioxide,

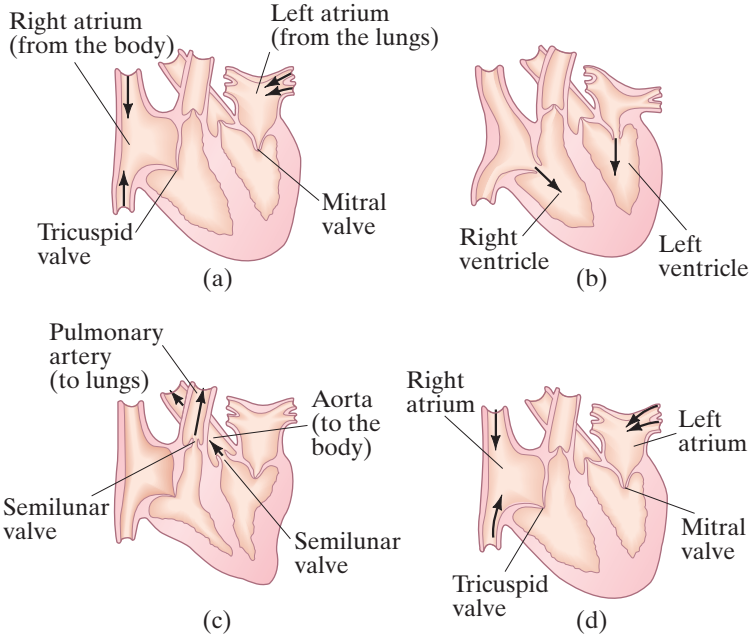


FIGURE 13–44 Pumping human heart. (a) In the diastole phase, the heart relaxes between beats. Blood moves into the heart; both atria fill rapidly. (b) When the atria contract, the systole or pumping phase begins. The contraction pushes the blood through the mitral and tricuspid valves into the ventricles. (c) The contraction of the ventricles forces the blood through the semilunar valves into the pulmonary artery, which leads to the lungs, and to the aorta (the body's largest artery), which leads to the arteries serving all the body. (d) When the heart relaxes, the semilunar valves close; blood fills the atria, beginning the cycle again.

which it carries back to the heart via veins. This blood is then pumped to the lungs (the second path), where the carbon dioxide is released and oxygen is taken up. The oxygen-laden blood is returned to the heart, where it is again pumped to the tissues of the body.

Blood pressure is measured using one of the types of gauge mentioned earlier (Section 13–6), and it is usually calibrated in mm-Hg. The gauge is attached to a closed, air-filled cuff that is wrapped around the upper arm at the level of the heart, Fig. 13–45. Two values of blood pressure are measured: the maximum pressure when the heart is pumping, called *systolic pressure*; and the pressure when the heart is in the resting part of the cycle, called *diastolic pressure*. Initially, the air pressure in the cuff is increased high above the systolic pressure by a pump, compressing the main (brachial) artery in the arm and briefly cutting off the flow of blood. The cuff pressure is then reduced slowly until blood again begins to flow into the arm; it can be detected by listening with a stethoscope to the characteristic tapping sound[†] of the blood returning to the forearm. At this point, systolic pressure is just equal to the air pressure in the arm cuff which can be read off the gauge. The cuff pressure is subsequently reduced further, and the tapping sound disappears when blood at low pressure can enter the artery. At this point, the gauge indicates the diastolic pressure. Normal systolic pressure is around 120 mm-Hg, whereas normal diastolic pressure is around 70 or 80 mm-Hg. Blood pressure is reported in the form 120/70. This is the technique that many doctors still use. Naturally there are digital devices, as readout and also to detect the turbulence or tapping sound.

PHYSICS APPLIED
Blood pressure

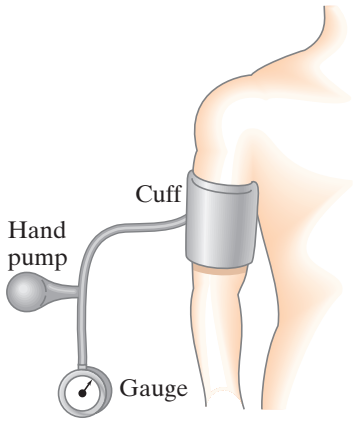


FIGURE 13–45 Device for measuring blood pressure.

[†]When the blood starts flowing through the constriction caused by the tight cuff, its velocity is high and the flow is turbulent. It is the turbulence that causes the tapping sound.

Summary

The three common phases of matter are **solid**, **liquid**, and **gas**. Liquids and gases are collectively called **fluids**, meaning they have the ability to flow. The **density** of a material is defined as its mass per unit volume:

$$\rho = \frac{m}{V}. \qquad \textbf{(13–1)}$$

Specific gravity (**SG**) is the ratio of the density of the material to the density of water (at 4°C).

Pressure is defined as force per unit area:

$$P = \frac{F}{A}. \qquad \textbf{(13–2)}$$

The pressure P at a depth h in a liquid of constant density ρ, due to the weight of the liquid, is given by

$$P = \rho g h, \qquad \textbf{(13–3)}$$

where g is the acceleration due to gravity. If P_0 = atmospheric pressure at the surface of the liquid, then the **absolute pressure** is $P = P_0 + \rho g h$, and $\rho g h$ represents the **gauge pressure**.

If the density of a fluid is not uniform, the pressure P varies with height y as

$$\frac{dP}{dy} = -\rho g. \qquad \textbf{(13–4)}$$

Pascal's principle says that an external pressure applied to a confined fluid is transmitted throughout the fluid.

Pressure is measured using a manometer or other type of gauge. A **barometer** is used to measure atmospheric pressure. **Standard atmospheric pressure** (average at sea level) is $1.013 \times 10^5\ \text{N/m}^2$.

Archimedes' principle states that an object submerged wholly or partially in a fluid is buoyed up by a force equal to the weight of fluid it displaces $(F_B = m_F g = \rho_F V_{\text{displ}} g)$.

Fluid flow can be characterized either as **streamline** (also called **laminar**), in which the layers of fluid move smoothly and regularly along paths called **streamlines**, or as **turbulent**, in which case the flow is not smooth and regular but is characterized by irregularly shaped whirlpools.

Fluid flow rate is the mass or volume of fluid that passes a given point per unit time. The **equation of continuity** states that for an incompressible fluid flowing in an enclosed tube, the product of the velocity of flow and the cross-sectional area of the tube remains constant:

$$Av = \text{constant}. \qquad \textbf{(13–7b)}$$

Bernoulli's principle tells us that where the velocity of a fluid is high, the pressure in it is low, and where the velocity is low, the pressure is high. For steady laminar flow of an incompressible and nonviscous fluid, **Bernoulli's equation**, which is based on the law of conservation of energy, is

$$P_2 + \tfrac{1}{2}\rho v_2^2 + \rho g y_2 = P_1 + \tfrac{1}{2}\rho v_1^2 + \rho g y_1, \qquad \textbf{(13–8)}$$

for two points along the flow.

Viscosity refers to friction within a fluid and is essentially a frictional force between adjacent layers of fluid as they move past one another.

[*Liquid surfaces hold together as if under tension (**surface tension**), allowing drops to form and objects like needles and insects to stay on the surface.]

Questions

1. If one material has a higher density than another, must the molecules of the first be heavier than those of the second? Explain.
2. Airplane travelers sometimes note that their cosmetics bottles and other containers have leaked during a flight. What might cause this?
3. The three containers in Fig. 13–46 are filled with water to the same height and have the same surface area at the base; hence the water pressure, and the total force on the base of each, is the same. Yet the total weight of water is different in each container. Explain this "hydrostatic paradox."



FIGURE 13–46
Question 3.

4. Consider what happens when you push both a pin and the blunt end of a pen against your skin with the same force. Decide what determines whether your skin is cut—the net force applied to it or the pressure. Explain.
5. A small amount of water is boiled in a 1-gallon metal can. The can is removed from the heat and the lid screwed on. As the can cools, it collapses and looks crushed. Explain.
6. When blood pressure is measured, why must the arm cuff be held at the level of the heart?
7. An ice cube floats in a glass of water filled to the brim. What can you say about the density of ice? As the ice melts, will the water overflow? Explain.
8. Will an ice cube float in a glass of alcohol? Why or why not?
9. A submerged can of Coke® will sink, but a can of Diet Coke® will float. (Try it!) Explain.
10. When submerged in water, two objects with different volumes have the same *apparent* weight. When taken out of water, compare their weights in air. Which object has the greater density? Explain.
11. Explain how the tube in Fig. 13–47, known as a **siphon**, can transfer liquid from one container to a lower one even though the liquid must flow uphill for part of its journey. (Note that the tube must be filled with liquid to start with.)

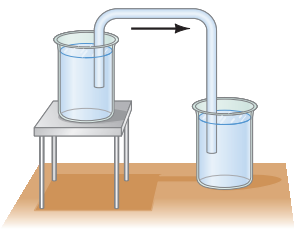


FIGURE 13–47
Question 11. A siphon.

12. A barge filled high with sand approaches a low bridge over the river and cannot quite pass under it. Should sand be added to, or removed from, the barge? [*Hint*: Consider Archimedes' principle.]
13. Explain why helium weather balloons, which are used to measure atmospheric conditions at high altitudes, are normally released from the ground while filled to only 10–20% of their maximum volume.
14. A row boat floats in a swimming pool, and the level of the water at the edge of the pool is marked. Consider the following situations and explain whether the *level of the water* will rise, fall, or stay the same. (*a*) The boat is removed from the water. (*b*) The boat in the water holds an iron anchor which is removed from the boat and placed on the shore. (*c*) The iron anchor is removed from the boat and dropped in the pool.
15. Will an empty balloon have precisely the same apparent weight on a scale as a balloon filled with air? Explain.
16. Why do you float higher in salt water than in fresh water?
17. If you dangle two pieces of paper vertically, a few inches apart (Fig. 13–48), and blow between them, how do you think the papers will move? Try it and see. Explain.



FIGURE 13–48
Question 17.

18. Why does the stream of water from a faucet become narrower as it falls (Fig. 13–49)?



FIGURE 13–49
Question 18. Water coming from a faucet.

19. Children are told to avoid standing too close to a rapidly moving train because they might get sucked under it. Is this possible? Explain.
20. What approximations are made in the derivation of Bernoulli's equation? Qualitatively, how do you think Bernoulli's equation would change if each of these approximations was not made?
21. A tall Styrofoam cup is filled with water. Two holes are punched in the cup near the bottom, and water begins rushing out. If the cup is dropped so it falls freely, will the water continue to flow from the holes? Explain.
22. Why do airplanes normally take off into the wind?
23. Two ships moving in parallel paths close to one another risk colliding. Why?
24. Why does the canvas top of a convertible bulge out when the car is traveling at high speed? [*Hint*: The windshield deflects air upward, pushing streamlines closer together.]
25. Roofs of houses are sometimes "blown" off (or are they pushed off?) during a tornado or hurricane. Explain using Bernoulli's principle.

MisConceptual Questions

1. You hold a piece of wood in one hand and a piece of iron in the other. Both pieces have the same volume, and you hold them fully under water at the same depth. At the moment you let go of them, which one experiences the greater buoyant force?
   (*a*) The piece of wood.
   (*b*) The piece of iron.
   (*c*) They experience the same buoyant force.
   (*d*) More information is needed.

2. Three containers are filled with water to the same height and have the same surface area at the base, but the total weight of water is different for each (Fig. 13–50). In which container does the water exert the greatest force on the bottom of the container?
   (*a*) Container A.
   (*b*) Container B.
   (*c*) Container C.
   (*d*) All three are equal.



FIGURE 13–50
MisConceptual Question 2.

3. A jar is completely filled with water, and then it is sealed with a flexible lid. If you push down on the lid with a force equal to 1000 Pa times the area of the lid, what happens to the water pressure at the bottom of the jar?
   (*a*) It remains unchanged.
   (*b*) It increases by 1000 Pa.
   (*c*) It increases by more than 1000 Pa.
   (*d*) It increases by an amount dependent on the height of the jar.
   (*e*) It increases by less than 1000 Pa.

4. Beaker A is filled to the brim with water. Beaker B is the same size and contains a small block of wood which floats when the beaker is filled with water to the brim. Which beaker weighs more?
   (*a*) Beaker A.
   (*b*) Beaker B.
   (*c*) The same for both.

5. Which of the following objects, submerged in water, experiences the largest magnitude of the buoyant force?
   (*a*) A 1-kg helium balloon.
   (*b*) 1 kg of wood.
   (*c*) 1 kg of ice.
   (*d*) 1 kg of iron.
   (*e*) All the same.

6. Why does an ocean liner float?
   (*a*) It is made of steel, which floats.
   (*b*) Its very big size means it is more buoyant.
   (*c*) It is held up in the water by large styrofoam compartments.
   (*d*) Air: the average density of the ocean liner is less than that of seawater.
   (*e*) Remember the *Titanic*—ocean liners do not float.

7. You put two ice cubes in a glass and fill the glass to the rim with water. As the ice melts, the water level
   (*a*) drops below the rim.
   (*b*) rises and water spills out of the glass.
   (*c*) remains the same.
   (*d*) drops at first, then rises until a little water spills out.

8. A mylar ballon (the kind that doesn't stretch much as more pressure is added) is filled with helium. If the amount of helium in the balloon is now doubled, how will the force needed to hold the balloon down change?
   (*a*) It will stay the same.
   (*b*) It will increase a bit.
   (*c*) It will decrease a bit.

9. Hot air is less dense than cold air. Could a hot-air balloon be flown on the Moon, where there is no atmosphere?
   (*a*) No, there is no air to displace, so no buoyancy force would exist.
   (*b*) Yes, warm air always rises, especially in a weak gravitational field like that of the Moon.
   (*c*) Yes, but the balloon would have to be filled with helium instead of hot air.

10. An object that can float in both water and in oil (whose density is less than that of water) experiences a buoyant force that is
    (*a*) greater when it is floating in oil than when floating in water.
    (*b*) greater when it is floating in water than when floating in oil.
    (*c*) the same when it is floating in water or in oil.

11. A piece of wood is floating in a tub of water. A second piece of wood rests on top of the first piece; it does not touch the water. If the second piece is removed from the first and placed in the water, what happens to the water level in the tub?
    (*a*) The water level cannot be determined without additional information.
    (*b*) The water level drops.
    (*c*) The water level rises.
    (*d*) The water level remains unchanged.

12. Water flows in a horizontal pipe that is narrow but then widens and the speed of the water becomes less. The pressure in the water moving in the pipe is
    (*a*) greater in the wide part.
    (*b*) greater in the narrow part.
    (*c*) the same in both parts.
    (*d*) greater where the speed is higher.
    (*e*) greater where the speed is lower.

13. When a baseball curves to the right (a curveball), air is flowing
    (*a*) faster over the left side than over the right side.
    (*b*) faster over the right side than over the left side.
    (*c*) faster over the top than underneath.
    (*d*) at the same speed all around the baseball, but the ball curves as a result of the way the wind is blowing on the field.

14. How is the smoke drawn up a chimney affected when a wind is blowing outside?
    (*a*) Smoke rises more rapidly in the chimney.
    (*b*) Smoke rises more slowly in the chimney.
    (*c*) Smoke is forced back down the chimney.
    (*d*) Smoke is unaffected.

Problems

13–2 Density and Specific Gravity

1. (I) The approximate volume of the granite monolith known as El Capitan in Yosemite National Park (Fig. 13–51) is about $10^8\ m^3$. What is its approximate mass?



FIGURE 13–51 Problem 1.

2. (I) What is the approximate mass of air in a living room $5.6\ m \times 4.6\ m \times 2.4\ m$?

3. (I) If you tried to smuggle gold bricks by filling your backpack, whose dimensions are $54\ cm \times 31\ cm \times 22\ cm$, what would its mass be?

4. (II) If 4.0 L of antifreeze solution (specific gravity = 0.80) is added to 5.0 L of water to make a 9.0-L mixture, what is the specific gravity of the mixture?

5. (II) A bottle has a mass of 32.00 g when empty and 98.44 g when filled with water. When filled with another fluid, the mass is 89.22 g. What is the specific gravity of this other fluid?

6. (III) The Earth is not a uniform sphere, but has regions of varying density. Consider a simple model of the Earth divided into three regions—inner core, outer core, and mantle. Let us assume each region has a constant density (the average density of that region in the real Earth):

Region	Radius (km)	Density (kg/m^3)
Inner Core	0–1220	13,000
Outer Core	1220–3480	11,100
Mantle	3480–6380	4400

(*a*) Use this model to predict the average density of the entire Earth. (*b*) If the radius of the Earth is 6380 km and its mass is 5.98×10^{24} kg, determine the actual average density of the Earth and compare it (as a percent difference) with the one you determined in (*a*).

13–3 to 13–6 Pressure; Pascal's Principle

7. (I) Estimate the pressure exerted on a floor by (*a*) one pointed heel of area $= 0.45\ cm^2$, and (*b*) one wide heel of area $18\ cm^2$, Fig. 13–52. The person wearing the shoes has a mass of 56 kg.



FIGURE 13–52 Problem 7.

8. (I) Estimate the pressure needed to raise a column of water to the same height as a 46-m-tall pine tree.

9. (I) (*a*) Calculate the total force of the atmosphere acting on the top of a table that measures $1.7\ m \times 2.6\ m$. (*b*) What is the total force acting upward on the underside of the table?

10. (I) What is the difference in blood pressure (mm-Hg) between the top of the head and bottom of the feet of a 1.75-m-tall person standing vertically?

11. (II) How high would the level be in an alcohol barometer at normal atmospheric pressure?

12. (II) In a movie, Tarzan evades his captors by hiding underwater for many minutes while breathing through a long, thin reed. Assuming the maximum pressure difference his lungs can manage and still breathe is −88 mm-Hg, calculate the deepest he could have been.

13. (II) The recommended tire pressure for a particular car is 220 kPa. If the mass of the car is 1250 kg, what area of each tire must be in contact with the road? Assume the car's weight is evenly distributed among all 4 tires.

14. (II) The gauge pressure in each of the four tires of an automobile is 240 kPa. If each tire has a "footprint" of $190\ cm^2$ (area touching the ground), estimate the mass of the car.

15. (II) (*a*) Determine the total force and the absolute pressure on the bottom of a swimming pool 24.0 m by 8.5 m whose uniform depth is 1.8 m. (*b*) What will be the absolute pressure against the *side* of the pool near the bottom?

16. (II) A house at the bottom of a hill is fed by a full tank of water 6.0 m deep and connected to the house by a pipe that is 75 m long at an angle of 61° from the horizontal (Fig. 13–53). (*a*) Determine the water gauge pressure at the house. (*b*) How high could the water shoot if it came vertically out of a broken pipe in front of the house?

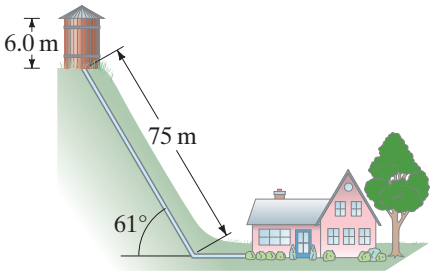


FIGURE 13–53 Problem 16.

17. (II) Water and then oil (which don't mix) are poured into a U-shaped tube, open at both ends. They come to equilibrium as shown in Fig. 13–54. What is the density of the oil? [*Hint*: Pressures at points a and b are equal. Why?]

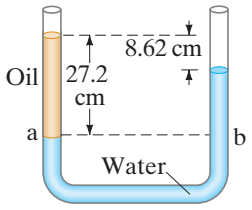


FIGURE 13–54 Problem 17.

18. (II) While exploring a sunken ocean liner, a small robot submarine measured the absolute pressure at the level of the ship to be 413 atmospheres. (*a*) Calculate the depth *D* of the sunken ocean liner. (*b*) Calculate the magnitude of the force due to the water on a viewing port of the submarine at this depth if the viewing port has a surface area of $0.0100\ m^2$.

19. (II) Determine the minimum gauge pressure needed in the water pipe leading into a building if water is to come out of a faucet on the fourteenth floor, 44 m above that pipe.

20. (II) A **hydraulic press** for compacting powdered samples has a large cylinder which is 10.0 cm in diameter, and a small cylinder with a diameter of 2.0 cm (Fig. 13–55). A lever is attached to the small cylinder as shown. The sample, which is placed on the large cylinder, has an area of 4.0 cm^2. What is the pressure on the sample if 320 N is applied to the lever?

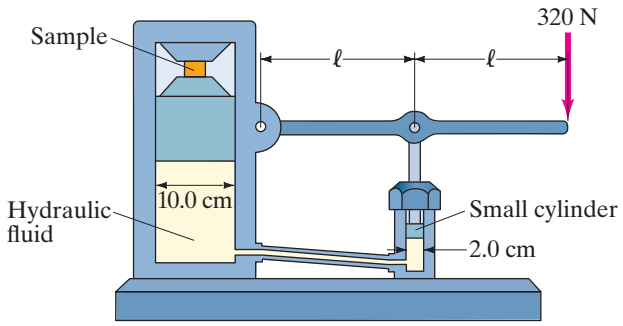


FIGURE 13–55 Problem 20.

21. (II) An open-tube mercury manometer is used to measure the pressure in an oxygen tank. When the atmospheric pressure is 1040 mbar, what is the absolute pressure (in Pa) in the tank if the height of the mercury in the open tube is (*a*) 18.5 cm higher, (*b*) 7.6 cm lower, than the mercury in the tube connected to the tank? See Fig. 13–10a.

22. (III) A beaker of liquid accelerates from rest, on a horizontal surface, with acceleration a to the right. (*a*) Show that the surface of the liquid makes an angle $\theta = \tan^{-1}(a/g)$ with the horizontal. (*b*) Which edge of the water surface is higher? (*c*) How does the pressure vary with depth below the surface?

23. (III) Water stands at a height h behind a vertical dam of uniform width b. (*a*) Use integration to show that the total force of the water on the dam is $F = \frac{1}{2}\rho g h^2 b$. (*b*) Show that the torque about the base of the dam due to this force can be considered to act with a lever arm equal to $h/3$. (*c*) A concrete dam of uniform thickness t and height h is unfortunately "freestanding"—that is, not anchored. What minimum thickness is needed to prevent overturning? Do you need to add in atmospheric pressure for this last part? Explain.

24. (III) Estimate the density of the water 5.4 km deep in the sea. (See Table 12–1 and Section 12–5 regarding bulk modulus.) By what fraction does it differ from the density at the surface?

25. (III) A cylindrical bucket of liquid (density ρ) is rotated about its symmetry axis, which is vertical. If the angular velocity is ω, show that the pressure at a distance r from the rotation axis is

$$P = P_0 + \tfrac{1}{2}\rho\omega^2 r^2,$$

where P_0 is the pressure at $r = 0$.

13–7 Buoyancy and Archimedes' Principle

26. (I) What fraction of a piece of iron will be submerged when it floats in mercury?

27. (II) A geologist finds that a Moon rock whose mass is 8.94 kg has an apparent mass of 6.18 kg when submerged in water. What is the density of the rock?

28. (II) A crane lifts the 21,000-kg steel hull of a sunken ship out of the water. Determine (*a*) the tension in the crane's cable when the hull is fully submerged in the water, and (*b*) the tension when the hull is completely out of the water.

29. (II) A spherical balloon has a radius of 7.15 m and is filled with helium. How large a cargo can it lift, assuming that the skin and structure of the balloon have a mass of 890 kg? Neglect the buoyant force on the cargo volume itself.

30. (II) A partially filled cylindrical container with an interior diameter of 52 mm is floating upright in water. The container is removed from the water and some of the contents removed. When the container is replaced in the water, it floats 28 mm higher than it did before. How much mass was removed from the container?

31. (II) The specific gravity of ice is 0.917, whereas that of seawater is 1.025. What percent of an iceberg is above the surface of the water?

32. (II) Calculate the true mass (in vacuum) of an aluminum sphere whose apparent mass is 4.0000 kg when weighed in air.

33. (II) A 32-kg child decides to make a raft out of empty 1.0-L water bottles and duct tape. Neglecting the mass of the duct tape and plastic in the bottles, what minimum number of water bottles will the child need to be able to stay dry on the raft?

34. (II) Archimedes' principle can be used not only to determine the specific gravity of a solid using a known liquid (Example 13–10). The reverse can be done as well. (*a*) As an example, a 3.80-kg aluminum ball has an apparent mass of 2.10 kg when submerged in a particular liquid: calculate the density of the liquid. (*b*) Determine a formula for finding the density of a liquid using this procedure.

35. (II) An undersea research chamber is spherical with an external diameter of 5.20 m. The mass of the chamber, when occupied, is 74,400 kg. It is anchored to the sea bottom by a cable. What is (*a*) the buoyant force on the chamber, and (*b*) the tension in the cable?

36. (II) A 0.48-kg piece of wood floats in water but is found to sink in alcohol (SG = 0.79), in which it has an apparent mass of 0.047 kg. What is the SG of the wood?

37. (II) A simple model used to determine percent body fat in a human body assumes that a fraction f (< 1) of the body's total mass m is composed of fat with a density of 0.90 g/cm^3, and that the remaining mass of the body is composed of fat-free tissue and bone with a density of 1.10 g/cm^3. If the specific gravity of the entire body's density is measured to be X, show that the percent body fat ($= f \times 100$) is given by

$$\%\ \text{Body fat} = \frac{495}{X} - 450.$$

38. (II) On dry land, an athlete weighs 70.2 kg. The same athlete, when submerged in a swimming pool and hanging from a scale, has an "apparent weight" of 3.4 kg. Using Example 13–10 as a guide, (*a*) find the total volume V of the submerged athlete. (*b*) Assume that when submerged, the athlete's body contains a residual volume $V_R = 1.3 \times 10^{-3}\ m^3$ of air (mainly in the lungs). Taking $V - V_R$ to be the actual volume of the athlete's body, find the body's specific gravity, SG. (*c*) What is the athlete's percent body fat assuming it is given by the formula (495/SG) − 450?

39. (II) How many helium-filled balloons would it take to lift a person? Assume the person has a mass of 72 kg and that each helium-filled balloon is spherical with a diameter of 36 cm.

40. (II) A cube of side length 10.0 cm and made of unknown material floats at the surface between water and oil. The oil has a density of 810 kg/m^3. If the cube floats so that it is 72% in the water and 28% in the oil, what is the mass of the cube and what is the buoyant force on the cube?

41. (II) A scuba tank, when fully submerged, displaces 15.7 L of seawater. The tank itself has a mass of 14.0 kg and, when "full," contains 3.00 kg of air. Assuming only its weight and the buoyant force act on the tank, determine the net force (magnitude and direction) on the fully submerged tank at the beginning of a dive (when it is full of air) and at the end of a dive (when it no longer contains any air).

42. (III) If an object floats in water, its density can be determined by tying a sinker to it so that both the object and the sinker are submerged. Show that the specific gravity is given by $w/(w_1 - w_2)$, where w is the weight of the object alone in air, w_1 is the apparent weight when a sinker is tied to it and the sinker only is submerged, and w_2 is the apparent weight when both the object and the sinker are submerged.

43. (III) A 3.75-kg block of wood (SG = 0.50) floats on water. What minimum mass of lead, hung from the wood by a string, will cause the block to sink?

13–8 to 13–10 Fluid Flow, Bernoulli's Equation

44. (I) Calculate the average speed of blood flow in the major arteries of the body, which have a total cross-sectional area of about 2.0 cm^2. Use the data of Example 13–14.

45. (I) A 12-cm-radius air duct is used to replenish the air of a room 8.2 m × 4.5 m × 3.5 m every 12 min. How fast does the air flow in the duct?

46. (I) How fast does water flow from a hole at the bottom of a very wide, 5.1-m-deep storage tank filled with water? Ignore viscosity.

47. (II) What is the volume rate of flow of water from a 1.85-cm-diameter faucet if the pressure head is 12.0 m?

48. (II) A fish tank has dimensions 36 cm wide by 1.0 m long by 0.60 m high. If the filter should process all the water in the tank once every 2.5 h, what should the flow speed be in the 3.0-cm-diameter input tube for the filter?

49. (II) A 180-km/h wind blowing over the flat roof of a house causes the roof to lift off the house. If the house is 6.2 m × 12.4 m in size, estimate the weight of the roof. Assume the roof is not nailed down.

50. (II) A $\frac{5}{8}$-in. (inside) diameter garden hose is used to fill a round swimming pool 6.1 m in diameter. How long will it take to fill the pool to a depth of 1.4 m if water flows from the hose at a speed of 0.40 m/s?

51. (II) Estimate the air pressure inside a category 5 hurricane, where the wind speed is 300 km/h (Fig. 13–56).



FIGURE 13–56
Problem 51.

52. (II) A 6.0-cm-diameter horizontal pipe gradually narrows to 4.5 cm. When water flows through this pipe at a certain rate, the gauge pressure in these two sections is 36.5 kPa and 22.6 kPa, respectively. What is the volume rate of flow?

53. (II) Show that the power needed to drive a fluid through a pipe with uniform cross-section is equal to the volume rate of flow, Q, times the pressure difference, $P_1 - P_2$. Ignore viscosity.

54. (II) What is the lift (in newtons) due to Bernoulli's principle on a wing of area 82 m^2 if the air passes over the top and bottom surfaces at speeds of 280 m/s and 150 m/s, respectively?

55. (II) Water at a gauge pressure of 3.8 atm at street level flows into an office building at a speed of 0.78 m/s through a pipe 5.0 cm in diameter. The pipe tapers down to 2.8 cm in diameter by the top floor, 16 m above (Fig. 13–57), where the faucet has been left open. Calculate the flow velocity and the gauge pressure in the pipe on the top floor. Assume no branch pipes and ignore viscosity.

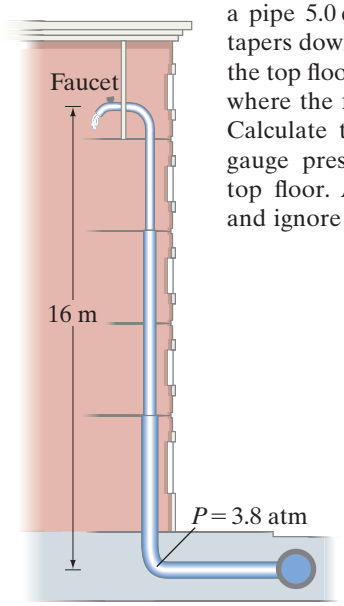


FIGURE 13–57
Problem 55.

56. (II) In Fig. 13–58, take into account the speed of the top surface of the tank and show that the speed of fluid leaving an opening near the bottom is (compare Eq. 13-9)

$$v_1 = \sqrt{\frac{2gh}{(1 - A_1^2/A_2^2)}},$$

where $h = y_2 - y_1$, and A_1 and A_2 are the areas of the opening and of the top surface, respectively. Assume $A_1 \ll A_2$ so that the flow remains nearly steady and laminar.

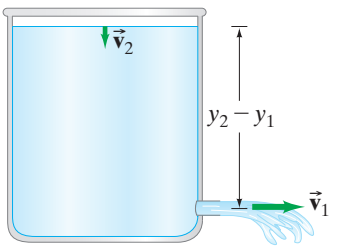


FIGURE 13–58
Problems 56, 57, and 58.

57. (II) Suppose the top surface of the vessel in Fig. 13–58 is subjected to an external pressure P_2 in addition to atmospheric pressure P_0. (*a*) Derive a formula for the speed, v_1, at which the liquid flows from the opening at the bottom into atmospheric pressure, P_0. Assume the velocity of the liquid surface, v_2, is approximately zero. (*b*) If $P_2 = 0.85$ atm and $y_2 - y_1 = 2.4$ m, determine v_1 for water.

58. (III) (*a*) In Fig. 13–58, show that Bernoulli's principle predicts that the level of the liquid, $h = y_2 - y_1$, drops at a rate

$$\frac{dh}{dt} = -\sqrt{\frac{2ghA_1^2}{A_2^2 - A_1^2}},$$

where A_1 and A_2 are the areas of the opening and the top surface, respectively, assuming $A_1 \ll A_2$, and viscosity is ignored. (*b*) Determine h as a function of time by integrating. Let $h = h_0$ at $t = 0$. (*c*) How long would it take to empty a 10.6-cm-tall cylinder filled with 1.3 L of water if the opening is at the bottom and has a 0.50-cm diameter? Do not assume $v_2 = 0$.

59. (III) *Thrust of a rocket.* (*a*) Use Bernoulli's equation and the equation of continuity to show that the emission speed of the propelling gases of a rocket is

$$v = \sqrt{2(P - P_0)/\rho},$$

where ρ is the density of the gas, P is the pressure of the gas inside the rocket, and P_0 is atmospheric pressure just outside the exit orifice. Assume that the gas density stays approximately constant, and that the area of the exit orifice, A_0, is much smaller than the cross-sectional area, A, of the inside of the rocket (take it to be a large cylinder). Assume also that the gas speed is not so high that significant turbulence or nonsteady flow sets in. (*b*) Show that the thrust force on the rocket due to the emitted gases is

$$F = 2A_0(P - P_0).$$

60. (III) (*a*) Show that the flow speed measured by a venturi meter (see Fig. 13–32) is given by the relation

$$v_1 = A_2\sqrt{\frac{2(P_1 - P_2)}{\rho(A_1^2 - A_2^2)}}.$$

(*b*) A venturi meter is measuring the flow of water; it has a main diameter of 3.5 cm tapering down to a throat diameter of 1.0 cm. If the pressure difference is measured to be 18 mm-Hg, what is the speed of the water entering the venturi throat?

61. (III) A fire hose exerts a force on the person holding it due to the water accelerating as it goes from the thicker hose out through the narrow nozzle. How much force is required to hold a 7.0-cm-diameter hose delivering 480 L/min through a 0.75-cm-diameter nozzle?

13–11 Viscosity

62. (II) A viscometer consists of two concentric cylinders, 10.20 cm and 10.60 cm in diameter. A liquid fills the space between them to a depth of 15.0 cm. The outer cylinder is fixed, and a torque of $0.024\ \text{m} \cdot \text{N}$ keeps the inner cylinder turning at a steady rotational speed of 57 rev/min. What is the viscosity of the liquid?

63. (III) A long vertical hollow tube with an inner diameter of 1.00 cm is filled with SAE 10 motor oil. A 0.900-cm-diameter, 30.0-cm-long 150-g rod is dropped vertically through the oil in the tube. What is the maximum speed attained by the rod as it falls?

*13–12 Flow in Tubes; Poiseuille's Equation

***64.** (I) A gardener feels it is taking too long to water a garden with a $\frac{3}{8}$-in.-diameter hose. By what factor will the time be cut using a $\frac{5}{8}$-in.-diameter hose instead? Assume nothing else is changed.

***65.** (I) Engine oil (assume SAE 10, Table 13–3) passes through a fine 1.80-mm-diameter tube that is 10.2 cm long. What pressure difference is needed to maintain a flow rate of 5.8 mL/min?

***66.** (II) Calculate the pressure drop per cm along the aorta using the data of Example 13–14 and Table 13–3.

***67.** (II) What diameter must an 18.0-m-long air duct have if the ventilation and heating system is to replenish the air in a room $8.0\ \text{m} \times 14.0\ \text{m} \times 4.0\ \text{m}$ every 15.0 min? Assume the pump can exert a gauge pressure of 0.710×10^{-3} atm.

***68.** (II) What must be the pressure difference between the two ends of a 1.6-km section of pipe, 29 cm in diameter, if it is to transport oil $(\rho = 950\ \text{kg/m}^3,\ \eta = 0.20\ \text{Pa} \cdot \text{s})$ at a rate of $650\ \text{cm}^3/\text{s}$?

***69.** (II) Poiseuille's equation does not hold if the flow velocity is high enough that turbulence sets in. The onset of turbulence occurs when the **Reynolds number**, *Re*, exceeds approximately 2000. *Re* is defined as

$$Re = \frac{2\bar{v}r\rho}{\eta},$$

where $\bar{v}$ is the average speed of the fluid, ρ is its density, η is its viscosity, and r is the radius of the tube in which the fluid is flowing. (*a*) Determine if blood flow through the aorta is laminar or turbulent when the average speed of blood in the aorta $(r = 1.2\ \text{cm})$ during the resting part of the heart's cycle is about 35 cm/s. (*b*) During exercise, the blood-flow speed approximately doubles. Calculate the Reynolds number in this case, and determine if the flow is laminar or turbulent. NOTE: There is a second **Reynolds** *number* used for assessing flow around a stationary object in a moving fluid (a large rock in a river), or an object moving in a stationary fluid (submarine moving under water or an object falling in air). It is

$$Re' = vL\rho/\eta$$

where v is the speed of the object relative to the fluid, L is the width of the object, and ρ and η are density and viscosity of the fluid. If Re' is less than 1, the flow around the object is essentially laminar. If $Re' \gtrsim 1$, the flow behind the object becomes turbulent (called the *wake*), as in Fig. 13-21b.

***70.** (II) If cholesterol buildup reduces the diameter of an artery by 25%, by what % will the blood flow rate be reduced, assuming the same pressure difference?

***71.** (III) A patient is to be given a *blood transfusion*. The blood is to flow through a tube from a raised bottle to a needle inserted in the vein (Fig. 13–59). The inside diameter of the 25-mm-long needle is 0.80 mm, and the required flow rate is $2.0\ \text{cm}^3$ of blood per minute. How high h should the bottle be placed above the needle? Obtain ρ and η from the Tables. Assume the blood pressure is 78 torr above atmospheric pressure.

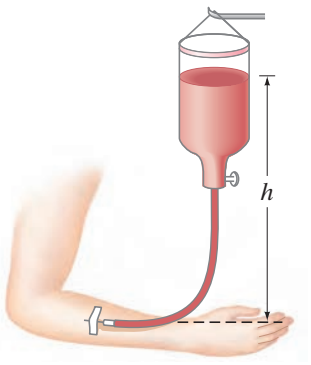


FIGURE 13–59
Problems 71, 79, and 104.

*13–13 Surface Tension and Capillarity

*72. (I) If the force F needed to move the wire in Fig. 13–37 is 3.1×10^{-3} N, calculate the surface tension γ of the enclosed fluid. Assume $\ell = 0.070$ m.

*73. (I) Calculate the force needed to move the wire in Fig. 13–37 if it holds a soapy solution (Table 13–4) and the wire is 21.5 cm long.

*74. (II) If the base of an insect's leg has a radius of about 3.0×10^{-5} m and the insect's mass is 0.016 g, would you expect the six-legged insect to remain on top of the water? Why or why not?

*75. (II) Estimate the diameter of a steel needle that can just barely remain on top of water due to surface tension. (See Figs. 13–38 and 13–39a, and Table 13–1.)

*76. (III) Show that inside a soap bubble, there must be a pressure ΔP in excess of that outside equal to $\Delta P = 4\gamma/r$, where r is the radius of the bubble and γ is the surface tension. [*Hint*: Think of the bubble as two hemispheres in contact with each other, one on the left and one on the right; and remember that there are two surfaces to the bubble, outside and inside. Note that this result applies to any kind of membrane, where 2γ is the tension per unit length in that membrane.]

*77. (III) A common effect of surface tension is the ability of a liquid to rise up a narrow tube due to capillary action. Show that for a narrow tube of radius r placed in a liquid of density ρ and surface tension γ, the liquid in the tube will reach a height $h = 2\gamma/\rho g r$ above the level of the liquid outside the tube, where g is the gravitational acceleration. Assume that the liquid "wets" the tube and that the liquid surface is vertical at the contact with the inside of the tube.

13–14 Pumps; the Heart

78. (II) A physician judges the health of a heart by measuring the pressure with which it pumps blood, as was shown in Fig. 13–45. If the pressurized cuff is attached around a standing person's calf (about 1 m below the heart) instead of the arm (Fig. 13–45), what error (in mm-Hg) would be introduced in the heart's blood pressure measurement?

General Problems

79. Intravenous transfusions are often made under gravity, as shown in Fig. 13–59. Assuming the fluid has a density of 1.00 g/cm^3, at what height h should the bottle be placed so the liquid pressure is (*a*) 52 mm-Hg, and (*b*) 680 mm-H_2O? (*c*) If the blood pressure is 75 mm-Hg above atmospheric pressure, how high should the bottle be placed so that the fluid just barely enters the vein?

80. A 3.2-N force is applied to the plunger of a hypodermic needle. If the diameter of the plunger is 1.3 cm and that of the needle is 0.20 mm, (*a*) with what force does the fluid leave the needle? (*b*) What force on the plunger would be needed to push fluid into a vein where the gauge pressure is 75 mm-Hg? Answer for the instant just before the fluid starts to move.

81. Estimate the difference in air pressure between the top and the bottom of the Empire State Building in New York City. It is 380 m tall and is located at sea level. Express as a fraction of atmospheric pressure at sea level.

82. A bicycle pump is used to inflate a tire. The initial tire (gauge) pressure is 210 kPa (30 psi). At the end of the pumping process, the final pressure is 310 kPa (45 psi). If the diameter of the plunger in the cylinder of the pump is 2.5 cm, what is the range of the force that needs to be applied to the pump handle from beginning to end?

83. A pump supplies water to a 1.59-cm inner diameter hose that tapers down to a 0.953-cm-diameter nozzle. The nozzle is aimed so water comes out at a 45° angle and lands 3.0 m away. The nozzle is 0.60 m above ground level, and the pump output is essentially at ground level. What pressure is supplied by the pump?

84. A hydraulic lift is used to jack a 960-kg car 52 cm off the floor. The diameter of the output piston is 18 cm, and the input force is 380 N. (*a*) What is the area of the input piston? (*b*) What is the work done in lifting the car 52 cm? (*c*) If the input piston moves 13 cm in each stroke, how high does the car move up for each stroke? (*d*) How many strokes are required to jack the car up 52 cm? (*e*) Show that energy is conserved.

85. When you ascend or descend a great deal when driving in a car, your ears "pop," which means that the pressure behind the eardrum is being equalized to that outside. If this did not happen, what would be the approximate force on an eardrum of area 0.20 cm^2 if a change in altitude of 1120 m takes place?

86. Airlines are not allowed to reduce the air pressure within the passenger cabin to less than the equivalent of an altitude of 8000 ft (2400 m) to reduce adverse health effects among passengers due to oxygen deprivation. Estimate this minimum pressure (in atm).

87. Giraffes are a wonder of cardiovascular engineering. Calculate the difference in pressure (in atmospheres) that the blood vessels in a giraffe's head must accommodate as the head is lowered from a full upright position to ground level for a drink. The height of an average giraffe is about 6 m.

88. A copper (Cu) weight is placed on top of a 0.40-kg block of wood (density = 0.60×10^3 kg/m^3) floating in water, as shown in Fig. 13–60. What is the mass of the copper if the top of the wood block is exactly at the water's surface?

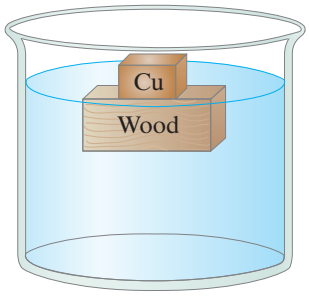


FIGURE 13–60
Problem 88.

89. When a person drives or hikes to a higher altitude, and even more during descent, volume changes of trapped air in the middle ear can cause ear discomfort until the middle-ear pressure and exterior pressure are equalized. (*a*) If a rapid descent at a rate of 7.0 m/s or faster commonly causes ear discomfort, what is the maximum rate of increase in atmospheric pressure (that is, dP/dt) tolerable to most people? (*b*) In a 350-m-tall building, what will be the fastest possible descent time for an elevator traveling from the top floor to the ground floor, assuming the elevator is properly designed to account for human physiology?

90. A tub of water rests on a scale as shown in Fig. 13–61. The weight of the tub plus water is 100 N. A 50-N concrete brick is tied by a cord to a fixed arm and lowered into the water but does not touch the bottom of the tub. What does the scale read now? [*Hint*: Draw two free-body diagrams, one for the brick and a second one for the tub + water + brick.]

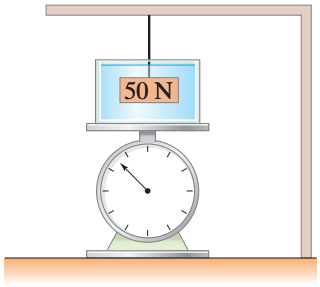


FIGURE 13–61 Problem 90.

91. A simple model (Fig. 13–62) considers a continent as a block $\left(\text{density} \approx 2800\ \text{kg/m}^3\right)$ floating in the mantle rock around it $\left(\text{density} \approx 3300\ \text{kg/m}^3\right)$. Assuming the continent is 35 km thick (the average thickness of the Earth's continental crust), estimate the height of the continent above the surrounding mantle rock.

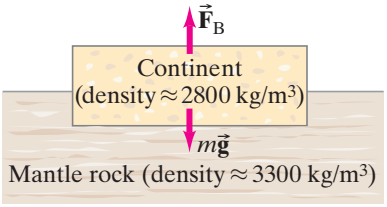


FIGURE 13–62 Problem 91.

92. A ship, carrying fresh water to a desert island in the Caribbean, has a horizontal cross-sectional area of 2240 m^2 at the waterline. When unloaded, the ship rises 8.55 m higher in the sea. How much water $\left(\text{m}^3\right)$ was delivered?

93. You are watering your lawn with a hose when you put your finger over the hose opening to increase the distance the water reaches. If you are holding the hose horizontally, and the distance the water reaches increases by a factor of 4, what fraction of the hose opening did you block?

94. A raft is made of 12 logs lashed together. Each is 45 cm in diameter and has a length of 6.5 m. How many people can the raft hold before they start getting their feet wet, assuming the average person has a mass of 68 kg? Do *not* neglect the weight of the logs. Assume the specific gravity of wood is 0.60.

95. The stream of water from a faucet decreases in diameter as it falls (Fig. 13–63). Derive an equation for the diameter of the stream as a function of the distance y below the faucet, given that the water has speed v_0 when it leaves the faucet, whose diameter is d.



FIGURE 13–63 Problem 95. Water coming from a faucet.

96. In working out his principle, Pascal showed dramatically how force can be multiplied with fluid pressure. He placed a long, thin tube of radius $r = 0.30\ \text{cm}$ vertically into a wine barrel of radius $R = 21\ \text{cm}$, Fig. 13–64. He found that when the barrel was filled with water and the tube filled to a height of 12 m, the barrel burst. Calculate (*a*) the mass of water in the tube, and (*b*) the net force exerted by the water in the barrel on the lid just before rupture.

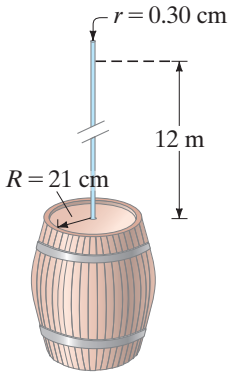


FIGURE 13–64
Problem 96. (Not to scale.)

97. (*a*) Show that the buoyant force F_B on a partially submerged object such as a ship acts at the center of gravity of the fluid before it was displaced, Fig. 13–65. This point is called the **center of buoyancy**. (*b*) To ensure that a ship is in stable equilibrium, would it be better if its center of buoyancy was above, below, or at the same point as its center of gravity? Explain.

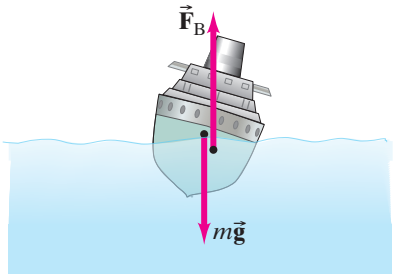


FIGURE 13–65 Problem 97.

98. An aluminum cube, 2.0 cm on a side, and an iron cube, 1.0 cm on a side, are connected by a thin, rigid rod (negligible mass) 55 cm long. The device is immersed in water in two different ways (Fig. 13–66). (*a*) To balance the rod horizontally on your finger underwater, how far from the iron block is your finger? (*b*) You place the rod vertically as shown in Fig. 13–66b, and release it. If the two cubes fall perfectly vertically, estimate their resulting vertical acceleration. (*c*) Describe the likely motion of the two cubes.

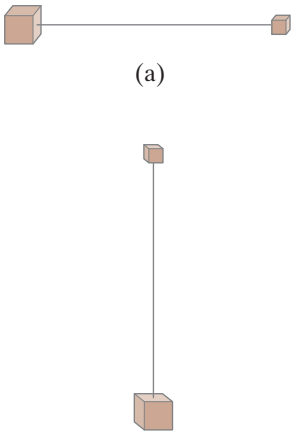


FIGURE 13–66
Problem 98.

99. Estimate the total mass of the Earth's atmosphere, using the known value of atmospheric pressure at sea level.

100. During each heartbeat, approximately 70 cm^3 of blood is pushed from the heart at an average pressure of 105 mm-Hg. Calculate the power output of the heart, in watts, assuming 70 beats per minute.

101. A hurricane-force wind of 180 km/h blows across the face of a storefront window. Estimate the force on the 3.0 m × 3.0 m window due to the difference in air pressure inside and outside the window. Assume the store is airtight so the inside pressure remains at 1.0 atm. (This is why you should not tightly seal a building in preparation for a hurricane.)

102. A drinking fountain shoots water about 12 cm up in the air from a nozzle of diameter 0.60 cm (Fig. 13–67). The pump at the base of the unit (1.1 m below the nozzle) pushes water into a 1.2-cm-diameter supply pipe that goes up to the nozzle. What gauge pressure does the pump have to provide? Ignore the viscosity; your answer will therefore be an underestimate.

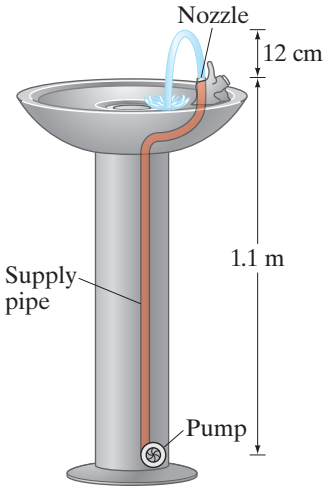


FIGURE 13–67
Problem 102.

103. Four lawn sprinkler heads are fed by a 1.9-cm-diameter pipe. The water comes out of the heads at an angle of 35° to the horizontal and covers a radius of 5.0 m. (*a*) What is the velocity of the water coming out of each sprinkler head? (Assume zero air resistance.) (*b*) If the output diameter of each head is 3.0 mm, how many liters of water do the four heads deliver per minute? (*c*) How fast is the water flowing inside the 1.9-cm-diameter pipe?

***104.** Blood is placed in a bottle 1.40 m above a 3.8-cm-long needle, of inside diameter 0.40 mm, from which it flows at a rate of 4.1 cm^3/min. What is the viscosity of this blood? (See Fig. 13–59 and Problem 71.)

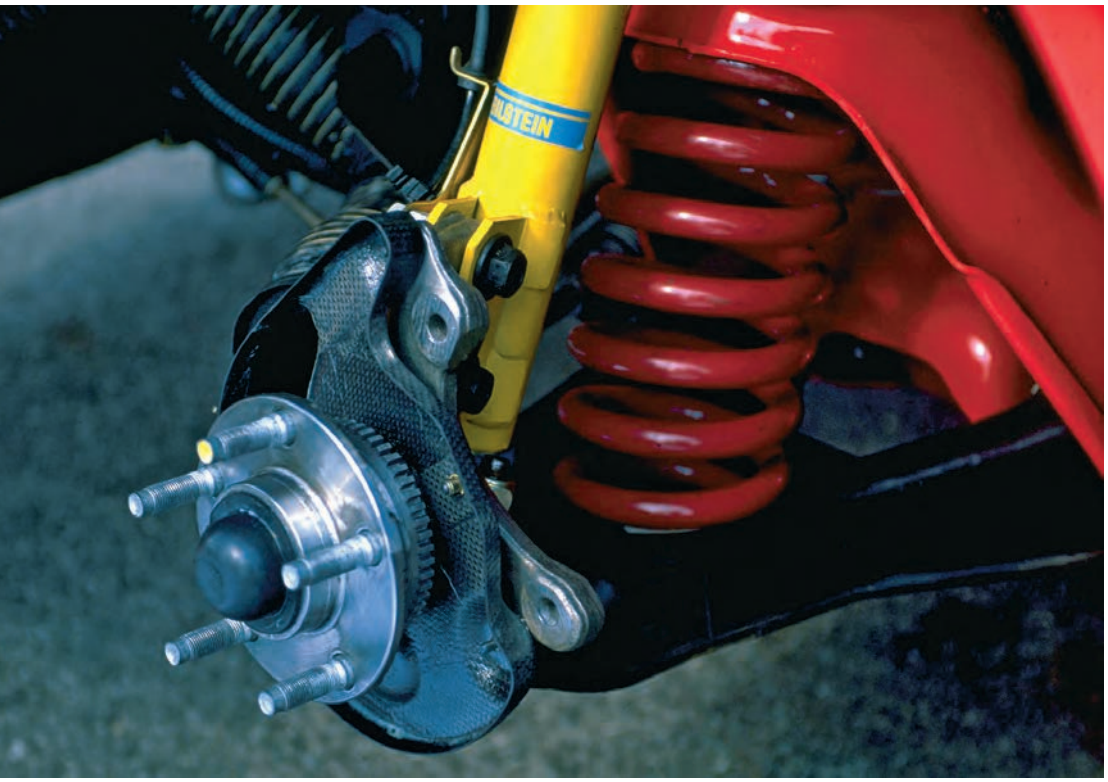
105. Three forces act significantly on a freely floating helium-filled balloon: gravity, air resistance (or drag force), and a buoyant force. Consider a spherical helium-filled balloon of radius $r = 15$ cm rising upward through 0°C air, and $m = 2.8$ g is the mass of the (deflated) balloon itself. For all speeds v, except the very slowest ones, the flow of air past a rising balloon is turbulent, and the drag force F_D is given by the relation

$$F_D = \tfrac{1}{2} C_D \rho_{air} \pi r^2 v^2$$

where the constant $C_D = 0.47$ is the "drag coefficient" for a smooth sphere of radius r. If this balloon is released from rest, it will accelerate very quickly (in a few tenths of a second) to its terminal velocity v_T, where the buoyant force is cancelled by the drag force and the balloon's total weight. Assuming the balloon's acceleration takes place over a negligible time and distance, how long does it take the released balloon to rise a distance $h = 15$ m?

ANSWERS TO EXERCISES

A: (*d*).
B: The same. Pressure depends on depth, not on length.
C: Lower.
D: (*e*).
E: The rowboat is shaped to have a lot of empty, air-filled space, so its "average" density is lower than that of water (unless the boat becomes full of water, in which case it sinks). Steel ships float for the same reason.
F: Increases.
G: (*b*).



An object attached to a coil spring can exhibit oscillatory motion. Many kinds of oscillatory motion are sinusoidal in time, or nearly so, and are referred to as simple harmonic motion. Real systems generally have at least some friction, causing the motion to be damped. The automobile spring shown here (red) has a shock absorber (yellow) that purposefully dampens the oscillation to make for a smooth ride.

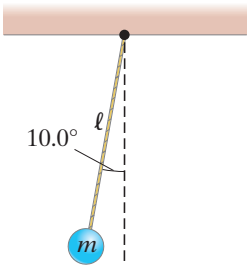
When an external sinusoidal force is exerted on a system able to oscillate, resonance occurs if the driving force is at or near the natural frequency of oscillation of the system.

CHAPTER 14

Oscillations

CHAPTER-OPENING QUESTION—Guess now!
A simple pendulum consists of a mass m (the "bob") hanging on the end of a thin string of length ℓ and negligible mass. The bob is pulled sideways so the string makes a 5.0° angle to the vertical. When released, it oscillates back and forth at a frequency f. If the pendulum is started at a 10.0° angle instead, its frequency would be

(a) twice as great.
(b) half as great.
(c) the same, or very close to it.
(d) not quite twice as great.
(e) a bit more than half as great.



Many objects vibrate or oscillate—an object connected to a spring, a tuning fork, a pendulum, a plastic ruler held firmly over the edge of a table and gently struck, the strings of a guitar or piano. Spiders detect prey by the vibrations of their webs; cars oscillate up and down when they hit a bump; buildings and bridges vibrate when heavy trucks pass or the wind is fierce. Indeed, because most solids are elastic (see Section 12–5), they vibrate (at least briefly) when given an impulse. Electrical oscillations occur in radio and television sets. At the atomic level, atoms oscillate within a molecule, and the atoms of a solid oscillate about their relatively fixed positions. Because it is so common in everyday life and occurs in so many areas of physics, oscillatory (or vibrational) motion is of great importance. Mechanical oscillations or vibrations are fully described on the basis of Newtonian mechanics.

CONTENTS

14–1 Oscillations of a Spring
14–2 Simple Harmonic Motion
14–3 Energy in the Simple Harmonic Oscillator
14–4 Simple Harmonic Motion Related to Uniform Circular Motion
14–5 The Simple Pendulum
*14–6 The Physical Pendulum and the Torsion Pendulum
14–7 Damped Harmonic Motion
14–8 Forced Oscillations; Resonance

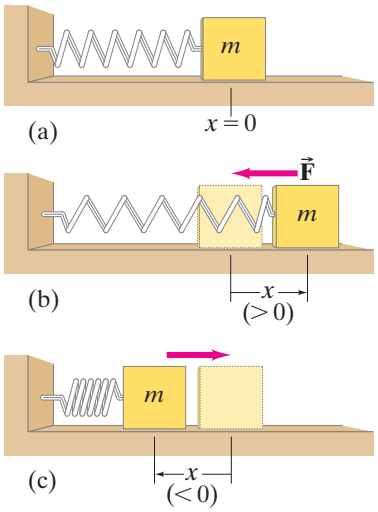
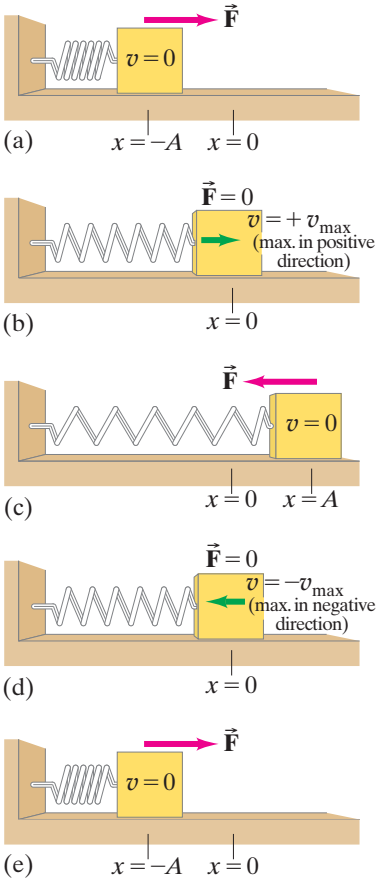


FIGURE 14–1 An object of mass m oscillating at the end of a uniform spring. The force $\vec{\mathbf{F}}$ on the object at the different positions is shown *above* the object. (The pale yellow box represents the object at the equilibrium position, $x = 0$.)

FIGURE 14–2 An object oscillating on a frictionless surface, indicating the force on the object and its velocity at different positions of its oscillation cycle.



14–1 Oscillations of a Spring

When an object **vibrates** or **oscillates** back and forth, over the same path, each oscillation taking the same amount of time, the motion is **periodic**. The simplest form of periodic motion is represented by an object oscillating on the end of a uniform coil spring. Because many other types of oscillatory motion closely resemble this system, we will look at it in detail. We assume that the mass of the spring can be ignored, and that the spring is mounted horizontally, as shown in Fig. 14–1a. We assume the object of mass m slides without friction on the horizontal surface. Any spring has a natural length at which it exerts no force on the mass m. The position of the mass and spring at this point is called the **equilibrium position**. If the mass is moved either to the left, which compresses the spring, or to the right, which stretches it, the spring exerts a force on the mass that acts in the direction of returning the mass to the equilibrium position; hence it is called a *restoring force*. We consider the common situation where we can assume the restoring force F is directly proportional to the displacement x the spring has been stretched (Fig. 14–1b) or compressed (Fig. 14–1c) from the equilibrium position:

$$F = -kx. \qquad \text{[force exerted by spring]} \quad \textbf{(14–1)}$$

Note that the equilibrium position has been chosen at $x = 0$ and the minus sign in Eq. 14–1 indicates that the restoring force is always in the direction opposite to the displacement x. For example, if we choose the positive direction to the right in Fig. 14–1, x is positive when the spring is stretched (Fig. 14–1b), but the direction of the restoring force is to the left (negative direction). If the spring is compressed, x is negative (to the left) but the force F acts toward the right (Fig. 14–1c).

Equation 14–1 is often referred to as Hooke's law (Sections 7–3, 8–2 and 12–5), and is accurate only if the spring is not compressed to where the coils are close to touching, or stretched beyond the elastic region (see Fig. 12–18). Hooke's law works not only for springs but for other oscillating solids as well; it thus has wide applicability, even though it is valid only over a certain range of F and x values.

The proportionality constant k in Eq. 14–1 is called the *force constant*, or more commonly the **spring constant** for that particular spring (or *spring stiffness constant*) with units = N/m. To stretch the spring a displacement x from equilibrium $(x = 0)$, an (external) force must be exerted on the free end of the spring with a magnitude at least equal to

$$F_{\text{ext}} = +kx. \qquad \text{[external force on spring]}$$

The greater the value of k, the greater the force needed to stretch a spring a given amount. That is, the stiffer the spring, the greater the spring constant k.

Note that the force F in Eq. 14–1 is *not* a constant, but varies with position. Therefore the acceleration of the mass m is not constant, so we *cannot* use the equations for constant acceleration developed in Chapter 2 (Eqs. 2–12).

Let us examine what happens when our uniform spring is initially compressed a displacement $x = -A$, as shown in Fig. 14–2a, and then our object of mass m is released on the frictionless surface. The spring exerts a force on the mass that accelerates it toward the equilibrium position. Because the mass has inertia, it passes the equilibrium position with considerable speed. Indeed, as the mass reaches the equilibrium position, the force on it decreases to zero, but its speed at this point is a maximum, v_{max} (Fig. 14–2b). As the mass moves farther to the right, the force on it acts to slow it down, and it stops for an instant at $x = A$ (Fig. 14–2c). It then begins moving back in the opposite direction, accelerating until it passes the equilibrium point (Fig. 14–2d), and then slows down until it reaches zero speed at the original starting point, $x = -A$ (Fig. 14–2e). It then repeats the motion, moving back and forth symmetrically between $x = A$ and $x = -A$.

EXERCISE A A mass is oscillating on a frictionless surface at the end of a horizontal spring. Where, if anywhere, is the acceleration of the mass zero (see Fig. 14–2)? (*a*) At $x = -A$; (*b*) at $x = 0$; (*c*) at $x = +A$; (*d*) at both $x = -A$ and $x = +A$; (*e*) nowhere.

To discuss oscillatory motion, we need to define a few terms. The distance x of the mass from the equilibrium point at any moment is the **displacement** (with a + or − sign). The maximum displacement—the greatest distance from the equilibrium point—is called the **amplitude**, A. One **cycle** refers to the complete to-and-fro motion from some initial point back to that same point—say, from $x = -A$ to $x = +A$ and back to $x = -A$. The **period**, T, is defined as the time required to complete one cycle. Finally, the **frequency**, f, is the number of complete cycles per second. Frequency is generally specified in hertz (Hz), where 1 Hz = 1 cycle per second (s^{-1}). Given their definitions, frequency and period are inversely related, as we saw earlier (Eqs. 5–2 and 10–8):

$$f = \frac{1}{T} \quad \text{and} \quad T = \frac{1}{f}. \tag{14–2}$$

For example, if the frequency is 2 cycles per second, then each cycle takes $\frac{1}{2}$ s.

EXERCISE B If an oscillating mass has a frequency of 1.25 Hz, it makes 100 oscillations in (*a*) 12.5 s, (*b*) 125 s, (*c*) 80 s, (*d*) 8.0 s.

The oscillation of a spring hung vertically is essentially the same as that of a horizontal spring. But because of gravity, the length of a vertical spring with a mass m on the end will be longer at equilibrium than when that same spring is horizontal; see Fig. 14–3. The spring is in equilibrium when $\Sigma F = 0 = mg - kx_0$, so the spring stretches an extra amount $x_0 = mg/k$ to be in equilibrium. If x is measured from this new equilibrium position, which we now call $x = 0$, Eq. 14–1 can be used directly with the same value of k.

CAUTION
For vertical spring, measure displacement (x or y) from the vertical equilibrium position

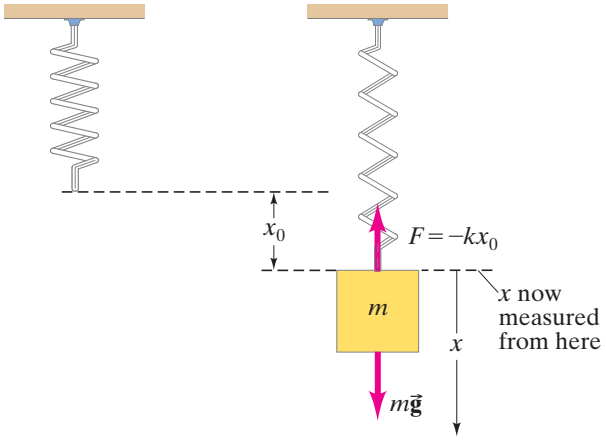


FIGURE 14–3
(a) Free spring, hung vertically.
(b) Mass m attached to spring in new equilibrium position, which occurs when $\Sigma F = 0 = mg - kx_0$.

EXAMPLE 14–1 **Car springs.** When a family of four with a total mass of $m = 200$ kg step into their 1200-kg car, the car's springs compress 3.0 cm (see Chapter-Opening photo, page 399). (*a*) What is the spring constant of the car's springs, assuming they act as a single spring? (*b*) How far will the car lower if loaded with 300 kg rather than 200 kg?

APPROACH We use Hooke's law: the weight of the people, mg, causes a 3.0-cm displacement.

SOLUTION (*a*) The added force of $mg = (200\text{ kg})(9.8\text{ m/s}^2) = 1960$ N causes the springs to compress 3.0×10^{-2} m. Therefore (Eq. 14–1), the spring constant is

$$k = \frac{F}{x} = \frac{1960\text{ N}}{3.0 \times 10^{-2}\text{ m}} = 6.5 \times 10^4\text{ N/m}.$$

(*b*) If the car is loaded with 300 kg, Hooke's law gives

$$x = \frac{F}{k} = \frac{(300\text{ kg})(9.8\text{ m/s}^2)}{(6.5 \times 10^4\text{ N/m})} = 4.5 \times 10^{-2}\text{ m},$$

or 4.5 cm.

NOTE In (*b*), we could have obtained x without solving for k: since x is proportional to F, if 200 kg compresses the spring 3.0 cm, then 1.5 times the force will compress the spring 1.5 times as much, or 4.5 cm.

14–2 Simple Harmonic Motion

Any oscillating system for which the net restoring force is directly proportional to the negative of the displacement (as in Eq. 14–1, $F = -kx$) is said to exhibit **simple harmonic motion** (SHM). Such a system is often called a **simple harmonic oscillator** (SHO). [The word "harmonic" refers to the motion being sinusoidal: see below. It is "simple" when the motion is sinusoidal of a single frequency. This can happen only if friction or other forces are not acting.]

We saw in Section 12–5 that most solid materials stretch or compress according to Eq. 14–1 as long as the displacement is not too great. Because of this, many natural oscillations are simple harmonic or sufficiently close to it that they can be treated using this SHM model.

CONCEPTUAL EXAMPLE 14–2 **Is the motion simple harmonic?** Which of the following forces would cause an object to move in simple harmonic motion? (*a*) $F = -0.5x^2$, (*b*) $F = -2.3y$, (*c*) $F = 8.6x$, (*d*) $F = -4\theta$.

RESPONSE Both (*b*) and (*d*) will give simple harmonic motion because they give the force as minus a constant times a displacement. The displacement need not be x, but the minus sign is required to restore the system to equilibrium, which is why (*c*) does not produce SHM. Nor does (*a*) because x is squared.

Let us now determine the position x as a function of time for a mass attached to the end of a simple spring with spring constant k. To do so, we make use of Newton's second law, $F = ma$. Since the acceleration $a = d^2x/dt^2$ (Chapter 2), we have

$$ma = \Sigma F$$
$$m\frac{d^2x}{dt^2} = -kx,$$

where m is the mass[†] which is oscillating. We rearrange this to obtain

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0, \qquad \text{[SHM]} \quad \textbf{(14–3)}$$

which is known as the **equation of motion** for the simple harmonic oscillator. Mathematically, this is a *second-order differential equation* because it involves second derivatives. We want to determine what function of time, $x(t)$, satisfies this equation. We might guess the form of the solution by noting that if a pen is attached to an oscillating mass (Fig. 14–4) and a sheet of paper is moved at a steady rate beneath it, the pen would trace the curve shown. The shape of this curve looks a lot like it might be **sinusoidal** (such as cosine or sine) as a function of time, and its height is the amplitude A. Let us then guess that the general solution to Eq. 14–3 can be written in a form such as

$$x = A\cos(\omega t + \phi), \qquad \textbf{(14–4)}$$

where ω is a constant to be determined (units rad/s) and we include the constant ϕ in the argument to be general.[‡] Let us now put this trial solution into Eq. 14–3 and see if it really works. We need to differentiate the $x = x(t)$ twice:

$$\frac{dx}{dt} = \frac{d}{dt}[A\cos(\omega t + \phi)] = -\omega A\sin(\omega t + \phi)$$
$$\frac{d^2x}{dt^2} = -\omega^2 A\cos(\omega t + \phi).$$

We next put this equation for d^2x/dt^2 into Eq. 14–3 along with Eq. 14–4 for x

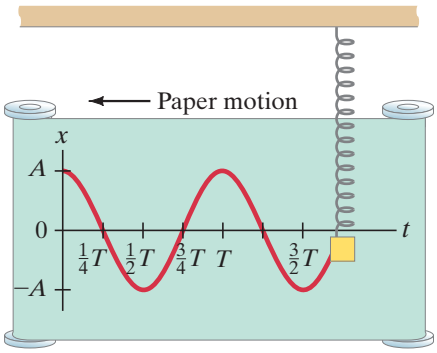


FIGURE 14–4 Sinusoidal nature of SHM: position as a function of time. In this case, $x = A\cos(2\pi t/T)$.

[†]We are assuming an ideal massless spring, but the spring itself also oscillates and a part of its mass should be included. Approximately one-third the mass of the spring, M_s, should be included, so for a mass m oscillating on the end of a spring, the total mass M is $M = m + \frac{1}{3}M_s$ in our equation. Often M_s is small enough to be ignored.

[‡]Another possible way to write the solution is the combination $x = a\cos\omega t + b\sin\omega t$, where a and b are constants. This is equivalent to Eq. 14–4 as can be seen using the trigonometric identity $\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B$.

to see if Eq. 14–4 was a good guess:

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0$$

$$-\omega^2 A\cos(\omega t + \phi) + \frac{k}{m}A\cos(\omega t + \phi) = 0$$

or

$$\left(\frac{k}{m} - \omega^2\right)A\cos(\omega t + \phi) = 0.$$

The left side of this equation must equal zero for *all* times t, because it is zero on the right side. This equation is satisfied for any time t if $(k/m - \omega^2) = 0$. So our solution, Eq. 14–4, does indeed satisfy the equation of motion (Eq. 14–3) for any time t, but only if

$$\omega^2 = \frac{k}{m}. \qquad \textbf{(14–5)}$$

Equation 14–4 is the general solution to Eq. 14–3, and we also know the value of the constant ω, i.e., $\omega = \sqrt{k/m}$. The solution is sinusoidal and it contains two arbitrary constants A and ϕ, which we should expect because the second derivative in Eq. 14–3 implies that two integrations are needed, each yielding a constant. They are "arbitrary" only in a calculus sense, in that they can be anything and still satisfy the differential equation, Eq. 14–3. In real physical situations, however, A and ϕ are determined by the **initial conditions**. Suppose, for example, that the mass is started at its maximum displacement A and is released from rest. This is, in fact, what is shown in Fig. 14–4, and for this case $x = A\cos\omega t$. Let us confirm it: we are given $v = 0$ at $t = 0$, where

$$v = \frac{dx}{dt} = \frac{d}{dt}[A\cos(\omega t + \phi)] = -\omega A\sin(\omega t + \phi) = 0. \qquad [\text{at } t = 0]$$

For v to be zero at $t = 0$, then $\sin(\omega t + \phi) = \sin(0 + \phi)$ can be zero if $\phi = 0$ (ϕ could also be π, 2π, etc.), and when $\phi = 0$, then

$$x = A\cos\omega t,$$

as we expected. We see immediately that A is the amplitude of the motion, and it is determined initially by how far you pull the mass m from equilibrium before releasing it.

Consider another interesting case: at $t = 0$ the mass m is at $x = 0$ and is struck, giving it an initial velocity toward increasing values of x. Given $x = 0$ at $t = 0$, we can write $x = A\cos(\omega t + \phi) = A\cos\phi = 0$, which can happen only if $\phi = \pm\pi/2$ (or $\pm 90°$). Whether $\phi = +\pi/2$ or $-\pi/2$ depends on $v = dx/dt = -\omega A\sin(\omega t + \phi) = -\omega A\sin\phi$ at $t = 0$, which we are given as positive ($v > 0$ at $t = 0$); hence $\phi = -\pi/2$ because $\sin(-90°) = -1$. Thus our solution for this case is

$$x = A\cos\left(\omega t - \frac{\pi}{2}\right)$$
$$= A\sin\omega t,$$

where we used $\cos(\theta - \pi/2) = \sin\theta$ (see Appendix A, Fig. A–5). The solution in this case is a pure sine function, Fig. 14–5, where A is still the amplitude.

Many other situations are possible, such as that shown in Fig. 14–6. The constant ϕ is called the **phase angle**, or **phase constant**, and it tells us how long after (or before) $t = 0$ the peak at $x = A$ is reached. Notice that the value of ϕ does not affect the shape of the $x(t)$ curve, but only affects the displacement at some arbitrary time, $t = 0$. Simple harmonic motion is thus always *sinusoidal*. Simple harmonic motion can in fact be *defined* as motion that is purely sinusoidal.

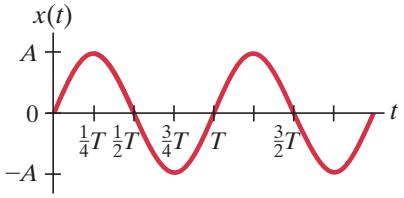
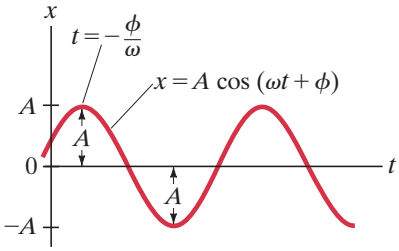


FIGURE 14–5 In this case of SHM, the mass m starts at $t = 0$ at the equilibrium position $x = 0$, and has initial velocity toward positive values of x ($v > 0$ at $t = 0$).

FIGURE 14–6 A plot of $x = A\cos(\omega t + \phi)$ when $\phi < 0$. Note: $A\cos(\omega t + \phi) = A$ only when $\cos\theta = 1$, which occurs (first) when $\theta = 0$, in which case $t = -\phi/\omega$ as shown.



Since our oscillating mass repeats its motion after a time equal to its period T, it must be at the same position and moving in the same direction at $t = T$ as it was at $t = 0$. Since a sine or cosine function repeats itself after every 2π radians, then from Eq. 14–4, we must have

$$\omega T = 2\pi.$$

Hence

$$\omega = \frac{2\pi}{T} = 2\pi f,$$

where f is the frequency of the motion. Strictly speaking, we call ω the **angular frequency** (units are rad/s) to distinguish it from the frequency f (units are s^{-1} = Hz); sometimes the word "angular" is dropped, so the symbol ω or f needs to be specified. Because $\omega = 2\pi f = 2\pi/T$, we can write Eq. 14–4 as

$$x = A\cos\left(\frac{2\pi t}{T} + \phi\right) \qquad \textbf{(14–6a)}$$

or

$$x = A\cos(2\pi f t + \phi). \qquad \textbf{(14–6b)}$$

CAUTION

t is a variable (time); T is a constant for a given situation

Notice in Eq. 14–6a that when $t = T$ (a time equal to one period), we have the cosine of $2\pi + \phi$ (or $360° + \phi$), which is the same as the cosine at $t = 0$. This makes sense because the motion repeats itself after a time $t = T$.

Because $\omega = 2\pi f = \sqrt{k/m}$ (Eq. 14–5), then

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}, \qquad \textbf{(14–7a)}$$

$$T = 2\pi\sqrt{\frac{m}{k}}. \qquad \textbf{(14–7b)}$$

Note that the *frequency and period do not depend on the amplitude*. Changing the amplitude of a simple harmonic oscillator does not affect its frequency. Equation 14–7a tells us that the greater the mass, the lower the frequency; and the stiffer the spring, the higher the frequency. This makes sense since a greater mass means more inertia and therefore a slower response (or acceleration); and larger k means greater force and therefore quicker response. The frequency f (Eq. 14–7a) at which a SHO oscillates naturally is called its **natural frequency** (to distinguish it from a frequency at which it might be forced to oscillate by an outside force, as discussed in Section 14–8).

The simple harmonic oscillator is important in physics because whenever we have a net restoring force proportional to the displacement $(F = -kx)$, which is at least a good approximation for a variety of systems, then the motion is simple harmonic—that is, sinusoidal.

PHYSICS APPLIED

Car springs

EXAMPLE 14–3 **Car springs again.** Determine the period and frequency of the car in Example 14–1a after hitting a bump. Assume the shock absorbers are poor, so the car really oscillates up and down.

APPROACH We put $m = 1400\,\text{kg}$ and $k = 6.5 \times 10^4\,\text{N/m}$ from Example 14–1a into Eqs. 14–7.

SOLUTION From Eq. 14–7b,

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{1400\,\text{kg}}{6.5 \times 10^4\,\text{N/m}}} = 0.92\,\text{s},$$

or slightly less than a second. The frequency $f = 1/T = 1.1\,\text{Hz}$.

EXERCISE C By how much should the mass on the end of a spring be changed to halve the frequency of its oscillations? (*a*) No change; (*b*) doubled; (*c*) quadrupled; (*d*) halved; (*e*) quartered.

EXAMPLE 14–4 **ESTIMATE** **Spider web.** A spider of mass 0.30 g waits in its web of negligible mass (Fig. 14–7). A slight movement causes the web to vibrate with a frequency of about 15 Hz. (*a*) Estimate the value of the spring stiffness constant k for the web. (*b*) At what frequency would you expect the web to vibrate if an insect of mass 0.10 g is trapped in addition to the spider?

APPROACH We can only make a rough estimate because a spider's web is in 3 dimensions and so is complicated and may vibrate with a mixture of frequencies. We use SHM as an approximate model.

SOLUTION (*a*) The frequency of SHM is given by Eq. 14–7a,

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

We solve for k:

$$k = (2\pi f)^2 m$$
$$= (2\pi)^2(15\,\text{s}^{-1})^2(3.0\times10^{-4}\,\text{kg}) = 2.7\,\text{N/m}.$$

(*b*) The total mass is now $0.10\,\text{g} + 0.30\,\text{g} = 4.0\times10^{-4}\,\text{kg}$. We could substitute $m = 4.0\times10^{-4}\,\text{kg}$ into Eq. 14–7a. Instead, we notice that the frequency decreases with the square root of the mass. Since the new mass is 4/3 times the first mass, the frequency changes by a factor of $1/\sqrt{4/3} = \sqrt{3/4}$. Thus $f = (15\,\text{Hz})(\sqrt{3/4}) \approx 13\,\text{Hz}$, and is only an approximation.

NOTE Check this result by direct substitution of k, found in part (*a*), and the new mass m into Eq. 14–7a.



FIGURE 14–7 Example 14–4. A spider waits for its prey (on the left).

EXERCISE D The position of a SHO is given by $x = (0.80\,\text{m})\cos(3.14t - 0.25)$. The frequency is (*a*) 3.14 Hz, (*b*) 1.0 Hz, (*c*) 0.50 Hz, (*d*) 9.88 Hz, (*e*) 19.8 Hz.

Let us continue our analysis of a simple harmonic oscillator. The velocity and acceleration of the oscillating mass can be obtained by differentiation of Eq. 14–4, $x = A\cos(\omega t + \phi)$:

$$v = \frac{dx}{dt} = -\omega A\sin(\omega t + \phi) \qquad \textbf{(14–8a)}$$

$$a = \frac{d^2x}{dt^2} = \frac{dv}{dt} = -\omega^2 A\cos(\omega t + \phi). \qquad \textbf{(14–8b)}$$

The velocity and acceleration of a SHO also vary sinusoidally. In Fig. 14–8 we plot the displacement, velocity, and acceleration of a SHO as a function of time for the case when $\phi = 0$. The speed reaches its maximum

$$v_{\text{max}} = \omega A = \sqrt{\frac{k}{m}}A \qquad \textbf{(14–9a)}$$

when the oscillating object is passing through its equilibrium point, $x = 0$, as seen in Fig. 14–8a and b. And the speed is zero at points of maximum displacement, $x = \pm A$. This is in accord with our discussion of Fig. 14–2. Similarly, the acceleration has its maximum value

$$a_{\text{max}} = \omega^2 A = \frac{k}{m}A \qquad \textbf{(14–9b)}$$

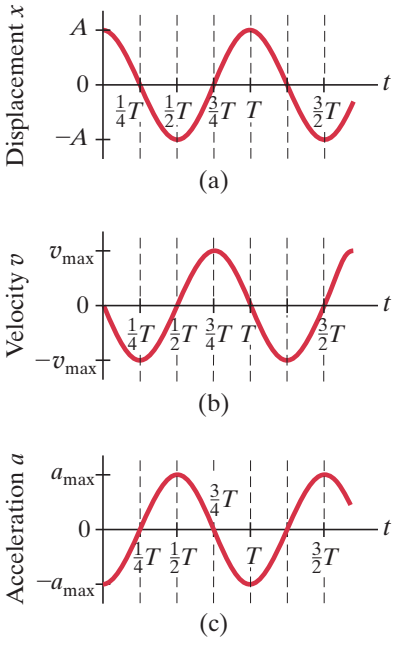
which occurs when $x = \pm A$; and a is zero at $x = 0$, as we expect, because $ma = F$ is proportional to x ($F = -kx$).

For the general case when $\phi \neq 0$, we can relate the constants A and ϕ to the initial values of x, v, and a by setting $t = 0$ in Eqs. 14–4, 14–8, and 14–9:

$$x_0 = x(t=0) = A\cos\phi$$
$$v_0 = v(t=0) = -\omega A\sin\phi = -v_{\text{max}}\sin\phi$$
$$a_0 = a(t=0) = -\omega^2 A\cos\phi = -a_{\text{max}}\cos\phi.$$

Note that v_0 is not the same as v_{max}, and a_0 is not the same as a_{max}.

FIGURE 14–8 Graphs showing (a) displacement x as a function of time t: $x = A\cos\omega t$; (b) velocity as a function of time: $v = -v_{\text{max}}\sin\omega t$, where $v_{\text{max}} = A\sqrt{k/m}$; (c) acceleration as a function of time: $a = -a_{\text{max}}\cos\omega t$, where $a_{\text{max}} = Ak/m$.



EXAMPLE 14–5 **ESTIMATE** **A vibrating floor.** A large motor in a factory causes the floor to vibrate at a frequency of 10 Hz. The amplitude of the floor's motion near the motor is about 3.0 mm. Estimate the maximum acceleration of the floor near the motor.

APPROACH Assuming the motion of the floor is roughly SHM we can make an estimate for the maximum acceleration using Eq. 14–9b.

SOLUTION Given $\omega = 2\pi f = (2\pi)(10\,\text{s}^{-1}) = 62.8\,\text{rad/s}$, then Eq. 14–9b gives

$$a_{\text{max}} = \omega^2 A = (62.8\,\text{rad/s})^2(0.0030\,\text{m}) = 12\,\text{m/s}^2.$$

PHYSICS APPLIED
Unwanted floor vibrations

NOTE The maximum acceleration is a little over g, so when the floor accelerates down, objects sitting on the floor will actually lose contact momentarily, which will cause noise and serious wear.

EXAMPLE 14–6 **Starting with $x = A\cos\omega t$.** The displacement of an object is described by the following equation, where x is in meters and t is in seconds:

$$x = (0.30\,\text{m})\cos(8.0\,t\ \text{rad}).$$

Determine the oscillating object's (*a*) amplitude, (*b*) frequency, (*c*) period, (*d*) maximum speed, and (*e*) maximum acceleration.

APPROACH We start by comparing the given equation for x with Eq. 14–6b, $x = A\cos(2\pi ft)$, with $\phi = 0$.

SOLUTION From $x = A\cos(2\pi ft)$, we see by inspection that (*a*) the amplitude $A = 0.30\,\text{m}$, and (*b*) $2\pi f = 8.0\,\text{rad/s}$; so $f = (8.0\,\text{rad/s}\,/2\pi\,\text{rad}) = 1.27\,\text{s}^{-1} = 1.27\,\text{Hz}$. (*c*) Then $T = 1/f = 0.79\,\text{s}$. (*d*) The maximum speed (Eq. 14–9a) is

$$v_{\text{max}} = \omega A = 2\pi f A$$
$$= (2\pi)(1.27\,\text{s}^{-1})(0.30\,\text{m}) = 2.4\,\text{m/s}.$$

(*e*) The maximum acceleration is given by Eq. 14–9b:

$$a_{\text{max}} = \omega^2 A = (2\pi f)^2 A$$
$$= (2\pi)^2(1.27\,\text{s}^{-1})^2(0.30\,\text{m}) = 19\,\text{m/s}^2.$$

EXAMPLE 14–7 **Loudspeaker.** The cone of a loudspeaker (Fig. 14–9) oscillates in SHM at a frequency of 262 Hz ("middle C"). The amplitude at the central part of the cone is $A = 1.5 \times 10^{-4}\,\text{m}$, and at $t = 0$, $x = A$. (*a*) What equation describes the motion of the center of the cone? (*b*) What are the velocity and acceleration as a function of time? (*c*) What is the position of the cone at $t = 1.00\,\text{ms}$ $(= 1.00 \times 10^{-3}\,\text{s})$?

FIGURE 14–9 Example 14–7. Loudspeaker cone.



APPROACH The motion begins $(t = 0)$ with the cone at its maximum displacement $(x = A$ at $t = 0)$. So we use the cosine function, $x = A\cos\omega t$, with $\phi = 0$.

SOLUTION (*a*) The amplitude $A = 1.5 \times 10^{-4}\,\text{m}$ and

$$\omega = 2\pi f = (6.28\,\text{rad})(262\,\text{s}^{-1}) = 1650\,\text{rad/s}.$$

The motion is described as

$$x = A\cos\omega t = (1.5 \times 10^{-4}\,\text{m})\cos(1650\,t\ \text{rad}),$$

where t is in seconds.

(*b*) The maximum velocity, from Eq. 14–9a, is

$$v_{\text{max}} = \omega A = (1650\,\text{rad/s})(1.5 \times 10^{-4}\,\text{m}) = 0.25\,\text{m/s},$$

so

$$v = -(0.25\,\text{m/s})\sin(1650\,t\ \text{rad}).$$

From Eq. 14–9b the maximum acceleration is $a_{\text{max}} = \omega^2 A = (1650\,\text{rad/s})^2(1.5 \times 10^{-4}\,\text{m}) = 410\,\text{m/s}^2$, which is more than 40 g's. Then

$$a = -(410\,\text{m/s}^2)\cos(1650\,t\ \text{rad}).$$

(*c*) At $t = 1.00 \times 10^{-3}\,\text{s}$,

$$x = A\cos\omega t = (1.5\times10^{-4}\,\text{m})\cos[(1650\,\text{rad/s})(1.00\times10^{-3}\,\text{s})]$$
$$= (1.5\times10^{-4}\,\text{m})\cos(1.65\,\text{rad}) = -1.2\times10^{-5}\,\text{m}.$$

NOTE Be sure your calculator is set in RAD mode, not DEG mode, for these $\cos\omega t$ calculations.

CAUTION
Always be sure your calculator is in the correct mode for angles

EXAMPLE 14–8 **Spring calculations.** A spring stretches 0.150 m when a 0.300-kg mass is gently suspended from it as in Fig. 14–3b. The spring is then set up horizontally with the 0.300-kg mass resting on a frictionless table as in Fig. 14–2. The mass is pulled so that the spring is stretched 0.100 m from the equilibrium point, and released from rest. Determine: (*a*) the spring stiffness constant k and angular frequency ω; (*b*) the amplitude of the horizontal oscillation A; (*c*) the magnitude of the maximum velocity v_{max}; (*d*) the magnitude of the maximum acceleration a_{max} of the mass; (*e*) the period T and frequency f; (*f*) the displacement x as a function of time; and (*g*) the velocity at $t = 0.150\,\text{s}$.

APPROACH When the 0.300-kg mass hangs at rest from the spring as in Fig. 14–3b, we apply Newton's second law for the vertical forces: $\Sigma F = 0 = mg - kx_0$, so $k = mg/x_0$. For the horizontal oscillations, the amplitude is given, and the other quantities can be found from Eqs. 14–4, 14–5, 14–7, and 14–9. We choose x positive to the right.

SOLUTION (*a*) The spring stretches 0.150 m due to the 0.300-kg load, so

$$k = \frac{F}{x_0} = \frac{mg}{x_0} = \frac{(0.300\,\text{kg})(9.80\,\text{m/s}^2)}{0.150\,\text{m}} = 19.6\,\text{N/m}.$$

From Eq. 14–5,

$$\omega = \sqrt{\frac{k}{m}} = \sqrt{\frac{19.6\,\text{N/m}}{0.300\,\text{kg}}} = 8.08\,\text{rad/s}.$$

(*b*) The spring is now horizontal (on a table). It is stretched 0.100 m from equilibrium and is given no initial speed, so $A = 0.100\,\text{m}$.

(*c*) From Eq. 14–9a, the maximum velocity has magnitude

$$v_{\text{max}} = \omega A = (8.08\,\text{rad/s})(0.100\,\text{m}) = 0.808\,\text{m/s}.$$

(*d*) Since $F = ma$, the maximum acceleration occurs where the force is greatest—that is, when $x = \pm A = \pm 0.100\,\text{m}$. Thus its magnitude is

$$a_{\text{max}} = \frac{F}{m} = \frac{kA}{m} = \frac{(19.6\,\text{N/m})(0.100\,\text{m})}{0.300\,\text{kg}} = 6.53\,\text{m/s}^2.$$

[This result could also have been obtained directly from Eq. 14–9b, but it is often useful to go back to basics as we did here.]

(*e*) Equations 14–7b and 14–2 give

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{0.300\,\text{kg}}{19.6\,\text{N/m}}} = 0.777\,\text{s}$$
$$f = \frac{1}{T} = 1.29\,\text{Hz}.$$

(*f*) The motion begins at a point of maximum stretch. If we take x positive to the right in Fig. 14–2, then at $t = 0$, $x = +A = 0.100\,\text{m}$. So we need a sinusoidal curve that has its maximum value at $t = 0$, which is a cosine:

$$x = A\cos\omega t,$$
$$= (0.100\,\text{m})\cos(8.08\,t\ \text{rad}),$$

where t is in seconds and x is in meters.

(*g*) The velocity at any time t is dx/dt (see also part *c*):

$$v = \frac{dx}{dt} = -\omega A\sin\omega t = -(0.808\,\text{m/s})\sin(8.08\,t\ \text{rad}).$$

At $t = 0.150\,\text{s}$, $v = -(0.808\,\text{m/s})\sin(1.21\,\text{rad}) = -0.756\,\text{m/s}$, and the object is moving to the left (the minus sign).

EXAMPLE 14–9 **Spring is started with a push.** Suppose the spring of Example 14–8 is compressed 0.100 m from equilibrium $(x_0 = -0.100\,\text{m})$ and is given a shove to create an initial velocity in the $+x$ direction of $v_0 = 0.400$ m/s. Determine (a) the phase angle ϕ, (b) the amplitude A, and (c) the displacement x as a function of time, $x(t)$.

APPROACH We use Eq. 14–8a, at $t = 0$, to write $v_0 = -\omega A \sin\phi$, and Eq. 14–4 to write $x_0 = A\cos\phi$. Combining these, we can obtain ϕ. We obtain A by using Eq. 14–4 again at $t = 0$. From Example 14–8, $\omega = 8.08$ rad/s.

SOLUTION (a) We combine Eqs. 14–8a and 14–4 at $t = 0$ and solve for the tangent:

$$\tan\phi = \frac{\sin\phi}{\cos\phi} = \frac{(v_0/(-\omega A))}{(x_0/A)} = -\frac{v_0}{\omega x_0} = -\frac{0.400\,\text{m/s}}{(8.08\,\text{rad/s})(-0.100\,\text{m})} = 0.495.$$

A calculator gives the angle as 26.3°, but we note from this equation that both the sine and cosine $(x_0 < 0)$ are negative at $t = 0$, so our angle is in the third quadrant. Hence

$$\phi = 26.3° + 180° = 206.3° = 3.60\,\text{rad}.$$

(b) Again using Eq. 14–4 at $t = 0$, as given in the Approach above,

$$A = \frac{x_0}{\cos\phi} = \frac{(-0.100\,\text{m})}{\cos(3.60\,\text{rad})} = 0.112\,\text{m}.$$

(c) $x = A\cos(\omega t + \phi) = (0.112\,\text{m})\cos(8.08\,t\ \text{rad} + 3.60\,\text{rad})$.

14–3 Energy in the Simple Harmonic Oscillator

When forces are not constant, such as here with simple harmonic motion, it is often convenient and useful to use the energy approach, as we saw in Chapters 7 and 8.

For a simple harmonic oscillator, such as a mass m oscillating on the end of a massless spring, the restoring force is given by

$$F = -kx.$$

The potential energy function, as we saw in Chapter 8, is given by

$$U = -\int F\,dx = \tfrac{1}{2}kx^2,$$

where we set the constant of integration equal to zero so $U = 0$ at $x = 0$ (the equilibrium position).

The total mechanical energy is the sum of the kinetic and potential energies,

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2,$$

where v is the velocity of the mass m at a distance x from the equilibrium position. SHM can occur only if friction is negligible, so that the total mechanical energy E remains constant. As the mass oscillates back and forth, the energy continuously changes from potential energy to kinetic energy, and back again (Fig. 14–10). At the extreme points, $x = -A$ and $x = +A$ (Figs. 14–10a, c), all the energy is stored in the spring as potential energy (and is the same whether the spring is compressed or stretched to the full amplitude). At these extreme points, the mass stops for an instant as it changes direction, so $v = 0$ and:

$$E = \tfrac{1}{2}m(0)^2 + \tfrac{1}{2}kA^2 = \tfrac{1}{2}kA^2. \qquad \textbf{(14–10a)}$$

Thus, the **total mechanical energy of a simple harmonic oscillator is proportional to the square of the amplitude**. At the equilibrium point, $x = 0$, all the energy is kinetic:

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}k(0)^2 = \tfrac{1}{2}mv_{\text{max}}^2, \qquad \textbf{(14–10b)}$$

where v_{max} is the maximum velocity during the motion (which occurs at $x = 0$). Note that $v_{\text{max}} \neq v_0$.

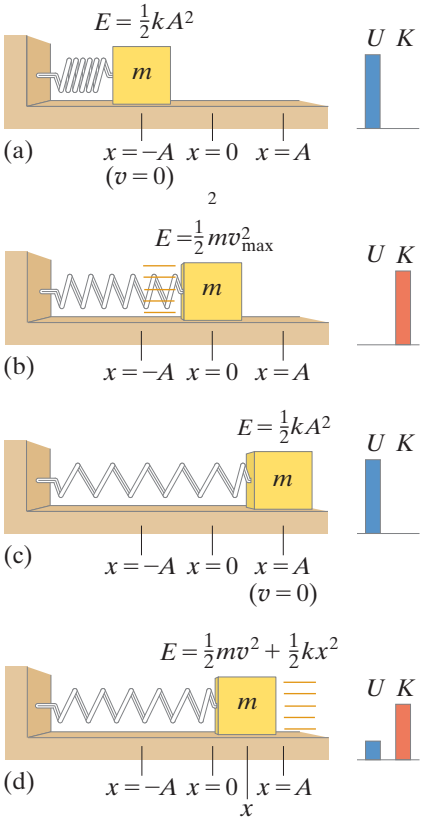


FIGURE 14–10 Energy changes from potential energy to kinetic energy and back again as the spring oscillates. Energy bar graphs (on the right) are described in Section 8–4 (Fig. 8–7).

At intermediate points (Fig. 14–10d), the energy is part kinetic and part potential. Energy conservation tells us

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 \qquad \textbf{(14–10c)}$$

Let us check Eqs. 14–10a and b by inserting Eqs. 14–8a and 14–4 into Eq. 14–10c:

$$E = \tfrac{1}{2}m\omega^2A^2\sin^2(\omega t + \phi) + \tfrac{1}{2}kA^2\cos^2(\omega t + \phi).$$

Using Eq. 14–5, $\omega^2 = k/m$, we have $m\omega^2 = k$, so

$$E = \tfrac{1}{2}kA^2\left[\sin^2(\omega t + \phi) + \cos^2(\omega t + \phi)\right].$$

Equation 14–9a tells us $kA^2 = mv_{\text{max}}^2$, and recalling the trigonometric identity $\sin^2(\omega t + \phi) + \cos^2(\omega t + \phi) = 1$, we obtain Eqs. 14–10a and b:

$$E = \tfrac{1}{2}kA^2 = \tfrac{1}{2}mv_{\text{max}}^2.$$

We can now obtain an equation for the velocity v as a function of x by solving for v in Eq. 14–10c, $v = \pm\sqrt{(2E - kx^2)/m}$, and then substituting (from Eq. 14–10a) $E = \frac{1}{2}kA^2$, to find

$$v = \pm\sqrt{\frac{k}{m}(A^2 - x^2)}. \qquad \textbf{(14–11a)}$$

Because $v_{\text{max}} = A\sqrt{k/m}$ (as in Eq. 14–9a), then

$$v = \pm v_{\text{max}}\sqrt{1 - \frac{x^2}{A^2}}. \qquad \textbf{(14–11b)}$$

Again we see that v is a maximum at $x = 0$, and is zero at $x = \pm A$.

The potential energy, $U = \frac{1}{2}kx^2$, is plotted in Fig. 14–11 (see also Section 8–9). The upper horizontal line represents a particular value of the total energy $E = \frac{1}{2}kA^2$. The distance between the E line and the U curve represents the kinetic energy, K, and the motion is restricted to x values between $-A$ and $+A$. These results are consistent with our full solution of the previous Section.

Energy conservation is a convenient way to obtain v, for example, if x is given (or vice versa), without having to deal with time t.

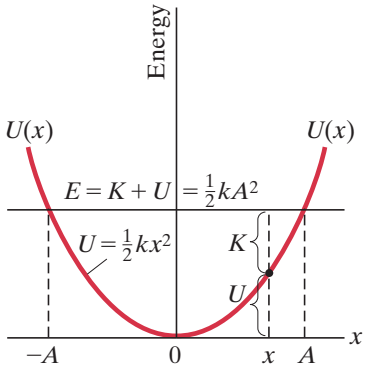


FIGURE 14–11 Graph of potential energy, $U = \frac{1}{2}kx^2$, as a function of x. $K + U = E =$ constant for any point x where $-A \le x \le A$. Values of K and U are indicated towards the right for an arbitrary position x.

EXAMPLE 14–10 **Energy calculations.** For the simple harmonic oscillator of Example 14–8, determine (*a*) the total energy, (*b*) the kinetic and potential energies as a function of time, (*c*) the velocity when the mass is 0.050 m from equilibrium, (*d*) the kinetic and potential energies at half amplitude $(x = \pm A/2)$.

APPROACH We use conservation of energy for a spring–mass system, Eqs. 14–10 and 14–11.

SOLUTION (*a*) From Example 14–8, $k = 19.6\,\text{N/m}$ and $A = 0.100\,\text{m}$, so the total energy E from Eq. 14–10a is

$$E = \tfrac{1}{2}kA^2 = \tfrac{1}{2}(19.6\,\text{N/m})(0.100\,\text{m})^2 = 0.0980\,\text{J}.$$

(*b*) We have, from parts (*f*) and (*g*) of Example 14–8, $x = (0.100\,\text{m})\cos(8.08\,t\ \text{rad})$ and $v = -(0.808\,\text{m/s})\sin(8.08\,t\ \text{rad})$, so

$$U = \tfrac{1}{2}kx^2 = \tfrac{1}{2}(19.6\,\text{N/m})(0.100\,\text{m})^2\cos^2(8.08\,t\ \text{rad}) = (0.0980\,\text{J})\cos^2(8.08\,t\ \text{rad})$$

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}(0.300\,\text{kg})(0.808\,\text{m/s})^2\sin^2(8.08\,t\ \text{rad}) = (0.0980\,\text{J})\sin^2(8.08\,t\ \text{rad}).$$

Note that $U + K = E$ because $\sin^2\theta + \cos^2\theta = 1$.

(*c*) We use Eq. 14–11b and find

$$v = v_{\text{max}}\sqrt{1 - x^2/A^2} = (0.808\,\text{m/s})\sqrt{1 - \left(\tfrac{1}{2}\right)^2} = 0.700\,\text{m/s}.$$

(*d*) At $x = A/2 = 0.050\,\text{m}$, we have

$$U = \tfrac{1}{2}kx^2 = \tfrac{1}{2}(19.6\,\text{N/m})(0.050\,\text{m})^2 = 2.5 \times 10^{-2}\,\text{J}$$

$$K = E - U = 7.3 \times 10^{-2}\,\text{J}.$$

CONCEPTUAL EXAMPLE 14–11 **Doubling the amplitude.** Suppose the spring in Fig. 14–10 is stretched twice as far (to $x = 2A$). What happens to (*a*) the energy of the system, (*b*) the maximum velocity of the oscillating mass, (*c*) the maximum acceleration of the mass?

RESPONSE (*a*) From Eq. 14–10a, the total energy is proportional to the square of the amplitude A, so stretching it twice as far quadruples the energy $(2^2 = 4)$. You may protest, "I did work stretching the spring from $x = 0$ to $x = A$. Don't I do the same work stretching it from A to $2A$?" No. The force you exert is proportional to the displacement x, so for the second displacement, from $x = A$ to $2A$, you do more work than for the first displacement $(x = 0$ to $A)$.

(*b*) From Eqs. 14–10a and b, we can see that when the amplitude is doubled, the maximum velocity must be doubled $(v_{\max} \propto \sqrt{E} \propto A)$.

(*c*) Since the force is twice as great when we stretch the spring twice as far, the acceleration is also twice as great: $a \propto F \propto x$.

EXERCISE E Suppose the spring in Fig. 14–10 is compressed to $x = -A$, but is given a push to the right so that the initial speed of the mass m is v_0. What effect does this push have on (*a*) the energy of the system, (*b*) the maximum velocity, (*c*) the maximum acceleration?

14–4 Simple Harmonic Motion Related to Uniform Circular Motion

Simple harmonic motion has a simple relationship to a particle rotating in a circle with uniform speed. Consider a mass m rotating in a circle of radius A with speed v_M on top of a table as shown in Fig. 14–12. As viewed from above, the motion is a circle. But a person who looks at the motion from the edge of the table sees an oscillatory motion back and forth, and this corresponds precisely to SHM as we shall now see. What the person sees, and what we are interested in, is the projection of the circular motion onto the x axis, Fig. 14–12. To see that this motion is analogous to SHM, let us calculate the x component of the velocity v_M which is labeled v in Fig. 14–12. The two right triangles indicated in Fig. 14–12 are similar, so

$$\frac{v}{v_M} = \frac{\sqrt{A^2 - x^2}}{A}$$

or

$$v = v_M \sqrt{1 - \frac{x^2}{A^2}}.$$

This is exactly the equation for the speed of a mass oscillating with SHM, Eq. 14–11b, where $v_M = v_{\max}$. Furthermore, we can see from Fig. 14–12 that if the angular displacement at $t = 0$ is ϕ, then after a time t the particle will have rotated through an angle $\theta = \omega t$, and so

$$x = A\cos(\theta + \phi) = A\cos(\omega t + \phi).$$

But what is ω here? The linear velocity v_M of our particle undergoing rotational motion is related to ω by $v_M = \omega A$, where A is the radius of the circle (see Eq. 10–4, $v = R\omega$). To make one revolution requires a time T, so we also have $v_M = 2\pi A/T$ where $2\pi A$ is the circle's circumference. Hence

$$\omega = \frac{v_M}{A} = \frac{2\pi A/T}{A} = 2\pi/T = 2\pi f$$

where T is the time required for one rotation and f is the frequency of rotation. This corresponds precisely to the back-and-forth motion of a simple harmonic oscillator. Thus, the projection on the x axis of a particle rotating in a circle has the same motion as a mass undergoing SHM. Indeed, we can say that the projection of circular motion onto a straight line is simple harmonic motion.

The projection of uniform circular motion onto the y axis is also simple harmonic. Thus uniform circular motion can be thought of as two simple harmonic motions operating at right angles.

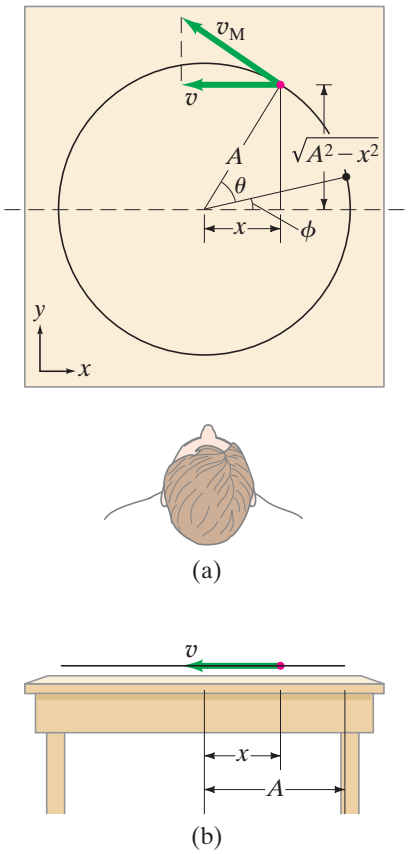


FIGURE 14–12 Analysis of simple harmonic motion as a side view (b) of circular motion (a).

14–5 The Simple Pendulum

A **simple pendulum** consists of a small object (the pendulum bob) suspended from the end of a lightweight cord, Fig. 14–13. We assume that the cord does not stretch and that its mass can be ignored relative to that of the bob. The motion of a simple pendulum moving back and forth with negligible friction resembles simple harmonic motion: the pendulum bob oscillates along the arc of a circle with equal amplitude on either side of its equilibrium point, and as it passes through the equilibrium point (where it would hang vertically) it has its maximum speed. But is it really undergoing SHM? That is, is the restoring force proportional to its displacement? Let us find out.

The displacement s of the pendulum along the arc is given by $s = \ell\theta$, where θ is the angle (in radians) that the cord makes with the vertical and ℓ is the length of the cord (Fig. 14–14). If the restoring force is proportional to s or to θ, the motion will be simple harmonic. The restoring force is the net force on the bob, which equals the component of the weight (mg) tangent to the arc:

$$F = -mg\sin\theta,$$

where g is the acceleration due to gravity. The minus sign here, as in Eq. 14–1, means the force is in the direction opposite to the angular displacement θ. Since F is proportional to the sine of θ and not to θ itself, the motion is *not* SHM. However, if θ is small, then $\sin\theta$ is very nearly equal to θ when the angle is specified in radians. This can be seen by looking at the series expansion[†] of $\sin\theta$, or simply by noting in Fig. 14–14 that the arc length $s\ (= \ell\theta)$ is nearly the same length as the chord $(= \ell\sin\theta)$ indicated by the horizontal straight dashed line, *if θ is small*. For angles less than 15°, the difference between θ (in radians) and $\sin\theta$ is less than 1%—see Table 14–1. Thus, to a very good approximation for small angles,

$$F = -mg\sin\theta \approx -mg\theta.$$

Substituting $s = \ell\theta$, or $\theta = s/\ell$, we have

$$F \approx -\frac{mg}{\ell}s.$$

Thus, for small displacements, the motion can be modeled as being approximately simple harmonic, because this approximate equation fits Hooke's law, $F = -kx$, where in place of x we have arc length s. The effective force constant is $k = mg/\ell$. Thus we can write

$$\theta = \theta_{\max}\cos(\omega t + \phi)$$

where $\theta_{\max}$ is the maximum angular displacement and $\omega = 2\pi f = 2\pi/T$. To obtain ω we use Eq. 14–5, where for k we substitute mg/ℓ: that is,[‡] $\omega = \sqrt{k/m} = \sqrt{(mg/\ell)/m}$, or

$$\omega = \sqrt{\frac{g}{\ell}}. \qquad [\theta \text{ small}] \quad \textbf{(14–12a)}$$

Then the frequency f and period T are

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{g}{\ell}}, \qquad [\theta \text{ small}] \quad \textbf{(14–12b)}$$

$$T = \frac{1}{f} = 2\pi\sqrt{\frac{\ell}{g}}. \qquad [\theta \text{ small}] \quad \textbf{(14–12c)}$$

The mass m of the pendulum bob does not appear in these formulas for T and f. Thus we have the surprising result that the period and frequency of a simple pendulum do not depend on the mass of the pendulum bob. You may have noticed this if you pushed a small child and then a large one on the same swing.

[†] $\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots$ (Appendix A).

[‡] We are using the symbol ω differently here; it is not $d\theta/dt$ as in rotational motion. In this Section θ is the angle of the pendulum at any instant (Fig. 14–14), but we use ω now to represent *not* the rate this angle θ changes, but rather a constant related to the period of oscillation, $\omega = 2\pi f = \sqrt{g/\ell}$.

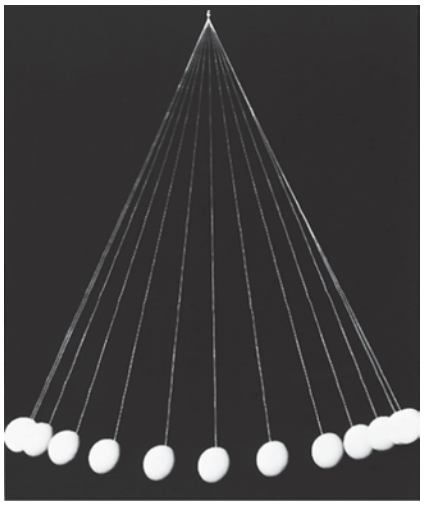


FIGURE 14–13 Strobe-light photo of an oscillating pendulum photographed at equal time intervals.

FIGURE 14–14 Simple pendulum, and a free-body diagram.

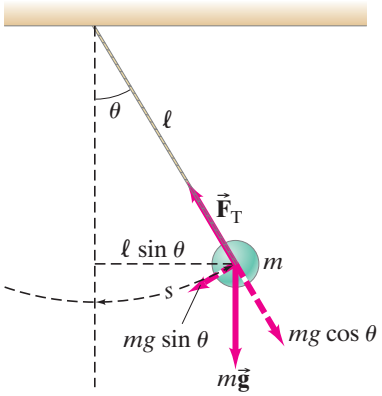


TABLE 14–1 Sin θ at Small Angles

θ (degrees)	θ (radians)	$\sin\theta$	% Difference
0	0	0	0
1°	0.01745	0.01745	0.005%
5°	0.08727	0.08716	0.1%
10°	0.17453	0.17365	0.5%
15°	0.26180	0.25882	1.1%
20°	0.34907	0.34202	2.0%
30°	0.52360	0.50000	4.5%

PHYSICS APPLIED
Pendulum clock

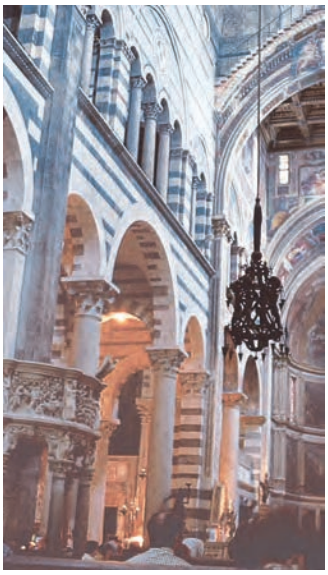


FIGURE 14–15 The swinging motion of this elaborate lamp, hanging by a very long cord from the ceiling of the cathedral at Pisa, Italy, is said to have been observed by Galileo and to have inspired him to the conclusion that the period of a pendulum does not depend on amplitude.

We also see from Eq. 14–12c that the period of a pendulum does not depend on the amplitude (like any SHM, Section 14–2), as long as the amplitude θ is small. Galileo is said to have first noted this fact while watching a swinging lamp in the cathedral at Pisa (Fig. 14–15). This discovery led to the invention of the pendulum clock, the first really precise timepiece, which became the standard for centuries.

Because a pendulum does not undergo *precisely* SHM, the period does depend slightly on the amplitude, the more so for large amplitudes. The accuracy of a pendulum clock would be affected, after many swings, by the decrease in amplitude due to friction. But a mainspring in a pendulum clock (or the falling weight in a grandfather clock) supplies energy to compensate for the friction and to maintain the amplitude constant, so that the timing remains precise.

EXERCISE F Return to the Chapter-Opening Question, page 399, and answer it again now. Try to explain why you may have answered differently the first time.

EXERCISE G If a simple pendulum is taken from sea level to the top of a high mountain and started at the same angle of 5°, it would oscillate at the top of the mountain (*a*) slightly slower; (*b*) slightly faster; (*c*) at exactly the same frequency; (*d*) not at all—it would stop; (*e*) none of these.

EXAMPLE 14–12 **Measuring *g*.** A geologist uses a simple pendulum that has a length of 37.10 cm and a frequency of 0.8190 Hz at a particular location on the Earth. What is the acceleration due to gravity at this location?

APPROACH We can use the length ℓ and frequency f of the pendulum in Eq. 14–12b, which contains our unknown, g.

SOLUTION We solve Eq. 14–12b for g and obtain

$$g = (2\pi f)^2\ell = (2\pi)^2(0.8190\,\text{s}^{-1})^2(0.3710\,\text{m}) = 9.824\,\text{m/s}^2.$$

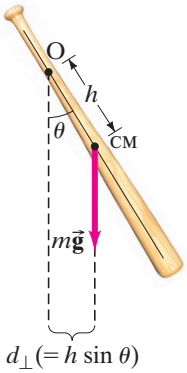
*14–6 The Physical Pendulum and the Torsion Pendulum

*Physical Pendulum

The term physical pendulum refers to any real extended object which oscillates back and forth, in contrast to the rather idealized simple pendulum where all the mass is assumed concentrated in the tiny pendulum bob. An example of a physical pendulum is a baseball bat suspended from the point O, as shown in Fig. 14–16. We can write the force of gravity as acting at the center of gravity or center of mass (CM) of the object located a distance h from the pivot point O. The physical pendulum is best analyzed using the equations of rotational motion. The torque on a physical pendulum, calculated about point O, is

$$\tau = -mgh\sin\theta.$$

FIGURE 14–16 A physical pendulum suspended from point O.



Newton's second law for rotational motion, Eq. 10–14, states that

$$\Sigma\tau = I\alpha = I\frac{d^2\theta}{dt^2},$$

where I is the moment of inertia of the object about the pivot point and $\alpha = d^2\theta/dt^2$ is the angular acceleration. Thus we have

$$I\frac{d^2\theta}{dt^2} = -mgh\sin\theta$$

or

$$\frac{d^2\theta}{dt^2} + \frac{mgh}{I}\sin\theta = 0,$$

where I is calculated about an axis through point O (Fig. 14–16).

For small angular amplitude, $\sin\theta \approx \theta$, so we have

$$\frac{d^2\theta}{dt^2} + \left(\frac{mgh}{I}\right)\theta = 0. \qquad \text{[small angular displacement]} \quad \textbf{(14–13)}$$

This is just the equation for SHM, Eq. 14–3, except that θ replaces x and mgh/I replaces k/m. Thus, for small angular displacements, a physical pendulum undergoes SHM, given by

$$\theta = \theta_{\max}\cos(\omega t + \phi),$$

where $\theta_{\max}$ is the maximum angular displacement and $\omega = 2\pi/T$. The period, T, is (see Eq. 14–7b, replacing m/k with I/mgh):

$$T = 2\pi\sqrt{\frac{I}{mgh}}. \qquad \text{[small angular displacement]} \quad \textbf{(14–14)}$$

EXAMPLE 14–13 **Moment of inertia measurement.** An easy way to measure the moment of inertia of an object about any axis is to measure the period of oscillation about that axis. (*a*) Suppose a nonuniform 1.0-kg stick can be balanced at a point 42 cm from one end. If it is pivoted about the end closest to that point (= CM), Fig. 14–17, it oscillates with a period of 1.6 s. What is its moment of inertia about this end? (*b*) What is its moment of inertia about an axis perpendicular to the stick through its center of mass?

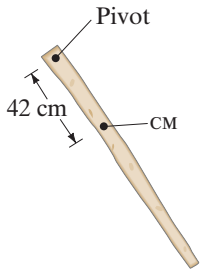


FIGURE 14–17 Example 14–13.

APPROACH We put the given values into Eq. 14–14 and solve for I. The distance from the pivot point at the rod's end to the CM is $h = 42$ cm. For (*b*) we use the parallel-axis theorem (Section 10–7).

SOLUTION (*a*) Given $T = 1.6$ s, and $h = 0.42$ m, Eq. 14–14 gives

$$I = mghT^2/4\pi^2 = 0.27\ \text{kg}\cdot\text{m}^2.$$

(*b*) We use the parallel-axis theorem, Eq. 10–17. The CM is where the stick balanced, 42 cm from the end, so

$$I_{CM} = I - mh^2 = 0.27\ \text{kg}\cdot\text{m}^2 - (1.0\ \text{kg})(0.42\ \text{m})^2 = 0.09\ \text{kg}\cdot\text{m}^2.$$

NOTE An object does not oscillate about its CM, so we cannot measure I_{CM} directly. But the parallel-axis theorem provides a convenient method to determine I_{CM}.

*Torsion Pendulum

Another type of oscillatory motion is by a **torsion pendulum**, in which a disk (Fig. 14–18) or a bar (as in Cavendish's apparatus, Fig. 6–3) is suspended from a wire. The twisting (torsion) of the wire serves as the elastic force. The motion here will be SHM since the restoring torque is very closely proportional to the negative of the angular displacement,

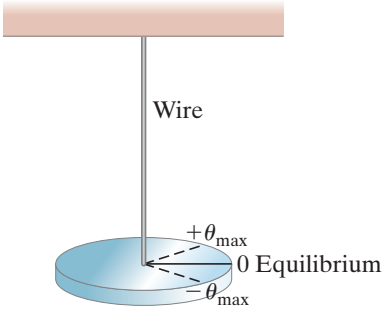
$$\tau = -k\theta,$$

where k is a constant that depends on the wire stiffness under rotation. Then, since $\tau = I\alpha = -k\theta$ corresponds to $F = ma = -kx$, so for a torsion pendulum

$$\omega = \sqrt{k/I}.$$

There is no small angle restriction here, as there is for the physical pendulum (where gravity acts), as long as the wire responds linearly in accordance with Hooke's law.

FIGURE 14–18 A torsion pendulum. The disk oscillates in SHM between $\theta_{\max}$ and $-\theta_{\max}$.



14–7 Damped Harmonic Motion

The amplitude of any real oscillating spring or swinging pendulum slowly decreases in time until the oscillations stop altogether. Figure 14–19 shows a typical graph of the displacement as a function of time. This is called **damped harmonic motion**. The damping[†] is generally due to the resistance of air and to friction within the oscillating system. The energy that is dissipated to thermal energy results in a decreased amplitude of oscillation.

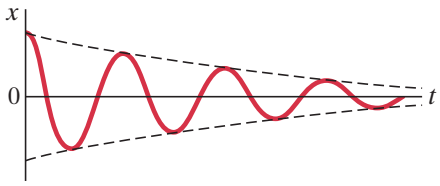


FIGURE 14–19 Damped harmonic motion. The solid red curve represents a cosine times a decreasing exponential (the dashed lines).

Since natural oscillating systems are damped in general, why do we even talk about (undamped) simple harmonic motion? The answer is that SHM is much easier to deal with mathematically. And if the damping is not large, the oscillations can be thought of as simple harmonic motion on which the damping is superposed, as represented by the dashed lines in Fig. 14–19. Although damping does alter the frequency of vibration, the effect can be small if the damping is small. Let us look at this in more detail.

The damping force can depend on the speed of the oscillating object, and opposes the motion. In some simple cases the damping force (Section 5–6) can be approximated as being directly proportional to the speed:

$$F_{\text{damping}} = -bv,$$

where b is a constant. For a mass oscillating on the end of a spring, the restoring force of the spring is $F = -kx$; so Newton's second law $(ma = \Sigma F)$ becomes

$$ma = -kx - bv.$$

We bring all terms to the left side of the equation and substitute $v = dx/dt$ and $a = d^2x/dt^2$ to obtain

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = 0, \qquad \textbf{(14–15)}$$

which is the *equation of motion* for damped harmonic motion. To solve this equation, we guess at a solution and then check to see if it works. If the damping constant b is small, x as a function of t is as plotted in Fig. 14–19, which looks like a cosine function times a factor (represented by the dashed lines) that decreases in time. A simple function that does this is the exponential, $e^{-\gamma t}$, and a solution that might satisfy Eq. 14–15 is

$$x = Ae^{-\gamma t}\cos\omega' t, \qquad \textbf{(14–16)}$$

where A, γ, and ω' are assumed to be constants, and $x = A$ at $t = 0$. If we substitute Eq. 14–16 into Eq. 14–15 (we do this in the optional subsection at the end of this Section, page 417), we find that Eq. 14–16 is indeed a solution if γ and ω' have the values

$$\gamma = \frac{b}{2m} \qquad \textbf{(14–17)}$$

$$\omega' = \sqrt{\frac{k}{m} - \frac{b^2}{4m^2}}. \qquad \textbf{(14–18)}$$

We have called the angular frequency ω' (and not ω) because it is not the same as the ω for SHM without damping $(\omega = \sqrt{k/m})$.

[†]To "damp" means to diminish, restrain, or extinguish, as to "dampen one's spirits."

Thus x as a function of time t for a (lightly) damped harmonic oscillator is

$$x = Ae^{(-b/2m)t}\cos\omega' t. \qquad \textbf{(14–19)}$$

Of course a phase constant, ϕ, can be added to the argument of the cosine in Eq. 14–19. As it stands with $\phi = 0$, it is clear that the constant A in Eq. 14–19 is simply the initial displacement, $x = A$ at $t = 0$. The frequency f' is

$$f' = \frac{\omega'}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m} - \frac{b^2}{4m^2}}. \qquad \textbf{(14–20)}$$

The frequency is lower, and the period longer, than for undamped SHM. (In many practical cases of light damping, however, ω' differs only slightly from $\omega = \sqrt{k/m}$.) This makes sense since we expect damping to slow down the motion. Equation 14–20 reduces to Eq. 14–7a, as it should, when there is no damping $(b = 0)$. The constant $\gamma = b/2m$ is a measure of how quickly the oscillations decrease toward zero (Fig. 14–19). The time $\tau = 2m/b$ is the time taken for the oscillations to drop to $1/e$ of the original amplitude, and is sometimes referred to as the **relaxation time** of the oscillations. Note that the larger b is, the more quickly the oscillations die away.

The solution, Eq. 14–19, is not valid if b is so large that

$$b^2 > 4mk$$

because then ω' (Eq. 14–18) would become imaginary. In this case the system does not oscillate at all but returns directly to its equilibrium position: the damping is so large that the motion no longer resembles simple harmonic motion.

Three common cases of *heavily damped* systems are shown in Fig. 14–20. Curve B represents an **underdamped** situation, in which the system makes several oscillations before coming to rest $(b^2 < 4mk)$; it corresponds to a more heavily damped version of Fig. 14–19. Curve D represents the **overdamped** situation, when the damping is so large $(b^2 > 4mk)$ that there is no oscillation and the system takes a long time to come to rest (equilibrium). Curve C represents **critical damping**: $(b^2 = 4mk)$; in this case the displacement approaches zero in the shortest time. These terms all derive from the use of practical damped systems such as door-closing mechanisms and **shock absorbers** in a car (Fig. 14–21), which are usually designed to give critical damping. But as they wear out, underdamping occurs: the door of a room slams and a car bounces up and down several times when it hits a bump.

In many systems, the oscillatory motion is what counts, as in clocks and musical instruments, and damping may need to be minimized. In other systems, oscillations are the problem, such as a car's springs, so a proper amount of damping (i.e., critical) is desired. Well-designed damping is needed for all kinds of applications. Large buildings, especially in California, are now built (or retrofitted) with huge **dampers** to reduce possible earthquake damage (Fig. 14–22).

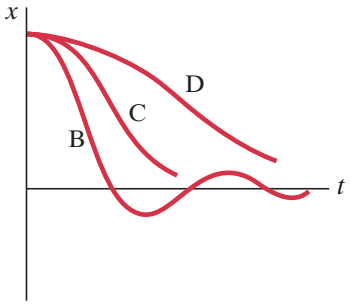


FIGURE 14–20 Underdamped (B), critically damped (C), and overdamped (D) motion.

FIGURE 14–21 Automobile spring with shock absorber which provides damping so that a car won't bounce up and down endlessly. See Chapter-Opening photograph, page 399.

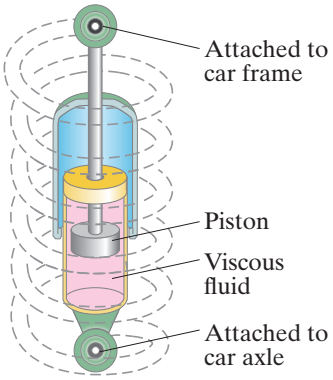




FIGURE 14–22 These huge dampers placed in a building look a lot like huge automobile shock absorbers, and they serve a similar purpose—to reduce the amplitude and the acceleration of movement when the shock of an earthquake hits.

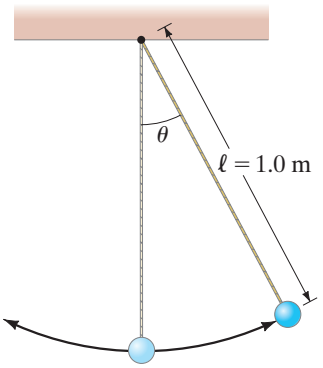


FIGURE 14–23 Example 14–14.

EXAMPLE 14–14 Simple pendulum with damping A simple pendulum has a length $\ell = 1.0\,\text{m}$ (Fig. 14–23). It is set swinging with small-amplitude oscillations. After 5.0 minutes, the amplitude is only 50% of what it was initially. (*a*) What is the value of γ for the motion? (*b*) By what factor does the frequency, f', differ from f, the undamped frequency?

APPROACH Lightly damped harmonic motion is described by (see Eqs. 14–16, 14–17, and 14–18)

$$x = Ae^{-\gamma t}\cos\omega' t, \quad \text{where} \quad \gamma = \frac{b}{2m} \quad \text{and} \quad \omega' = \sqrt{\frac{k}{m} - \frac{b^2}{4m^2}},$$

for damped motion of a mass on the end of a spring. For the simple pendulum without damping, we saw in Section 14–5 that, for small θ,

$$F = -mg\theta.$$

We start with $F = ma$, where a can be written in terms of the angular acceleration $\alpha = d^2\theta/dt^2$; that is, $a = \ell\alpha = \ell\, d^2\theta/dt^2$ so $F = ma = m\ell\, d^2\theta/dt^2$. Then, $F = ma$ means $ma - F = 0$ where $F = -mg\theta$, so (cancelling out m)

$$\ell\frac{d^2\theta}{dt^2} + g\theta = 0.$$

We introduce a damping term, $b(d\theta/dt)$, assuming the damping force is proportional to *angular* speed, $d\theta/dt$ (instead of plain v as before):

$$\ell\frac{d^2\theta}{dt^2} + b\frac{d\theta}{dt} + g\theta = 0,$$

which is the same as Eq. 14–15 with θ replacing x, and ℓ and g replacing m and k.

SOLUTION (*a*) We compare Eq. 14–15 with our equation just above and see that Eq. 14–16, $x = Ae^{-\gamma t}\cos\omega' t$, becomes an equation for θ,

$$\theta = Ae^{-\gamma t}\cos\omega' t,$$

where

$$\gamma = \frac{b}{2\ell} \quad \text{and} \quad \omega' = \sqrt{\frac{g}{\ell} - \frac{b^2}{4\ell^2}}.$$

At $t = 0$,

$$\theta_0 = Ae^{-\gamma\cdot 0}\cos\omega'\cdot 0 = A.$$

Then at $t = 5.0\,\text{min} = 300\,\text{s}$, the amplitude θ has fallen to $0.50A$, so

$$0.50A = Ae^{-\gamma(300\,\text{s})}.$$

We solve this for γ and obtain $\gamma = \ln 2.0/(300\,\text{s}) = 2.3\times10^{-3}\,\text{s}^{-1}$. This means the oscillations fall to $1/e$ of their initial amplitude in a time $t = 1/\gamma = 1/(2.3\times10^{-3}\,\text{s}) = 430\,\text{s}$ or about 7 minutes.

(*b*) We have $\ell = 1.0\,\text{m}$, so $b = 2\gamma\ell = 2(2.3\times10^{-3}\,\text{s}^{-1})(1.0\,\text{m}) = 4.6\times10^{-3}\,\text{m/s}$. Thus $(b^2/4\ell^2)$ is very much less than g/ℓ $(= 9.8\,\text{s}^{-2})$, and the angular frequency of the motion remains almost the same as that of the undamped motion. Specifically (see Eq. 14–20),

$$f' = \frac{1}{2\pi}\sqrt{\frac{g}{\ell}}\left[1 - \frac{\ell}{g}\left(\frac{b^2}{4\ell^2}\right)\right]^{\frac{1}{2}} \approx \frac{1}{2\pi}\sqrt{\frac{g}{\ell}}\left[1 - \frac{1}{2}\frac{\ell}{g}\left(\frac{b^2}{4\ell^2}\right)\right]$$

where we have used the binomial expansion (see Appendix A–2). Then, with $f = (1/2\pi)\sqrt{g/\ell}$ (Eq. 14–12b),

$$\frac{f - f'}{f} \approx \frac{1}{2}\frac{\ell}{g}\left(\frac{b^2}{4\ell^2}\right) = 2.7\times10^{-7}.$$

So f' differs from f by less than one part in a million.

*Showing $x = Ae^{-\gamma t}\cos\omega' t$ Is a Solution

We start with Eq. 14–16, to see if it is a solution to Eq. 14–15. First we take the first and second derivatives

$$\frac{dx}{dt} = -\gamma Ae^{-\gamma t}\cos\omega' t - \omega' Ae^{-\gamma t}\sin\omega' t$$

$$\frac{d^2x}{dt^2} = \gamma^2 Ae^{-\gamma t}\cos\omega' t + \gamma A\omega' e^{-\gamma t}\sin\omega' t + \omega'\gamma Ae^{-\gamma t}\sin\omega' t - \omega'^2 Ae^{-\gamma t}\cos\omega' t.$$

We next substitute these relations back into Eq. 14–15 and reorganize to obtain

$$Ae^{-\gamma t}\left[(m\gamma^2 - m\omega'^2 - b\gamma + k)\cos\omega' t + (2\omega'\gamma m - b\omega')\sin\omega' t\right] = 0. \qquad \textbf{(i)}$$

The left side of this equation must equal zero for all times t, because it is zero on the right. But this can only be so for certain values of γ and ω'. We choose two values of t that will make their evaluation easy. At $t = 0$, $\sin\omega' t = 0$, so the above relation reduces to $A(m\gamma^2 - m\omega'^2 - b\gamma + k) = 0$, which means[†] that

$$m\gamma^2 - m\omega'^2 - b\gamma + k = 0. \qquad \textbf{(ii)}$$

Then at $t = \pi/2\omega'$, $\cos\omega' t = 0$, so Eq. (i) can be valid only if

$$2\gamma m - b = 0, \quad \text{or} \quad \gamma = \frac{b}{2m}. \qquad [= \text{Eq. 14–17}]$$

From Eq. (ii)

$$\omega' = \sqrt{\gamma^2 - \frac{b\gamma}{m} + \frac{k}{m}} = \sqrt{\frac{k}{m} - \frac{b^2}{4m^2}}. \qquad [= \text{Eq. 14–18}]$$

We see Eq. 14–16 is a solution to the equation of motion for damped harmonic motion as long as γ and ω' satisfy Eqs. 14–17 and 14–18.

14–8 Forced Oscillations; Resonance

When an oscillating system is set into motion, it oscillates at its natural frequency $f = 1/2\pi\sqrt{k/m}$, Eq. 14–7a (or, for a simple pendulum, Eq. 14–12b). However, a system may have an external force applied to it that has its own particular frequency. Then we have a **forced oscillation**.

For example, we might pull the mass on the spring of Fig. 14–1 back and forth at an externally applied frequency f. The mass then oscillates at the external frequency f of the external force, even if this frequency is different from the **natural frequency** of the spring, which we will now denote by f_0, where (see Eq. 14–7a)

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

For a forced oscillation with only light damping, the amplitude of oscillation is found to depend on the difference between f and f_0, and is a maximum when the frequency of the external force equals the natural frequency of the system—that is, when $f = f_0$. The amplitude is plotted in Fig. 14–24 as a function of the external frequency f. Curve B represents light damping and curve C heavy damping. When the external driving frequency f is near the natural frequency, $f \approx f_0$, the amplitude can become large if the damping is small. This effect of increased amplitude at $f = f_0$ is known as **resonance**. The natural oscillation frequency f_0 of a system is also called its **resonant frequency**.

A simple illustration of resonance is pushing a child on a swing. A swing, like any pendulum, has a natural frequency of oscillation. If you push on the swing at a random frequency, the swing bounces around and reaches no great amplitude. But if you push with a frequency equal to the natural frequency of the swing, the amplitude increases greatly. At resonance, relatively little effort is required to obtain and maintain a large amplitude.

The great tenor Enrico Caruso was said to be able to shatter a crystal goblet by singing a note of just the right frequency at full voice. This is an example of resonance, for the sound waves emitted by the voice act as a forced oscillation on the glass. At resonance, the resulting oscillation of the goblet may be large enough in amplitude that the glass exceeds its elastic limit and breaks (Fig. 14–25).

[†]It would also be satisfied by $A = 0$, but this gives the trivial and uninteresting solution $x = 0$ for all t—that is, no oscillation.

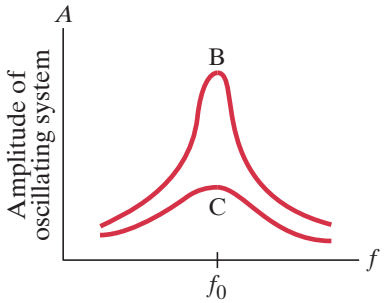


FIGURE 14–24 Amplitude as a function of driving frequency f, showing resonance for lightly damped (B) and heavily damped (C) systems.

PHYSICS APPLIED

Child on a swing

FIGURE 14–25 A goblet breaks as it vibrates in resonance to a trumpet call.





(a)



(b)

FIGURE 14–26 (a) Large-amplitude oscillations of the Tacoma Narrows Bridge, due to gusty winds, led to its collapse (1940). (b) Collapse of a freeway in California, due to resonance in the soft ground below, caused by the 1989 earthquake.

Since material objects are, in general, elastic, resonance is an important phenomenon in a variety of situations. It is particularly important in structural engineering, although the effects are not always foreseen. For example, it has been reported that a railway bridge collapsed because a nick in one of the wheels of a crossing train set up a resonant oscillation in the bridge. Marching soldiers break step when crossing a bridge to avoid the possibility that their normal rhythmic march might match a resonant frequency of the bridge. The famous collapse of the Tacoma Narrows Bridge (Fig. 14–26a) in 1940 occurred (at least in part) as a result of winds gusting at just the right frequency, driving the span into large-amplitude oscillatory motion. The Oakland freeway collapse in the 1989 California earthquake (Fig. 14–26b) involved resonant oscillation of a section built on mudfill that readily transmitted that frequency.

Resonance can be useful, too, and we will meet important examples later, such as in musical instruments and tuning a radio. We will also see that vibrating objects often have not one, but many resonant frequencies.

Equation of Motion and Its Solution

We now look at the equation of motion for a forced oscillation and its solution. Suppose the external force is sinusoidal and can be represented by

$$F_{\text{ext}} = F_0 \cos \omega t,$$

where $\omega = 2\pi f$ is the angular frequency applied externally to the oscillator. Then the equation of motion (with damping) is

$$ma = -kx - bv + F_0 \cos \omega t,$$

or

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = F_0 \cos \omega t. \qquad \textbf{(14–21)}$$

The external force, on the right of the equation, is the only term that does not involve x or one of its derivatives. Problem 72 asks you to show that

$$x = A \sin(\omega t + \phi) \qquad \textbf{(14–22)}$$

is a solution to Eq. 14–21, by direct substitution, where

$$A = \frac{F_0}{m\sqrt{(\omega^2 - \omega_0^2)^2 + b^2\omega^2/m^2}} \qquad \textbf{(14–23)}$$

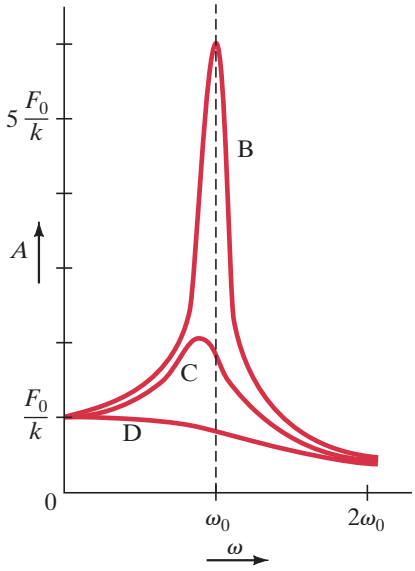
and

$$\phi = \tan^{-1}\frac{\omega_0^2 - \omega^2}{\omega(b/m)}. \qquad \textbf{(14–24)}$$

Actually, the general solution to Eq. 14–21 is Eq. 14–22 plus another term of the form of Eq. 14–19 for the natural damped motion of an oscillator. This second term approaches zero in time, so let us concentrate on the long-term steady state solution, Eq. 14–22.

The amplitude of forced harmonic motion, A, depends strongly on the difference between the applied and the natural frequency. A plot of A (Eq. 14–23) as a function of the applied frequency, ω, is shown in Fig. 14–27 (a more detailed version of Fig. 14–24) for three specific values of the damping constant b. Curve B $\left(b = \frac{1}{6}m\omega_0\right)$ represents light damping, curve C $\left(b = \frac{1}{2}m\omega_0\right)$ fairly heavy damping, and curve D $\left(b = \sqrt{2}m\omega_0\right)$ overdamped motion. The amplitude can become large when the driving frequency is near the natural frequency, $\omega \approx \omega_0$, as long as the damping is not too large. When the damping is small, the increase in amplitude near $\omega = \omega_0$ is very large and, as we saw, is known as *resonance*. The natural oscillating frequency $f_0\,(=\omega_0/2\pi)$ of a system is its *resonant frequency*.[†]

FIGURE 14–27 Amplitude of a forced harmonic oscillator as a function of ω. Curves B, C, and D correspond to light, heavy, and overdamped systems, respectively $\left(Q = m\omega_0/b = 6, 2, 0.71\right)$.



[†]Sometimes the resonant frequency is defined as the actual value of ω at which the amplitude has its maximum value, and this depends somewhat on the damping constant. Except for very heavy damping, this value is quite close to ω_0.

If $b = 0$, resonance occurs at $\omega = \omega_0$ and the resonant peak (of A) becomes infinite; in such a case, energy is being continuously transferred into the system and none is dissipated. For real systems, b is never precisely zero, and the resonant peak is finite. The peak does not occur precisely at $\omega = \omega_0$ (because of the term $b^2\omega^2/m^2$ in the denominator of Eq. 14–23), although it is quite close to ω_0 unless the damping is very large. If the damping is large, there is little or no peak (curve D in Fig. 14–27).

Q Value

The height and narrowness of a resonant peak is often specified by its **quality factor** or **Q value**, defined as

$$Q = \frac{m\omega_0}{b}. \tag{14–25}$$

In Fig. 14–27, curve B has $Q = 6$, curve C has $Q = 2$, and curve D has $Q = 1/\sqrt{2}$. The smaller the damping constant b, the larger the Q value becomes, and the higher the resonance peak. The Q value is also a measure of the width of the peak. To see why, let us consider a graph like Fig. 14–27 but instead of the amplitude A on the vertical axis, we put A^2. The graphs will look a lot like Fig. 14–27, but now they will essentially be graphs of energy vs. ω (recall Eq. 14–10a, $E \propto A^2$). On these energy graphs, let ω_1 and ω_2 be the frequencies where the square of the amplitude A^2 has *half* its maximum value. Then $\Delta\omega = \omega_1 - \omega_2$, which is called the *width* of the resonance peak (or **full width at half maximum**, or FWHM), is related to Q by

$$\frac{\Delta\omega}{\omega_0} = \frac{1}{Q}. \tag{14–26}$$

This relation is accurate only for weak damping. The larger the Q value, the narrower will be the resonance peak relative to its height. Thus a large Q value, representing a system of high quality, has a *high*, *narrow* resonance peak. To show the validity of Eq. 14–26, square Eq. 14–23 for A and determine ω_1 and ω_2 when A^2 is $\frac{1}{2}A^2_{\text{max}}$. Or wait and do it in Problem 73.

Summary

An oscillating (or vibrating) object undergoes **simple harmonic motion** (SHM) if the restoring force is proportional to the displacement,

$$F = -kx. \tag{14–1}$$

The maximum displacement from equilibrium is called the **amplitude**.

The **period**, T, is the time required for one complete cycle (back and forth), and the **frequency**, f, is the number of cycles per second; they are related by

$$f = \frac{1}{T}. \tag{14–2}$$

The period of oscillation for a mass m on the end of an ideal massless spring is given by

$$T = 2\pi\sqrt{\frac{m}{k}}. \tag{14–7b}$$

SHM is **sinusoidal**, which means that the displacement as a function of time follows a sine or cosine curve. The general solution can be written

$$x = A\cos(\omega t + \phi) \tag{14–4}$$

where A is the amplitude, ϕ is the **phase angle**, and

$$\omega = 2\pi f = \sqrt{\frac{k}{m}}. \tag{14–5}$$

The values of A and ϕ depend on the **initial conditions** (x and v at $t = 0$).

During SHM, the total energy

$$E = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2$$

is continually changing from potential to kinetic and back again.

A **simple pendulum** of length ℓ approximates SHM if its amplitude is small and friction can be ignored. For small amplitudes, its period is given by

$$T = 2\pi\sqrt{\frac{\ell}{g}}, \tag{14–12c}$$

where g is the acceleration of gravity.

When friction is present (for all real springs and pendulums), the motion is said to be **damped**. The maximum displacement decreases in time, and the mechanical energy is eventually all transformed to thermal energy. If the friction is very large, so no oscillations occur, the system is said to be **overdamped**. If the friction is small enough that oscillations occur, the system is **underdamped**, and the displacement is given by

$$x = Ae^{-\gamma t}\cos\omega' t, \tag{14–16}$$

where γ and ω' are constants. For a **critically damped** system, no oscillations occur and equilibrium is reached in the shortest time.

If a varying force of frequency f is applied to a system capable of oscillating, the amplitude of oscillation can be very large if the frequency of the applied force is near the **natural** (or **resonant**) **frequency** of the oscillator. This is called **resonance**.

Questions

1. Give some examples of everyday vibrating objects. Which exhibit SHM, at least approximately?
2. Is the acceleration of a simple harmonic oscillator ever zero? If so, where?
3. Real springs have mass. Will the true period and frequency be larger or smaller than given by the equations for a mass oscillating on the end of an idealized massless spring? Explain.
4. How could you double the maximum speed of a simple harmonic oscillator (SHO)?
5. A 5.0-kg trout is attached to the hook of a vertical spring scale, and then is tugged and released. Describe the scale reading as a function of time.
6. If a pendulum clock is accurate at sea level, will it gain or lose time when taken to high altitude? Why?
7. A tire swing hanging from a branch reaches nearly to the ground (Fig. 14–28). How could you estimate the height of the branch using only a stopwatch?



FIGURE 14–28
Question 7.

8. For a simple harmonic oscillator, when (if ever) are the displacement and velocity vectors in the same direction? When are the displacement and acceleration vectors in the same direction?
9. Two equal masses are attached to separate identical springs next to one another. One mass is pulled so its spring stretches 20 cm and the other is pulled so its spring stretches only 10 cm. The masses are released simultaneously. Which mass reaches the equilibrium point first? Explain.
10. Does a car bounce on its springs more rapidly when it is empty or when it is fully loaded?
11. What happens to the period of a playground swing if you rise up from sitting and stand on the swing seat? Explain.
12. * A thin uniform rod of mass m is suspended from one end and oscillates with a frequency f. If a small sphere of mass $2m$ is attached to the other end, does the frequency increase or decrease? Explain.
13. What is the approximate natural period of your walking step?
14. A tuning fork of natural frequency 264 Hz sits on a table at the front of a room. At the back of the room, two tuning forks, one of natural frequency 260 Hz and one of 420 Hz are initially silent, but when the tuning fork at the front of the room is set into vibration, the 260-Hz fork spontaneously begins to vibrate but the 420-Hz fork does not. Explain.
15. Why can you make water slosh back and forth in a pan only if you shake the pan at a certain frequency?
16. Give several everyday examples of resonance.
17. Sometimes a car develops a pronounced *rattle* or vibration at a particular speed, especially if the road has regularly spaced bumps, as between slabs of concrete. Which of the following is a factor and, if so, how: underdamping, overdamping, critical damping, and forced resonance?
18. Over the years, buildings have been able to be built out of lighter and lighter materials. How has this affected the natural oscillation frequencies of buildings and the problems of resonance due to passing trucks, airplanes, or by wind and other natural sources of vibration?

MisConceptual Questions

1. A mass on a spring in SHM (Fig. 14–1) has amplitude A and period T. At what point in the motion is the velocity zero and the acceleration zero simultaneously?
   (*a*) $x = A$.
   (*b*) $x > 0$ but $x < A$.
   (*c*) $x = 0$.
   (*d*) $x < 0$.
   (*e*) None of the above.
2. An object oscillates back and forth on the end of a spring. Which of the following statements are true at some time during the course of the motion?
   (*a*) The object can have zero velocity and nonzero acceleration simultaneously.
   (*b*) The object can have zero velocity and zero acceleration simultaneously.
   (*c*) The object can have zero acceleration and nonzero velocity simultaneously.
   (*d*) The object can have nonzero velocity and nonzero acceleration simultaneously.
3. An object of mass M oscillates on the end of a spring. To double the period, replace the object with one of mass:
   (*a*) $2M$.
   (*b*) $M/2$.
   (*c*) $4M$.
   (*d*) $M/4$.
   (*e*) None of the above.
4. An object of mass m rests on a frictionless surface and is attached to a horizontal ideal spring with spring constant k. The system oscillates with amplitude A. The oscillation frequency of this system can be increased by
   (*a*) decreasing k.
   (*b*) decreasing m.
   (*c*) increasing A.
   (*d*) More than one of the above.
   (*e*) None of the above will work.

5. A block fastened to the end of a horizontal spring rests on a frictionless surface. When the block is pulled back a distance A from equilibrium, the spring–block system has energy E. If the block is pulled back a distance $2A$ from equilibrium, what is the energy of the spring–block system now?
(*a*) Still E, by energy conservation.
(*b*) Work was done on the block, increasing the energy to $2E$.
(*c*) Work was done by the block, decreasing the energy to $\frac{1}{2}E$.
(*d*) Work was done by the block, decreasing the energy to $\frac{1}{4}E$.
(*e*) Work was done on the block, increasing the energy to $4E$.

6. A 100-g mass hangs from a long cord forming a pendulum. The mass is pulled a short distance to one side and released from rest. The time to swing over and back is carefully measured to be 2.0 s. If the 100-g mass is replaced by a 200-g mass, which is then pulled over the same distance and released from rest, the time will be
(*a*) 1.0 s.
(*b*) 1.41 s.
(*c*) 2.0 s.
(*d*) 2.82 s.
(*e*) 4.0 s.

7. A mass m oscillates at the end of a spring with spring constant k_1. A second spring has the same amplitude but is twice as stiff $(k_2 = 2k_1)$ and the mass oscillating at its end is $2m$. Which quantities are the same for the two systems? (Choose all that apply.)
(*a*) The total energy.
(*b*) The period.
(*c*) The maximum speed of the mass.
(*d*) The maximum acceleration of the mass.
(*e*) None of the above.

8. When you use the approximation $\sin\theta \approx \theta$ for a pendulum, you must specify the angle θ in
(*a*) radians only.
(*b*) degrees only.
(*c*) revolutions or radians.
(*d*) degrees or radians.

9. At a playground, two young children are on identical swings. One child appears to be about twice as heavy as the other. If you pull them back together the same distance and release them to start them swinging, what will you notice about the oscillations of the two children?
(*a*) The heavier child swings with a period twice that of the lighter one.
(*b*) The lighter child swings with a period twice that of the heavier one.
(*c*) Both children swing with the same period.

10. A *grandfather clock* is "losing" time because its *pendulum* moves too slowly. Assume that the pendulum is a massive bob at the end of a string. The motion of this pendulum can be sped up by (list all that work):
(*a*) shortening the string.
(*b*) lengthening the string.
(*c*) increasing the mass of the bob.
(*d*) decreasing the mass of the bob.

11. Two simple pendulums swing at a small angle with the same period. The mass of one bob is twice the mass of the other. What can you say about the length of the pendulum with the heavier bob?
(*a*) It must be longer.
(*b*) It must be be shorter.
(*c*) It can be either longer or shorter as long as the amplitudes are adjusted appropriately.
(*d*) The lengths should be the same.

Problems

14–1 and 14–2 Simple Harmonic Motion

1. (I) If a particle undergoes SHM with amplitude 0.27 m, what is the total distance it travels in one period?

2. (I) The springs of a 1700-kg car compress 5.0 mm when its 72-kg driver gets into the driver's seat. If the car goes over a bump, what will be the frequency of oscillations? Ignore damping.

3. (I) (*a*) What is the equation describing the motion of a mass on the end of a spring which is stretched 7.8 cm from equilibrium and then released from rest, and whose period is 0.66 s? (*b*) What will be its displacement after 1.8 s?

4. (I) An elastic cord is 61 cm long when a weight of 68 N hangs from it but is 85 cm long when a weight of 210 N hangs from it. What is the "spring" constant k of this elastic cord?

5. (II) Estimate the spring constant in a child's pogo stick if the child has a mass of 32 kg and bounces once every 1.6 seconds.

6. (II) A fisherman's scale stretches 3.2 cm when a 2.4-kg fish hangs from it. (*a*) What is the spring constant and (*b*) what will be the amplitude and frequency of oscillation if the fish is pulled down 2.1 cm more and released so that it oscillates up and down?

7. (II) A small fly of mass 0.28 g is caught in a spider's web. The web oscillates predominantly with a frequency of 4.0 Hz. (*a*) Estimate the value of the effective force constant k for the web. (*b*) At what frequency would you expect the web to oscillate if an insect of mass 0.46 g is trapped?

8. (II) Construct a Table indicating the position x of the mass in Fig. 14–2 at times $t = 0, \frac{1}{4}T, \frac{1}{2}T, \frac{3}{4}T, T,$ and $\frac{5}{4}T$, where T is the period of oscillation. On a graph of x vs. t, plot these six points. Now connect these points with a smooth curve. Based on these simple considerations, does your curve resemble that of a cosine or sine wave?

9. (II) A mass m at the end of a spring oscillates with a frequency of 0.83 Hz. When an additional 830-g mass is added to m, the frequency is 0.60 Hz. What is the value of m?

10. (II) A wood block of mass 52 g floats on a lake, bobbing up and down at a frequency of 3.5 Hz. (*a*) Estimate the effective force constant of this motion. (*b*) A partially filled water bottle of mass 0.28 kg and almost the same size and shape of the wood block is tossed into the water. At what frequency would you expect the bottle to bob up and down? Assume SHM.

11. (II) A 1.45-kg mass stretches a vertical spring 0.245 m. If the spring is stretched an additional 0.130 m and released, how long does it take to reach the (new) equilibrium position again?

12. (II) An object of unknown mass m is hung from a vertical spring of unknown spring constant k, and the object is observed to be at rest when the spring has extended by 14 cm. The object is then given a slight push and executes SHM. Determine the period T of this oscillation.

13. (II) Figure 14–29 shows two examples of SHM, labeled A and B. For each, what is (*a*) the amplitude, (*b*) the frequency, and (*c*) the period? (*d*) Write the equations for both A and B in the form of a sine or cosine.

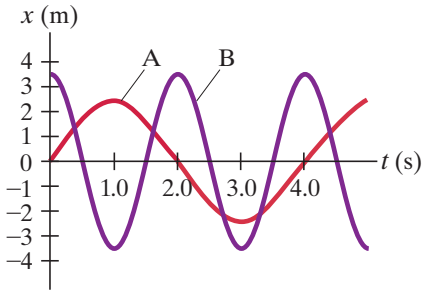


FIGURE 14–29 Problem 13.

14. (II) The graph of displacement vs. time for a small mass m at the end of a spring is shown in Fig. 14–30. At $t = 0$, $x = 0.43$ cm. (*a*) If $m = 7.7$ g, find the spring constant, k. (*b*) Write the equation for displacement x as a function of time.

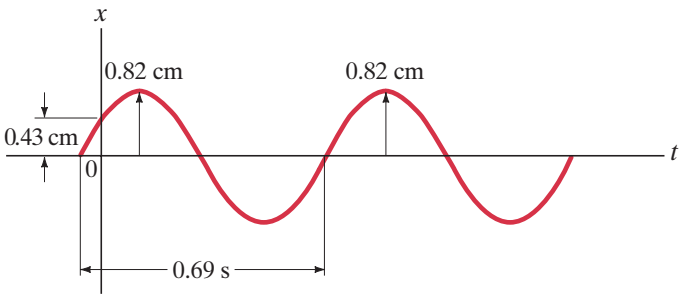


FIGURE 14–30 Problem 14.

15. (II) A 0.350-kg mass oscillates on the end of a thin vertical spring (spring constant 265 N/m) with an amplitude of 28.0 cm. The mass passes through the equilibrium point ($y = 0$) with positive velocity at $t = 0$. (*a*) What equation describes this motion as a function of time? (*b*) At what times will the length of the spring be longest and shortest?

16. (II) Determine the phase constant ϕ in Eq. 14–4 if, at $t = 0$, the oscillating mass is at (*a*) $x = -A$, (*b*) $x = 0$, (*c*) $x = A$, (*d*) $x = \frac{1}{2}A$, (*e*) $x = -\frac{1}{2}A$, (*f*) $x = A/\sqrt{2}$.

17. (II) A tuning fork oscillates at a frequency of 441 Hz and the tip of each prong moves 1.8 mm to either side of center. Calculate (*a*) the maximum speed and (*b*) the maximum acceleration of the tip of a prong.

18. (II) The position of a SHO as a function of time is given by $x = 3.8\cos(5\pi t/4 + \pi/6)$ where t is in seconds and x in meters. Find (*a*) the period and frequency, (*b*) the position and velocity at $t = 0$, and (*c*) the velocity and acceleration at $t = 2.5$ s.

19. (II) A 1.60-kg object oscillates at the end of a vertically hanging light spring once every 0.68 s. (*a*) Write down the equation giving its position y (+ upward) as a function of time t. Assume the spring was initially compressed 16 cm from the equilibrium position (where $y = 0$), and then ($t = 0$) was released. (*b*) How long will it take the object to get to the equilibrium position for the first time? (*c*) What will be its maximum speed? (*d*) What will be its maximum acceleration, and where will it first be attained?

20. (II) Consider two objects, A and B, both undergoing SHM, but with different frequencies, as described by the equations $x_A = (2.0\,\text{m})\sin(4.0\,t)$ and $x_B = (5.0\,\text{m})\sin(3.0\,t)$, where t is in seconds. After $t = 0$, find the next three times t at which both objects simultaneously pass through the origin.

21. (II) A block of mass m is supported by two identical parallel vertical springs, each with spring stiffness constant k (Fig. 14–31). What will be the frequency of vertical oscillation?

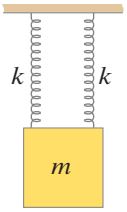


FIGURE 14–31 Problem 21.

22. (II) A *bungee jumper* with mass 65.0 kg jumps from a high bridge. After arriving at his lowest point, he oscillates up and down, reaching a low point five more times in 28.0 s. He eventually comes to rest 25.0 m below the level of the bridge. Estimate the spring constant and the unstretched length of the bungee cord assuming SHM. Assume very little damping.

23. (III) A uniform 1.25-m-long stick of mass M can rotate easily about a hinge at one end and is held horizontal by a spring with spring constant k attached at the other end (Fig. 14–32). If that end of the stick oscillates up and down slightly, what is its frequency? [*Hint*: Write a torque equation about the hinge.]

24. (III) A mass m is at rest on the end of a spring of spring constant k. At $t = 0$ it is given an impulse J by a hammer. Write the formula for the subsequent motion in terms of m, k, J, and t.

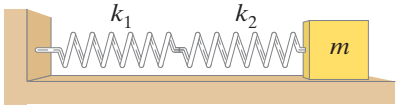
25. (III) A mass m is connected to two springs, with spring constants k_1 and k_2, in two different ways as shown in Fig. 14–33a and b. Show that the period for the configuration shown in part (a) is given by

$$T = 2\pi\sqrt{m\left(\frac{1}{k_1} + \frac{1}{k_2}\right)}$$

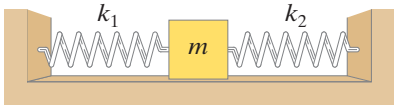
and for that in part (b) is given by

$$T = 2\pi\sqrt{\frac{m}{k_1 + k_2}}.$$

Ignore friction.



(a)



(b)

FIGURE 14–33 Problem 25.

14–3 Energy in SHM

26. (I) A 1.50-kg mass on a particular spring oscillates with amplitude 10.0 cm. The spring is replaced by another resulting in oscillation with twice the energy for the same amplitude. How do the spring constants compare?

27. (I) A 1.28-kg mass oscillates according to the equation $x = 0.650\cos 7.40\,t$ where x is in meters and t in seconds. Determine (*a*) the amplitude, (*b*) the frequency, (*c*) the total energy, and (*d*) the kinetic energy and potential energy when $x = 0.260$ m.

28. (I) A 5.5-kg mass is found to cause a spring to stretch a distance of 25 cm when hung vertically from it. If this mass is displaced a further 12 cm and released, find its velocity when it returns to its equilibrium position.

29. (I) If one oscillation has 3.0 times the energy of a second one of equal frequency and mass, what is the ratio of their amplitudes?

30. (II) (*a*) At what displacement of a SHO is the energy half kinetic and half potential? (*b*) What fraction of the total energy of a SHO is kinetic and what fraction potential when the displacement is one third the amplitude?

31. (II) Draw a graph like Fig. 14–11 for a horizontal spring whose spring constant is 95 N/m and which has a mass of 75 g on the end of it. Assume the spring was started with an initial amplitude of 2.0 cm. Neglect the mass of the spring and any friction with the horizontal surface. Use your graph to estimate (*a*) the potential energy, (*b*) the kinetic energy, and (*c*) the speed of the mass, for $x = 1.5$ cm.

32. (II) A 0.25-kg mass at the end of a spring oscillates 3.2 times per second with an amplitude of 0.15 m. Determine (*a*) the speed when it passes the equilibrium point, (*b*) the speed when it is 0.10 m from equilibrium, (*c*) the total energy of the system, and (*d*) the equation describing the motion of the mass, assuming that at $t = 0$, x was a maximum.

33. (II) A mass of 210 g oscillates on a horizontal frictionless surface at a frequency of 2.5 Hz and with amplitude of 4.5 cm. (*a*) What is the effective spring constant for this motion? (*b*) How much energy is involved in this motion?

34. (II) An object with mass 2.7 kg is executing simple harmonic motion, attached to a spring with spring constant $k = 310$ N/m. When the object is 0.020 m from its equilibrium position, it is moving with a speed of 0.60 m/s. (*a*) Calculate the amplitude of the motion. (*b*) Calculate the maximum speed attained by the object.

35. (II) A mass resting on a horizontal, frictionless surface is attached to one end of a spring; the other end of the spring is fixed to a wall. It takes 3.2 J of work to compress the spring by 0.13 m. The mass is then released from rest and experiences a maximum acceleration of 12 m/s^2. Find the value of (*a*) the spring constant and (*b*) the mass.

36. (II) Agent Arlene devised the following method of measuring the muzzle velocity of a rifle (Fig. 14–34). She fires a bullet into a 4.148-kg wooden block resting on a smooth surface, and attached to a spring of spring constant $k = 162.7$ N/m. The bullet, whose mass is 7.450 g, remains embedded in the wooden block. She measures the maximum distance that the block compresses the spring to be 9.460 cm. What is the speed v of the bullet?

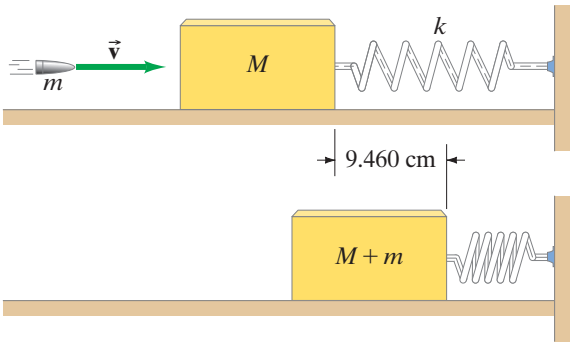


FIGURE 14–34 Problem 36.

37. (II) At $t = 0$, an 885-g mass at rest on the end of a horizontal spring ($k = 184$ N/m) is struck by a hammer which gives it an initial speed of 2.12 m/s. Determine (*a*) the period and frequency of the motion, (*b*) the amplitude, (*c*) the maximum acceleration, (*d*) the position as a function of time, (*e*) the total energy, and (*f*) the kinetic energy when $x = 0.40\,A$ where A is the amplitude.

38. (II) Obtain the displacement x as a function of time for the simple harmonic oscillator using the conservation of energy, Eqs. 14–10. [*Hint*: Integrate Eq. 14–11a with $v = dx/dt$. See Appendix B.]

39. (III) A *pinball machine* uses a spring launcher that is compressed 6.0 cm to launch a ball up a 22° ramp. Assume that the pinball is a solid uniform sphere of radius $r = 1.0$ cm and mass $m = 25$ g. If it is rolling without slipping (Section 10–9) at a speed of 3.0 m/s when it leaves the launcher, what is the spring constant of the spring launcher?

14–5 Simple Pendulum

40. (I) How long must a simple pendulum be if it is to make exactly one swing per second? (That is, one complete oscillation takes exactly 2.0 s.)

41. (I) A pendulum has a period of 2.15 s on Earth. What is its period on Mars, where the acceleration of gravity is about 0.37 that on Earth?

42. (I) A pendulum makes 23 oscillations in exactly 50 s. What is its (*a*) period and (*b*) frequency?

43. (II) A musician wishes to construct a *metronome* based on a simple pendulum design where a "beat" is sounded when the bob reaches each opposite extreme of its swing. How long must the pendulum be so that this metronome is set for 104 bpm (bpm = beats per minute)?

44. (II) A simple pendulum is swinging with amplitude equal to 8.5°. With what angle should it swing if its total energy is to be doubled?

45. (II) A simple pendulum is 0.30 m long. At $t = 0$ it is released from rest starting at an angle of 15°. Ignoring friction, what will be the angular position of the pendulum at (*a*) $t = 0.35$ s, (*b*) $t = 3.45$ s, and (*c*) $t = 6.00$ s?

46. (II) What is the period of a simple pendulum 47 cm long (*a*) on the Earth, and (*b*) when it is in a freely falling elevator?

47. (II) Your grandfather clock's pendulum has a length of 0.9930 m. If the clock loses 18 s per day, how should you adjust the length of the pendulum?

48. (II) Derive a formula for the maximum speed v_{max} of a simple pendulum bob in terms of g, the length ℓ, and the maximum angle of swing θ_{max}.

49. (II) A clock pendulum oscillates at a frequency of 2.5 Hz. At $t = 0$, it is released from rest starting at an angle of 12° to the vertical. Ignoring friction, what will be the position (angle in radians) of the pendulum at (*a*) $t = 0.25$ s, (*b*) $t = 1.60$ s, and (*c*) $t = 500$ s?

50. (III) A simple pendulum oscillates with an amplitude of 10.0°. What fraction of a period does it spend between +8.0° and −8.0°? Assume SHM.

*14–6 Physical Pendulum and Torsion Pendulum

***51.** (II) The balance wheel of a *mechanical (analog) watch* is a thin ring of radius 0.95 cm attached to a coil spring (roughly like Fig. 14–35). It oscillates with a frequency of 1.55 Hz. If a torque of 1.1×10^{-5} m · N causes the wheel to rotate 45°, calculate the mass of the balance wheel.

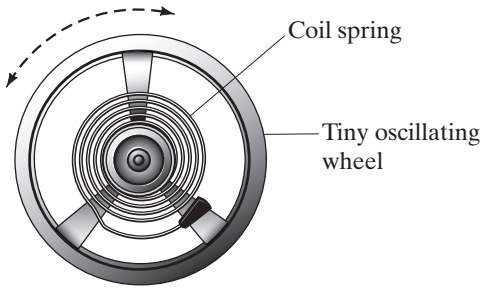


FIGURE 14–35 Problem 51.

***52.** (II) (*a*) Determine the equation of motion and its solution (θ as a function of time) for a torsion pendulum, Fig. 14–18, and show that the motion is simple harmonic. (*b*) Show that the period T is $T = 2\pi\sqrt{I/k}$. [The balance wheel of a mechanical watch is an example of a torsion pendulum in which the restoring torque is applied by a coil spring, Fig. 14–35.]

***53.** (II) A pendulum consists of a tiny bob of mass M and a thin uniform rod of mass m and length ℓ. (*a*) Determine a formula for the period using the small angle approximation. (*b*) What would be the fractional error if you use the formula for a simple pendulum, Eq. 14–12c? (*c*) What ratio of m/M would produce a 10% error?

***54.** (II) A meter stick is hung at its center from a thin wire (Fig. 14–36a). It is twisted and oscillates with a period of 5.0 s. The meter stick is sawed off to a length of 70.0 cm and is again balanced at its center and set in oscillation (Fig. 14–36b). With what period does it oscillate?

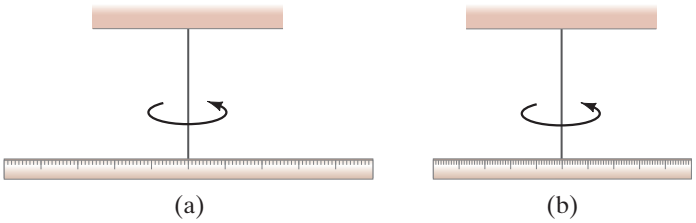


FIGURE 14–36 Problem 54.

***55.** (II) The *human leg* can be compared to a *physical pendulum*, with a "natural" swinging period at which walking is easiest. Consider the leg as two rods joined rigidly together at the knee; the axis for the leg is the hip joint. The length of each rod is about the same: assume 55 cm. Let the upper rod have a mass of 7.0 kg and the lower rod a mass of 4.0 kg. (*a*) Calculate the natural swinging period of the system. (*b*) Check your answer by standing on a chair and measuring the time for one or more complete back-and-forth swings of your leg. The effect of a shorter leg is a shorter swinging period, enabling a faster *natural stride*.

***56.** (II) A student wants to use a meter stick as a pendulum. She plans to drill a small hole through the meter stick and suspend it from a smooth pin attached to the wall (Fig. 14–37). Where in the meter stick should she drill the hole to obtain the shortest possible period? How short an oscillation period can she obtain with a meter stick in this way?

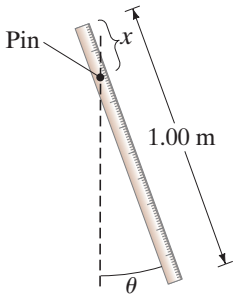


FIGURE 14–37 Problem 56.

***57.** (II) A plywood disk of radius 20.0 cm and mass 2.20 kg has a small hole drilled through it, 2.00 cm from its edge (Fig. 14–38). The disk is hung from the wall by means of a metal pin through the hole, and is used as a pendulum. What is the period of this pendulum for small oscillations?

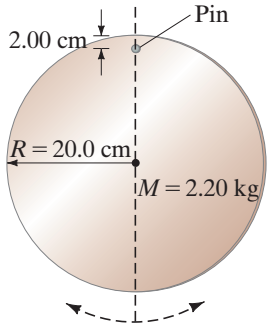


FIGURE 14–38 Problem 57.

***58.** (II) An aluminum disk, 15.4 cm in diameter and 375 g in mass, is mounted on a vertical shaft with very low friction (Fig. 14–39). One end of a flat coil spring is attached to the disk, the other end to the base of the apparatus. The disk is set into rotational oscillation and the frequency is 0.331 Hz. What is the torsional spring constant k ($\tau = -k\theta$)?

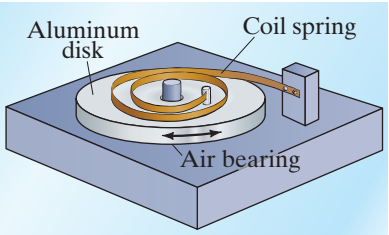


FIGURE 14–39 Problem 58.

14–7 Damping

59. (II) Estimate how the damping constant changes when a car's shock absorbers get old and the car bounces three times after going over a speed bump.

60. (II) A 0.735-kg block oscillates on the end of a spring whose spring constant is $k = 41.0\ \text{N/m}$. The mass moves in a fluid which offers a resistive force $F = -bv$, where $b = 0.662\ \text{N}\cdot\text{s/m}$. (*a*) What is the period of the motion? (*b*) What is the fractional decrease in amplitude per cycle? (*c*) Write the displacement as a function of time if at $t = 0$, $x = 0$, and at $t = 1.00\ \text{s}$, $x = 0.120\ \text{m}$.

61. (II) A damped harmonic oscillator loses 8.0% of its mechanical energy per cycle. (*a*) By what percentage does its frequency differ from the natural frequency $f_0 = (1/2\pi)\sqrt{k/m}$? (*b*) After how many periods will the amplitude have decreased to $1/e$ of its original value?

***62.** (II) A physical pendulum consists of an 85-cm-long uniform wooden rod of mass 280 g hung from a nail near one end (Fig. 14–40). The motion is damped because of friction in the pivot; the damping force is approximately proportional to $d\theta/dt$. The rod is set in oscillation by displacing it 15° from its equilibrium position and releasing it. After 8.0 s, the amplitude of the oscillation has been reduced to 5.5°. If the angular displacement can be written as $\theta = Ae^{-\gamma t}\cos\omega' t$, find (*a*) γ, (*b*) the approximate period of the motion, and (*c*) how long it takes for the amplitude to be reduced to $\frac{1}{2}$ of its original value.

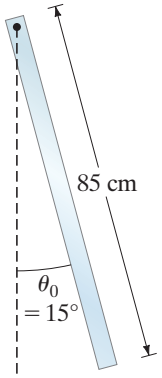


FIGURE 14–40
Problem 62.

63. (II) A vertical spring of spring constant 115 N/m supports a mass of 58 g. The mass oscillates in a tube of liquid. If the mass is initially given an amplitude of 5.0 cm, the mass is observed to have an amplitude of 2.0 cm after 3.5 s. Estimate the damping constant b. Neglect buoyant forces.

64. (III) A glider on an air track is connected by springs to either end of the track (Fig. 14–41). Both springs have the same spring constant, k, and the glider has mass M. (*a*) Determine the frequency of the oscillation, assuming no damping, if $k = 125\ \text{N/m}$ and $M = 215\ \text{g}$. (*b*) It is observed that after 68 oscillations, the amplitude of the oscillation has dropped to one-half of its initial value. Estimate the value of γ, using Eq. 14–16. (*c*) How long does it take the amplitude to decrease to one-quarter of its initial value?

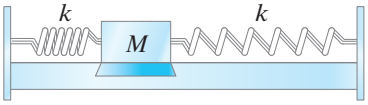


FIGURE 14–41 Problem 64.

65. (III) (*a*) Show that the total mechanical energy, $E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2$, for a lightly damped harmonic oscillator as a function of time, is

$$E = \tfrac{1}{2}kA^2e^{-(b/m)t} = E_0e^{-(b/m)t},$$

where E_0 is the total mechanical energy at $t = 0$. (Assume $\omega' \gg b/2m$.) (*b*) Show that the fractional energy lost per period is

$$\frac{\Delta E}{E} = \frac{2\pi b}{m\omega_0} = \frac{2\pi}{Q},$$

where $\omega_0 = \sqrt{k/m}$ and $Q = m\omega_0/b$ is called the **quality factor** or ***Q* value** of the system. A larger Q value means the system can undergo oscillations for a longer time.

66. (III) Consider a simple pendulum (point mass bob) 0.50 m long with a Q of 350 (as defined in Problem 65). (*a*) How long does it take for the amplitude (assumed small) to decrease by two-thirds? (*b*) If the amplitude is 2.0 cm and the bob has mass 0.27 kg, what is the initial energy loss rate of the pendulum in watts?

14–8 Forced Oscillations; Resonance

67. (II) An 1150-kg automobile has springs with $k = 14{,}000\ \text{N/m}$. One of the tires is not properly balanced; it has a little extra mass on one side compared to the other, causing the car to shake at certain speeds. If the tire radius is 42 cm, at what speed will the wheel shake most?

68. (III) (*a*) For a forced oscillation at resonance $(\omega = \omega_0)$, what is the value of the phase angle ϕ in Eq. 14–22? (*b*) What, then, is the displacement at a time when the driving force F_{ext} is a maximum, and at a time when $F_{\text{ext}} = 0$? (*c*) What is the phase difference (in degrees) between the driving force and the displacement in this case?

69. (III) Differentiate Eq. 14–23 to show that the resonant amplitude peaks at

$$\omega = \sqrt{\omega_0^2 - \frac{b^2}{2m^2}}.$$

70. (III) The amplitude of a driven harmonic oscillator reaches a value of $23.7F_0/k$ at a resonant frequency of 418 Hz. What is the Q value of this system?

71. (III) Construct an accurate resonance curve, from $\omega = 0$ to $\omega = 2\omega_0$, for $Q = 5.0$, (*a*) for amplitude A, and (*b*) for A^2 $(\propto E)$.

72. (III) By direct substitution, show that Eq. 14–22, with Eqs. 14–23 and 14–24, is a solution of the equation of motion (Eq. 14–21) for the forced oscillator. [*Hint*: To find $\sin\phi$ and $\cos\phi$ from $\tan\phi$, draw a right triangle.]

73. (III) Derive Eq. 14–26. Assume b is small, but not zero.

74. (III) *Power transferred to driven oscillator.* (*a*) Show that the power input to a forced oscillator due to the external force $F_{\text{ext}} = F_0\cos\omega t$ is (see Eq. 14–21 through Eq. 14–24):

$$P = F_{\text{ext}}v = \frac{F_0^2\,\omega\cos\phi\,\cos^2\omega t - \frac{1}{2}F_0^2\,\omega\sin\phi\,\sin 2\omega t}{m\sqrt{(\omega^2 - \omega_0^2)^2 + \omega^2b^2/m^2}}.$$

(*b*) Show that the average power input, averaged over one (or many) full cycles, is

$$\overline{P} = \frac{\omega F_0^2\cos\phi}{2m\sqrt{(\omega^2 - \omega_0^2)^2 + \omega^2b^2/m^2}} = \tfrac{1}{2}F_0\,v_{\text{max}}\cos\phi,$$

where v_{max} is the maximum value of dx/dt. (*c*) Also show:

$$\overline{P} = \frac{bF_0^2\,(\omega^2/2m^2)}{(\omega^2 - \omega_0^2)^2 + \omega^2b^2/m^2}.$$

(*d*) Plot $\overline{P}$ versus ω from $\omega = 0$ to $\omega = 2\omega_0$ for $Q = 6.0$. Note that although the amplitude is not zero at $\omega = 0$, $\overline{P}$ is zero at $\omega = 0$.

General Problems

75. An energy-absorbing car bumper has a spring constant of 410 kN/m. Find the maximum compression of the bumper if the car, with mass 1300 kg, collides with a wall at a speed of 2.0 m/s (approximately 5 mi/h).

76. A 62-kg person jumps from a window to a fire net 20.0 m directly below, which stretches the net 1.4 m. Assume that the net behaves like a simple spring. (*a*) Calculate how much it would stretch if the same person were lying in it. (*b*) How much would it stretch if the person jumped from 38 m?

77. A simple pendulum oscillates with frequency f. What is its frequency if the entire pendulum accelerates at $0.35\,g$ (*a*) upward, and (*b*) downward?

78. The length of a simple pendulum is 0.72 m, the pendulum bob has a mass of 295 g, and it is released at an angle of 12° to the vertical. Assume SHM. (*a*) With what frequency does it oscillate? (*b*) What is the pendulum bob's speed when it passes through the lowest point of the swing? (*c*) What is the total energy stored in this oscillation assuming no losses?

79. A 250-g mass attached to a spring oscillates with a frequency of 3.0 Hz. At time $t = 4.10\,\text{s}$, the mass is at $x = -7.0\,\text{cm}$ and is moving to the right at 1.80 m/s. Find an expression for x as a function of time.

80. A block with mass $M = 6.0\,\text{kg}$ rests on a frictionless table and is attached by a horizontal spring $(k = 130\,\text{N/m})$ to a wall. A second block, of mass $m = 1.25\,\text{kg}$, rests on top of M. The coefficient of static friction between the two blocks is 0.30. What is the maximum possible amplitude of oscillation such that m will not slip off M?

81. (*a*) A crane has hoisted a 1550-kg car at the junkyard. The crane's steel cable is 20.0 m long and has a diameter of 6.4 mm. If the car starts bouncing at the end of the cable, what is the period of the bouncing? [*Hint*: Refer to Table 12–1.] (*b*) What amplitude of bouncing will likely cause the cable to snap? (See Table 12–2, and assume Hooke's law holds all the way up to the breaking point.)

82. A 0.650-kg mass oscillates according to the equation $x = 0.25 \sin(4.70\,t)$ where x is in meters and t is in seconds. Determine (*a*) the amplitude, (*b*) the frequency, (*c*) the period, (*d*) the total energy, and (*e*) the kinetic energy and potential energy when x is 15 cm.

83. *Tall buildings* are designed to *sway* in the wind. In a 100-km/h wind, suppose the top of a 110-story building oscillates horizontally with an amplitude of 15 cm at its natural frequency, which corresponds to a period of 7.0 s. Assuming SHM, find the maximum horizontal velocity and acceleration experienced by an employee as she sits working at her desk located on the top floor. Compare the maximum acceleration (as a percentage) with the acceleration due to gravity.

84. An oxygen atom at a particular site within a DNA molecule can be made to execute simple harmonic motion when illuminated by infrared light. The oxygen atom is bound with a spring-like chemical bond to a phosphorus atom, which is rigidly attached to the DNA backbone. The oscillation of the oxygen atom occurs with frequency $f = 3.7 \times 10^{13}\,\text{Hz}$. If the oxygen atom at this site is chemically replaced with a sulfur atom, the spring constant of the bond is unchanged (sulfur is just below oxygen in the Periodic Table). Predict the frequency after the sulfur substitution.

85. A "seconds" pendulum has a period of exactly 2.000 s. That is, each one-way swing takes 1.000 s. What is the length of a seconds pendulum in Austin, Texas, where $g = 9.793\,\text{m/s}^2$? If the pendulum is moved to Paris, where $g = 9.809\,\text{m/s}^2$, by how many millimeters must we lengthen the pendulum? What is the length of a seconds pendulum on the Moon, where $g = 1.62\,\text{m/s}^2$?

86. A 280-kg wooden raft floats on a lake. When a 68-kg man stands on the raft, it sinks 3.5 cm deeper into the water. When he steps off, the raft oscillates for a while. (*a*) What is the frequency of oscillation? (*b*) What is the total energy of oscillation (ignoring damping)?

87. A diving board oscillates with simple harmonic motion of frequency 3.0 cycles per second. What is the maximum amplitude with which the end of the board can oscillate in order that a pebble placed there (Fig. 14–42) does not lose contact with the board during the oscillation?

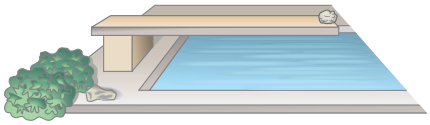


FIGURE 14–42 Problem 87.

88. A 950-kg car strikes a huge spring at a speed of 25 m/s (Fig. 14–43), compressing the spring 4.0 m. (*a*) What is the spring stiffness constant of the spring? (*b*) How long is the car in contact with the spring before it bounces off in the opposite direction?

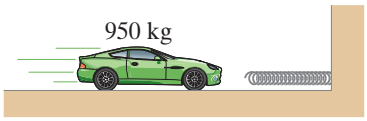


FIGURE 14–43 Problem 88.

89. A 1.80-kg table is supported on four springs. A 0.60-kg chunk of modeling clay is held above the table and dropped so that it hits the center of the table with a speed of 1.65 m/s (Fig. 14–44). The clay makes an inelastic collision with the table, and the table and clay oscillate up and down. After a long time the table comes to rest 6.0 cm below its original position. (*a*) What is the effective spring constant of all four springs taken together? (*b*) With what maximum amplitude does the platform oscillate?

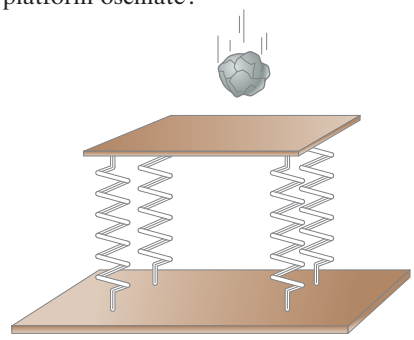


FIGURE 14–44 Problem 89.

90. Estimate the effective spring constant of a trampoline. [*Hint*: Go and jump, or watch, and give your data.]

91. A particular unbalanced wheel of a car shakes when the car moves at 90.0 km/h. The wheel plus tire has mass 17.0 kg and diameter 0.58 m. By how much will the springs of this car compress when it is loaded with 280 kg? (Assume the 280 kg is split evenly among all four springs, which are identical.)

92. Carbon dioxide is a linear molecule. The carbon–oxygen bonds in this molecule act very much like springs. Figure 14–45 shows one possible way the oxygen atoms in this molecule can oscillate: the oxygen atoms oscillate symmetrically in and out, while the central carbon atom remains at rest. Hence each oxygen atom acts like a simple harmonic oscillator with a mass equal to the mass of an oxygen atom. It is observed that this oscillation occurs with a frequency of $f = 2.83 \times 10^{13}$ Hz. What is the spring constant of the C—O bond?

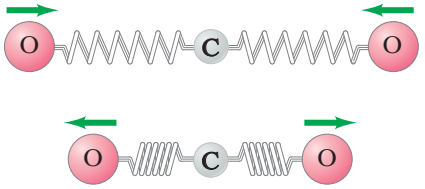


FIGURE 14–45 Problem 92, the CO_2 molecule.

93. A mass attached to the end of a spring is stretched a distance x_0 from equilibrium and released. At what distance from equilibrium will it have (*a*) velocity equal to half its maximum velocity, and (*b*) acceleration equal to half its maximum acceleration?

94. A thin, straight, uniform rod of length $\ell = 1.00$ m and mass $m = 185$ g hangs from a pivot at one end. (*a*) What is its period for small-amplitude oscillations? (*b*) What is the length of a simple pendulum that will have the same period?

95. Imagine that a 10-cm-diameter circular hole was drilled all the way through the center of the Earth (Fig. 14–46). At one end of the hole, you drop an apple into the hole. Show that, if you assume that the Earth has a constant density, the subsequent motion of the apple is simple harmonic. How long will the apple take to return? Assume that we can ignore all frictional effects. [*Hint*: See Appendix E.]



FIGURE 14–46 Problem 95.

96. We usually neglect the *mass of a spring* if it is small compared to the mass attached to it. But in some applications, the mass of the spring must be taken into account. Consider a spring of unstretched length ℓ and mass M_S uniformly distributed along the length of the spring. A mass m is attached to the end of the spring. One end of the spring is fixed and the mass m is allowed to vibrate horizontally without friction (Fig. 14–47). Each point on the spring moves with a velocity proportional to the distance from that point to the fixed end. For example, if the mass on the end moves with speed v_0, the midpoint of the spring moves with speed $v_0/2$. Show that the kinetic energy of the mass plus spring when the mass m is moving with velocity v is

$$K = \tfrac{1}{2}Mv^2$$

where $M = m + \frac{1}{3}M_S$ is the "effective mass" of the system. [*Hint*: Let D be the total length of the stretched spring. Then the velocity of an infinitesimal length dx of spring, of mass dM, located at x is $v(x) = v_0(x/D)$. Note also that $dM = dx\,(M_S/D)$.]

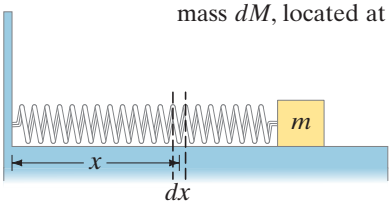


FIGURE 14–47 Problem 96.

97. A mass m is gently placed on the end of a freely hanging spring. The mass then falls 27.0 cm before it stops and begins to rise. What is the frequency of the oscillation?

98. Describe a procedure to measure the spring constant k of a car's springs. Assume that the owner's manual gives the car's mass M and that the shock absorbers are worn out so that the springs are underdamped.

99. In Section 14–5, the oscillation of a simple pendulum (Fig. 14–48) is viewed as linear motion along the arc length x and analyzed via $F = ma$. Alternatively, the pendulum's movement can be regarded as rotational motion about its point of support and analyzed using $\tau = I\alpha$. Carry out this alternative analysis and show that

$$\theta(t) = \theta_{\max}\cos\left(\sqrt{\frac{g}{\ell}}\,t + \phi\right),$$

where $\theta(t)$ is the angular displacement of the pendulum from the vertical at time t, as long as its maximum value is less than about 15°.

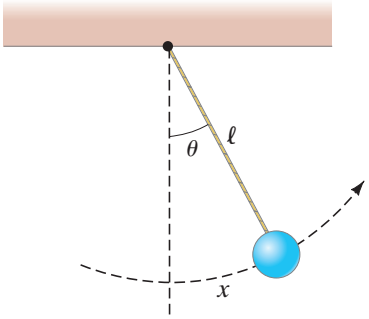


FIGURE 14–48 Problem 99.

ANSWERS TO EXERCISES

A: (*b*).
B: (*c*).
C: (*c*).
D: (*c*).
E: All are larger.
F: (*c*).
G: (*a*).



Waves—such as these water waves—spread outward from a source. The source in this case is a small spot of water oscillating up and down briefly where a rock, or piece of wood or bark, was thrown in (left photo). Other kinds of waves include waves on a cord or string, which also are produced by a vibration. Waves move away from their source, but we also study waves that seem to stand still ("standing waves"). Waves reflect, and they can interfere with each other when they pass through any point at the same time.

CHAPTER 15

Wave Motion

CONTENTS

15 – 1 Characteristics of Wave Motion
15 – 2 Types of Waves: Transverse and Longitudinal
15 – 3 Energy Transported by Waves
15 – 4 Mathematical Representation of a Traveling Wave
*15 – 5 The Wave Equation
15 – 6 The Principle of Superposition
15 – 7 Reflection and Transmission
15 – 8 Interference
15 – 9 Standing Waves; Resonance
15–10 Refraction
15–11 Diffraction

CHAPTER-OPENING QUESTION—Guess now!

You drop a rock into a pond, and water waves spread out in circles.

(a) The waves carry water outward, away from where the rock hit. That moving water carries energy outward.

(b) The waves only make the water move up and down. No energy is carried outward from where the rock hit.

(c) The waves only make the water move up and down, but the waves do carry energy outward, away from where the rock hit.

When you throw a stone into a lake or pool of water, circular waves form and move outward, as shown in the photos above. Waves will also travel along a rope or cord that is stretched out flat on a table if you vibrate one end back and forth as shown in Fig. 15–1. Water waves and waves on a cord are two common examples of **mechanical waves**, which propagate as oscillations of matter. We will discuss other kinds of waves in later Chapters, including electromagnetic waves and light.

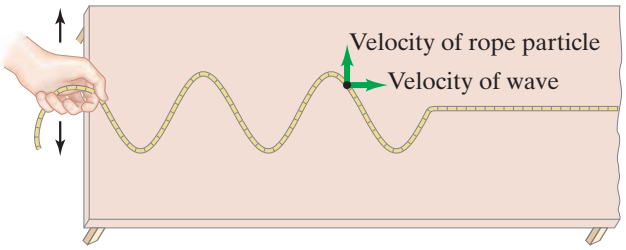


FIGURE 15–1 Wave traveling on a rope or cord. The wave travels to the right along the rope. Particles of the rope oscillate back and forth on the tabletop.

If you have ever watched ocean waves moving toward shore before they break, you may have wondered if the waves were carrying water from far out at sea onto the beach. They don't.[†] Water waves move with a recognizable velocity. But each particle (or molecule) of the water itself merely oscillates about an equilibrium point. This is clearly demonstrated by observing leaves on a pond as waves move by. The leaves (or a cork) are not carried forward by the waves, but oscillate more or less up and down about an equilibrium point because this is the motion of the water itself.

CONCEPTUAL EXAMPLE 15–1 **Wave vs. particle velocity.** Is the velocity of a wave moving along a rope the same as the velocity of a particle of the rope? See Fig. 15–1.

RESPONSE No. The two velocities are different, both in magnitude and direction. The wave on the rope of Fig. 15–1 moves to the right along the tabletop, but each piece of the rope only vibrates to and fro, perpendicular to the traveling wave. (The rope does not travel in the direction that the wave on it does, and the rope does not fall off the edge of the table.)

Waves can move over large distances, but the medium (the water or the rope) itself has only a limited movement, oscillating about an equilibrium point as in simple harmonic motion. Thus, although a wave is not itself matter, the wave pattern can travel in matter. A wave consists of oscillations that move without carrying matter with them.

Waves carry energy from one place to another. Energy is given to a water wave, for example, by a rock thrown into the water, or by wind far out at sea. The energy is transported by waves in all directions and eventually they reach the shore. The oscillating hand in Fig. 15–1 transfers energy to the rope, and that energy is transported down the rope and can be transferred to an object at the other end. All forms of traveling waves transport energy.

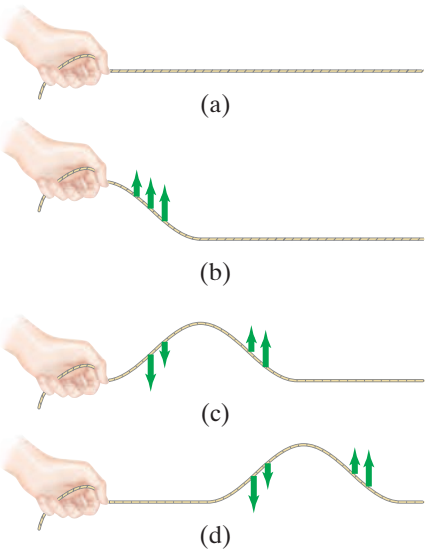
EXERCISE A Return to the Chapter-Opening Question, page 428, and answer it again now. Try to explain why you may have answered differently the first time.

15–1 Characteristics of Wave Motion

Let us look a little more closely at how a wave is formed and how it comes to "travel." We first look at a single wave bump, or **pulse**. A single pulse can be formed on a cord by a quick up-and-down motion of the hand, Fig. 15–2. The hand pulls up on one end of the cord. Because the end section is attached to adjacent sections, these also feel an upward force and they too begin to move upward. As each succeeding section of cord moves upward, the crest of the wave moves outward along the cord. Meanwhile, the end section of cord has been returned to its original position by the hand. As each succeeding section of cord reaches its peak position, it too is pulled back down again by tension from the adjacent section of cord. Thus the source of a traveling wave pulse is a disturbance (or vibration), and cohesive forces between adjacent sections of cord cause the pulse to travel.

[†]Do not be confused by the "breaking" of ocean waves, which occurs when a wave interacts with the ground in shallow water and hence is no longer a simple wave.

FIGURE 15–2 A wave pulse is generated by a hand holding the end of a cord and moving up and down once. Motion of the wave pulse is to the right. Arrows indicate velocity of cord particles.



Waves in other media are created and propagate outward in a similar fashion. A dramatic example of a wave pulse is a tsunami or tidal wave that is created by an earthquake in the Earth's crust under the ocean. The bang you hear when a door slams is a sound wave pulse.

A **continuous** or **periodic wave**, such as that shown in Fig. 15–1, has as its source a disturbance that is continuous and oscillating; that is, the source is a *vibration* or *oscillation*. In Fig. 15–1, a hand oscillates one end of the rope. Water waves may be produced by any vibrating object at the surface, such as your hand; or the water itself is made to vibrate when wind blows across it or a rock is thrown into it. A vibrating tuning fork or drum membrane gives rise to sound waves in air. We will see later that oscillating electric charges give rise to light waves. Indeed, almost any vibrating object sends out waves.

The source of any wave, then, is a vibration. And it is a *vibration* that propagates outward and thus constitutes the wave. If the source vibrates sinusoidally in SHM, then the wave itself—if the medium is elastic—will have a *sinusoidal* shape both in space and in time. (1) In space: if you take a picture of the wave in space at a given instant of time, the wave will have the shape of a sine or cosine as a function of position. (2) In time: if you look at the motion of the medium at one place over a long period of time—for example, if you look between two closely spaced posts of a pier or out of a ship's porthole as water waves pass by—the up-and-down motion of that small segment of water will be simple harmonic motion. The water moves up and down sinusoidally in time.

Some of the important quantities used to describe a periodic sinusoidal wave are shown in Fig. 15–3. The high points on a wave are called **crests**; the low points, **troughs**. The **amplitude**, A, is the maximum height of a crest, or depth of a trough, relative to the normal (or equilibrium) level. The total swing from a crest to a trough is twice the amplitude $(= 2A)$. The distance between two successive crests is the **wavelength**, λ (the Greek letter lambda). The wavelength is also equal to the distance between *any* two successive identical points on the wave. The **frequency**, f, is the number of crests—or complete cycles—that pass a given point per unit time. The **period**, T, equals $1/f$ and is the time elapsed between two successive crests passing by the same point in space.

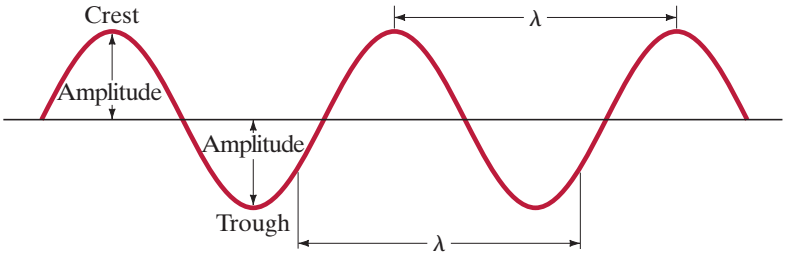


FIGURE 15–3 Characteristics of a single-frequency continuous wave moving through space.

The **wave speed**, v, is the speed at which wave crests (or any other fixed point on the wave shape) move forward. The wave speed must be distinguished from the speed of a particle of the medium itself as we saw in Example 15–1.

A wave crest travels a distance of one wavelength, λ, in a time equal to one period, T. (That is what we mean by the period T.) Thus the wave speed is $v = \lambda/T$. Then, since $1/T = f$,

$$v = \lambda f. \qquad \textbf{(15–1)}$$

For example, suppose a wave has a wavelength of 5 m and a frequency of 3 Hz. Since three crests pass a given point per second, and the crests are 5 m apart, the first crest (or any other part of the wave) must travel a distance of 15 m during the 1 s. So the wave speed is 15 m/s, as Eq. 15–1 predicts.

EXERCISE B You notice a water wave pass by the end of a pier, with about 0.5 s between crests. Therefore (*a*) the frequency is 0.5 Hz; (*b*) the speed is 0.5 m/s; (*c*) the wavelength is 0.5 m; (*d*) the period is 0.5 s.

15–2 Types of Waves: Transverse and Longitudinal

When a wave travels down a cord—say, from left to right as in Fig. 15–1—the particles of the cord vibrate back and forth in a direction transverse (that is, perpendicular) to the motion of the wave itself. Such a wave is called a **transverse wave** (Fig. 15–4a). There exists another type of wave known as a **longitudinal wave**. In a longitudinal wave, the vibration of the particles of the medium is *along* the direction of the wave's motion. Longitudinal waves are readily formed on a stretched spring or Slinky by alternately compressing and expanding one end. This is shown in Fig. 15–4b, and can be compared to the transverse wave in Fig. 15–4a.

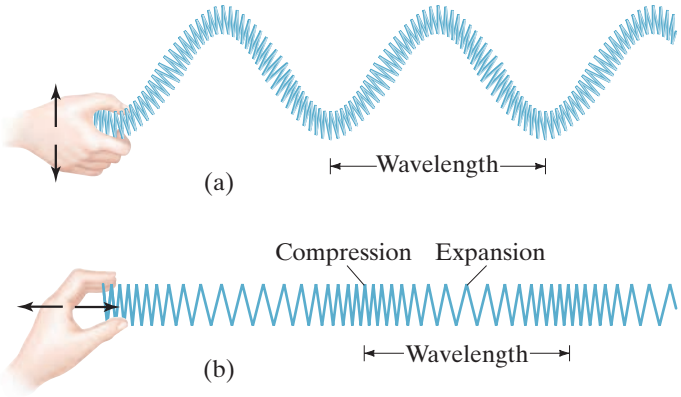


FIGURE 15–4 (a) A transverse wave, and (b) a longitudinal wave, on a Slinky.

A series of compressions and expansions travel along the spring in Fig. 15–4b. The **compressions** are those areas where the coils are momentarily close together. **Expansions** (sometimes called **rarefactions**) are regions where the coils are momentarily far apart. Compressions and expansions correspond to the crests and troughs of a transverse wave.

An important example of a longitudinal wave is a sound wave in air. A vibrating drumhead, for instance, alternately compresses and expands the air in contact with it, producing a longitudinal wave that travels outward in the air, as shown in Fig. 15–5.

As in the case of transverse waves, each section of the medium in which a longitudinal wave passes oscillates over a very small distance, whereas the wave itself can travel large distances. Wavelength, frequency, and wave speed all have meaning for a longitudinal wave. The wavelength is the distance between successive compressions (or between successive expansions), and frequency is the number of compressions that pass a given point per second. The wave speed is the speed with which each compression appears to move; it is equal to the product of wavelength and frequency, $v = \lambda f$ (Eq. 15–1).

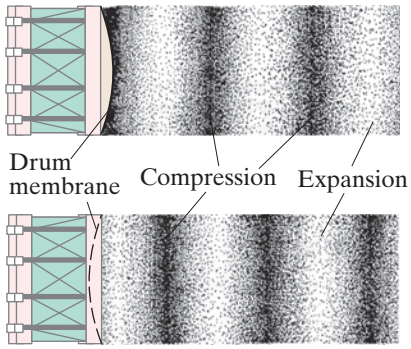


FIGURE 15–5 Production of a sound wave, which is longitudinal, shown at two moments in time about a half period $(\frac{1}{2}T)$ apart. (An "expansion" is often called a *rarefaction* of the air.) The dots represent air molecules.

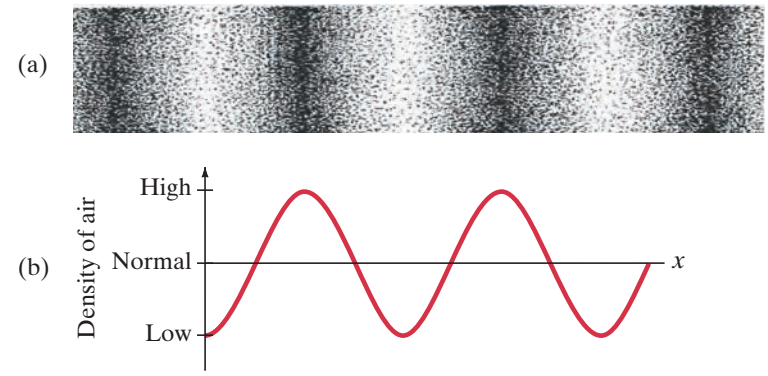


FIGURE 15–6 (a) A longitudinal wave with (b) its graphical representation at a particular instant in time.

A longitudinal wave can be represented graphically by plotting the density of air molecules (or coils of a Slinky) versus position at a given instant, as shown in Fig. 15–6. Such a graphical representation helps make it clearer what is happening. Note that the graph looks much like a transverse wave.

Speed of Transverse Waves

The speed of a wave depends on the properties of the medium in which it travels. The speed of a transverse wave on a stretched string or cord, for example, depends on the tension in the cord, F_T, and on the mass per unit length of the cord, μ (the Greek letter mu). If m is the mass of a length ℓ of wire, then $\mu = m/\ell$. For waves of small amplitude, the wave speed is

$$v = \sqrt{\frac{F_T}{\mu}}. \qquad \left[\begin{array}{l}\text{transverse}\\ \text{wave on a cord}\end{array}\right] \quad \textbf{(15–2)}$$

This formula makes sense qualitatively on the basis of Newtonian mechanics. That is, we do expect the tension to be in the numerator and the mass per unit length in the denominator. Why? Because when the tension is greater, we expect the speed to be greater since each segment of cord is in tighter contact with its neighbor. And, the greater the mass per unit length, the more inertia the cord has and the more slowly the wave would be expected to propagate. (*Propagate* = travel, spread out.)

EXERCISE C A wave starts at the left end of a long cord (see Fig. 15–1) when someone shakes the cord back and forth at the rate of 2.0 Hz. The wave is observed to move to the right at 4.0 m/s. If the frequency is increased from 2.0 to 3.0 Hz, the new speed of the wave is (*a*) 1.0 m/s, (*b*) 2.0 m/s, (*c*) 4.0 m/s, (*d*) 8.0 m/s, (*e*) 16.0 m/s.

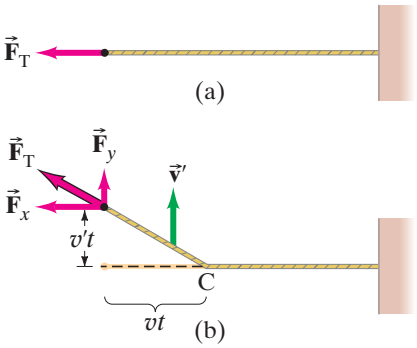


FIGURE 15–7 Diagram of simple wave pulse on a cord for derivation of Eq. 15–2. (Diagram is not to scale: we assume $v' \ll v$, and $F_x \approx F_T$. The upward angle of the cord is exaggerated for visibility.)

We can make a simple *derivation* of Eq. 15–2 using a simple model of a cord under a tension F_T as shown in Fig. 15–7a. The cord is initially pulled quickly upward at a speed v' by the force F_y. As shown in Fig. 15–7b all points of the cord to the left of point C move upward at the speed v', and those to the right are still at rest. The speed of propagation, v, of this wave pulse is the speed of point C, the leading edge of the pulse. Point C moves to the right a distance vt in a time t, whereas the end of the cord moves upward a distance $v't$. By similar triangles we have the approximate relation

$$\frac{F_T}{F_y} = \frac{vt}{v't} = \frac{v}{v'},$$

which works for small amplitudes so that $F_x \approx F_T$, meaning that $v't \ll vt$. As we saw in Chapter 9, the impulse given to an object is equal to its change in momentum. During the time t the total upward impulse is $F_y t = (v'/v)F_T t$. The change in momentum of the cord, Δp, is the mass of cord moving upward times its speed v'. Since the upward moving segment of cord has mass equal to the mass per unit length μ times its length vt we have $\Delta p = (\mu vt)v'$, and so

$$F_y t = \Delta p$$
$$\frac{v'}{v} F_T t = (\mu vt)v'.$$

Solving for v we find $v = \sqrt{F_T/\mu}$ which is Eq. 15–2. Although it was derived for a

special case, it is valid for any wave shape since other shapes can be considered to be made up of many tiny such lengths. But it is valid only for small displacements (as was our derivation). Experiment is in accord with this result derived from Newtonian mechanics.

EXAMPLE 15–2 **Wave along a wire.** A wave whose wavelength is 0.30 m is traveling down a 300-m-long wire whose total mass is 15 kg. If the wire is under a tension of 1000 N, what are the speed and frequency of this wave?

APPROACH We assume the speed of this wave on a wire is given by Eq. 15–2. We get the frequency from Eq. 15–1, $f = v/\lambda$.

SOLUTION From Eq. 15–2, the speed is

$$v = \sqrt{\frac{1000\,\text{N}}{(15\,\text{kg})/(300\,\text{m})}} = \sqrt{\frac{1000\,\text{N}}{(0.050\,\text{kg/m})}} = 140\,\text{m/s}.$$

The frequency is

$$f = \frac{v}{\lambda} = \frac{140\,\text{m/s}}{0.30\,\text{m}} = 470\,\text{Hz}.$$

NOTE A higher tension would increase both v and f, whereas a thicker, denser wire would reduce v and f.

EXAMPLE 15–3 **Pulse on a wire.** An 80.0-m-long, 2.10-mm-diameter copper wire is stretched between two poles. A bird lands at the center point of the wire, sending a small wave pulse out in both directions. The pulses reflect at the ends and arrive back at the bird's location 0.750 seconds after it landed. Determine the tension in the wire.

APPROACH From Eq. 15–2, the tension is given by $F_T = \mu v^2$. The speed v is distance divided by the time. The mass per unit length μ is calculated from the density of copper and the dimensions of the wire.

SOLUTION Each wave pulse travels 40.0 m to the pole and back again (= 80.0 m) in 0.750 s. Thus their speed is $v = (80.0\,\text{m})/(0.750\,\text{s}) = 107\,\text{m/s}$. We take the density of copper (Table 13–1) as $8.90 \times 10^3\,\text{kg/m}^3$. The volume of copper in the wire is the cross-sectional area (πr^2) times the length ℓ, and the mass of the wire is the density times the volume: $m = \rho(\pi r^2)\ell$ for a wire of radius r and length ℓ. Then $\mu = m/\ell$ is

$$\mu = \rho\pi r^2\ell/\ell = \rho\pi r^2 = (8.90 \times 10^3\,\text{kg/m}^3)\pi(1.05 \times 10^{-3}\,\text{m})^2 = 0.0308\,\text{kg/m}.$$

Thus, the tension is $F_T = \mu v^2 = (0.0308\,\text{kg/m})\,(107\,\text{m/s})^2 = 353\,\text{N}$.

Speed of Longitudinal Waves

The speed of a longitudinal wave has a form similar to that for a transverse wave on a cord (Eq. 15–2); that is,

$$v = \sqrt{\frac{\text{elastic force factor}}{\text{inertia factor}}}.$$

In particular, for a longitudinal wave traveling down a long solid rod,

$$v = \sqrt{\frac{E}{\rho}}, \qquad \left[\begin{array}{l}\text{longitudinal}\\ \text{wave in a long rod}\end{array}\right] \quad \textbf{(15–3)}$$

where E is the elastic modulus (Section 12–5) of the material and ρ is its density. For a longitudinal wave traveling in a liquid or gas,

$$v = \sqrt{\frac{B}{\rho}}, \qquad \left[\begin{array}{l}\text{longitudinal wave}\\ \text{in a fluid}\end{array}\right] \quad \textbf{(15–4)}$$

where B is the bulk modulus (Section 12–5) and ρ again is the density. We will derive this, starting on the next page.

PHYSICS APPLIED
Space perception by animals, using sound waves

EXAMPLE 15–4 Echolocation. Echolocation is a form of perception used by animals such as bats, dolphins and toothed whales. The animal emits a pulse of sound (a longitudinal wave) which, after reflection from objects, returns and is detected by the animal. Echolocation waves can have frequencies of about 100,000 Hz. (*a*) Estimate the wavelength of a sea animal's echolocation wave. (*b*) If an obstacle is 100 m from the animal, how long after the animal emits a wave is its reflection detected?

APPROACH We first compute the speed of longitudinal (sound) waves in sea water, using Eq. 15–4 and Tables 12–1 and 13–1. The wavelength is $\lambda = v/f$.

SOLUTION (*a*) The speed of longitudinal waves in sea water, which is slightly more dense than pure water, is

$$v = \sqrt{\frac{B}{\rho}} = \sqrt{\frac{2.0 \times 10^9\,\text{N/m}^2}{1.025 \times 10^3\,\text{kg/m}^3}} = 1.4 \times 10^3\,\text{m/s}.$$

Then, using Eq. 15–1, we find

$$\lambda = \frac{v}{f} = \frac{(1.4 \times 10^3\,\text{m/s})}{(1.0 \times 10^5\,\text{Hz})} = 14\,\text{mm}.$$

(*b*) The time required for the round-trip between the animal and the object is

$$t = \frac{\text{distance}}{\text{speed}} = \frac{2(100\,\text{m})}{1.4 \times 10^3\,\text{m/s}} = 0.14\,\text{s}.$$

A pretty fast response for this underwater "radar".

NOTE We shall see later that waves can be used to "resolve" (or detect) objects whose size is comparable to or larger than the wavelength. Thus, a dolphin can resolve objects on the order of a centimeter or larger in size.

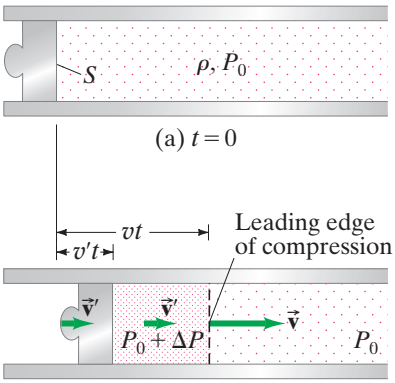


FIGURE 15–8 Determining the speed of a one-dimensional longitudinal wave in a fluid contained in a long narrow tube.

*Deriving Speed of Wave in a Fluid

We now derive Eq. 15–4. Consider a wave pulse traveling in a fluid in a long tube, so that the wave motion is one dimensional. The tube is fitted with a piston at the end and is filled with a fluid which, at $t = 0$, is of uniform density ρ and at uniform pressure P_0, Fig. 15–8a. At this moment the piston is abruptly made to start moving to the right with speed v', compressing the fluid in front of it. In the (short) time t the piston moves a distance $v't$. The compressed fluid itself also moves with speed v', but the leading edge of the compressed region moves to the right at the characteristic speed v of compression waves in that fluid; we assume the wave speed v is much larger than the piston speed v'. The leading edge of the compression (which at $t = 0$ was at the piston face) thus moves a distance vt in time t as shown in Fig. 15–8b. Let the pressure in the compressed fluid be $P_0 + \Delta P$, which is ΔP higher than in the uncompressed fluid. To move the piston to the right requires an external force $(P_0 + \Delta P)S$ acting to the right, where S is the cross-sectional area of the tube. (S for "surface area"; we save A for amplitude.) The *net* force on the compressed region of the fluid is

$$F_{\text{net}} = (P_0 + \Delta P)S - P_0 S = S\,\Delta P$$

since the uncompressed fluid exerts a force $P_0 S$ to the left at the leading edge. Hence the impulse $(= F_{\text{net}}\,t)$ given to the compressed fluid, which equals its change in momentum, is

$$F_{\text{net}}\,t = \Delta m v'$$
$$S\,\Delta P\,t = (\rho S v t)v',$$

where $\Delta m = \rho V = \rho S v t$ represents the mass of fluid which is given the speed v' (the compressed fluid of area S moves a distance vt, Fig. 15–8, so the volume moved is $V = Svt$). Hence we have

$$\Delta P = \rho v v'.$$

From the definition of the bulk modulus, B (Eq. 12–7):

$$B = -\frac{\Delta P}{\Delta V/V_0} = -\frac{\rho v v'}{\Delta V/V_0},$$

where $\Delta V/V_0$ is the fractional change in volume due to compression. The original volume (i.e., before compression) of the compressed fluid is $V_0 = Svt$ (see Fig. 15–8), and it has been compressed by an amount $\Delta V = -Sv't$ (Fig. 15–8b). Thus

$$B = -\frac{\rho v v'}{\Delta V/V_0} = -\frac{\rho v v' V_0}{\Delta V} = -\rho v v'\left(\frac{Svt}{-Sv't}\right) = \rho v^2.$$

Solving this for v gives

$$v = \sqrt{\frac{B}{\rho}},$$

which is what we set out to show, Eq. 15–4.

The derivation of Eq. 15–3 follows similar lines, but takes into account the expansion of the sides of a rod when the end of the rod is compressed.

Other Waves

PHYSICS APPLIED
Earthquake waves

Both transverse and longitudinal waves are produced when an **earthquake** occurs. The transverse waves that travel through the body of the Earth are called S waves (S for shear), and the longitudinal waves are called P waves (P for pressure) or *compression* waves. Both longitudinal and transverse waves can travel through a solid since the atoms or molecules can vibrate about their relatively fixed positions in any direction. But only longitudinal waves can propagate through a fluid, because any transverse motion would not experience any restoring force since a fluid is readily deformable. This fact was used by geophysicists to infer that a portion of the Earth's core must be liquid: after an earthquake, longitudinal waves are detected diametrically across the Earth, but not transverse waves.

Besides these two types of waves that can pass through the body of the Earth (or other substance), there can also be **surface waves** that travel along the boundary between two materials, such as earth and air.

A wave on water is actually a surface wave that moves on the boundary between water and air. The motion of each particle of water at the surface is circular or elliptical (Fig. 15–9), so it is a combination of horizontal and vertical motions. Below the surface, there is also horizontal and vertical motion, as shown. At the bottom, the motion is only horizontal.

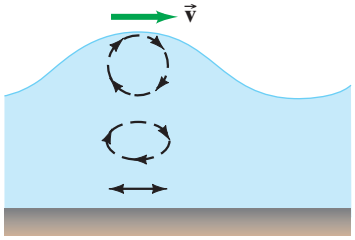


FIGURE 15–9 A shallow water wave is an example of a *surface wave*, which is a combination of vertical and horizontal motions.

When an ocean wave approaches shore, the water drags at the bottom and is slowed down, while the crests move ahead at higher speed (Fig. 15–10) and "spill" over the top.

Surface waves are also set up on the Earth when an earthquake occurs. The waves that travel along the surface are mainly responsible for the damage caused by earthquakes.

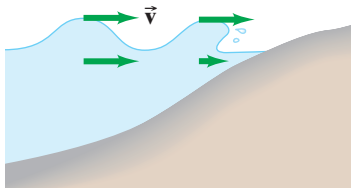


FIGURE 15–10 How a water wave breaks. The green arrows represent the local velocity of water molecules.

Waves which travel along a line in one dimension, such as transverse waves on a stretched string, or longitudinal waves in a rod or fluid-filled tube, are *linear* or *one-dimensional waves*. Surface waves, such as the water waves shown at the start of this Chapter, page 428, are *two-dimensional waves*. Finally, waves that move out from a source in all directions, such as sound from a loudspeaker or earthquake waves through the Earth, are *three-dimensional waves*.

15–3 Energy Transported by Waves

Waves transport energy from one place to another. As waves travel through a medium, the energy is transferred as vibrational energy from particle to particle of the medium. For a sinusoidal wave of frequency f, the particles move in simple harmonic motion (Chapter 14) as a wave passes, and each particle has energy $E = \frac{1}{2}kA^2$ where A is the amplitude (maximum displacement) of its motion, either transversely or longitudinally (Eq. 14–10a).

Using Eq. 14–10a for the energy of a particle undergoing SHM, $E = \frac{1}{2}kA^2$, as well as Eq. 14–7a for k, we can write $k = 4\pi^2 mf^2$, where m is the mass of a particle (or small volume) of the medium. Then in terms of the frequency f and amplitude A,

$$E = \tfrac{1}{2}kA^2 = 2\pi^2 mf^2A^2.$$

For a simple wave of frequency f traveling in an elastic medium, a mass m of the medium is $m = \rho V$, where ρ is the density of the medium and V is the volume of a small slice of the medium. The volume $V = S\ell$ where S is the cross-sectional area through which the wave travels (Fig. 15–11), and we can write ℓ as the distance the wave travels in a time t as $\ell = vt$, where v is the speed of the wave. Thus $m = \rho V = \rho S\ell = \rho Svt$ and

$$E = 2\pi^2 \rho Svtf^2A^2. \qquad \textbf{(15–5)}$$

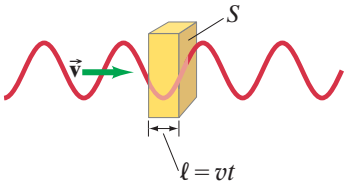


FIGURE 15–11 Calculating the energy carried by a wave moving with speed v.

From this equation we have the important result that **the energy transported by a wave is proportional to the square of the amplitude, and also to the square of the frequency**. The average *rate* of energy transferred is the average power P:

$$\overline{P} = \frac{E}{t} = 2\pi^2 \rho Svf^2A^2. \qquad \textbf{(15–6)}$$

Finally, the **intensity**, I, of a wave is defined as the average power (energy per unit time) transported across unit area perpendicular to the direction of energy flow:

$$I = \frac{\overline{P}}{S} = 2\pi^2 v\rho f^2A^2. \qquad \textbf{(15–7a)}$$

The intensity too is proportional to amplitude A squared,

$$I \propto A^2, \qquad \textbf{(15–7b)}$$

and its SI unit is watts per square meter $(\mathrm{W/m^2})$.

If a wave flows out from the source in all directions, it is a three-dimensional wave. Examples are sound traveling in the open air, earthquake waves, and light waves. If the medium is isotropic (same in all directions), the wave from a point source is a *spherical wave* (Fig. 15–12). As the wave moves outward, the energy it carries is spread over a larger and larger area since the surface area of a sphere of radius r is $4\pi r^2$. Thus the intensity of a wave is

$$I = \frac{\overline{P}}{S} = \frac{\overline{P}}{4\pi r^2}.$$

If the power output $\overline{P}$ of the source is constant, then the intensity decreases as the inverse square of the distance from the source:

$$I \propto \frac{1}{r^2}. \qquad \text{[spherical wave]} \quad \textbf{(15–8a)}$$

This is often called the **inverse square law**, or "one over r^2 law."

If we consider two points at distances r_1 and r_2 from the source, as in Fig. 15–12, then $I_1 = \overline{P}/4\pi r_1^2$ and $I_2 = \overline{P}/4\pi r_2^2$, so

$$\frac{I_2}{I_1} = \frac{r_1^2}{r_2^2}. \qquad \textbf{(15–8b)}$$

Thus, for example, when the distance doubles $(r_2/r_1 = 2)$, then the intensity is reduced to $\frac{1}{4}$ of its earlier value: $I_2/I_1 = (\frac{1}{2})^2 = \frac{1}{4}$.

The amplitude of a wave also decreases with distance. Since the intensity is proportional to the square of the amplitude (Eq. 15–7), $I \propto A^2$, the amplitude A must decrease as $1/r$ (so that $I \propto A^2$ will be proportional to $1/r^2$, as in Eq. 15–8a). Hence

$$A \propto \frac{1}{r}.$$

For two different distances from the source, r_1 and r_2, then

$$\frac{A_2}{A_1} = \frac{r_1}{r_2}.$$

When the wave is twice as far from the source, the amplitude is half as large, and so on (ignoring damping due to friction).

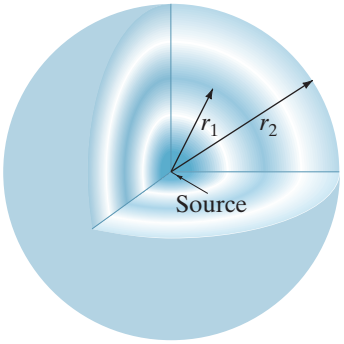


FIGURE 15–12 A wave traveling uniformly outward in three dimensions from a source is spherical. Two crests (or compressions) are shown, of radii r_1 and r_2.

EXAMPLE 15–5 **Earthquake intensity.** The intensity of an earthquake P wave traveling through the Earth and detected 100 km from the source is $1.0 \times 10^6\ \text{W/m}^2$. What is the intensity of that wave if detected 400 km from the source?

APPROACH We assume the wave is spherical, so the intensity decreases as the square of the distance from the source.

SOLUTION At 400 km the distance is 4 times greater than at 100 km, so the intensity will be $(\frac{1}{4})^2 = \frac{1}{16}$ of its value at 100 km, or $(1.0 \times 10^6\ \text{W/m}^2)/16 = 6.3 \times 10^4\ \text{W/m}^2$.

NOTE Using Eq. 15–8b directly gives

$$I_2 = I_1 r_1^2/r_2^2 = (1.0 \times 10^6\ \text{W/m}^2)(100\ \text{km})^2/(400\ \text{km})^2 = 6.3 \times 10^4\ \text{W/m}^2.$$

The situation is different for a one-dimensional wave, such as a transverse wave on a string or a longitudinal wave pulse traveling down a thin uniform metal rod. The area remains constant, so the amplitude A also remains constant (ignoring friction). Thus the amplitude and the intensity do not decrease with distance.

In practice, frictional damping is generally present, and some of the energy is transformed into thermal energy. Thus the amplitude and intensity of a one-dimensional wave will decrease with distance from the source. For a three-dimensional wave, the decrease will be greater than that discussed above, more than $1/r^2$, although the effect may often be small.

15–4 Mathematical Representation of a Traveling Wave

Consider a one-dimensional wave traveling along the x axis. It could be a transverse wave on a cord or a longitudinal wave traveling in a solid rod or in a fluid-filled tube. Let us assume the wave shape is sinusoidal and has a particular wavelength λ and frequency f. At $t = 0$, suppose the wave shape is given by

$$D(x) = A \sin \frac{2\pi}{\lambda} x, \qquad \textbf{(15–9)}$$

as shown by the solid curve in Fig. 15–13. $D(x)$ is the **displacement** of the wave (either longitudinal or transverse) at position x, and A is the **amplitude** (maximum displacement) of the wave. Equation 15–9 works because it repeats itself every wavelength: the displacement is the same at $x = 0$, $x = \lambda$, $x = 2\lambda$, and so on because $\sin 4\pi = \sin 2\pi = \sin 0$.

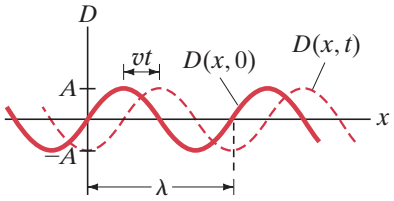


FIGURE 15–13 A traveling wave. In time t, the wave moves a distance vt. The solid line represents the wave at $t = 0$. The dashed line is $D(x, t)$, the wave displacement as a function of x at time t.

Suppose the wave is moving to the right with speed v. After a time t, each part of the wave (indeed, the whole wave "shape") has moved to the right a distance vt; see the dashed curve in Fig. 15–13. Consider any point on the wave at $t = 0$: say, a crest at some position x. After a time t, that crest will have traveled a distance vt so its new position is a distance vt larger than its old position. To describe this crest (or other point on the wave shape), the argument of the sine function must have the same numerical value; so we replace x in Eq. 15–9 by $(x - vt)$, and write the displacement as $D(x, t)$ since it depends on both x and t:

$$D(x, t) = A \sin\left[\frac{2\pi}{\lambda}(x - vt)\right]. \quad \begin{bmatrix}\text{1-D wave moving} \\ \text{in } +x \text{ direction}\end{bmatrix} \qquad \textbf{(15–10a)}$$

Said another way: if you are riding on a crest, as t increases, x must increase at the same rate so that the value of $(x - vt)$ remains constant.

Equation 15–10a gives the displacement $D(x, t)$ of a wave traveling at speed v along the x axis to the right (increasing x) at any point x at any time t.

The function $D(x,t)$ describes a curve that represents the actual shape of the wave in space at time t. Since $v = \lambda f$ (Eq. 15–1) we can write Eq. 15–10a in other ways that are often convenient:

$$D(x,t) = A\sin\left(\frac{2\pi x}{\lambda} - \frac{2\pi t}{T}\right), \qquad \textbf{(15–10b)}$$

where $T = 1/f = \lambda/v$ is the period; and more simply as

$$D(x,t) = A\sin(kx - \omega t), \qquad \textbf{(15–10c)}$$

where $\omega = 2\pi f = 2\pi/T$ is the angular frequency and

$$k = \frac{2\pi}{\lambda} \qquad \textbf{(15–11)}$$

CAUTION

Don't confuse wave number k with spring constant k

is called the **wave number**. (Do not confuse the wave number k with the spring constant k; they are very different quantities.) All three forms, Eqs. 15–10a, b, and c, are equivalent; Eq. 15–10c is the simplest to write and is perhaps the most common. The quantity $(kx - \omega t)$, and its equivalent in the other two equations, is called the **phase** of the wave. The speed of the wave v is often referred to as the **phase velocity**, since it describes the velocity of the phase (or shape) of the wave, and it can be written in terms of ω and k:

$$v = \lambda f = \left(\frac{2\pi}{k}\right)\left(\frac{\omega}{2\pi}\right) = \frac{\omega}{k}. \qquad \textbf{(15–12)}$$

Equations 15–10 apply for a wave moving to the right along the x axis. For a wave traveling along the x axis to the *left*, toward decreasing values of x, v becomes $-v$. So

$$D(x,t) = A\sin\left[\frac{2\pi}{\lambda}(x + vt)\right] \qquad \left[\begin{array}{l}\text{1-D wave moving}\\ \text{in } -x \text{ direction}\end{array}\right] \qquad \textbf{(15–13a)}$$

$$= A\sin\left(\frac{2\pi x}{\lambda} + \frac{2\pi t}{T}\right) \qquad \textbf{(15–13b)}$$

$$= A\sin(kx + \omega t). \qquad \textbf{(15–13c)}$$

Let us look at Eq. 15–13c (or, just as well, at Eq. 15–10c). At $t = 0$ we have

$$D(x,0) = A\sin kx,$$

which is what we started with, a sinusoidal wave shape. If we look at the wave shape in space at a particular later time t_1, then we have

$$D(x,t_1) = A\sin(kx + \omega t_1).$$

That is, if we took a picture of the wave at $t = t_1$, we would still see a sine wave in space but shifted by an amount ωt_1. On the other hand, if we consider a fixed point in space, say $x = 0$ in Eq. 15–13c, we can see how the wave varies in time:

$$D(0,t) = A\sin\omega t.$$

This is the equation for simple harmonic motion (Section 14–2). For any other fixed value of x, say $x = x_1$, $D = A\sin(\omega t + kx_1)$ which differs only by a constant kx_1. Thus, at any point in space, the displacement undergoes the oscillations of simple harmonic motion in time. Equations 15–10 and 15–13 combine both these aspects to give us the representation for a **traveling sinusoidal wave** (also called a **harmonic wave**).

The argument of the sine in Eqs. 15–10 and 15–13 can in general contain a **phase constant** or **phase angle**, ϕ, which for Eq. 15–10c is

$$D(x,t) = A\sin(kx - \omega t + \phi),$$

to adjust for the position of the wave at $t = 0$, $x = 0$, just as in Section 14–2 (see Fig. 14–6). If the displacement is zero at $t = 0$, $x = 0$, as in Fig. 14–5 (or Fig. 15–13), then $\phi = 0$.

EXERCISE D A wave is given by $D(x,t) = (5.0\,\text{mm})\sin(2.0x - 20.0t)$ where x is in meters and t is in seconds. What is the speed of the wave? (*a*) 10 m/s, (*b*) 0.10 m/s, (*c*) 40 m/s, (*d*) 0.005 m/s, (*e*) 2.5×10^{-4} m/s.

EXAMPLE 15–6 **A traveling wave.** The left-hand end of a long horizontal stretched cord oscillates transversely in SHM with frequency $f = 25$ Hz and amplitude 2.6 cm. The cord is under a tension of 140 N and has a linear density $\mu = 0.12$ kg/m. At $t = 0$, the end of the cord has a positive displacement of 1.6 cm and is falling (Fig. 15–14). Determine (a) the wavelength of waves produced, (b) the equation for the traveling wave, and (c) the speed v' of a particle of cord at $x = 1.00$ m.

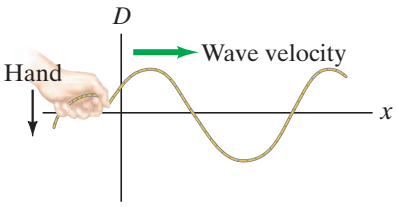


FIGURE 15–14 Example 15–6. The wave at $t = 0$ (the hand is falling). Not to scale.

APPROACH We first find the speed of the transverse wave from Eq. 15–2; then we calculate $\lambda = v/f$. In (b), we need to find the phase ϕ using the initial conditions.

SOLUTION (a) The wave speed (or phase velocity) is, from Eq. 15–2,

$$v = \sqrt{\frac{F_T}{\mu}} = \sqrt{\frac{140\text{ N}}{0.12\text{ kg/m}}} = 34\text{ m/s}.$$

Then

$$\lambda = \frac{v}{f} = \frac{34\text{ m/s}}{25\text{ Hz}} = 1.4\text{ m}.$$

(b) At the left-hand end of the cord $x = 0$. The phase of the wave at $t = 0$ is not zero in general as was assumed in Eqs. 15–9, 10, and 13. The general form for a wave traveling to the right is

$$D(x, t) = A\sin(kx - \omega t + \phi),$$

where ϕ is the phase angle. In our case, the amplitude $A = 2.6$ cm; and at $t = 0$, $x = 0$, we are given $D = 1.6$ cm. Thus

$$1.6\text{ cm} = (2.6\text{ cm})\sin\phi,$$

so $\phi = \sin^{-1}(1.6/2.6) = 38° = 0.66$ rad. We also have $\omega = 2\pi f = (2\pi\text{ rad})(25\text{ s}^{-1}) = 157$ rad/s and $k = 2\pi/\lambda = 2\pi\text{ rad}/1.4\text{ m} = 4.5$ rad/m. Hence

$$D = (0.026\text{ m})\sin[(4.5\text{ rad/m})x - (157\text{ rad/s})t + 0.66\text{ rad}]$$

which we can write more simply without detailed units as

$$D = 0.026\sin(4.5x - 157t + 0.66),$$

and we specify clearly that D and x are in meters, t in seconds, and trig functions using rads.

(c) The velocity v' of a particle of cord at $x = 1.00$ m as a function of time is

$$v' = \frac{d}{dt}[(0.026\text{ m})\sin(4.5 - (157\text{ rad/s})t + 0.66)]$$
$$= -(0.026\text{ m})(157\text{ s}^{-1})\cos(4.5 - (157\text{ rad/s})t + 0.66).$$

The speed of this particle varies from zero at times when the cosine = 0, to a maximum (when cos = 1) of (0.026)(157) m/s = 4.1 m/s (vertically). The wave speed, on the other hand, is a constant 34 m/s horizontally.

*A General Wave (non-sinusoidal)

Now let us consider a general wave (or wave pulse) of any shape. If frictional losses are small, experiment shows that the wave maintains its shape as it travels. Thus we can make the same arguments as we did right after Eq. 15–9. Suppose our wave has some shape at $t = 0$, given by

$$D(x, 0) = D(x)$$

where $D(x)$ is the displacement of the wave at x and is not necessarily sinusoidal. At some later time, if the wave is traveling to the right along the x axis, the wave will have the same shape but all parts will have moved a distance vt where v is the phase velocity of the wave. Hence we replace x by $x - vt$ to obtain the displacement at time t:

$$D(x, t) = D(x - vt). \qquad \textbf{(15–14)}$$

Similarly, if the wave moves to the left, we replace x by $x + vt$, so

$$D(x, t) = D(x + vt). \qquad \textbf{(15–15)}$$

Thus, any wave traveling along the x axis has the form of Eq. 15–14 or 15–15.

*15–5 The Wave Equation

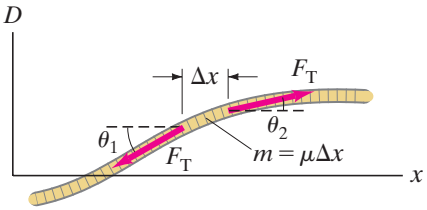


FIGURE 15–15 Deriving the wave equation from Newton's second law: a segment of string under tension F_T.

Many types of waves satisfy an important general equation that is the equivalent of Newton's second law of motion for particles. This "equation of motion for a wave" is called the **wave equation**, and we derive it now for a one-dimensional wave traveling on a stretched horizontal string.

We assume the amplitude of the wave is small compared to the wavelength so that each point on the string can be assumed to move only vertically and the tension in the string, F_T, does not vary during a vibration. We apply Newton's second law, $\Sigma F = ma$, to the vertical motion of a tiny section of the string as shown in Fig. 15–15. The amplitude of the wave is small, so the angles θ_1 and θ_2 that the string makes with the horizontal are small. The length of this section is then approximately Δx, and its mass is $\mu\,\Delta x$, where μ is the mass per unit length of the string. The net vertical force on this section of string is $F_T \sin\theta_2 - F_T \sin\theta_1$ (see Fig. 15–15). So Newton's second law applied to the vertical direction gives

$$\Sigma F = ma$$

$$F_T \sin\theta_2 - F_T \sin\theta_1 = (\mu\,\Delta x)\frac{\partial^2 D}{\partial t^2}. \qquad \textbf{(i)}$$

We have written the acceleration as $a = \partial^2 D/\partial t^2$ and $m = \mu\,\Delta x$. We use the partial derivative notation because the displacement D is a function of both x and t: that is, $D = D(x, t)$. On the left of Eq. (i) the angles θ_1 and θ_2 are assumed small, so $\sin\theta \approx \tan\theta$ and $\tan\theta$ is equal to the slope s of the string at each point:

$$\sin\theta \approx \tan\theta = \frac{\partial D}{\partial x} = s.$$

Thus our equation (i) becomes

$$F_T(s_2 - s_1) = \mu\,\Delta x\frac{\partial^2 D}{\partial t^2}$$

or

$$F_T\frac{\Delta s}{\Delta x} = \mu\frac{\partial^2 D}{\partial t^2}, \qquad \textbf{(ii)}$$

where $\Delta s = s_2 - s_1$ is the difference in the slope between the two ends of our tiny section. Now we take the limit of $\Delta x \to 0$, so that

$$F_T \lim_{\Delta x\to 0}\frac{\Delta s}{\Delta x} = F_T\frac{\partial s}{\partial x}$$

$$= F_T\frac{\partial}{\partial x}\left(\frac{\partial D}{\partial x}\right) = F_T\frac{\partial^2 D}{\partial x^2}$$

since the slope $s = \partial D/\partial x$, as we wrote above. Substituting this into the equation labeled (ii) above gives

$$F_T\frac{\partial^2 D}{\partial x^2} = \mu\frac{\partial^2 D}{\partial t^2}$$

or

$$\frac{\partial^2 D}{\partial x^2} = \frac{\mu}{F_T}\frac{\partial^2 D}{\partial t^2}.$$

We saw earlier in this Chapter (Eq. 15–2) that the speed of waves on a string is given by $v = \sqrt{F_T/\mu}$, so we can write this last equation as

$$\frac{\partial^2 D}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 D}{\partial t^2}. \qquad \textbf{(15–16)}$$

This is the **one-dimensional wave equation**, and it can describe not only small amplitude waves on a stretched string, but also small amplitude longitudinal waves (such as sound waves) in gases, liquids, and elastic solids, in which case D can refer to the pressure variations. In this case, the wave equation is a direct consequence of Newton's second law applied to a continuous elastic medium.

The wave equation also describes electromagnetic waves for which D refers to the electric or magnetic field, as we shall see in Chapter 31. Equation 15–16 applies to waves traveling in one dimension only. For waves spreading out in three dimensions, the wave equation is the same, with the addition of $\partial^2 D/\partial y^2$ and $\partial^2 D/\partial z^2$ to the left side of Eq. 15–16.

The wave equation is a *linear* equation: the displacement D appears singly in each term. There are no terms that contain D^2, or $D(\partial D/\partial x)$, or the like in which D appears more than once. Thus, if $D_1(x, t)$ and $D_2(x, t)$ are two different solutions of the wave equation, then the linear combination

$$D_3(x, t) = aD_1(x, t) + bD_2(x, t),$$

where a and b are constants, is also a solution. This is readily seen by direct substitution into the wave equation. This is the essence of the *superposition principle*, which we discuss in the next Section. Basically it says that if two waves pass through the same region of space at the same time, the actual displacement is the sum of the separate displacements. For waves on a string, or for sound waves, this is valid only for small-amplitude waves. If the amplitude is not small enough, the equations for wave propagation may become nonlinear and the principle of superposition would not hold and more complicated effects may occur.

EXAMPLE 15–7 **Wave equation solution.** Verify that the sinusoidal wave of Eq. 15–10c, $D(x, t) = A \sin(kx - \omega t)$, satisfies the wave equation.

APPROACH We substitute Eq. 15–10c into the wave equation, Eq. 15–16.

SOLUTION We take the derivative of Eq. 15–10c twice with respect to t:

$$\frac{\partial D}{\partial t} = -\omega A \cos(kx - \omega t)$$

$$\frac{\partial^2 D}{\partial t^2} = -\omega^2 A \sin(kx - \omega t).$$

With respect to x, the derivatives are

$$\frac{\partial D}{\partial x} = kA \cos(kx - \omega t)$$

$$\frac{\partial^2 D}{\partial x^2} = -k^2 A \sin(kx - \omega t).$$

If we now divide the second derivatives we get

$$\frac{\partial^2 D/\partial t^2}{\partial^2 D/\partial x^2} = \frac{-\omega^2 A \sin(kx - \omega t)}{-k^2 A \sin(kx - \omega t)} = \frac{\omega^2}{k^2}.$$

From Eq. 15–12 we have $\omega^2/k^2 = v^2$, so we see that Eq. 15–10 does satisfy the wave equation (Eq. 15–16).

15–6 The Principle of Superposition

When two or more waves pass through the same region of space at the same time, it is found that for many waves *the actual displacement is the vector* (or *algebraic*) *sum of the separate displacements*. This is called the **principle of superposition**. It is valid for mechanical waves as long as the displacements are not too large and there is a linear relationship between the displacement and the restoring force of the oscillating medium.[†] If the amplitude of a mechanical wave, for example, is so large that it goes beyond the elastic region of the medium, and Hooke's law is no longer operative, the superposition principle is no longer accurate.[‡] For the most part, we will consider systems for which the superposition principle can be assumed to hold.

[†]For electromagnetic waves in vacuum, Chapter 31, the superposition principle always holds.

[‡]Intermodulation distortion in audio equipment is an example of the superposition principle not holding when two frequencies do not combine linearly in the electronics.

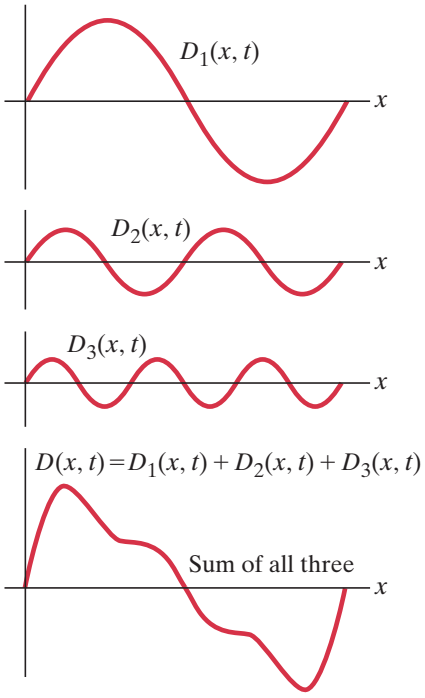


FIGURE 15–16 The superposition principle for one-dimensional waves. The composite wave formed from three sinusoidal waves of different amplitudes and frequencies $(f_0, 2f_0, 3f_0)$ is shown at the bottom at a certain instant in time. The displacement of the composite wave at each point in space, at any time, is the algebraic sum of the displacements of the component waves. Amplitudes are shown exaggerated; for the superposition principle to hold, they must be small compared to the wavelengths.

One result of the superposition principle is that if two waves pass through the same region of space, they continue to move independently of one another. You may have noticed, for example, that the ripples on the surface of water (two-dimensional waves) that form from two rocks striking the water at different places will pass right through each other.

Figure 15–16 shows an example of the superposition principle. In this case there are three sinusoidal waves present on a stretched string, each of different amplitude and frequency. At any time, such as at the instant shown, the actual displacement at any position x is the algebraic sum of the displacements of the three waves at that position. The actual wave is not a simple sinusoidal wave and is called a *composite* (or *complex*) *wave*. (Amplitudes are exaggerated in Fig. 15–16.)

It can be shown that any complex wave can be considered as being composed of many simple sinusoidal waves of different amplitudes, wavelengths, and frequencies. This is known as *Fourier's theorem*. A complex periodic wave of period T can be represented as a sum of pure sinusoidal terms whose frequencies are integral multiples of $f = 1/T$. If the wave is not periodic, the sum becomes an integral (called a *Fourier integral*). Although we will not go into the details here, we see the importance of considering sinusoidal waves (and simple harmonic motion): because any other wave shape can be considered a sum of such pure sinusoidal waves.

CONCEPTUAL EXAMPLE 15–8 **Making a square wave.** At $t = 0$, three waves are given by $D_1 = A\cos kx$, $D_2 = -\frac{1}{3}A\cos 3kx$, and $D_3 = \frac{1}{5}A\cos 5kx$, where $k = 10\,\text{m}^{-1}$ and $A = 1.0\,\text{mm} = 1.0 \times 10^{-3}\,\text{m}$. Plot the sum of the three waves from $x = -0.4\,\text{m}$ to $+0.4\,\text{m}$. (These three waves are the first three Fourier components of a "square wave.")

RESPONSE The first wave, D_1, has amplitude of 1.0 mm and wavelength $\lambda = 2\pi/k = (2\pi/10)\,\text{m} = 0.628\,\text{m}$. The second wave, D_2, has amplitude of 0.33 mm and wavelength $\lambda = 2\pi/3k = (2\pi/30)\,\text{m} = 0.209\,\text{m}$. The third wave, D_3, has amplitude of 0.20 mm and wavelength $\lambda = 2\pi/5k = (2\pi/50)\,\text{m} = 0.126\,\text{m}$. Each wave is plotted in Fig. 15–17a. The sum of the three waves is shown in Fig. 15–17b. The sum begins to resemble a "square wave," shown in blue in Fig. 15–17b.

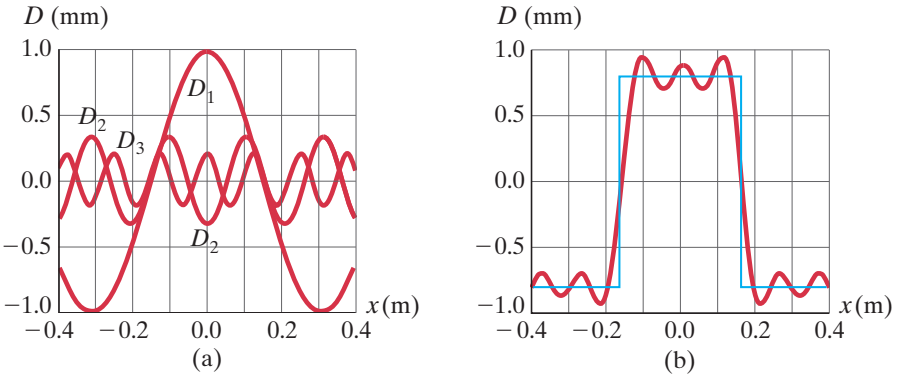


FIGURE 15–17 Example 15–8. Making a square wave.

PHYSICS APPLIED
Square wave (useful in electronics)

When the restoring force is not precisely proportional to the displacement for mechanical waves in some continuous medium, the speed of sinusoidal waves depends on the frequency. The variation of speed with frequency is called **dispersion**. The different sinusoidal waves that compose a complex wave will travel with slightly different speeds in such a case. Consequently, a complex wave will change shape as it travels if the medium is "dispersive." A pure sine wave will not change shape under these conditions, however, except by the influence of friction or dissipative forces. If there is no dispersion (or friction), even a complex linear wave does not change shape.

PHYSICS APPLIED
When wave speed depends on frequency, a composite wave changes shape over time

15–7 Reflection and Transmission

When a wave strikes an obstacle, or comes to the end of the medium in which it is traveling, at least a part of the wave is reflected. You have probably seen water waves reflect off a rock or the side of a swimming pool. And you may have heard a shout reflected from a distant cliff—which we call an "echo."

A wave pulse traveling along a cord is reflected as shown in Fig. 15–18 (time increasing going downward in both a and b). The reflected pulse returns inverted as in Fig. 15–18a if the end of the cord is fixed; it returns right side up if the end is free as in Fig. 15–18b. When the end is fixed to a support, as in Fig. 15–18a, the pulse reaching that fixed end exerts a force (upward) on the support. The support exerts an equal but opposite force downward on the cord (Newton's third law). This downward force on the cord is what causes the inverted reflected pulse.

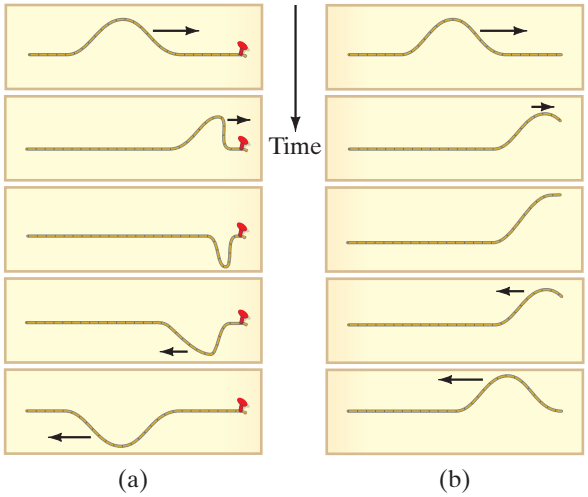
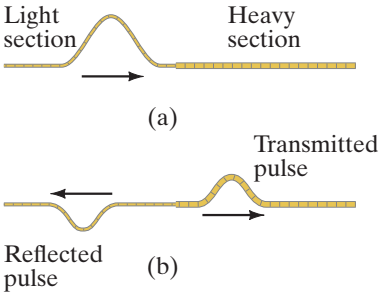


FIGURE 15–18 Reflection of a wave pulse traveling along a cord lying on a table. (Time increases going down.) (a) The end of the cord is fixed to a peg. (b) The end of the cord is free to move.

Consider next a pulse that travels along a cord which consists of a light section and a heavy section, as shown in Fig. 15–19. When the wave pulse reaches the boundary between the two sections, part of the pulse is reflected and part is transmitted, as shown. The heavier the second section of the cord, the less the energy that is transmitted. Indeed, when the second section is a wall or rigid support, very little is transmitted and most is reflected, as in Fig. 15–18a. For a sinusoidal wave, the *frequency of the transmitted wave does not change* as it crosses the boundary (the boundary point oscillates at that frequency). Thus if the transmitted wave has a lower speed, its wavelength is also less $(\lambda = v/f)$.

FIGURE 15–19 When a wave pulse traveling to the right along a thin cord (a) reaches a discontinuity where the cord becomes thicker and heavier, then part is reflected and part is transmitted (b).



For a two or three dimensional wave, such as a water wave, we are concerned with **wave fronts**, by which we mean all the points along the wave forming the wave crest (what we usually refer to simply as a "wave" at the seashore). A line drawn in the direction of wave motion, perpendicular to the wave front, is called a **ray**, as shown in Fig. 15–20. Wave fronts far from the source have lost almost all their curvature (Fig. 15–20b) and are nearly straight, as ocean waves often are. They are then called **plane waves**.

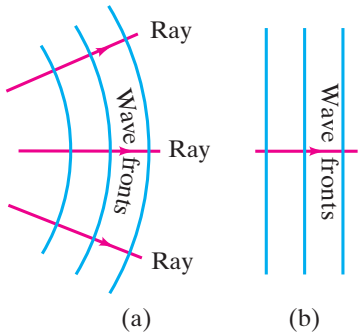


FIGURE 15–20 Rays, signifying the direction of wave motion, are always perpendicular to the wave fronts (wave crests). (a) Circular or spherical waves near the source. (b) Far from the source, the wave fronts are nearly straight or flat, and are called plane waves.

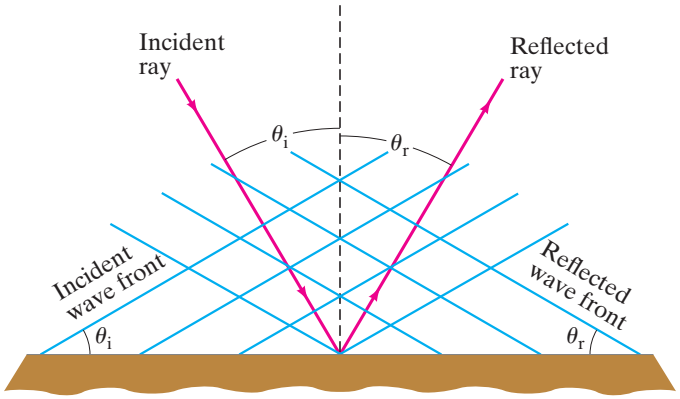


FIGURE 15–21
Law of reflection: $\theta_r = \theta_i$.

For reflection of a two- or three-dimensional plane wave, as shown in Fig. 15–21, the angle that the incoming or *incident wave* makes with the reflecting surface is equal to the angle made by the reflected wave. This is the **law of reflection**:

the angle of reflection equals the angle of incidence.

The **angle of incidence** is defined as the angle (θ_i) the incident ray makes with the perpendicular to the reflecting surface (or the wave front makes with a tangent to the surface). The **angle of reflection** is the corresponding angle (θ_r) for the reflected wave.

15–8 Interference

Interference refers to what happens when two waves pass through the same region of space at the same time. Consider, for example, the two wave pulses on a cord traveling toward each other as shown in Fig. 15–22 (time increases downward in both a and b). In Fig. 15–22a the two pulses have the same amplitude, but one is a crest and the other a trough; in Fig. 15–22b they are both crests. In both cases, the waves meet and pass right by each other. However, in the region where they overlap, the resultant displacement is the *algebraic sum of their separate displacements* (a crest is considered positive and a trough negative). This is another example of the principle of superposition. In Fig. 15–22a, the two waves have opposite displacements at the instant they pass one another, and they add to zero. The result is called **destructive interference**. In Fig. 15–22b, at the instant the two pulses overlap, they produce a resultant displacement that is greater than the displacement of either separate pulse, and the result is **constructive interference**.

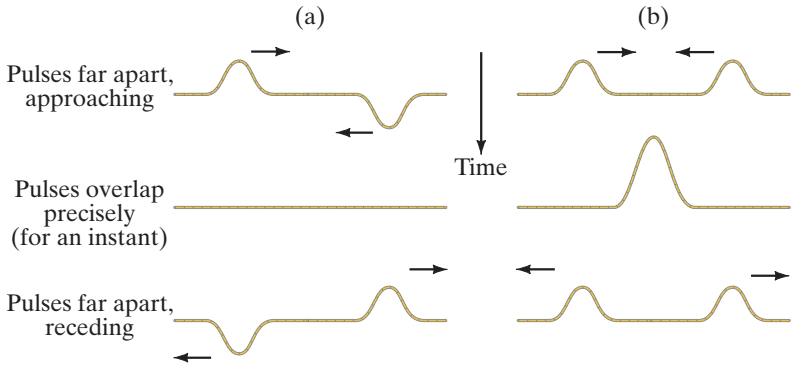


FIGURE 15–22 Two wave pulses pass each other. Where they overlap, interference occurs: (a) destructive, and (b) constructive. Read (a) and (b) downward (increasing time).

You may wonder where the energy is at the moment of destructive interference in Fig. 15–22a; the cord may be straight at this instant, but the central parts of it are still moving up or down (kinetic energy).

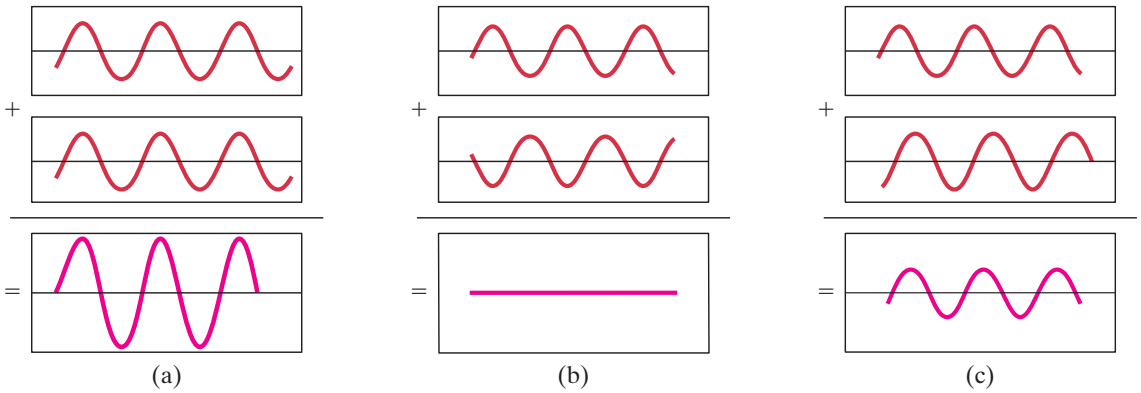


FIGURE 15–23 (a) Interference of water waves. (b) Constructive interference occurs where one wave's maximum (a crest) meets the other's maximum. Destructive interference ("flat water") occurs where one wave's maximum (a crest) meets the other's minimum (a trough).

When two rocks are thrown into a pond simultaneously, the two sets of circular waves that move outward interfere with one another as shown in Fig. 15–23a. In some areas of overlap, crests of one wave repeatedly meet crests of the other (and troughs meet troughs), Fig. 15–23b. Constructive interference is occurring at these points, and the water continuously oscillates up and down with greater amplitude than either wave separately. In other areas, destructive interference occurs where the water does not move up and down at all over time. This is where crests of one wave meet troughs of the other, and vice versa. Figure 15–24a shows the displacement of two identical waves graphically as a function of time, as well as their sum, for the case of constructive interference. For any two such waves, we use the term **phase** to describe the relative positions of their crests. When the crests and troughs are aligned as in Fig. 15–24a, for constructive interference, the two waves are **in phase**. At points where destructive interference occurs (Fig. 15–24b), crests of one wave repeatedly meet troughs of the other wave and the two waves are said to be completely **out of phase** or, more precisely, out of phase by one-half wavelength (or π radians, or 180°).[†] That is, the crests of one wave occur a half wavelength behind the crests of the other wave. The relative phase of the two water waves in Fig. 15–23 in most areas is intermediate between these two extremes, resulting in *partially* destructive interference, as illustrated in Fig. 15–24c. If the amplitudes of two interfering waves are not equal, fully destructive interference (as in Fig. 15–24b) does not occur.

[†]One wavelength, or one full oscillation, corresponds to 360° or 2π radians.

FIGURE 15–24 Graphs showing two identical waves, and their sum, as a function of time at three locations. In (a) the two waves interfere constructively, in (b) destructively, and in (c) partially destructively.



15–9 Standing Waves; Resonance

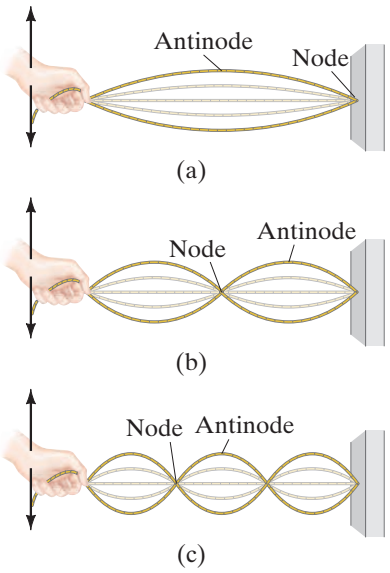


FIGURE 15–25 Standing waves corresponding to three resonant frequencies.

If you shake one end of a cord and the other end is kept fixed, a continuous wave will travel down to the fixed end and be reflected back, inverted, as we saw in Fig. 15–18a. As you continue to oscillate the cord, waves will travel in both directions, and the wave traveling along the cord, away from your hand, will interfere with the reflected wave coming back. Usually there will be quite a jumble. But if you oscillate the cord at just the right frequency, the two traveling waves will interfere in such a way that a large-amplitude **standing wave** will be produced, Fig. 15–25. It is called a "standing wave" because it does not appear to be traveling. The cord simply appears to have segments that oscillate up and down in a fixed pattern. The points of destructive interference, where the cord remains still at all times, are called **nodes**. Points of constructive interference, where the cord oscillates with maximum amplitude, are called **antinodes**. The nodes and antinodes remain in fixed positions for a particular frequency.

Standing waves can occur at more than one frequency. The lowest frequency of oscillation that produces a standing wave gives rise to the pattern shown in Fig. 15–25a. The standing waves shown in Figs. 15–25b and 15–25c are produced at precisely twice and three times the lowest frequency, respectively, assuming the tension in the cord is the same. The cord can also oscillate with four loops (four antinodes) at four times the lowest frequency, and so on.

The frequencies at which standing waves are produced are the **natural frequencies** or **resonant frequencies** of the cord, and the different standing wave patterns shown in Fig. 15–25 are different "resonant modes of vibration." A standing wave on a cord is the result of the interference of two waves traveling in opposite directions. A standing wave can also be considered a vibrating object at resonance. Standing waves represent the same phenomenon as the resonance of an oscillating spring or pendulum, which we discussed in Section 14–8. However, a spring or pendulum has only one resonant frequency, whereas the cord has an infinite number of resonant frequencies, each of which is a whole-number multiple of the lowest resonant frequency.

Consider a string stretched between two supports that is plucked like a guitar or violin string, Fig. 15–26a. Waves of a great variety of frequencies will travel in both directions along the string, will be reflected at the ends, and will travel back in the opposite direction. Most of these waves interfere with each other and quickly die out. However, those waves that correspond to the resonant frequencies of the string will persist. The ends of the string, since they are fixed, will be nodes. There may be other nodes as well. Some of the possible *resonant modes of vibration* (standing waves) are shown in Fig. 15–26b. Generally, the motion will be a combination of these different resonant modes, but only those frequencies that correspond to a resonant frequency will be present.

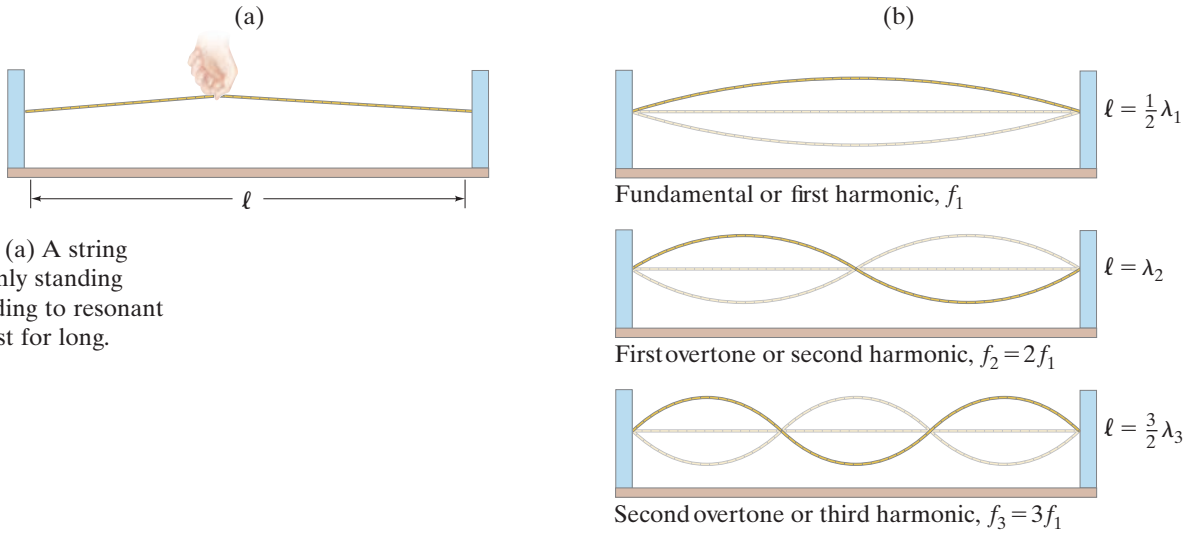


FIGURE 15–26 (a) A string is plucked. (b) Only standing waves corresponding to resonant frequencies persist for long.

To determine the resonant frequencies, we first note that the wavelengths of the standing waves bear a simple relationship to the length ℓ of the string. The lowest frequency, called the **fundamental frequency**, corresponds to one antinode (or loop). And as can be seen in Fig. 15–26b, the whole length corresponds to one-half wavelength. Thus $\ell = \frac{1}{2}\lambda_1$, where λ_1 stands for the wavelength of the fundamental frequency. The other natural frequencies are called **overtones**; for a vibrating string they are whole-number (integral) multiples of the fundamental, and then are also called **harmonics**, with the fundamental being referred to as the **first harmonic**.[†] The next mode of vibration after the fundamental has two loops and is called the **second harmonic** (or first overtone), Fig. 15–26b. The length of the string ℓ at the second harmonic corresponds to one complete wavelength: $\ell = \lambda_2$. For the third and fourth harmonics, $\ell = \frac{3}{2}\lambda_3$, and $\ell = 2\lambda_4$, respectively, and so on. In general, we can write

$$\ell = \frac{n\lambda_n}{2}, \qquad \text{where } n = 1, 2, 3, \cdots.$$

The integer n labels the number of the harmonic: $n = 1$ for the fundamental, $n = 2$ for the second harmonic, and so on. We solve for λ_n and find

$$\lambda_n = \frac{2\ell}{n}, \qquad n = 1, 2, 3, \cdots. \qquad \begin{bmatrix}\text{string fixed}\\ \text{at both ends}\end{bmatrix} \qquad \textbf{(15–17a)}$$

To find the frequency f of each vibration we use Eq. 15–1, $f = v/\lambda$, and see that

$$f_n = \frac{v}{\lambda_n} = n\frac{v}{2\ell} = nf_1, \qquad n = 1, 2, 3, \cdots, \qquad \textbf{(15–17b)}$$

where $f_1 = v/\lambda_1 = v/2\ell$ is the fundamental frequency. We see that each resonant frequency is an *integer multiple of the fundamental frequency*.

Because a standing wave is equivalent to two traveling waves moving in opposite directions, the concept of wave speed still makes sense and is given by Eq. 15–2 in terms of the tension F_T in the string and its mass per unit length ($\mu = m/\ell$). That is, $v = \sqrt{F_T/\mu}$ for waves traveling in both directions.

EXAMPLE 15–9 **Piano string.** A piano string is 1.10 m long and has mass 9.00 g. (*a*) How much tension must the string be under if it is to vibrate at a fundamental frequency of 131 Hz? (*b*) What are the frequencies of the first four harmonics?

APPROACH To determine the tension, we need to find the wave speed using Eq. 15–1 ($v = \lambda f$), and then use Eq. 15–2, solving it for F_T.

SOLUTION (*a*) The wavelength of the fundamental is $\lambda = 2\ell = 2.20\,\text{m}$ (Eq. 15–17a with $n = 1$). The speed of the wave on the string is $v = \lambda f = (2.20\,\text{m})(131\,\text{s}^{-1}) = 288\,\text{m/s}$. Then we have (Eq. 15–2)

$$F_T = \mu v^2 = \frac{m}{\ell}v^2 = \left(\frac{9.00\times10^{-3}\,\text{kg}}{1.10\,\text{m}}\right)(288\,\text{m/s})^2 = 679\,\text{N}.$$

(*b*) The first harmonic (the fundamental) has a frequency $f_1 = 131\,\text{Hz}$. The frequencies of the second, third, and fourth harmonics are two, three, and four times the fundamental frequency: 262, 393, and 524 Hz, respectively.

NOTE The speed of the wave on the string is *not* the same as the speed of the sound wave that the piano string produces in the air (as we shall see in Chapter 16).

A standing wave does appear to be standing in place (and a traveling wave appears to move). The term "standing" wave is also meaningful from the point of view of energy. Since the string is at rest at the nodes, no energy flows past these points. Hence the energy is not transmitted down the string but "stands" in place in the string.

Standing waves are produced not only on strings, but on any object that is struck, such as a drum membrane or an object made of metal or wood. The resonant frequencies depend on the dimensions of the object, just as for a string they depend on its length. Large objects have lower resonant frequencies than small objects. All musical instruments, from stringed instruments to wind instruments (in which a column of air vibrates as a standing wave) to drums and other percussion instruments, depend on standing waves to produce their particular musical sounds, as we shall see in Chapter 16.

[†]The term "harmonic" comes from music, because such integral multiples of frequencies "harmonize."

Mathematical Representation of a Standing Wave

As we saw, a standing wave can be considered to consist of two traveling waves that move in opposite directions. These can be written (see Eqs. 15–10c and 15–13c)

$$D_1(x,t) = A\sin(kx - \omega t) \quad \text{and} \quad D_2(x,t) = A\sin(kx + \omega t)$$

since, assuming no damping, the amplitudes are equal as are the frequencies and wavelengths. The sum of these two traveling waves produces a standing wave which can be written mathematically as

$$D = D_1 + D_2 = A\big[\sin(kx - \omega t) + \sin(kx + \omega t)\big].$$

From the trigonometric identity $\sin\theta_1 + \sin\theta_2 = 2\sin\frac{1}{2}(\theta_1 + \theta_2)\cos\frac{1}{2}(\theta_1 - \theta_2)$, Appendix A–9, we can rewrite this as

$$D = 2A\sin kx\cos\omega t. \qquad \textbf{(15–18)}$$

If we let $x = 0$ at the left-hand end of the string, then the right-hand end is at $x = \ell$ where ℓ is the length of the string. Since the string is fixed at its two ends (Fig. 15–26), $D(x,t)$ must be zero at $x = 0$ and at $x = \ell$. Equation 15–18 already satisfies the first condition ($D = 0$ at $x = 0$) and satisfies the second condition if $\sin k\ell = 0$ which means

$$k\ell = \pi, 2\pi, 3\pi, \cdots, n\pi, \cdots$$

where n = an integer. Since $k = 2\pi/\lambda$, then $\lambda = 2\pi/k = 2\pi\ell/n\pi = 2\ell/n$, which is just Eq. 15–17a.

Equation 15–18, with the condition $\lambda = 2\ell/n$, is the mathematical representation of a standing wave. We see that a particle at any position x vibrates in simple harmonic motion (because of the factor $\cos\omega t$). All particles of the string vibrate with the same frequency $f = \omega/2\pi$, but the amplitude depends on x and equals $2A\sin kx$. (Compare this to a traveling wave for which all particles vibrate with the same amplitude.) The amplitude has a maximum, equal to $2A$, when $kx = \pi/2, 3\pi/2, 5\pi/2$, and so on—that is, at

$$x = \frac{\lambda}{4}, \quad \text{and for overtones also at} \quad \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots.$$

These are the positions of the antinodes, as we saw in Fig. 15–26.

EXAMPLE 15–10 **Wave forms.** Two waves traveling in opposite directions on a string fixed at $x = 0$ are described by the functions

$$D_1 = (0.20\,\text{m})\sin(2.0x - 4.0t) \quad \text{and} \quad D_2 = (0.20\,\text{m})\sin(2.0x + 4.0t)$$

(where x is in m, t is in s), and they produce a standing wave pattern. Determine: (*a*) the function for the standing wave, (*b*) the amplitude at $x = 0.45$ m, (*c*) the length ℓ of the string, (*d*) the maximum amplitude and where it occurs.

APPROACH We use the principle of superposition to add the two waves. The given waves have the form we used to obtain Eq. 15–18, which we thus can use.

SOLUTION (*a*) The two waves are of the form $D = A\sin(kx \pm \omega t)$, so

$$k = 2.0\,\text{rad/m} \quad \text{and} \quad \omega = 4.0\,\text{rad/s}.$$

These combine to form a standing wave of the form of Eq. 15–18:

$$D = 2A\sin kx\cos\omega t = (0.40\,\text{m})\sin(2.0x)\cos(4.0t),$$

where x is in meters and t in seconds, and in trig functions we have rads.

(*b*) At $x = 0.45$ m,

$$D = (0.40\,\text{m})\sin(0.90\,\text{rad})\cos(4.0t) = (0.31\,\text{m})\cos(4.0t).$$

The amplitude at this point is 0.31 m and occurs when $\cos(4.0t) = 1$.

(*c*) These waves make a standing wave pattern, so both ends of the string must be nodes. Nodes occur every half wavelength, which for our string is (see Eq. 15–11)

$$\frac{\lambda}{2} = \frac{1}{2}\frac{2\pi}{k} = \frac{\pi}{2.0}\,\text{m} = 1.57\,\text{m}.$$

If the string includes only one loop (first harmonic), its length is $\ell = 1.57$ m. But without more information, it could be twice as long, $\ell = 3.14$ m, or any integral number times 1.57 m, and still provide a standing wave pattern, Fig. 15–27.

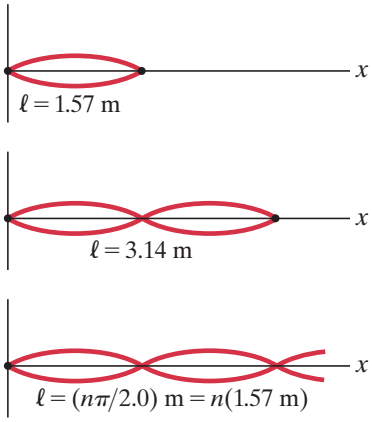


FIGURE 15–27 Example 15–10: possible lengths for the string.

(*d*) The nodes occur at $x = 0$, $x = 1.57\,\text{m}$, and, if the string is longer than $\ell = 1.57\,\text{m}$, at $x = 3.14\,\text{m}$, 4.71 m, and so on. The maximum amplitude (antinode) is $2A = 0.40\,\text{m}$ [see part (*b*)] and occurs midway between the nodes. For $\ell = 1.57\,\text{m}$, there is only one antinode, at $x = 0.79\,\text{m}$.

15–10 Refraction†

When any wave strikes a boundary, some of the energy is reflected and some is transmitted or absorbed. When a two- or three-dimensional wave traveling in one medium crosses a boundary into a medium where its speed is different, the transmitted wave may move in a different direction than the incident wave, as shown in Fig. 15–28. This phenomenon is known as **refraction**. One example is a water wave; the speed decreases in shallow water and the waves refract, as shown in Fig. 15–29. [When the wave speed changes gradually as in Fig. 15–29, without a sharp boundary, the waves change direction (refract) gradually.]

In Fig. 15–28, the speed of the wave in medium 2 is less than in medium 1. In this case, the wave front bends so that it travels more nearly parallel to the boundary. That is, the *angle of refraction*, θ_r, is less than the *angle of incidence*, θ_i. To see why this is so, and to help us get a quantitative relation between θ_r and θ_i, let us think of each wave front as a row of soldiers. The soldiers are marching from firm ground (medium 1) into mud (medium 2) and hence are slowed down after the boundary. The soldiers that reach the mud first are slowed down first, and the row bends as shown in Fig. 15–30a. Let us consider the wave front (or row of soldiers) labeled A in Fig. 15–30b. In the same time t that point A_1 moves a distance $\ell_1 = v_1 t$, we see that point A_2 moves a distance $\ell_2 = v_2 t$. The two right triangles in Fig. 15–30b, shaded yellow and green, have side a in common. Thus

$$\sin\theta_1 = \frac{\ell_1}{a} = \frac{v_1 t}{a}$$

since a is the hypotenuse, and

$$\sin\theta_2 = \frac{\ell_2}{a} = \frac{v_2 t}{a}.$$

Dividing these two equations, we obtain the **law of refraction**:

$$\frac{\sin\theta_2}{\sin\theta_1} = \frac{v_2}{v_1}. \qquad \textbf{(15–19)}$$

Since θ_1 is the angle of incidence (θ_i), and θ_2 is the angle of refraction (θ_r), Eq. 15–19 gives the quantitative relation between the two. If the wave were going in the opposite direction, the geometry would not change; only θ_1 and θ_2 would change roles: θ_2 would be the angle of incidence and θ_1 the angle of refraction. Thus, if the wave travels into a medium where it can move faster, it will bend the opposite way, $\theta_r > \theta_i$. We see from Eq. 15–19 that if the speed increases, the angle increases, and vice versa.

Earthquake waves refract within the Earth as they travel through rock layers of different densities (which have different speeds) just as water waves do. Light waves refract as well, and when we discuss light, we shall find Eq. 15–19 very useful.

†This Section and the next are covered in more detail in Chapters 32 to 35 on optics.

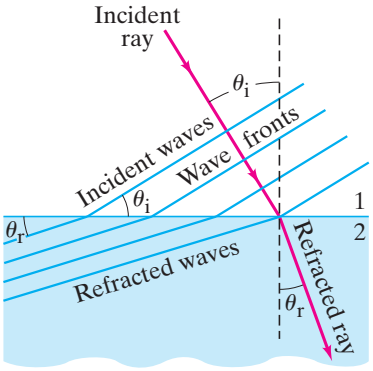
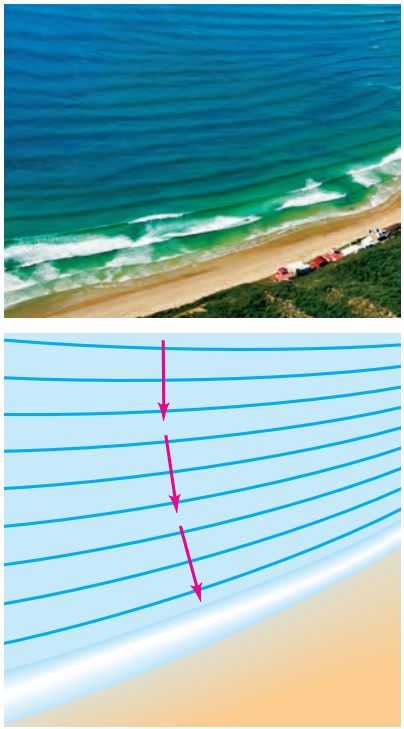


FIGURE 15–28 Refraction of waves passing a boundary.

FIGURE 15–29 Water waves refract gradually as they approach the shore, because their speed decreases. There is no distinct boundary, as in Fig. 15–28, because the wave speed changes gradually.



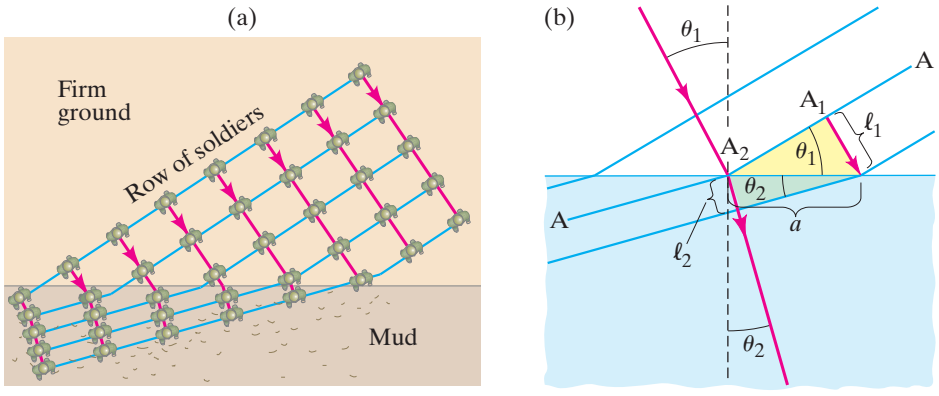


FIGURE 15–30 (a) Marching soldier analogy to derive (b) the law of refraction for waves.

PHYSICS APPLIED
Earthquake wave refraction

EXAMPLE 15–11 Refraction of an earthquake wave. An earthquake P wave (page 435) passes across a boundary in rock where its speed increases from 6.5 km/s to 8.0 km/s. If it strikes this boundary at 30°, what is the angle of refraction?

APPROACH We apply the law of refraction, Eq. 15–19, $\sin\theta_2/\sin\theta_1 = v_2/v_1$.

SOLUTION Since $\sin 30° = 0.50$, Eq. 15–19 yields

$$\sin\theta_2 = \frac{(8.0\,\text{m/s})}{(6.5\,\text{m/s})}(0.50) = 0.62.$$

So $\theta_2 = \sin^{-1}(0.62) = 38°$.

NOTE Be careful with angles of incidence and refraction. As we discussed in Section 15–7 (Fig. 15–21), these angles are between the wave front and the boundary line, or—equivalently—between the ray (direction of wave motion) and the line perpendicular to the boundary. Inspect Fig. 15–30b carefully.

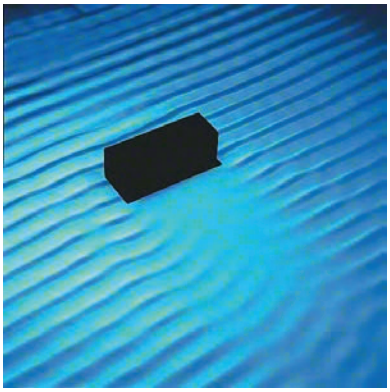


FIGURE 15–31 Wave diffraction. The waves are coming from the upper left. Note how the waves, as they pass the obstacle, bend around it, into the "shadow region" behind the obstacle.

15–11 Diffraction

Waves spread as they travel. When they encounter an obstacle, they bend around it somewhat and pass into the region behind it, as shown in Fig. 15–31 for water waves. This phenomenon is called **diffraction**.

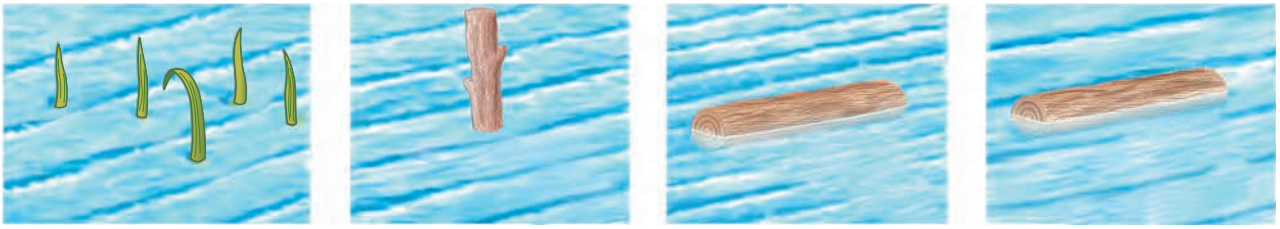
The amount of diffraction depends on the wavelength of the wave and on the size of the obstacle, as shown in Fig. 15–32. If the wavelength is much larger than the object, as with the grass blades of Fig. 15–32a, the wave bends around them almost as if they are not there. For larger objects, parts (b) and (c), there is more of a "shadow" region behind the obstacle where we might not expect the waves to penetrate—but they do, at least a little. Then notice in part (d), where the obstacle is the same as in part (c) but the wavelength is longer, that there is more diffraction into the shadow region. As a rule of thumb, *only if the wavelength is smaller than the size of the object will there be a significant shadow region*. This rule applies to *reflection* from an obstacle as well. Very little of a wave is reflected unless the wavelength is smaller than the size of the obstacle.

A rough guide to the amount of diffraction is

$$\theta(\text{radians}) \approx \frac{\lambda}{\ell},$$

where θ is roughly the angular spread of waves after they have passed through an opening of width ℓ or around an obstacle of width ℓ.

FIGURE 15–32 Water waves coming from upper left, pass objects of various sizes. Note that the longer the wavelength compared to the size of the object, the more diffraction there is into the "shadow region."



(a) Water waves passing blades of grass

(b) Stick in water

(c) Short-wavelength waves passing log

(d) Long-wavelength waves passing log

That waves can bend around obstacles, and thus can carry energy to areas behind obstacles, is very different from energy carried by material particles. A clear example is the following: if you are standing around a corner on one side of a building, you cannot be hit by a baseball thrown from the other side, but you can hear a shout or other sound because the sound waves diffract around the edges of the building.

CONCEPTUAL EXAMPLE 15–12 **Cell phones.** Cellular phones operate by radio waves with frequencies of about 1 or 2 GHz (1 gigahertz = 10^9 Hz). These waves cannot penetrate objects that conduct electricity, such as a sheet of metal or a tree trunk.† The sound quality is best if the transmitting antenna is within clear view of the handset. Yet it is possible to carry on a phone conversation even if the tower is blocked by trees, or if the handset is inside a car. Why?

RESPONSE If the radio waves have a frequency of about 2 GHz, and the speed of propagation is equal to the speed of light, 3×10^8 m/s (Section 1–4), then the wavelength is $\lambda = v/f = (3 \times 10^8 \text{ m/s})/(2 \times 10^9 \text{ Hz}) = 0.15$ m. The waves can diffract readily around objects 15 cm in diameter or smaller.

†Inside the bark are layers of ion-containing water carrying nutrients up to the leaves. The ions are electrically charged and so act as an electrical conductor. Dry wood is not an electrical conductor, but a good insulator.

Summary

Vibrating objects act as sources of **waves** that travel outward from the source. Waves on water and on a cord are examples. The wave may be a **pulse** (a single crest) or it may be continuous (many crests and troughs).

The **wavelength** of a continuous sinusoidal wave is the distance between two successive crests.

The **frequency** is the number of full wavelengths (or crests) that pass a given point per unit time.

The **amplitude** of a wave is the maximum height of a crest, or depth of a trough, relative to the normal (or equilibrium) level.

The **wave speed** (how fast a crest moves) is equal to the product of wavelength and frequency,

$$v = \lambda f. \qquad \textbf{(15–1)}$$

In a **transverse wave**, the oscillations are perpendicular to the direction in which the wave travels. An example is a wave on a cord.

In a **longitudinal wave**, the oscillations are along (parallel to) the line of travel; sound is an example.

The speed of both longitudinal and transverse waves in matter is proportional to the square root of an elastic force factor divided by an inertia factor (or density).

After an earthquake, both transverse and longitudinal waves (called S and P) propagate through the Earth, plus there are **surface waves**.

Waves carry energy from place to place without matter being carried. The **intensity** I of a wave is the energy per unit time transported across unit area (in watts/m^2) and is proportional to the *square of the amplitude, A,* of the wave: $I \propto A^2$.

For a wave traveling uniformly outward in three dimensions from a point source, the intensity decreases inversely with the square of the distance from the source (ignoring damping):

$$I \propto \frac{1}{r^2}, \qquad \textbf{(15–8a)}$$

which is referred to as an **inverse square law**. The amplitude $A(\propto \sqrt{I})$ decreases linearly with distance from the source.

A one-dimensional transverse wave traveling in a medium to the right along the x axis (x increasing) can be represented by a formula for the displacement of the medium from equilibrium at any point x as a function of time as

$$D(x,t) = A \sin\left[\left(\frac{2\pi}{\lambda}\right)(x - vt)\right] \qquad \textbf{(15–10a)}$$

$$= A \sin(kx - \omega t) \qquad \textbf{(15–10c)}$$

where

$$k = \frac{2\pi}{\lambda} \qquad \textbf{(15–11)}$$

is the **wave number**, and $\omega = 2\pi f$ is the *angular frequency*. The **phase velocity** v is

$$v = \lambda f = \omega/k. \qquad \textbf{(15–12)}$$

If a wave is traveling toward decreasing values of x, to the left along the x axis, then

$$D(x,t) = A \sin(kx + \omega t). \qquad \textbf{(15–13c)}$$

[*Waves can be described by the **wave equation**, which in one dimension is

$$\partial^2 D/\partial x^2 = (1/v^2)\,\partial^2 D/\partial t^2. \qquad \textbf{(15–16)}]$$

When two or more waves pass through the same region of space at the same time, the displacement at any given point will be the vector sum of the displacements of the separate waves. This is the **principle of superposition**. It is valid for mechanical waves if the amplitudes are small enough that the restoring force of the medium is proportional to displacement. A *complex wave* is one made up of two or more frequencies.

Waves reflect off objects in their path. When the **wave front** of a two- or three-dimensional wave strikes an object, the **angle of reflection** equals the **angle of incidence**, which is the **law of reflection**. When a wave strikes a boundary between two materials in which it can travel, part of the wave is reflected and part is transmitted.

When two waves pass through the same region of space at the same time, they **interfere**. The resultant displacement at any point and time is the sum of their separate displacements (the superposition principle). This can result in **constructive interference**, **destructive interference**, or something in between depending on the amplitudes and relative phases of the waves.

Waves traveling on a cord of fixed length interfere with waves that have reflected off the end and are traveling back in the opposite direction. At certain frequencies, **standing waves** can be produced in which the waves seem to be standing still rather than traveling. The cord (or other medium) is vibrating as a whole. This is a resonance phenomenon, and the frequencies at which standing waves occur are called **resonant frequencies**. Points of destructive interference (no oscillation) are called **nodes**. Points of constructive interference (maximum amplitude of vibration) are called **antinodes**. On a cord of length ℓ fixed at both ends, the wavelengths of standing waves are given by

$$\lambda_n = 2\ell/n \qquad \textbf{(15–17a)}$$

where n is an integer.

Waves change direction, or **refract**, when traveling from one medium into a second medium where their speed is different. Waves spread, or **diffract**, as they travel and encounter obstacles. A rough guide to the amount of diffraction is $\theta \approx \lambda/\ell$, where λ is the wavelength of the wave, and ℓ the width of an opening or obstacle. There is a significant "shadow region" only if the wavelength λ is smaller than the size of the obstacle.

Questions

1. Is the frequency of a simple periodic wave equal to the frequency of its source? Why or why not?
2. Explain the difference between the speed of a transverse wave traveling along a cord and the speed of a tiny piece of the cord.
3. You are finding it a challenge to climb from one boat up onto a higher boat in heavy waves. If the climb varies from 2.5 m to 4.3 m, what is the amplitude of the wave? Assume the centers of the two boats are a half wavelength apart.
4. What kind of waves do you think will travel along a horizontal metal rod if you strike its end (*a*) vertically from above and (*b*) horizontally parallel to its length?
5. Since the density of air decreases with an increase in temperature, but the bulk modulus B is nearly independent of temperature, how would you expect the speed of sound waves in air to vary with temperature?
6. Describe how you could estimate the speed of water waves across the surface of a pond.
7. The speed of sound in most solids is somewhat greater than in air, yet the density of solids is much greater (10^3 to 10^4 times). Explain.
8. Give two reasons why circular water waves decrease in amplitude as they travel away from the source.
9. Two linear waves have the same amplitude and speed, and otherwise are identical, except one has half the wavelength of the other. Which transmits more energy? By what factor?
10. Will any function of $(x - vt)$—see Eq. 15–14—represent a wave motion? Why or why not? If not, give an example.
11. When a sinusoidal wave crosses the boundary between two sections of cord as in Fig. 15–19, the frequency does not change (although the wavelength and speed do change). Explain why.
12. If a sinusoidal wave on a two-section cord (Fig. 15–19) is inverted upon reflection, does the transmitted wave have a longer or shorter wavelength?
13. Is energy always conserved when two waves interfere? Explain.
14. If a string is vibrating as a standing wave in three segments, are there any places you could touch it with a knife blade without disturbing the motion?
15. If a rope has a free end, a pulse sent down the rope behaves differently on reflection than if the rope has that end fixed in position. What is this difference, and why does it occur?
16. How did geophysicists determine that part of the Earth's interior is liquid? (See Section 15–2.)
17. "In a round bowl of water, waves move from the center to the rim, or from the rim to the center, depending on whether you strike at the center or at the rim." So wrote Dante Alighieri 700 years ago in his great poem *Paradiso* (Canto 14), the last part of his famous *Divine Comedy*. Try this experiment and discuss your results.
18. Can the amplitude of the standing waves in Fig. 15–25 be greater than the amplitude of the vibrations that cause them (up and down motion of the hand)?
19. When a standing wave exists on a string, the vibrations of incident and reflected waves cancel at the nodes. Does this mean that energy was destroyed? Explain.
20. When a cord is vibrated as in Fig. 15–25 by hand or by a mechanical oscillator, the "nodes" are not quite true nodes (at rest). Explain. [*Hint*: Consider damping and energy flow from hand or oscillator.]

*21. AM *radio signals* can usually be heard behind a hill, but FM often cannot. That is, AM signals bend more than FM. Explain. (Radio signals, as we shall discuss in Chapter 31, are carried by electromagnetic waves whose wavelength for AM is typically 200 to 600 m and for FM about 3 m.)

*22. If we knew that energy was being transmitted from one place to another, how might we determine whether the energy was being carried by particles (material objects) or by waves?

MisConceptual Questions

1. A transverse wave is traveling down a cord. Which of the following is true about the transverse motion of a small piece of the cord?
   (*a*) The speed of the wave must be the same as the speed of a small piece of the cord.
   (*b*) The frequency of the wave must be the same as the frequency of a small piece of the cord.
   (*c*) The amplitude of the wave must be the same as the amplitude of a small piece of the cord.
   (*d*) All of the above are true.
   (*e*) Both (*b*) and (*c*) are true.
2. Two waves are traveling toward each other along a rope. When they meet, the waves
   (*a*) pass through each other.
   (*b*) bounce off of each other.
   (*c*) disappear.
3. Which of the following increases the speed of waves in a stretched elastic cord?
   (*a*) Increasing the wave amplitude.
   (*b*) Increasing the wave frequency.
   (*c*) Increasing the wavelength.
   (*d*) Stretching the elastic cord further.
4. Consider a wave on a string moving to the right, as shown in Fig. 15–33. What is the direction of the velocity of a particle of string at point B?
   (*a*) →
   (*b*) ↘
   (*c*) ↓
   (*d*) ↑
   (*e*) $\vec{\mathbf{v}} = 0$, so no direction.

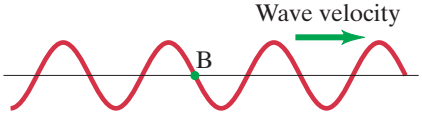


FIGURE 15–33
MisConceptual Question 4.

5. What happens when two waves, such as waves on a lake, come from different directions and run into each other?
(*a*) They cancel each other out and disappear.
(*b*) If they are the same size, they cancel each other out and disappear. If one wave is larger than the other, the smaller one disappears and the larger one shrinks but continues.
(*c*) They get larger where they run into each other; then they continue in a direction between the direction of the two original waves and larger than either original wave.
(*d*) They may have various patterns where they overlap, but each wave continues with its original pattern away from the region of overlap.
(*e*) Waves cannot run into each other; they always come from the same direction and so are parallel.

6. A student fixes one end of a Slinky on a table. She holds the other end in her hand, stretches it to a length ℓ, and then moves it back and forth to send a transverse wave down the Slinky. If she next moves her hand faster while keeping the length of the Slinky the same, how does the wavelength along the Slinky change?
(*a*) It increases. (*b*) It stays the same. (*c*) It decreases.

7. A standing wave on a string is the result of
(*a*) two waves of different frequency but equal amplitude, traveling in the same direction.
(*b*) two equal amplitude, equal frequency waves traveling in opposite directions.
(*c*) two equal amplitude, equal frequency waves traveling in the same direction.
(*d*) two waves of different frequency but equal amplitude traveling in opposite directions.
(*e*) two waves of different amplitude but equal wavelength traveling in the same direction.

8. A wave traveling along a thin cord will pass onto a heavier section of cord as shown in Fig. 15–34.



FIGURE 15–34 MisConceptual Question 8.

Which quantities will change?
(*a*) Amplitude.
(*b*) Frequency.
(*c*) Wavelength.
(*d*) Wave speed.
(*e*) All of the above.

9. A wave transports
(*a*) energy but not matter.
(*b*) matter but not energy.
(*c*) both energy and matter.

10. When both the frequency and the amplitude of a wave are doubled, the energy carried by the wave
(*a*) is unchanged.
(*b*) increases by a factor of 2.
(*c*) increases by a factor of 4.
(*d*) increases by a factor of 8.
(*e*) increases by a factor of 16.
(*f*) None of the above.

11. A wave is traveling along a cord stretched with tension F_T. To double the speed of the wave, the tension must be
(*a*) increased by a factor of 2.
(*b*) decreased by a factor of 2.
(*c*) increased by a factor of $\sqrt{2}$.
(*d*) decreased by a factor of $\sqrt{2}$.
(*e*) increased by a factor of 4.
(*f*) decreased by a factor of 4.
(*g*) None of the above.

Problems

15–1 and 15–2 Characteristics of Waves

1. (I) A fisherman notices that water wave crests pass the bow of his anchored boat every 3.0 s. He measures the distance between two crests to be 7.5 m. How fast are the waves traveling?

2. (I) A sound wave in air has a frequency of 265 Hz and travels with a speed of 343 m/s. How far apart are the wave crests (compressions)?

3. (I) Calculate the speed of longitudinal waves in (*a*) water, (*b*) granite, and (*c*) steel, using Tables in Chapters 12 and 13.

4. (I) AM radio signals have frequencies between 550 kHz and 1600 kHz (kilohertz) and travel with a speed of 3.0×10^8 m/s. What are the wavelengths of these signals? On FM the frequencies range from 88 MHz to 108 MHz (megahertz) and travel at the same speed. What are their wavelengths?

5. (II) Determine the wavelength of a 6200-Hz longitudinal wave traveling along an iron rod. Use Tables 12–1 and 13–1.

6. (II) A cord of mass 0.65 kg is stretched between two supports 7.2 m apart. If the tension in the cord is 120 N, how long will it take a pulse to travel from one support to the other?

7. (II) A 0.40-kg cord is stretched between two supports 8.7 m apart. When one support is struck by a hammer, a transverse wave travels down the cord and reaches the other support in 0.85 s. What is the tension in the cord?

8. (II) A sailor strikes the side of his ship just below the surface of the sea. He hears the echo of the wave reflected from the ocean floor directly below 2.4 s later. How deep is the ocean at this point? (Use Tables 12–1 and 13–1.)

9. (II) A ski gondola [pronounced gon′do-la] is connected to the top of a hill by a steel cable of length 710 m and diameter 1.5 cm. As the gondola comes to the end of its run, it bumps into the terminal and sends a transverse wave pulse along the cable. It is observed that it took 17 s for the pulse to return. (*a*) What is the speed of the pulse? (*b*) What is the tension in the cable?

10. (II) P and S waves from an earthquake travel at different speeds, and this difference helps locate the earthquake "epicenter" (where the slippage took place). (*a*) Assuming typical speeds of 8.5 km/s and 5.5 km/s for P and S waves, respectively, how far away did the earthquake occur if a particular seismic station detects the arrival of these two types of waves 1.5 min apart? (*b*) Is one seismic station sufficient to determine the position of the epicenter? Explain.

11. (II) The wave on a string shown in Fig. 15–35 is moving to the right with a speed of 1.10 m/s. (*a*) Draw the shape of the string 1.00 s later and indicate which parts of the string are moving up and which down at that instant. (*b*) Estimate the vertical speed of point A on the string at the instant shown in Fig. 15–35.

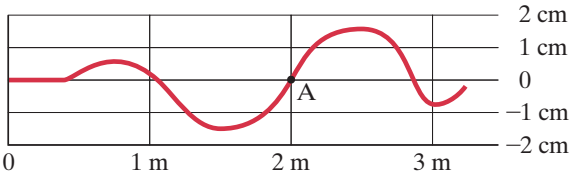


FIGURE 15–35 Problem 11.

12. (II) A 4.4-kg ball hangs from a steel wire 1.00 mm in diameter and 5.00 m long. What would be the speed of a transverse wave in the steel wire?

13. (II) Two children are sending signals along a nylon cord tied between tin cans 18 m apart. How long does it take the longitudinal vibrations in the string to go from one child to the other? Nylon has density $\rho = 1.15 \times 10^3\ \text{kg/m}^3$.

14. (III) A uniform cord of length ℓ and mass m is hung vertically from a support. (*a*) Show that the speed of transverse waves in this cord is $\sqrt{gh}$, where h is the height above the lower end. (*b*) How long does it take for a pulse to travel upward from one end to the other?

15–3 Energy Transported by Waves

15. (I) Sound waves are produced by a 15-watt output loudspeaker. Compare the intensity of the sound wave heard by a listener 12 m away to the wave intensity heard by a listener 24 m away. Assume spherical waves.

16. (I) Two earthquake waves of the same frequency travel through the same portion of the Earth, but one is carrying 3.5 times the energy. What is the ratio of the amplitudes of the two waves?

17. (II) What is the ratio of (*a*) the intensities, and (*b*) the amplitudes, of an earthquake P wave passing through the Earth and detected at two points 15 km and 55 km from the source?

18. (II) The intensity of an earthquake wave passing through the Earth is measured to be $2.4 \times 10^6\ \text{J/m}^2 \cdot \text{s}$ at a distance of 48 km from the source. (*a*) What was its intensity when it passed a point only 1.0 km from the source? (*b*) At what rate did energy pass through an area of $2.0\ \text{m}^2$ at 1.0 km?

19. (II) A thin steel wire of diameter 1.0 mm is connected to an oscillator and is under a tension of 7.5 N. The frequency of the oscillator is 60.0 Hz and it is observed that the amplitude of the wave on the steel wire is 0.40 cm. (*a*) What is the power output of the oscillator, assuming that the wave is not reflected back? (*b*) If the power output stays constant but the frequency is doubled, what is the amplitude of the wave?

20. (II) Show that the intensity of a wave (Eq. 15–7a) is equal to the energy density (energy per unit volume) in the wave times the wave speed.

21. (II) (*a*) Show that the average rate with which energy is transported along a cord by a transverse wave of frequency f and amplitude A is

$$\overline{P} = 2\pi^2 \mu v f^2 A^2,$$

where v is the speed of the wave and μ is the mass per unit length of the cord. (*b*) If the cord is under a tension $F_T = 135\ \text{N}$ and has mass per unit length 0.10 kg/m, what power is required to transmit 120-Hz transverse waves of amplitude 2.4 cm?

22. (III) Show that if damping is ignored, the amplitude A of circular waves on the surface of water decreases as the square root of the distance r from the source: $A \propto 1/\sqrt{r}$.

15–4 Mathematical Representation of Traveling Wave

23. (I) A transverse wave on a wire is given by $D(x, t) = 0.015 \sin(35x - 1200t)$ where D and x are in meters and t is in seconds. (*a*) Write an expression for a wave with the same amplitude, wavelength, and frequency but traveling in the opposite direction. (*b*) What is the speed of either wave?

24. (II) Suppose at $t = 0$, a wave shape is represented by $D = A \sin(2\pi x/\lambda + \phi)$; that is, it differs from Eq. 15–9 by a constant phase factor ϕ. What will be the equation for such a wave as a function of x and t if it travels to the left along the x axis?

25. (II) A transverse traveling wave on a cord is represented by $D = 0.22 \sin(5.6x + 34t)$ where D and x are in meters and t is in seconds. For this wave determine (*a*) the wavelength, (*b*) frequency, (*c*) amplitude, (*d*) wave velocity (magnitude and direction), and (*e*) maximum and minimum speeds of particles of the cord at $x = 0$.

26. (II) Consider the point $x = 1.00\ \text{m}$ on the cord of Example 15–6. Determine (*a*) the maximum acceleration of this point. (*b*) What is the speed and acceleration of this point at $t = 2.15\ \text{s}$?

27. (II) A transverse wave pulse travels to the right along a string with a speed $v = 2.0\ \text{m/s}$. At $t = 0$ the shape of the pulse is given by the function

$$D = 0.45 \cos(2.6x + 1.2),$$

where D and x are in meters. (*a*) Plot D vs. x at $t = 0$. (*b*) Determine a formula for the wave pulse at any time t assuming there are no frictional losses. (*c*) Plot $D(x, t)$ vs. x at $t = 1.0\ \text{s}$. (*d*) Repeat parts (*b*) and (*c*) assuming the pulse is traveling to the left. Plot all 3 graphs on the same axes for easy comparison.

28. (II) A transverse wave with a frequency of 220 Hz and a wavelength of 10.0 cm is traveling along a cord. The maximum speed of particles on the cord is 0.10 the wave speed. What is the amplitude of the wave?

29. (II) A 572-Hz longitudinal wave in air has a speed of 345 m/s. (*a*) What is the wavelength? (*b*) How much time is required for the phase to change by 90° at a given point in space? (*c*) At a particular instant, what is the phase difference (in degrees) between two points 4.4 cm apart?

30. (II) Write the equation for the wave in Problem 29 traveling to the right, if its amplitude is 0.020 cm, and $D = -0.020$ cm, at $t = 0$ and $x = 0$.

31. (II) A sinusoidal wave traveling on a cord in the negative x direction has amplitude 1.00 cm, wavelength 3.00 cm, and frequency 245 Hz. At $t = 0$, the particle of string at $x = 0$ is displaced a distance $D = 0.80$ cm above the origin and is moving upward. (*a*) Sketch the shape of the wave at $t = 0$ and (*b*) determine the function of x and t that describes the wave.

32. (III) A transverse wave on a cord is given by $D(x, t) = 0.12 \sin(3.0x - 15.0t)$, where D and x are in meters and t is in seconds. (*a*) What are the period and frequency of this wave? (*b*) At $t = 0.20$ s, what are the displacement, velocity, and acceleration of a point on the cord where $x = 0.60$ m? (*c*) What are the maximum values of the speed and acceleration of a particle of the cord? (*d*) Determine the velocity of a particle of the cord as a function of x at $t = 0$. (*e*) What is the velocity of a particle of cord at $x = 0.60$ m as a function of time?

*15–5 The Wave Equation

***33.** (I) Show by direct substitution that $D = A\cos(\omega t - kx)$ is a solution of the wave equation.

***34.** (II) Determine if the function $D = A \sin kx \cos \omega t$ is a solution of the wave equation.

***35.** (II) Show by direct substitution that the following functions satisfy the wave equation: (*a*) $D(x, t) = A \ln(x + vt)$; (*b*) $D(x, t) = (x - vt)^4$.

***36.** (II) Show that the wave forms of Eqs. 15–13c and 15–15 satisfy the wave equation, Eq. 15–16.

***37.** (II) Let two linear waves be represented by $D_1 = f_1(x, t)$ and $D_2 = f_2(x, t)$. If both these waves satisfy the wave equation (Eq. 15–16), show that any combination $D = C_1 D_1 + C_2 D_2$ does as well, where C_1 and C_2 are constants.

***38.** (II) Does the function $D(x, t) = e^{-(kx-\omega t)^2}$ satisfy the wave equation? Why or why not?

***39.** (II) In deriving Eq. 15–2, $v = \sqrt{F_T/\mu}$, for the speed of a transverse wave on a string, it was assumed that the wave's amplitude A is much less than its wavelength λ. Assuming a sinusoidal wave shape $D = A \sin(kx - \omega t)$, show via the partial derivative $v' = \partial D/\partial t$ that the assumption $A \ll \lambda$ implies that the maximum transverse speed v'_{max} of the string itself is much less than the wave speed. If $A = \lambda/100$ determine the ratio v'_{max}/v.

15–7 Reflection and Transmission

40. (II) Consider a sine wave traveling down the stretched two-part cord of Fig. 15–19. Determine a formula (*a*) for the ratio of the speeds of the wave in the two sections, v_H/v_L, and (*b*) for the ratio of the wavelengths in the two sections. (The frequency is the same in both sections. Why?) (*c*) Is the wavelength larger in the heavier cord or the lighter?

41. (II) A cord has two sections with linear densities of 0.10 kg/m and 0.20 kg/m, Fig. 15–36. An incident wave, given by $D = (0.050\text{ m})\sin(7.5x - 12.0t)$, where x is in meters and t in seconds, travels along the lighter cord. (*a*) What is the wavelength on the lighter section of the cord? (*b*) What is the tension in the cord? (*c*) What is the wavelength when the wave travels on the heavier section?

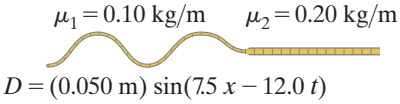


FIGURE 15–36
Problem 41.

42. (III) A cord stretched to a tension F_T consists of two sections (as in Fig. 15–19) whose linear densities are μ_1 and μ_2. Take $x = 0$ to be the point (a tiny knot) where they are joined, with μ_1 referring to that section of cord to the left and μ_2 that to the right. A sinusoidal wave, $D = A\sin[k_1(x - v_1 t)]$, starts at the left end of the cord. When it reaches the knot, part of it is *reflected* and part is *transmitted*. Let the equation of the reflected wave be $D_R = A_R \sin[k_1(x + v_1 t)]$ and that for the transmitted wave be $D_T = A_T \sin[k_2(x - v_2 t)]$. Since the frequency must be the same in both sections, we have $\omega_1 = \omega_2$ or $k_1 v_1 = k_2 v_2$. (*a*) Because the cord is continuous, a point an infinitesimal distance to the left of the knot has the same displacement at any moment (due to incident plus reflected waves) as a point just to the right of the knot (due to the transmitted wave). Thus show that $A = A_T + A_R$. (*b*) Assuming that the slope (derivative of D with respect to x at a fixed t) of the cord just to the left of the knot is the same as the slope just to the right of the knot, show that the amplitude of the reflected wave is given by

$$A_R = \left(\frac{v_1 - v_2}{v_1 + v_2}\right)A = \left(\frac{k_2 - k_1}{k_2 + k_1}\right)A.$$

(*c*) What is A_T in terms of A?

15–8 Interference

43. (I) The two pulses shown in Fig. 15–37 are moving toward each other. (*a*) Sketch the shape of the string at the moment they directly overlap. (*b*) Sketch the shape of the string a few moments later. (*c*) In Fig. 15–37, at the moment the pulses pass each other, the string is straight. What has happened to the energy at this moment?

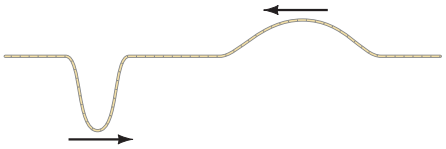


FIGURE 15–37
Problem 43.

44. (II) Suppose two linear waves of equal amplitude and frequency have a phase difference ϕ as they travel in the same medium. They can be represented by

$$D_1 = A\sin(kx - \omega t)$$
$$D_2 = A\sin(kx - \omega t + \phi).$$

(*a*) Use the trigonometric identity $\sin\theta_1 + \sin\theta_2 = 2\sin\frac{1}{2}(\theta_1 + \theta_2)\cos\frac{1}{2}(\theta_1 - \theta_2)$ to show that the resultant wave is given by

$$D = \left(2A\cos\frac{\phi}{2}\right)\sin\left(kx - \omega t + \frac{\phi}{2}\right).$$

(*b*) What is the amplitude of this resultant wave? Is the wave purely sinusoidal, or not? (*c*) Show that constructive interference occurs if $\phi = 0, 2\pi, 4\pi$, and so on, and destructive interference occurs if $\phi = \pi, 3\pi, 5\pi$, etc. (*d*) Describe the resultant wave, by equation and in words, if $\phi = \pi/2$.

15–9 Standing Waves; Resonance

45. (I) The fundamental frequency of a violin string is 441 Hz when unfingered. What is its fundamental frequency if it is fingered one-third of the way down from the end? (That is, only two-thirds of the string vibrates as a standing wave.)

46. (I) If a violin string vibrates at 294 Hz as its fundamental frequency, what are the frequencies of the first four harmonics?

47. (I) A particular string resonates in four loops at a frequency of 320 Hz. Name at least three other frequencies at which it will resonate. What is each called?

48. (II) In an earthquake, it is noted that a footbridge oscillated up and down in a one-loop (fundamental standing wave) pattern once every 1.3 s (= the period). What other possible resonant periods of motion are there for this bridge? What frequencies do they correspond to?

49. (II) If two successive overtones of a vibrating string are 240 Hz and 320 Hz, what is the frequency of the fundamental?

50. (II) The speed of waves on a string is 96 m/s. If the frequency of standing waves is 435 Hz, how far apart are two adjacent nodes?

51. (II) Show that the frequency of standing waves on a cord of length ℓ and linear density μ, which is stretched to a tension F_T, is given by

$$f = \frac{n}{2\ell}\sqrt{\frac{F_T}{\mu}}$$

where n is an integer.

52. (II) A guitar string is 91 cm long and has a mass of 3.2 g. The vibrating portion of the string from the bridge to the support post is $\ell = 64$ cm and the string is under a tension of 520 N. What are the frequencies of the fundamental and first two overtones?

53. (II) One end of a horizontal string is attached to a small-amplitude mechanical 60.0-Hz oscillator. The string's mass per unit length is 3.9×10^{-4} kg/m. The string passes over a pulley, a distance $\ell = 1.50$ m away, and weights are hung from this end, Fig. 15–38. What mass m must be hung from this end of the string to produce (*a*) one loop, (*b*) two loops, and (*c*) five loops of a standing wave? Assume the string at the oscillator is a node, which is nearly true.

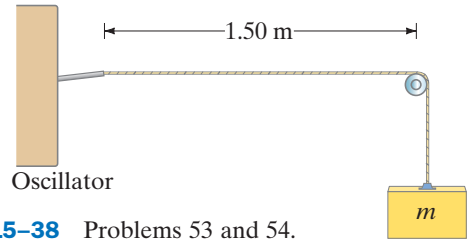


FIGURE 15–38 Problems 53 and 54.

54. (II) In Problem 53, Fig. 15–38, the length of the string may be adjusted by moving the pulley. If the hanging mass m is fixed at 0.070 kg, how many different standing wave patterns may be achieved by varying ℓ between 10 cm and 1.5 m?

55. (II) The displacement of a standing wave on a string is given by $D = 2.4\sin(0.60x)\cos(42t)$, where x and D are in centimeters and t is in seconds. (*a*) What is the distance (cm) between nodes? (*b*) Give the amplitude, frequency, and speed of each of the component waves. (*c*) Find the speed of a particle of the string at $x = 3.80$ cm when $t = 2.5$ s.

56. (II) The displacement of a transverse wave traveling on a string is represented by $D_1 = 4.2\sin(0.84x - 47t + 2.1)$, where D_1 and x are in cm and t in s. (*a*) Find an equation that represents a wave which, when traveling in the opposite direction, will produce a standing wave when added to this one. (*b*) What is the equation describing the standing wave?

57. (II) A particular violin string plays at a frequency of 294 Hz. If the tension is increased 22%, what will the new frequency be?

58. (II) Two traveling waves are described by the functions

$$D_1 = A\sin(kx - \omega t)$$
$$D_2 = A\sin(kx + \omega t),$$

where $A = 0.15$ m, $k = 3.5\ \text{m}^{-1}$, and $\omega = 1.6\ \text{s}^{-1}$. (*a*) Plot these two waves, from $x = 0$ to a point x (> 0) that includes one full wavelength. Choose $t = 1.5$ s. (*b*) Plot the sum of the two waves as a function of x. Mark the nodes and antinodes in the plot. Are they where we would calculate them to be?

59. (II) Plot the two waves given in Problem 58 and their sum, as a function of time from $t = 0$ to $t = T$ (one period). Choose (*a*) $x = 0$ and (*b*) $x = \lambda/4$. Interpret your results.

60. (II) A standing wave on a 1.85-m-long horizontal string displays three loops when the string vibrates at 125 Hz. The maximum swing of the string (top to bottom) at the center of each loop is 8.00 cm. (*a*) What is the function describing the standing wave? (*b*) What are the functions describing the two equal-amplitude waves traveling in opposite directions that make up the standing wave?

61. (II) On an electric guitar, a "pickup" under each string transforms the string's vibrations directly into an electrical signal. If a pickup is placed 16.25 cm from one of the fixed ends of a 65.00-cm-long string, which of the harmonics from $n = 1$ to $n = 12$ will not be "picked up" by this pickup?

62. (II) Two oppositely directed traveling waves given by $D_1 = (5.0\ \text{mm})\cos[(2.0\ \text{m}^{-1})x - (3.0\ \text{rad/s})t]$ and $D_2 = (5.0\ \text{mm})\cos[(2.0\ \text{m}^{-1})x + (3.0\ \text{rad/s})t]$ form a standing wave. Determine the position of nodes along the x axis.

63. (II) When you slosh the water back and forth in a tub at just the right frequency, the water alternately rises and falls at each end, remaining relatively calm at the center. Suppose the frequency to produce such a standing wave in a 45-cm-wide tub is 0.85 Hz. What is the speed of the water wave?

15–10 Refraction

64. (I) Water waves approach an underwater "shelf" where the speed changes from 2.9 m/s to 2.4 m/s. If the incident wave fronts make a 35° angle with the shelf, what will be the angle of refraction?

65. (I) An earthquake P wave traveling 8.0 km/s strikes a boundary within the Earth between two kinds of material. If it approaches the boundary at an incident angle of 52° and the angle of refraction is 31°, what is the speed in the second medium?

66. (II) For any type of wave that reaches a boundary beyond which its speed is increased, there is a maximum incident angle if there is to be a transmitted refracted wave. This maximum incident angle θ_{iM} corresponds to an angle of refraction equal to 90°. If $\theta_i > \theta_{iM}$, all the wave is reflected at the boundary and none is refracted, because this would correspond to $\sin\theta_r > 1$ (where θ_r is the angle of refraction), which is impossible. (*a*) Find a formula for θ_{iM} using the law of refraction, Eq. 15–19. (*b*) How far from the bank should a trout fisherman stand (Fig. 15–39) so trout won't be frightened by his voice (1.8 m above the ground)? The speed of sound is about 343 m/s in air and 1440 m/s in water. [This phenomenon is referred to as *total internal reflection* in optics.]

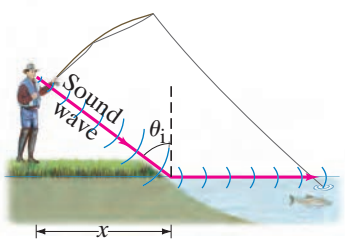


FIGURE 15–39 Problem 66b.

67. (II) A sound wave is traveling in warm air (15°C) when it hits a layer of cold (−15°C) denser air. If the sound wave hits the cold air interface at an angle of 28°, what is the angle of refraction? The speed of sound as a function of temperature can be approximated by $v = (331 + 0.60\,T)$ m/s, where T is in °C.

68. (II) A longitudinal earthquake wave strikes a boundary between two types of rock at a 41° angle. As the wave crosses the boundary, the specific gravity of the rock changes from 3.6 to 2.8. Assuming that the elastic modulus (Section 15–2) is the same for both types of rock, determine the angle of refraction.

15–11 Diffraction

69. (II) A longitudinal wave is traveling upwards in water when it encounters a layer of gasoline floating on the water. The wave is incident on the boundary at an angle of 45 degrees. The bulk modulus of the gasoline is about 65% that of water; the densities of both liquids are given in Table 13-1. Determine the angle of refraction of the wave in the gasoline.

70. (II) What frequencies of sound would easily diffract through an open window 0.85 m wide? (The speed of sound is 344 m/s at 20°C.)

General Problems

71. A sinusoidal traveling wave has frequency 880 Hz and speed 440 m/s. (*a*) At a given time, find the distance between two positions that correspond to a difference in phase of $\pi/6$ rad. (*b*) At a fixed location, by how much does the phase change during a time interval of 1.0×10^{-4} s?

72. A guitar string is supposed to vibrate at 247 Hz, but is measured to actually vibrate at 262 Hz. By what percentage should the tension in the string be changed to get the frequency to the correct value?

73. *Seismic reflection* prospecting is commonly used to map deeply buried formations containing oil. In this technique, a seismic wave generated at the Earth's surface (for example, by an explosion or falling weight) reflects from the subsurface formation and is detected upon its return to ground level. By placing ground-level detectors at a variety of locations relative to the source, and observing the variation in the source-to-detector travel times, the depth of the subsurface formation can be determined. (*a*) Assume a ground-level detector is placed a horizontal distance x away from the seismic-wave source and that a horizontal boundary between subsurface rock and a cavity (hopefully containing oil) exists at depth D (Fig. 15–40a). Determine an expression for the time t taken by the reflected wave to travel from source to detector, assuming the seismic wave propagates at constant speed v. (*b*) Suppose several detectors are placed along a line at different distances x from the source as in Fig. 15–40b. Then, when a seismic wave is generated, the different travel times t for each detector are measured. Starting with your result from part (*a*), explain how a graph of t^2 vs. x^2 can be used to determine D.

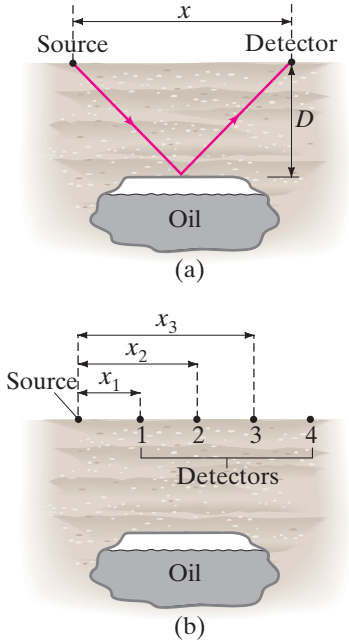


FIGURE 15–40 Problem 73.

74. A 65-cm guitar string is fixed at both ends. In the frequency range between 1.0 and 2.0 kHz, the string is found to resonate only at frequencies 1.2, 1.5, and 1.8 kHz. What is the speed of traveling waves on this string?

75. A bug on the surface of a pond is observed to move up and down a total vertical distance of 0.10 m, lowest to highest point, as a wave passes. (*a*) What is the amplitude of the wave? (*b*) If the amplitude increases to 0.15 m, by what factor does the bug's maximum kinetic energy change?

76. Destructive interference occurs where two overlapping waves are $\frac{1}{2}$ wavelength or 180° out of phase. Explain why 180° is equivalent to $\frac{1}{2}$ wavelength.

77. An earthquake-produced surface wave can be approximated by a sinusoidal transverse wave. Assuming a frequency of 0.60 Hz (typical of earthquakes, which actually include a mixture of frequencies), what amplitude is needed so that objects begin to leave contact with the ground? [*Hint*: Set the acceleration $a > g$. Why?]

78. When you walk with a cup of coffee (diameter 8 cm) at just the right pace of about one step per second, the coffee sloshes higher and higher in your cup until eventually it starts to spill over the top, Fig 15–41. Estimate the speed of the waves in the coffee.



FIGURE 15–41 Problem 78.

79. Two solid rods have the same bulk modulus but one is 2.1 times as dense as the other. In which rod will the speed of longitudinal waves be greater, and by what factor?

80. Two waves traveling along a stretched string have the same frequency, but one transports 2.5 times the power of the other. What is the ratio of the amplitudes of the two waves?

81. A transverse wave pulse travels to the right along a string with a speed $v = 2.4\,\text{m/s}$. At $t = 0$ the shape of the pulse is given by the function

$$D = \frac{4.0\,\text{m}^3}{x^2 + 2.0\,\text{m}^2},$$

where D and x are in meters. (*a*) Plot D vs. x at $t = 0$ from $x = -10\,\text{m}$ to $x = +10\,\text{m}$. (*b*) Determine a formula for the wave pulse at any time t assuming there are no frictional losses. (*c*) Plot $D(x, t)$ vs. x at $t = 1.00\,\text{s}$. (*d*) Repeat parts (*b*) and (*c*) assuming the pulse is traveling to the left.

82. (*a*) Show that if the tension in a stretched string is changed by a small amount ΔF_T, the frequency of the fundamental is changed by an amount $\Delta f = \frac{1}{2}(\Delta F_\text{T}/F_\text{T})f$. (*b*) By what percent must the tension in a piano string be increased or decreased to raise the frequency from 436 Hz to 448 Hz? (*c*) Does the formula in part (*a*) apply to the overtones as well? Explain.

83. A cord of length 1.0 m has two equal-length sections with linear densities of 0.50 kg/m and 1.00 kg/m. The tension in the entire cord is constant. The ends of the cord are oscillated so that a standing wave is set up in the cord with a single node where the two sections meet. What is the ratio of the oscillatory frequencies?

84. A wire is composed of aluminum with length $\ell_1 = 0.600\,\text{m}$ and mass per unit length $\mu_1 = 2.70\,\text{g/m}$ joined to a steel section with length $\ell_2 = 0.882\,\text{m}$ and mass per unit length $\mu_2 = 7.80\,\text{g/m}$. This composite wire is fixed at both ends and held at a uniform tension of 135 N. Find the lowest frequency standing wave that can exist on this wire, assuming there is a node at the joint between aluminum and steel. How many nodes (including the two at the ends) does this standing wave possess?

85. Two strings on a musical instrument are tuned to play at 392 Hz (G) and 494 Hz (B). (*a*) What are the frequencies of the first two overtones for each string? (*b*) If the two strings have the same length and are under the same tension, what must be the ratio of their masses (m_G/m_B)? (*c*) If the strings, instead, have the same mass per unit length and are under the same tension, what is the ratio of their lengths $(\ell_\text{G}/\ell_\text{B})$? (*d*) If their masses and lengths are the same, what must be the ratio of the tensions in the two strings?

86. The ripples in a certain groove 10.8 cm from the center of a 33-rpm vinyl phonograph record have a wavelength of 1.55 mm. What will be the frequency of the sound emitted?

87. A 10.0-m-long wire of mass 138 g is stretched under a tension of 255 N. A pulse is generated at one end, and 20.0 ms later a second pulse is generated at the opposite end. Where will the two pulses first meet?

88. A uniform cable 2.0-m long hangs vertically under its own weight. (*a*) Determine the speed of waves on the cable as a function the distance y (positive down from the top of the cable, where $y = 0$). (*b*) Plot v vs. y.

89. For a spherical wave traveling uniformly away from a point source, show that the displacement can be represented by

$$D = \left(\frac{A}{r}\right)\sin(kr - \omega t),$$

where r is the radial distance from the source and A is a constant.

90. A string can have a "free" end if that end is attached to a ring that can slide without friction on a vertical pole (Fig. 15–42). Determine the wavelengths of the resonant vibrations of such a string with one end fixed and the other free.

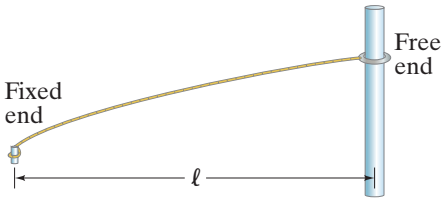
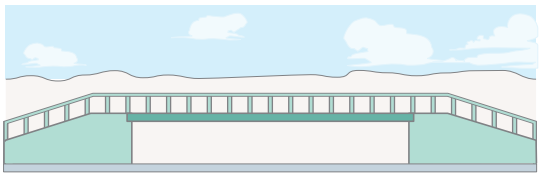
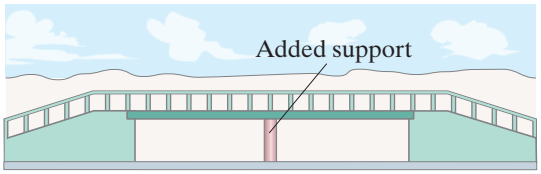


FIGURE 15–42 Problem 90.

91. A highway overpass was observed to resonate as one full loop $(\frac{1}{2}\lambda)$ when a small earthquake shook the ground vertically at 3.0 Hz. The highway department put a support at the center of the overpass, anchoring it to the ground as shown in Fig. 15–43. What resonant frequency would you now expect for the overpass? It is noted that earthquakes rarely do significant shaking above 5 or 6 Hz. Did the modifications do any good? Explain.



Before modification



After modification

FIGURE 15–43 Problem 91.

92. Figure 15–44 shows the wave shape at two instants of time for a sinusoidal wave traveling to the right. What is the mathematical representation of this wave?

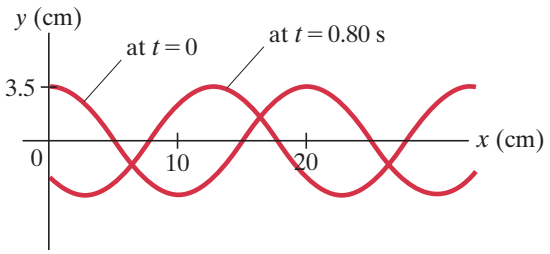


FIGURE 15–44 Problem 92.

93. Estimate the average power of a moving water wave that strikes the chest of an adult standing in the water at the seashore. Assume that the amplitude of the wave is 0.50 m, the wavelength is 2.5 m, and the period is 4.0 s.

94. A *tsunami* is a sort of pulse or "wave packet" consisting of several crests and troughs that become dramatically large as they enter shallow water at the shore. Suppose a tsunami of wavelength 215 km and velocity 550 km/h travels across the Pacific Ocean. As it approaches Hawaii, people observe an unusual decrease of sea level in the harbors. Approximately how much time do they have to run to safety? (In the absence of knowledge and warning, people have died during tsunamis, some of them attracted to the shore to see stranded fishes and boats.)

95. Two wave pulses are traveling in opposite directions with the same speed of 7.0 cm/s as shown in Fig. 15–45. At $t = 0$, the leading edges of the two pulses are 15 cm apart. Sketch the wave pulses at $t = 1.0\,\text{s}, 2.0\,\text{s}$, and 3.0 s.

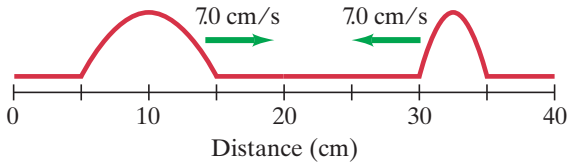


FIGURE 15–45 Problem 95.

***96. Dimensional analysis.** Waves on the surface of the ocean do not depend significantly on the properties of water such as density or surface tension. The primary "return force" for water piled up in the wave crests is due to the gravitational attraction of the Earth. Thus the speed $v\,(\text{m/s})$ of ocean waves depends on the acceleration due to gravity g. It is reasonable to expect that v might also depend on water depth h and the wave's wavelength λ. Assume the wave speed is given by the functional form $v = Cg^{\alpha}h^{\beta}\lambda^{\gamma}$, where α, β, γ, and C are numbers without dimension. (*a*) In *deep* water, the water deep below the surface does not affect the motion of waves at the surface. Thus v should be independent of depth h (i.e., $\beta = 0$). Using only dimensional analysis (Section 1–7 and Appendix D), determine the formula for the speed of surface ocean waves in deep water. (*b*) In *shallow* water, the speed of surface waves is found experimentally to be independent of the wavelength (i.e., $\gamma = 0$ in our assumed equation above for v). Using only dimensional analysis, determine the formula for the speed of waves in shallow water.

ANSWERS TO EXERCISES

A: (*c*).
B: (*d*).
C: (*c*).
D: (*a*).



"If music be the food of physics, play on." [Paraphrase of Shakespeare's *Twelfth Night*, line 1.]

Stringed instruments depend on transverse standing waves on strings to produce their harmonious sounds. The sound of wind instruments originates in longitudinal standing waves of an air column. Trumpets and percussion instruments create more complicated standing waves.

Besides examining sources of sound, we also study the decibel scale of sound level, sound wave interference and beats, the Doppler effect, shock waves and sonic booms, and ultrasound imaging.

CHAPTER 16

Sound

CONTENTS

16–1 Characteristics of Sound
16–2 Mathematical Representation of Longitudinal Waves
16–3 Intensity of Sound: Decibels
16–4 Sources of Sound: Vibrating Strings and Air Columns
*16–5 Quality of Sound, and Noise; Superposition
16–6 Interference of Sound Waves; Beats
16–7 Doppler Effect
*16–8 Shock Waves and the Sonic Boom
*16–9 Applications: Sonar, Ultrasound, and Medical Imaging

CHAPTER-OPENING QUESTION—Guess now!

A pianist plays the note "middle C." The sound is made by the vibration of the piano string and is propagated outward as a vibration of the air (which can reach your ear). How does the vibration on the string compare to the vibration in the air?

(a) The vibration on the string and the vibration in the air have the same wavelength.
(b) They have the same frequency.
(c) They have the same speed.
(d) Neither wavelength, frequency, nor speed are the same in the air as on the string.

Sound is associated with our sense of hearing and, therefore, with the physiology of our ears and the psychology of our brain, which interprets the sensations that reach our ears. The term *sound* also refers to the physical sensation that stimulates our ears: namely, longitudinal pressure waves.

We can distinguish three aspects of any sound. First, there must be a *source* for a sound; as with any mechanical wave, the source of a sound wave is a vibrating object. Second, the energy is transferred from the source in the form of longitudinal sound *waves* in air or other material. And third, the sound is *detected*, usually by an ear or by a microphone. We start by looking at sound waves themselves.

16–1 Characteristics of Sound

We saw in Chapter 15, Fig. 15–5, how a vibrating drumhead produces a sound wave in air. Indeed, we usually think of sound waves traveling in the air, because normally it is the vibrations of the air that force our eardrums to vibrate.

But sound waves can also travel in other materials. Two stones struck together under water can be heard by a swimmer beneath the surface, for the vibrations are carried to the ear by the water. When you put your ear flat against the ground, you can hear an approaching train or truck. In this case the ground does not actually touch your eardrum, but the longitudinal wave transmitted by the ground is called a sound wave just the same, since its vibrations cause the outer ear and the air within it to vibrate. Sound cannot travel in the absence of matter. For example, a bell ringing inside an evacuated jar cannot be heard, and sound cannot travel through the empty reaches of outer space.

The **speed of sound** is different in different materials. In air at 0°C and 1 atm, sound travels at a speed of 331 m/s. We saw in Eq. 15–4 $\left(v = \sqrt{B/\rho}\right)$ that the speed depends on the elastic modulus, B, and the density, ρ, of the material. Thus for helium, whose density is much less than that of air but whose elastic modulus is not greatly different, the speed is about three times as great as in air. In liquids and solids, which are much less compressible and therefore have much greater elastic moduli, the speed is larger still. The speed of sound in various materials is given in Table 16–1. The values depend somewhat on temperature, especially for gases. For example, in air near room temperature, the speed increases approximately 0.60 m/s for each Celsius degree increase in temperature:

$$v \approx (331 + 0.60\,T)\ \text{m/s}, \qquad \text{[speed of sound in air]}$$

where T is the temperature in °C. Unless stated otherwise, we will assume in this Chapter that $T = 20°\text{C}$ precisely, so that $v = [331 + (0.60)(20)]\ \text{m/s} = 343\ \text{m/s}$.

TABLE 16–1 Speed of Sound in Various Materials
(20°C and 1 atm)

Material	Speed (m/s)
Air	343
Air (0°C)	331
Helium	1005
Hydrogen	1300
Water	1440
Sea water	1560
Iron and steel	≈ 5000
Glass	≈ 4500
Aluminum	≈ 5100
Hardwood	≈ 4000
Concrete	≈ 3000

CONCEPTUAL EXAMPLE 16–1 **Distance from a lightning strike.** A rule of thumb that tells how close lightning has struck is "one mile for every five seconds before the thunder is heard." Explain why this works, noting that the speed of light is so high (3×10^8 m/s, almost a million times faster than sound) that the time for light to travel to us is negligible compared to the time for the sound.

PHYSICS APPLIED
How far away is the lightning?

RESPONSE The speed of sound in air is about 340 m/s, so to travel 1 km = 1000 m takes about 3 seconds. One mile is about 1.6 kilometers, so the time for the thunder to travel a mile is about $(1.6)(3) \approx 5$ seconds.

Two aspects of any sound are immediately evident to a human listener: "loudness" and "pitch." Each refers to a sensation in the consciousness of the listener. But to each of these subjective sensations there corresponds a physically measurable quantity. **Loudness** is related to the intensity (energy per unit time crossing unit area) in the sound wave, and we shall discuss it in Section 16–3.

The **pitch** of a sound refers to whether it is high, like the sound of a piccolo or violin, or low, like the sound of a bass drum or string bass. The physical quantity that determines pitch is the frequency, as was first noted by Galileo. The lower the frequency, the lower the pitch; the higher the frequency, the higher the pitch.[†] The best human ears can respond to frequencies from about 20 Hz to almost 20,000 Hz. (Recall that 1 Hz is 1 cycle per second.) This frequency range is called the **audible range**. These limits vary somewhat from one individual to another. One general trend is that as people age, they are less able to hear high frequencies, so the high-frequency limit may be 10,000 Hz or less.

[†]Although pitch is determined mainly by frequency, it also depends to a slight extent on loudness. For example, a very loud sound may seem slightly lower in pitch than a quiet sound of the same frequency.

CAUTION

Do not confuse ultrasonic (high frequency) with supersonic (high speed)

Sound waves whose frequencies are outside the audible range may reach the ear, but we are not generally aware of them. Frequencies above 20,000 Hz are called **ultrasonic**. [Do not confuse with **supersonic**, which is used for objects moving with a speed faster than the speed of sound.] Many animals can hear ultrasonic frequencies; dogs, for example, can hear sounds as high as 50,000 Hz, and bats can detect frequencies as high as 100,000 Hz. Ultrasonic waves have many useful applications in medicine and other fields, which we discuss later in this Chapter.

PHYSICS APPLIED

Autofocusing camera

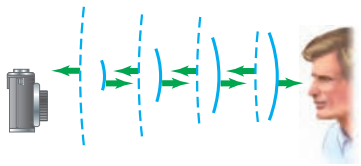


FIGURE 16–1 Example 16–2. Autofocusing camera that emits an ultrasonic pulse. Solid curved lines represent the wave front of the outgoing wave pulse moving to the right; dashed lines represent the wave front of the pulse reflected off the person's face, returning to the camera. The time between emission and reception by the camera of these waves provides the object distance and allows the camera mechanism to adjust the lens to focus at the proper distance.

EXAMPLE 16–2 Autofocusing with sound waves. Autofocusing cameras can use very high frequency (ultrasonic) sound that is emitted and travels to the object being photographed. A sensor on the camera detects the returning reflected sound, as shown in Fig. 16–1. To get an idea of the time sensitivity of the detector, calculate the travel time of the pulse for an object (*a*) 1.0 m away, and (*b*) 20 m away.

APPROACH If we assume the temperature is about 20°C, then the speed of sound is 343 m/s. Using this speed v and the total distance d back and forth in each case, we can obtain the time $(v = d/t)$.

SOLUTION (*a*) The pulse travels 1.0 m to the object and 1.0 m back, for a total of 2.0 m. We solve for t in $v = d/t$:

$$t = \frac{d}{v} = \frac{2.0\text{ m}}{343\text{ m/s}} = 0.0058\text{ s} = 5.8\text{ ms}.$$

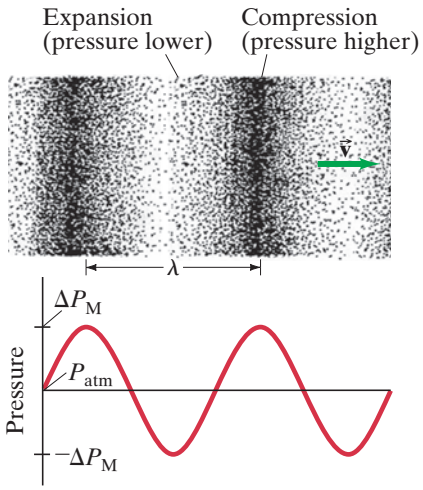
(*b*) The total distance now is $2 \times 20\text{ m} = 40\text{ m}$, so

$$t = \frac{40\text{ m}}{343\text{ m/s}} = 0.12\text{ s} = 120\text{ ms}.$$

NOTE Newer autofocus cameras use infrared light $(v = 3 \times 10^8\text{ m/s})$ instead of ultrasound, and/or a digital sensor array that detects light intensity differences between adjacent receptors as the lens is automatically moved back and forth, choosing the lens position that provides maximum intensity differences (sharpest focus).

Sound waves whose frequencies are below the audible range (that is, less than 20 Hz) are called **infrasonic**. Sources of infrasonic waves include earthquakes, thunder, volcanoes, and waves produced by vibrating heavy machinery. This last source can be particularly troublesome to workers, because infrasonic waves—even though inaudible—can cause damage to the human body. These low-frequency waves act in a resonant fashion, causing motion and irritation of the body's organs.

FIGURE 16–2 Longitudinal sound wave traveling to the right, and its graphical representation in terms of pressure (ΔP relative to normal air pressure).



16–2 Mathematical Representation of Longitudinal Waves

In Section 15–4, we saw that a one-dimensional sinusoidal wave traveling along the x axis can be represented by the relation (Eq. 15–10c)

$$D = A \sin(kx - \omega t), \qquad \textbf{(16–1)}$$

where D is the displacement of the wave at position x and time t, and A is its *amplitude* (maximum value). The wave number k is related to the wavelength λ by $k = 2\pi/\lambda$, and $\omega = 2\pi f$ where f is the frequency. For a transverse wave—such as a wave on a string—the displacement D is perpendicular to the direction of wave propagation along the x axis. But for a longitudinal wave the displacement D is *along the direction of wave propagation*. That is, D is parallel to x and represents the displacement of a tiny volume element of the medium from its equilibrium position.

Longitudinal (sound) waves can also be analyzed from the point of view of variations in pressure rather than displacement. Indeed, longitudinal waves are often called **pressure waves**. The pressure variation is usually easier to measure than the displacement (see Example 16–7). As Fig. 16–2 shows, in a wave "compression" (where molecules are closest together), the pressure is higher than normal, whereas in an expansion (or rarefaction) the pressure is less than normal.

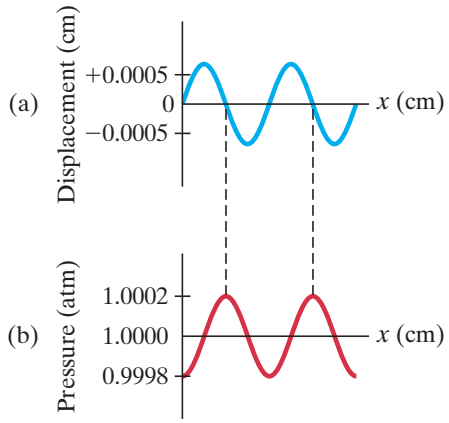


FIGURE 16–3 Representation of a sound wave in space at a given instant in terms of (a) displacement, and (b) pressure.

Figure 16–3 shows a graphical representation of a sound wave in air in terms of (a) displacement and (b) pressure. Note that the displacement wave is a quarter wavelength, or 90° ($\pi/2$ rad), out of phase with the pressure wave: where the pressure is a maximum or minimum, the displacement from equilibrium is zero; and where the pressure variation is zero, the displacement is a maximum or minimum.

Pressure Wave Derivation

Let us now derive the mathematical representation of the pressure variation in a traveling longitudinal wave. From the definition of the bulk modulus, B (Eq. 12–7),

$$\Delta P = -B(\Delta V/V),$$

where ΔP represents the pressure difference from the normal pressure P_0 (no wave present) and $\Delta V/V$ is the fractional change in volume of the medium due to the pressure change ΔP. The negative sign reflects the fact that the volume decreases ($\Delta V < 0$) if the pressure is increased. Consider now a layer of fluid, with a fixed number of molecules, through which the longitudinal wave is passing (Fig. 16–4). If this layer has thickness Δx and area S, then its volume is $V = S\,\Delta x$. As a result of pressure variation in the wave, the volume will change by an amount $\Delta V = S\,\Delta D$, where ΔD is the change in thickness of this layer as it compresses or expands. (Remember that D represents the displacement of the medium.) Thus we have

$$\Delta P = -B\frac{S\,\Delta D}{S\,\Delta x}.$$

To be precise, we take the limit of $\Delta x \to 0$, so we obtain

$$\Delta P = -B\frac{\partial D}{\partial x}, \qquad \textbf{(16–2)}$$

where we use the partial derivative notation since D is a function of both x and t. If the displacement D is sinusoidal as given by Eq. 16–1, then we have from Eq. 16–2 that

$$\Delta P = -(BAk)\cos(kx - \omega t). \qquad \textbf{(16–3)}$$

(Here A is the displacement amplitude, not area which is S.) Thus the pressure varies sinusoidally as well, but is out of phase from the displacement by 90° or a quarter wavelength (cosine vs. sine), as in Fig. 16–3. The quantity BAk is called the **pressure variation amplitude**, ΔP_{M}. It represents the maximum and minimum amounts by which the pressure varies from the normal ambient pressure. We can thus write

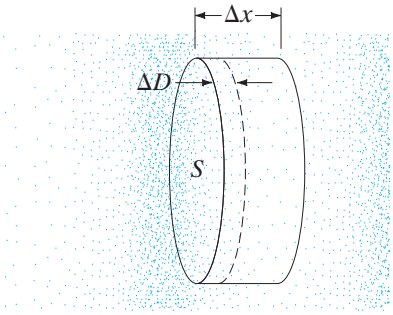
$$\Delta P = -\Delta P_{\text{M}}\cos(kx - \omega t), \qquad \textbf{(16–4)}$$

where, using $v = \sqrt{B/\rho}$ (Eq. 15–4), and $k = \omega/v = 2\pi f/v$ (Eq. 15–12), then

$$\begin{aligned}\Delta P_{\text{M}} &= BAk \\ &= \rho v^2 Ak \\ &= 2\pi\rho v A f. \end{aligned} \qquad \textbf{(16–5)}$$

We will use this connection between ΔP_{M} and A shortly.

FIGURE 16–4 Longitudinal wave in a fluid moves to the right. A thin layer of fluid, in a thin imaginary cylinder of area S and thickness Δx, changes in volume as a result of pressure variation as the wave passes. At the moment shown, the pressure will increase as the wave moves to the right, so the thickness of our layer will decrease, by an amount ΔD.



16–3 Intensity of Sound: Decibels

Loudness is a sensation in the consciousness of a human being and is related to a physically measurable quantity, the **intensity** of the wave. Intensity is defined as the energy transported by a wave per unit time across a unit area perpendicular to the energy flow. As we saw in Chapter 15, intensity is proportional to the square of the wave amplitude. Intensity has units of power per unit area, or watts/meter^2 (W/m^2).

PHYSICS APPLIED
Wide range of human hearing

An average human ear can detect sounds with an intensity as low as 10^{-12} W/m^2 and as high as 1 W/m^2 (and even higher, although above this it is painful). This is an incredibly wide range of intensity, spanning a factor of 10^{12} from lowest to highest. Presumably because of this wide range, what we perceive as loudness is not directly proportional to the intensity. To produce a sound that we hear as about twice as loud requires a sound wave that has about 10 times the intensity. This is roughly valid at any sound level for frequencies near the middle of the audible range. For example, a sound wave of intensity 10^{-2} W/m^2 sounds to an average human being like it is about twice as loud as one with intensity of 10^{-3} W/m^2, and four times as loud as 10^{-4} W/m^2.

Sound Level

Because of this relationship between the subjective sensation of loudness and the physically measurable quantity "intensity," sound intensity levels are usually specified on a logarithmic scale. The unit on this scale is a **bel**, after the inventor Alexander Graham Bell, or much more commonly, the **decibel** (dB), which is $\frac{1}{10}$bel (10 dB = 1 bel).[†] The **sound level**, β, of any sound is defined in terms of its intensity, I, as

$$\beta\,(\text{in dB}) = 10 \log \frac{I}{I_0}, \qquad \textbf{(16–6)}$$

where I_0 is the intensity of a chosen reference level, and the logarithm is to the base 10. I_0 is usually taken as the minimum intensity audible to a good ear—the "threshold of hearing," which is $I_0 = 1.0 \times 10^{-12}$ W/m^2. Thus, for example, the sound level of a sound whose intensity $I = 1.0 \times 10^{-10}$ W/m^2 will be

$$\beta = 10 \log\left(\frac{1.0 \times 10^{-10}\ \text{W/m}^2}{1.0 \times 10^{-12}\ \text{W/m}^2}\right) = 10 \log 100 = 20\ \text{dB},$$

since log 100 is equal to 2.0. (Appendix A has a brief review of logarithms.) Notice that the sound level at the threshold of hearing is 0 dB. That is, $\beta = 10 \log 10^{-12}/10^{-12} = 10 \log 1 = 0$ since $\log 1 = 0$. Notice too that an increase in intensity by a factor of 10 corresponds to a sound level increase of 10 dB. An increase in intensity by a factor of 100 corresponds to a sound level increase of 20 dB. Thus a 50-dB sound is 100 times more intense than a 30-dB sound, and so on.

Intensities and sound levels for some common sounds are listed in Table 16–2.

CAUTION
0 dB does not mean zero intensity

TABLE 16–2
Intensity of Various Sounds

Source of the Sound	Sound Level (dB)	Intensity (W/m^2)
Jet plane at 30 m	140	100
Threshold of pain	120	1
Loud rock concert	120	1
Siren at 30 m	100	1×10^{-2}
Busy street traffic	80	1×10^{-4}
Noisy restaurant	70	1×10^{-5}
Talk, at 50 cm	65	3×10^{-6}
Quiet radio	40	1×10^{-8}
Whisper	30	1×10^{-9}
Rustle of leaves	10	1×10^{-11}
Threshold of hearing	0	1×10^{-12}

EXAMPLE 16–3 **Sound intensity on the street.** At a busy street corner, the sound level is 75 dB. What is the intensity of sound there?

APPROACH We have to solve Eq. 16–6 for intensity I, remembering that $I_0 = 1.0 \times 10^{-12}$ W/m^2.

SOLUTION From Eq. 16–6

$$\log \frac{I}{I_0} = \frac{\beta}{10}.$$

Recalling that $x = \log y$ is the same as $y = 10^x$ (Appendix A–7), then

$$\frac{I}{I_0} = 10^{\beta/10}.$$

With $\beta = 75$, then

$$I = I_0 10^{\beta/10} = (1.0 \times 10^{-12}\ \text{W/m}^2)(10^{7.5}) = 3.2 \times 10^{-5}\ \text{W/m}^2.$$

[†]The dB is dimensionless and so does not have to be included in calculations.

EXAMPLE 16–4 **Loudspeaker output.** A high-quality loudspeaker is advertised to reproduce, at full volume, frequencies from 30 Hz to 18,000 Hz with uniform sound level ± 3 dB. That is, over this frequency range, the sound level output does not vary by more than 3 dB for a given input level. By what factor does the intensity change for a 3 dB change in output level?

PHYSICS APPLIED
Loudspeaker response (± 3 dB)

APPROACH Let us call the desired "correct" output intensity I_1 (for a given input) and the corresponding output sound level β_1. Then the maximum error in output intensity, I_2, corresponds to a level $\beta_2 = \beta_1 + 3$ dB. We then use the relation between intensity and sound level, Eq. 16–6.

SOLUTION Equation 16–6 gives

$$\beta_2 - \beta_1 = 10 \log \frac{I_2}{I_0} - 10 \log \frac{I_1}{I_0}$$

$$3 \text{ dB} = 10\left(\log \frac{I_2}{I_0} - \log \frac{I_1}{I_0}\right) = 10 \log \frac{I_2}{I_1}$$

because $(\log a - \log b) = \log a/b$, as discussed in Appendix A–7. This last equation gives

$$\log \frac{I_2}{I_1} = 0.30,$$

or

$$\frac{I_2}{I_1} = 10^{0.30} = 2.0.$$

So ± 3 dB corresponds to a doubling or halving of the intensity.

EXERCISE A If an increase of 3 dB means "twice as intense," what does an increase of 6 dB mean?

It is worth noting that a sound-level difference of 3 dB (which corresponds to a doubled intensity, as we just saw) corresponds to only a very small change in the subjective sensation of *apparent loudness*. Indeed, the average human can distinguish a difference in sound level of only about 1 or 2 dB.

Normally, the loudness or intensity of a sound decreases as you get farther from the source of the sound. Indoors, this effect is altered because of reflections from the walls. However, if a source is in the open so that sound can radiate out freely in all directions, the intensity decreases as the inverse square of the distance,

$$I \propto \frac{1}{r^2},$$

as we saw in Section 15–3. Over large distances, the intensity decreases faster than $1/r^2$ because some of the energy is transferred into irregular motion of air molecules. This loss happens more for higher frequencies, so any sound of mixed frequencies will be less "bright" at a distance. Far from an outdoor concert, you hear mainly the boom of the drums.

CONCEPTUAL EXAMPLE 16–5 **Trumpet players.** A trumpeter plays at a sound level of 75 dB. Three equally loud trumpet players join in. What is the new sound level?

RESPONSE The intensity of four trumpets is four times the intensity of one trumpet $(= I_1)$ or $4I_1$. The sound level of the four trumpets would be

$$\beta = 10 \log \frac{4I_1}{I_0} = 10 \log 4 + 10 \log \frac{I_1}{I_0}$$

$$= 6.0 \text{ dB} + 75 \text{ dB} = 81 \text{ dB},$$

since $\log 4 = 0.60$ (again Appendix A–7; and $\log ab = \log a + \log b$).

EXERCISE B From Table 16–2, we see that *ordinary conversation* corresponds to a sound level of about 65 dB. If two people are talking at once, the sound level would be about (*a*) 65 dB, (*b*) 68 dB, (*c*) 75 dB, (*d*) 130 dB, (*e*) 59 dB.

PHYSICS APPLIED
Jet plane noise



FIGURE 16–5 Example 16–6. Airport worker with sound-intensity-reducing ear covers (headphones).

EXAMPLE 16–6 **Airplane roar.** The sound level measured 30 m from a jet plane is 140 dB. Estimate the sound level at 300 m. (Ignore reflections from the ground.)

APPROACH Given the sound level, we can determine the intensity at 30 m using Eq. 16–6. Because intensity decreases as the square of the distance, ignoring reflections, we can find I at 300 m and again apply Eq. 16–6 to obtain the sound level.

SOLUTION The intensity I at 30 m is (Eq. 16–6)

$$140\,\text{dB} = 10\log\left(\frac{I}{10^{-12}\,\text{W/m}^2}\right)$$

or, dividing through by 10,

$$14 = \log\left(\frac{I}{10^{-12}\,\text{W/m}^2}\right).$$

Recall that $y = \log x$ means $10^y = x$ (Appendix A–7). Then

$$10^{14} = \frac{I}{10^{-12}\,\text{W/m}^2},$$

so $I = (10^{14})(10^{-12}\,\text{W/m}^2) = 10^2\,\text{W/m}^2$. At 300 m, 10 times as far, the intensity is assumed to decrease approximately as $1/r^2$, and so will be $(\frac{1}{10})^2 = 1/100$ as much, or $1\,\text{W/m}^2$. Hence, the sound level is

$$\beta = 10\log\left(\frac{1\,\text{W/m}^2}{10^{-12}\,\text{W/m}^2}\right) = 120\,\text{dB}.$$

Even at 300 m, the sound is at the threshold of pain. This is why workers at airports wear ear covers to protect their ears from damage (Fig. 16–5).

NOTE Here is a simpler approach that avoids Eq. 16–6. Because the intensity decreases as the square of the distance, at 10 times the distance the intensity decreases by $(\frac{1}{10})^2 = \frac{1}{100}$. We can use the result that 10 dB corresponds to an intensity change by a factor of 10 (see just before Example 16–3). Then an intensity change by a factor of 100 corresponds to a sound-level change of (2)(10 dB) = 20 dB. This confirms our result above: 140 dB − 20 dB = 120 dB.

Intensity Related to Amplitude

The intensity I of a wave is proportional to the square of the wave amplitude, A, as we saw in Section 15–3. We can therefore relate the amplitude quantitatively to the intensity I or level β, as the following Example shows.

PHYSICS APPLIED
Incredible sensitivity of the ear

EXAMPLE 16–7 **How tiny the displacement is.** (*a*) Calculate the displacement of air molecules for a sound having a frequency of 1000 Hz at the threshold of hearing. (*b*) Determine the maximum pressure variation in such a sound wave.

APPROACH In Section 15–3 we found a relation between intensity I and displacement amplitude A of a wave, Eq. 15–7. The amplitude of oscillation of air molecules is what we want to solve for, given the intensity. Assume the temperature is 20°C so the speed of sound is 343 m/s. The pressure variation is found from Eq. 16–5.

SOLUTION At the threshold of hearing, $I = 1.0 \times 10^{-12}\,\text{W/m}^2$ (Table 16–2). We solve for the amplitude A in Eq. 15–7:

$$A = \frac{1}{\pi f}\sqrt{\frac{I}{2\rho v}}$$

$$= \frac{1}{(3.14)(1.0\times10^3\,\text{s}^{-1})}\sqrt{\frac{1.0\times10^{-12}\,\text{W/m}^2}{(2)(1.29\,\text{kg/m}^3)(343\,\text{m/s})}} = 1.1\times10^{-11}\,\text{m},$$

where we have taken the density of air to be $1.29\,\text{kg/m}^3$ (Table 13–1).

NOTE We see how incredibly sensitive the human ear is: it can detect displacements of air molecules which are less than the diameter of atoms (about 10^{-10} m).

(*b*) Now we are dealing with sound as a pressure wave (Section 16–2). From Eq. 16–5,

$$\Delta P_\text{M} = 2\pi\rho v A f$$

$$= 2\pi(1.29\,\text{kg/m}^3)(343\,\text{m/s})(1.1\times10^{-11}\,\text{m})(1.0\times10^3\,\text{s}^{-1}) = 3.1\times10^{-5}\,\text{Pa}$$

or 3.1×10^{-10} atm. Again we see that the human ear is incredibly sensitive.

By combining Eqs. 15–7 and 16–5, we can write the intensity in terms of the pressure amplitude, ΔP_M:

$$I = 2\pi^2 v\rho f^2 A^2 = 2\pi^2 v\rho f^2 \left(\frac{\Delta P_M}{2\pi\rho v f}\right)^2$$

$$I = \frac{(\Delta P_M)^2}{2v\rho}. \qquad \textbf{(16–7)}$$

The intensity in terms of pressure amplitude does not depend on frequency.

The Ear's Response, and Loudness

The ear is not equally sensitive to all frequencies. To hear the same loudness for sounds of different frequencies requires different intensities. Studies on large numbers of people have produced the curves shown in Fig. 16–6. On this graph, each curve represents sounds that seemed to be equally loud. The number labeling each curve represents the **loudness level** (the units are called *phons*), which is numerically equal to the sound level in dB at 1000 Hz. For example, the curve labeled 40 represents sounds that are heard by an average person to have the same loudness as a 1000-Hz sound with a sound level of 40 dB. From this 40-phon curve, we see that a 100-Hz tone must be at a level of about 62 dB to be perceived as loud as a 1000-Hz tone of only 40 dB.

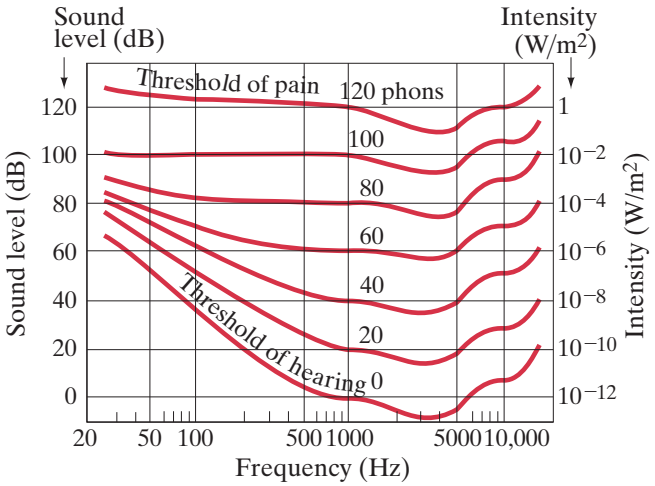


FIGURE 16–6 Sensitivity of the human ear as a function of frequency (see text). Note that the frequency scale is "logarithmic" in order to cover a wide range of frequencies. We can see that at lower sound levels our ears are less sensitive to high and low frequencies relative to mid-range frequencies (which we hear better).

The lowest curve in Fig. 16–6 (labeled 0) represents the sound level, as a function of frequency, for the *threshold of hearing*, the softest sound that is just audible by a very good ear. Note that the ear is most sensitive to sounds of frequency between 2000 and 4000 Hz, which are common in speech and music. Note too that whereas a 1000-Hz sound is audible at a level of 0 dB, a 100-Hz sound must be nearly 40 dB to be heard. The top curve in Fig. 16–6, labeled 120 phons, represents the *threshold of pain*. Sounds above this level can actually be felt and cause pain.

16–4 Sources of Sound: Vibrating Strings and Air Columns

The source of any sound is a vibrating object. Almost any object can vibrate and hence be a source of sound. We now discuss some simple sources of sound, particularly musical instruments. In musical instruments, the source is set into vibration by striking, plucking, bowing, or blowing. Standing waves are produced and the source vibrates at its natural resonant frequencies. The vibrating source is in contact with the air (or other medium) and pushes on it to produce sound waves that travel outward. The frequencies of the waves are the same as those of the source, but the speed and wavelengths can be different. A drum has a stretched membrane that vibrates. Xylophones and marimbas have metal or wood bars that can be set into vibration. Bells, cymbals, and gongs also make use of a vibrating metal. Many instruments make use of vibrating strings, such as the violin, guitar, and piano, or make use of vibrating columns of air, such as the flute, trumpet, and pipe organ. We have already seen that the pitch of a pure sound is determined by the frequency.

TABLE 16–3 Equally Tempered Chromatic Scale[†]

Note (letter)	Note (name)	Frequency (Hz)
C	do	262
C♯ or D♭		277
D	re	294
D♯ or E♭		311
E	mi	330
F	fa	349
F♯ or G♭		370
G	sol	392
G♯ or A♭		415
A	la	440
A♯ or B♭		466
B	ti	494
C′	do	524

[†]Only one octave is included.

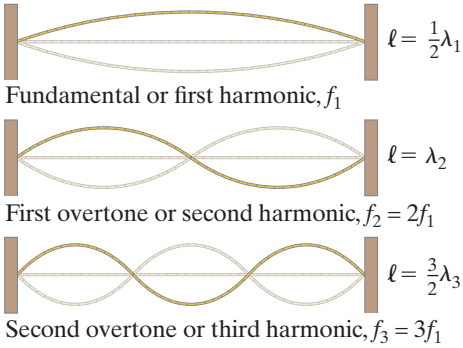
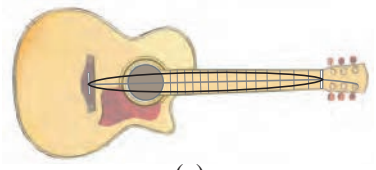
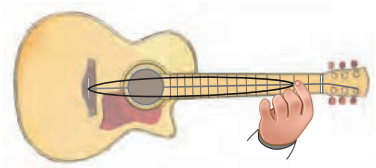


FIGURE 16–7 Standing waves on a string—only the lowest three frequencies are shown.

FIGURE 16–8 The wavelength of (a) an unfingered string is longer than that of (b) a fingered string. Hence, the frequency of the fingered string is higher. Only one string is shown on this guitar, and only the simplest standing wave, the fundamental, is shown.



(a)



(b)

Typical frequencies for musical notes on the *equally tempered chromatic scale* are given in Table 16–3 for the octave beginning with middle C. Note that one octave corresponds to a doubling of frequency. For example, middle C has a frequency of 262 Hz whereas C′ (C above middle C) has twice that frequency, 524 Hz. [Middle C is the C or "do" note at the middle of a piano keyboard.][†]

Stringed Instruments

We saw in Chapter 15, Fig. 15–26b, how standing waves are established on a string, and we show this again here in Fig. 16–7. Such standing waves are the basis for all stringed instruments. The pitch is normally determined by the lowest resonant frequency, the **fundamental**, which corresponds to nodes occurring only at the ends. The string vibrating up and down as a whole corresponds to a half wavelength as shown at the top of Fig. 16–7; so the wavelength of the fundamental on the string is equal to twice the length ℓ of the string. Therefore, the fundamental frequency is $f_1 = v/\lambda = v/2\ell$, where v is the velocity of the wave on the string (*not* in the air). The possible frequencies for standing waves on a stretched string are whole-number multiples of the fundamental frequency:

$$f_n = nf_1 = n\frac{v}{2\ell}, \qquad n = 1, 2, 3, \cdots$$

(just as in Eq. 15–17b), where $n = 1$ refers to the fundamental and $n = 2, 3, \cdots$ are the **overtones**. All of the standing waves, $n = 1, 2, 3, \cdots$, are called **harmonics**,[‡] as we saw in Section 15–9.

When a finger is placed on the string of a guitar or violin, the effective length of the vibrating string is shortened. So its fundamental frequency, and pitch, is higher since the wavelength of the fundamental is shorter (Fig. 16–8). You move your fingers to play *different notes* on a string.

The strings on a guitar or violin are all the same length. They sound at a different pitch because the strings have different mass per unit length, $\mu = m/\ell$, which affects the velocity on the string, Eq. 15–2,

$$v = \sqrt{F_T/\mu}. \qquad \text{[stretched string]}$$

Thus the velocity on a heavier string is lower and the frequency will be lower for the same wavelength or string length. The tension F_T also effects the velocity and therefore the frequency. Indeed, adjusting the tension is the means for **tuning** the pitch of each string, to make it precise by turning a knob at the end (6 knobs are visible in Fig. 16–8 but only one string is shown for clarity). In pianos and harps the strings are of different lengths. For the lower notes the strings are not only longer, but heavier as well, and the reason is illustrated in the following Example.

EXAMPLE 16–8 **Piano strings.** The highest key on a piano corresponds to a frequency about 150 times that of the lowest key. If the string for the highest note is 5.0 cm long, how long would the string for the lowest note have to be if it had the same mass per unit length and was under the same tension?

APPROACH Since $v = \sqrt{F_T/\mu}$, the velocity would be the same on each string. So the frequency is inversely proportional to the length ℓ of the string ($f = v/\lambda = v/2\ell$).

SOLUTION We can write, for the fundamental frequencies of each string, the ratio

$$\frac{\ell_L}{\ell_H} = \frac{f_H}{f_L},$$

where the subscripts L and H refer to the lowest and highest notes, respectively. Thus $\ell_L = \ell_H(f_H/f_L) = (5.0\text{ cm})(150) = 750\text{ cm}$, or 7.5 m. This would be ridiculously long (≈ 25 ft) for a piano.

NOTE The longer strings of lower frequency are made heavier (higher mass per unit length), so even on grand pianos the strings are less than 3 m long.

EXERCISE C How many octaves does the piano of Example 16–8 cover?

[†]Some musicians, bands, or orchestras may use a different middle C. Tuning an orchestra is often done at A, which may be as low as 415 Hz or as high as 445 Hz.

[‡]When overtones are integral multiples of the fundamental, they are harmonics. If the overtones are not integral multiples of the fundamental, as for a vibrating drumhead, they are not harmonics.

EXAMPLE 16–9 **Frequencies and wavelengths in the violin.** A 0.32-m-long violin string is tuned to play A above middle C at 440 Hz. (*a*) What is the wavelength of the fundamental string vibration, and (*b*) what are the frequency and wavelength of the sound wave produced? (*c*) Why is there a difference?

APPROACH The wavelength of the fundamental string vibration equals twice the length of the string (Fig. 16–7). As the string vibrates, it pushes on the air, which is thus forced to oscillate at the same frequency as the string. The speed of sound in air (assumed at 20°C) is 343 m/s, Section 16–1.

SOLUTION (*a*) From Fig. 16–7 the wavelength of the fundamental is

$$\lambda = 2\ell = 2(0.32\,\text{m}) = 0.64\,\text{m} = 64\,\text{cm}.$$

This is the wavelength of the standing wave on the string.

(*b*) The sound wave that travels outward in the air (to reach our ears) has the same frequency, 440 Hz. Its wavelength is

$$\lambda = \frac{v}{f} = \frac{343\,\text{m/s}}{440\,\text{Hz}} = 0.78\,\text{m} = 78\,\text{cm}.$$

(*c*) The wavelength of the sound wave is different from that of the standing wave on the string because the speed of sound in air (343 m/s at 20°C) is different from the speed of the wave on the string $(= f\lambda = 440\,\text{Hz} \times 0.64\,\text{m} = 280\,\text{m/s})$ which depends on the tension in the string and its mass per unit length.

NOTE The frequencies on the string and in the air are the same. But the wavelengths are different because the wave speed on the string is different than the speed of sound waves in air (280 m/s vs. 343 m/s at 20°C).

EXERCISE D Return to Chapter-Opening Question, page 460, and answer it again now. Try to explain why you may have answered differently the first time.

Stringed instruments would not be very loud if they relied on their vibrating strings to produce the sound waves since the strings are too thin to compress and expand much air. Stringed instruments therefore make use of a kind of mechanical amplifier known as a *sounding board* (piano) or *sounding box* (guitar, violin), which acts to amplify the sound by putting a greater surface area in contact with the air (Fig. 16–9). When the strings are set into vibration, the sounding board or box is set into vibration as well. Since it has much greater area in contact with the air, it can produce a more intense sound wave. On an electric guitar, the sounding box is not so important since the vibrations of the strings are amplified electronically.

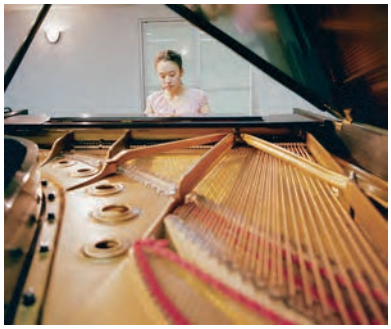
Wind Instruments

Instruments such as woodwinds, the brasses, and the pipe organ produce sound from the vibrations of standing waves in a column of air within a tube (Fig. 16–10). Standing waves can occur in the air of any cavity, including the human throat, but the frequencies present are complicated for any but very simple shapes such as the uniform, narrow tube of a flute or an organ pipe. In some instruments, a vibrating reed or the vibrating lip of the player helps to set up vibrations of the air column. In others, a stream of air is directed against one edge of the opening or mouthpiece, leading to turbulence which sets up the vibrations. Because of the disturbance, whatever its source, the air within the tube vibrates with a variety of frequencies, but only frequencies that correspond to standing waves will persist.

Recall that for a string fixed at both ends, Fig. 16–7, the standing waves have nodes (no movement) at the two ends, and one or more antinodes (large amplitude of vibration) in between. A node separates successive antinodes. The lowest-frequency standing wave, the *fundamental*, corresponds to a single antinode. The higher-frequency standing waves are called **overtones** or **harmonics**, as we saw in Section 15–9. Specifically, the first harmonic is the fundamental, the second harmonic (= first overtone) has twice the frequency of the fundamental, and so on.

⚠ CAUTION

Speed of standing wave on string ≠ speed of sound wave in air



(a)



(b)

FIGURE 16–9 (a) Piano, showing sounding board to which strings are attached; (b) sounding box (guitar).

FIGURE 16–10 Wind instruments: flute (right) and clarinet.



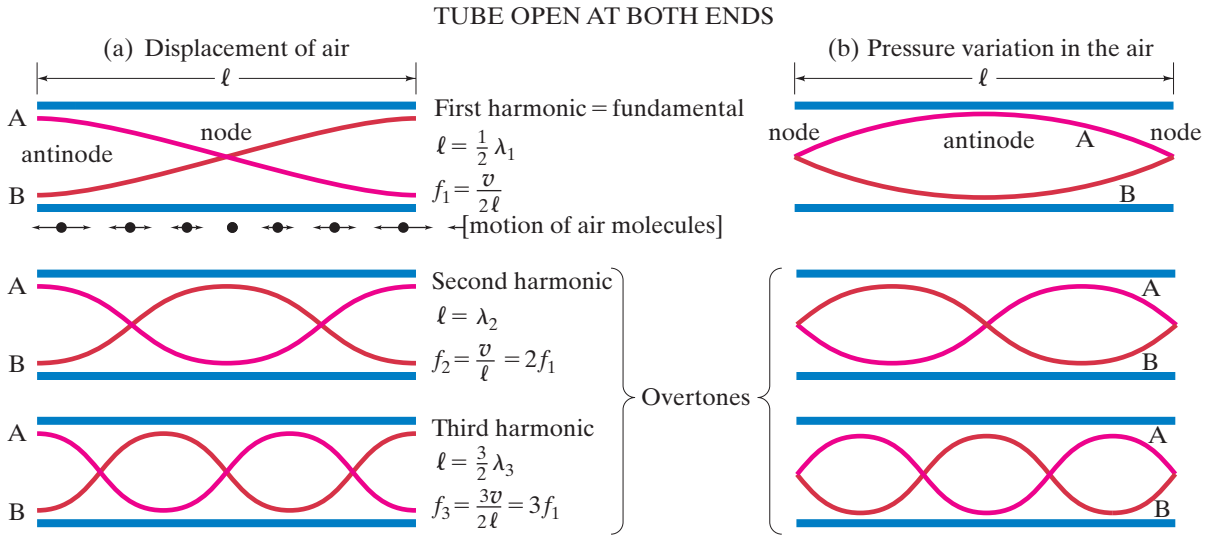


FIGURE 16–11 Graphs of the three simplest modes of vibration (standing waves) for a uniform tube open at both ends ("open tube"). These simplest modes of vibration are graphed in (a), on the left, in terms of the motion of the air (displacement), and in (b), on the right, in terms of air pressure. Each graph shows the wave form at two times, A and B, a half period apart. The actual motion of molecules for one case, the fundamental, is shown just below the tube at top left.

PHYSICS APPLIED
Wind instruments

The situation is similar for a column of air in a tube of uniform diameter, but we must remember that it is now air itself that is vibrating. We can describe the waves either in terms of the flow of the air—that is, in terms of the *displacement* of air—or in terms of the *pressure* in the air (see Figs. 16–2 and 16–3). In terms of displacement, the air at the closed end of a tube is a displacement node since the air is not free to move there, whereas near the open end of a tube there will be an antinode because the air can move freely in and out. The air within the tube vibrates in the form of longitudinal standing waves. A few of the possible modes of vibration for a tube open at both ends (called an **open tube**) are shown graphically in Fig. 16–11. Possible vibration modes for a tube that is open at one end but closed at the other (called a **closed tube**) are shown in Fig. 16–12. [A tube closed at *both* ends, having no connection to the outside air, would be useless as an instrument; but see Problem 43.] The graphs in part (a) of each Figure (left-hand sides) represent the displacement amplitude of the vibrating air in the tube. Note that these are graphs, and that the air molecules themselves oscillate *horizontally*, parallel to the tube length, as shown by the small arrows in the top diagram of Fig. 16–11a (on the left). The exact position of the antinode near the open end of a tube depends on the diameter of the tube, but if the diameter is small compared to the length, which is the usual case, the antinode occurs very close to the end as shown.[†] We assume this is the case in the following analysis. (The position of the antinode may also depend slightly on the wavelength and other factors.)

Let us look in detail at the open tube, in Fig. 16–11a, which might be an organ pipe or a flute. An open tube has displacement antinodes at both ends since the air is free to move at open ends. There must be at least one node within an open tube if there is to be a standing wave at all. A single node corresponds to the *fundamental frequency* of the tube. Since the distance between two successive nodes, or between two successive antinodes, is $\frac{1}{2}\lambda$, there is one-half of a wavelength within the length of the tube for the simplest case of the fundamental (top diagram in Fig. 16–11a): $\ell = \frac{1}{2}\lambda$, or $\lambda = 2\ell$. So the fundamental frequency is $f_1 = v/\lambda = v/2\ell$, where v is the velocity of sound in air (the air in the tube). The standing wave with two nodes is the *first overtone* or *second harmonic* and has half the wavelength, $\ell = \lambda$, and twice the frequency of the fundamental. Indeed, in a uniform tube open at both ends, the frequency of each overtone is an integral multiple of the fundamental frequency, as shown in Fig. 16–11a. This is just what is found for a string.

[†]The diameter D of a tube does affect the antinode at the open end of a tube. The end correction can be roughly approximated by adding $D/3$ to ℓ to give us an *effective length* for the tube in calculations.

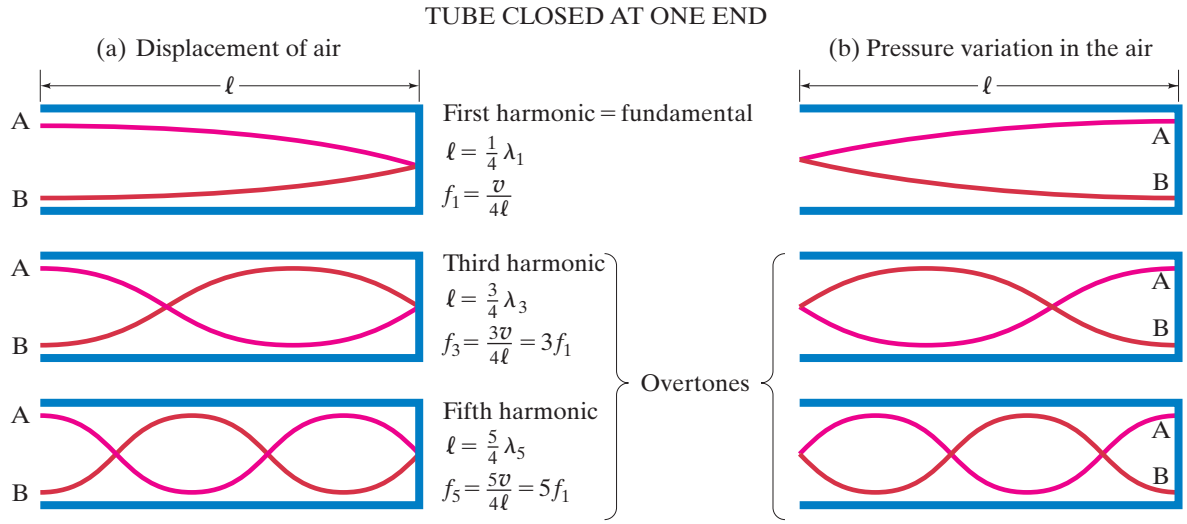


FIGURE 16–12 Modes of vibration (standing waves) for a tube closed at one end ("closed tube"). See caption for Fig. 16–11.

For a closed tube, shown in Fig. 16–12a, which could be an organ pipe, there is always a displacement node at the closed end (because the air is not free to move) and an antinode at the open end (where the air can move freely). Since the distance between a node and the nearest antinode is $\frac{1}{4}\lambda$, we see that the fundamental in a closed tube corresponds to only one-fourth of a wavelength within the length of the tube: $\ell = \lambda/4$, and $\lambda = 4\ell$. The fundamental frequency is thus $f_1 = v/4\ell$, or half that for an open pipe of the same length. There is another difference, for as we can see from Fig. 16–12a, only the odd harmonics are present in a closed tube: the overtones have frequencies equal to $3, 5, 7, \cdots$ times the fundamental frequency. There is no way for waves with $2, 4, 6, \cdots$ times the fundamental frequency to have a node at one end and an antinode at the other, and thus they cannot exist as standing waves in a closed tube.

Another way to analyze the vibrations in a uniform tube is to consider a description in terms of the *pressure* in the air, shown in part (b) of Figs. 16–11 and 16–12 (right-hand sides). Where the air in a wave is compressed, the pressure is higher, whereas in a wave expansion (or rarefaction), the pressure is less than normal. The open end of a tube is open to the atmosphere. Hence the pressure variation at an open end must be a *node*: the pressure does not alternate, but remains at the outside atmospheric pressure. If a tube has a closed end, the pressure at that closed end can readily alternate to be above or below atmospheric pressure. Hence there is a pressure *antinode* at a closed end of a tube. There can be pressure nodes and antinodes within the tube. Some of the possible vibrational modes in terms of pressure are shown in Fig. 16–11b for an open tube, and in Fig. 16–12b for a closed tube.

EXAMPLE 16–10 **Organ pipes.** What will be the fundamental frequency and first three overtones for a 26-cm-long organ pipe at 20°C if it is (*a*) open and (*b*) closed?

APPROACH All our calculations can be based on Figs. 16–11a and 16–12a.

SOLUTION (*a*) For the open pipe, Fig. 16–11a, the fundamental frequency is

$$f_1 = \frac{v}{2\ell} = \frac{343\ \text{m/s}}{2(0.26\ \text{m})} = 660\ \text{Hz}.$$

The speed v is the speed of sound in air (the air vibrating in the pipe). The overtones include all harmonics: 1320 Hz, 1980 Hz, 2640 Hz, and so on.

(*b*) For a closed pipe, Fig. 16–12a, the fundamental frequency is

$$f_1 = \frac{v}{4\ell} = \frac{343\ \text{m/s}}{4(0.26\ \text{m})} = 330\ \text{Hz}.$$

Only odd harmonics are present: the first three overtones are 990 Hz, 1650 Hz, and 2310 Hz.

NOTE The closed pipe plays 330 Hz, which, from Table 16–3, is E above middle C, whereas the open pipe of the same length plays 660 Hz, an octave higher.

Pipe organs use both open and closed pipes, with lengths from a few centimeters to 5 m or more. A flute acts as an open tube, for it is open not only where you blow into it, but is open also at the opposite end. The different notes on a flute are obtained by shortening the length of the vibrating air column, by uncovering holes along the tube (so a displacement antinode can occur at the hole). The shorter the length of the vibrating air column, the higher the fundamental frequency. A clarinet, on the other hand, acts as a closed tube (closed at the reed end where you blow), so will sound lower than a flute of similar length.

EXAMPLE 16–11 **Flute.** A flute is designed to play middle C (262 Hz) as the fundamental frequency when all the holes are covered. Approximately how long should the distance be from the mouthpiece to the far end of the flute? (This is only approximate since the antinode does not occur precisely at the mouthpiece.) Assume the temperature is 20°C.

APPROACH When all holes are covered, the length of the vibrating air column is the full length. The speed of sound in air at 20°C is 343 m/s. Because a flute is open at both ends, we use Fig. 16–11: the fundamental frequency f_1 is related to the length ℓ of the vibrating air column by $f = v/2\ell$.

SOLUTION Solving for ℓ, we find

$$\ell = \frac{v}{2f} = \frac{343\ \text{m/s}}{2(262\ \text{s}^{-1})} = 0.655\ \text{m}.$$

EXERCISE E To see why players of wind instruments "warm up" their instruments (so they will be in tune), determine the fundamental frequency of the flute of Example 16–11 when all holes are covered and the temperature is 10°C instead of 20°C.

*16–5 Quality of Sound, and Noise; Superposition

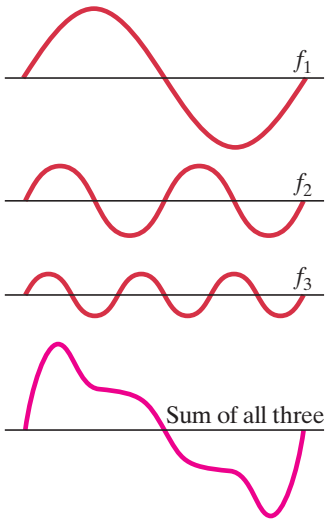


FIGURE 16–13 In the example shown here, the displacements of the fundamental and first two overtones are added at each point to get the "sum," or composite waveform. Normally many overtones are present.

Whenever we hear a sound, particularly a musical sound, we are aware of its loudness, its pitch, and also of a third aspect called "quality." For example, when a piano and then a flute play a note of the same loudness and pitch (say, middle C), there is a clear difference in the overall sound. We would never mistake a piano for a flute. This is what is meant by the **quality** of a sound. For musical instruments, the terms *timbre* and *tone color* are also used.

Just as loudness and pitch can be related to physically measurable quantities, so too can quality. The quality of a sound depends on the presence of overtones—their number and their relative amplitudes. Generally, when a note is played on a musical instrument, the fundamental as well as overtones are present simultaneously. Figure 16–13 illustrates how the *principle of superposition* (Section 15–6) applies to three wave forms, in this case the fundamental and first two overtones (with particular amplitudes): they add together at each point to give a composite **waveform**. By "waveform" we mean the shape of the wave in space at a given moment. Normally, more than two overtones are present. [Any complex wave can be analyzed into a superposition of sinusoidal waves of appropriate amplitudes, wavelengths, and frequencies—see Section 15–6. Such an analysis is called a *Fourier analysis*.]

The relative amplitudes of the overtones for a given note are different for different musical instruments, which is what gives each instrument its characteristic quality or timbre. A bar graph showing the relative amplitudes of the overtones for a given note produced by an instrument is called a *sound spectrum*. Several typical examples for different musical instruments are shown in Fig. 16–14. The fundamental usually has the greatest amplitude, and its frequency is what is heard as the pitch.

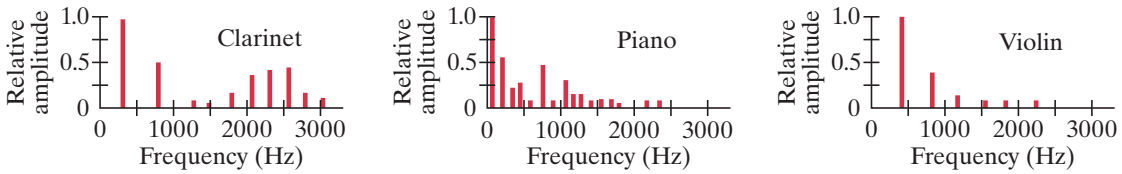


FIGURE 16–14 Sound spectra for different instruments for a specific note in each case. The spectra will change when the instruments play other notes. The clarinet is a bit complicated: it acts like a closed tube at lower frequencies, having only odd harmonics, but at higher frequencies all harmonics occur as for an open tube.

The manner in which an instrument is played strongly influences the sound quality. Plucking a violin string, for example, makes a very different sound than pulling a bow across it. The sound spectrum at the very start (or end) of a note (as when a hammer strikes a piano string) can be very different from the subsequent sustained tone. This too affects the subjective tone quality of an instrument.

An ordinary sound, like that made by striking two stones together, is a noise that has a certain quality, but a clear pitch is not discernible. Such a noise is a mixture of many frequencies which bear little relation to one another. A sound spectrum made of that noise would not show discrete lines like those of Fig. 16–14. Instead it would show a continuous, or nearly continuous, spectrum of frequencies. Such a sound we call "noise" in comparison with the more harmonious sounds which contain frequencies that are simple multiples of the fundamental.

16–6 Interference of Sound Waves; Beats

Interference in Space

We saw in Section 15–8 that when two waves simultaneously pass through the same region of space, they interfere with one another. Interference also occurs with sound waves.

Consider two large loudspeakers, A and B, a distance d apart as shown in Fig. 16–15. Let us assume the two speakers are emitting sound waves of the same single frequency and that they are in phase: that is, when one speaker is forming a compression, so is the other. (We ignore reflections from walls, floor, etc.) The curved lines in the diagram represent the crests of sound waves from each speaker at one instant in time. We must remember that for a sound wave, a crest is a compression in the air whereas a trough—which falls between two crests—is a rarefaction. A human ear or detector at a point such as C, which is the same distance from each speaker, will experience a loud sound because the interference will be constructive—two crests reach it at one moment, two troughs reach it a moment later. On the other hand, at a point such as D in the diagram, little if any of that single frequency sound will be heard because destructive interference occurs—compressions of one wave meet rarefactions of the other and vice versa (see Figs. 15–23 and 15–24 and the related discussion in Section 15–8).

An analysis where we graphically represent the waveforms of a single frequency is shown in Fig. 16–16. In Fig. 16–16a it can be seen that at point C, constructive interference occurs because both waves simultaneously have crests or simultaneously have troughs when they arrive at C. In Fig. 16–16b we see that, to reach point D, the wave from speaker B must travel a greater distance than the wave from A. Thus the wave from B lags behind that from A. In this diagram, point E is chosen so that the distance ED is equal to AD. Thus we see that if the distance BE is equal to precisely one-half the wavelength of the sound, the two waves will be exactly out of phase when they reach D, and destructive interference occurs.

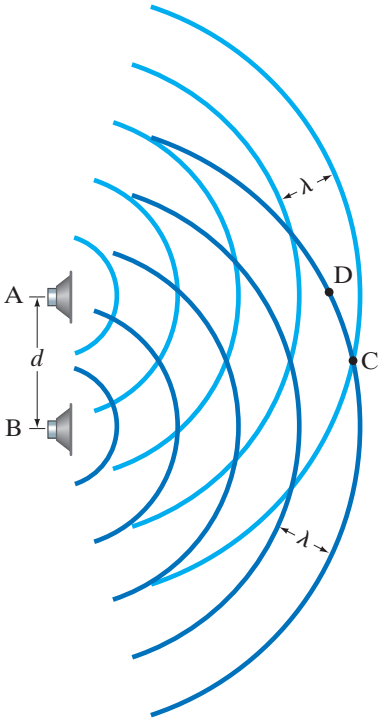


FIGURE 16–15 Sound waves of a single frequency emitted by two loudspeakers interfere.

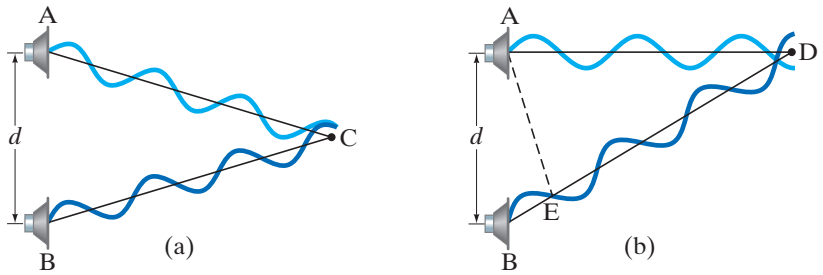


FIGURE 16–16 Sound waves of a single frequency from loudspeakers A and B (see Fig. 16–15) constructively interfere at C and destructively interfere at D. Shown here are graphical representations, not the actual longitudinal sound waves.

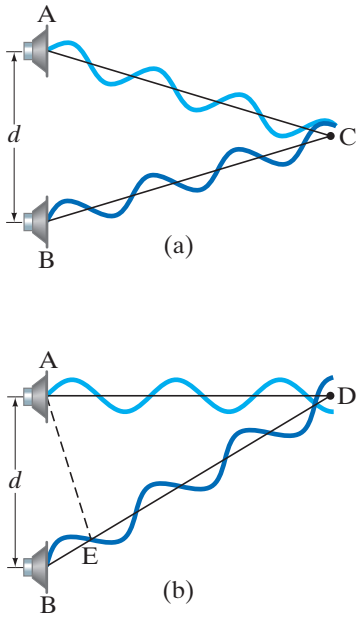


FIGURE 16–16 (Repeated.) Sound waves of a single frequency from loudspeakers A and B (see Fig. 16–15) constructively interfere at C and destructively interfere at D. Shown here are graphical representations, not the actual longitudinal sound waves.

This then is the criterion for determining at what points destructive interference occurs: destructive interference occurs at any point whose distance from one speaker is one-half wavelength greater than its distance from the other speaker. Notice that if this extra distance (BE in Fig. 16–16b) is equal to a whole wavelength (or 2, 3, $\cdots$ wavelengths), then the two waves will be in phase and *constructive interference* occurs. If the distance BE equals $\frac{1}{2}$, $1\frac{1}{2}$, $2\frac{1}{2}$, $\cdots$ wavelengths, *destructive interference* occurs.

It is important to realize that a person at point D in Fig. 16–15 or 16–16 hears nothing at all (or nearly so), yet sound is coming from both speakers. But if one of the speakers is turned off, the sound from the other speaker will be heard clearly. To confirm this, it would be easier to use two tuning forks at A and B because they emit (very nearly) a single frequency. It is harder to observe with musical instruments because they emit a range of frequencies (overtones, Fig. 16–14). Thus when a loudspeaker emits a whole range of frequencies, only specific wavelengths will destructively interfere completely at a given point. (If you move around near two loudspeakers you might hear a slight change in overall sound quality or timbre.)

EXAMPLE 16–12 **Loudspeakers' interference.** Two loudspeakers are 1.00 m apart. A person stands 4.00 m from one speaker. How far must this person be from the second speaker to detect destructive interference when the speakers emit an 1150-Hz sound? Assume the temperature is 20°C.

APPROACH To sense destructive interference, the person must be one-half wavelength closer to or farther from one speaker than from the other—that is, at a distance = 4.00 m $\pm$ $\lambda/2$. We can determine λ because we know f and v.

SOLUTION The speed of sound at 20°C is 343 m/s, so the wavelength of this sound is (Eq. 15–1)

$$\lambda = \frac{v}{f} = \frac{343\ \text{m/s}}{1150\ \text{Hz}} = 0.30\ \text{m}.$$

For destructive interference to occur, the person must be one-half wavelength farther from one loudspeaker than from the other, or 0.15 m. Thus the person must be 3.85 m or 4.15 m from the second speaker.

NOTE If the speakers are less than 0.15 m apart, there will be no location that is 0.15 m farther from one speaker than the other, and there will be no place where destructive interference could occur.

Beats—Interference in Time

We have been discussing interference of sound waves that takes place in *space*. An interesting and important example of interference that occurs in *time* is the phenomenon known as *beats*: If two sources of sound—say, two tuning forks—are close in frequency but not exactly the same, sound waves from the two sources interfere with each other in an interesting way. The sound level at a given position alternately rises and falls in time, because the two waves are sometimes in phase and sometimes out of phase due to their slightly different frequencies. The regularly spaced intensity changes are called **beats**.

To see how beats arise, consider two equal-amplitude sound waves of frequency $f_A = 50$ Hz and $f_B = 60$ Hz, respectively. In 1.00 s, the first source makes 50 vibrations whereas the second makes 60. We now examine the waves at one point in space equidistant from the two sources. The waveforms for each wave as a function of time, at a fixed position, are shown on the top graph of Fig. 16–17; the red line represents the 50-Hz wave, and the blue line represents the 60-Hz wave. The lower graph in Fig. 16–17 shows the sum of the two waves as a function of time. At time $t = 0$ the two waves are shown to be in phase and interfere constructively. Because the two waves vibrate at different rates, at time $t = 0.05$ s they are completely out of phase and interfere destructively.

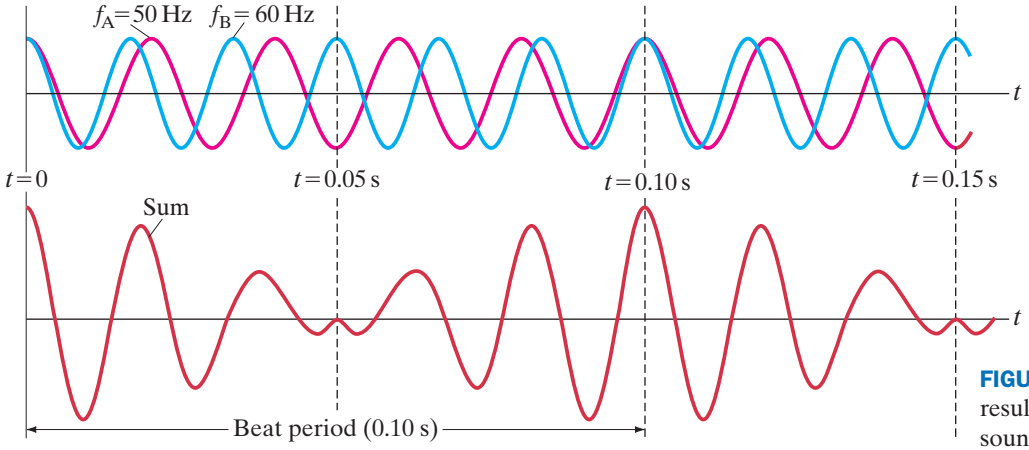


FIGURE 16–17 Beats occur as a result of the superposition of two sound waves of slightly different frequency.

At $t = 0.10\,\text{s}$, they are again in phase and the resultant amplitude again is large. Thus the resultant amplitude is large every 0.10 s and drops drastically in between. This rising and falling of the intensity[†] is what is heard as beats. In this case the beats are 0.10 s apart. That is, the **beat frequency** is ten per second, or 10 Hz. This result, that the beat frequency equals the difference in frequency of the two waves $(60\,\text{s}^{-1} - 50\,\text{s}^{-1})$, is valid in general, as we now show.

Let the two waves, of frequencies f_1 and f_2, be represented at a fixed point in space by

$$D_1 = A \sin 2\pi f_1 t$$

and

$$D_2 = A \sin 2\pi f_2 t.$$

The resultant displacement, by the principle of superposition, is

$$D = D_1 + D_2 = A(\sin 2\pi f_1 t + \sin 2\pi f_2 t).$$

Using the trigonometric identity $\sin\theta_1 + \sin\theta_2 = 2\sin\frac{1}{2}(\theta_1 + \theta_2)\cos\frac{1}{2}(\theta_1 - \theta_2)$, we have

$$D = \left[2A\cos 2\pi\left(\frac{f_1 - f_2}{2}\right)t\right]\sin 2\pi\left(\frac{f_1 + f_2}{2}\right)t. \qquad \textbf{(16–8)}$$

We can interpret Eq. 16–8 as follows. The superposition of the two waves results in a wave that vibrates at the average frequency of the two components, $(f_1 + f_2)/2$. This vibration has an amplitude given by the expression in brackets, and this amplitude varies in time, from zero to a maximum of $2A$ (the sum of the separate amplitudes), with a frequency of $(f_1 - f_2)/2$. A beat occurs whenever $\cos 2\pi[(f_1 - f_2)/2]t$ equals $+1$ or -1, as we can see in Fig. 16–17; that is, two beats occur per cycle, so the beat frequency is twice $(f_1 - f_2)/2$ which equals $f_1 - f_2$, the difference in frequency of the component waves.

PHYSICS APPLIED
Tuning a piano

The phenomenon of beats can occur with any kind of wave and is a very sensitive method for comparing frequencies. For example, to tune a piano, a piano tuner begins by listening for beats produced between a standard tuning fork and the frequency of a particular string on the piano, and knows that string is in tune when the beats disappear. The members of an orchestra tune up by listening for beats between a standard tone, usually A above middle C at 440 Hz produced by a piano or an oboe, and an A on their instrument.

Humans hear individual beats if they are only a few per second. For higher beat frequencies, they run together and up to about 20 Hz or so you can still hear an intensity variation like beats (maybe like a buzz). Above 20 Hz you will mainly hear the two original tones, but if the tones are loud enough you may also hear a separate low tone equal to their frequency difference.

[†]Beats will be heard even if the amplitudes are not equal, as long as the difference in amplitude is not great.

EXAMPLE 16–13 **Beats.** A tuning fork produces a steady 494-Hz tone. When this tuning fork is struck and held near a vibrating guitar string, twenty beats are counted in five seconds. What are the possible frequencies produced by the guitar string?

APPROACH For beats to occur, the string must vibrate at a frequency different from 494 Hz by whatever the beat frequency is.

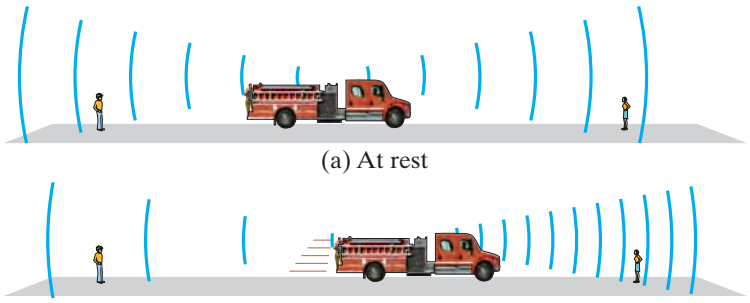
SOLUTION The beat frequency is

$$f_{\text{beat}} = 20 \text{ vibrations}/5 \text{ s} = 4 \text{ Hz}.$$

This is the difference of the frequencies of the two waves. Because one wave is known to be 494 Hz, the other must be either 490 Hz or 498 Hz.

16–7 Doppler Effect

You may have noticed that you hear the pitch of the siren on a speeding fire truck drop abruptly as it passes you. Or you may have noticed the change in pitch of a blaring horn on a fast-moving car as it passes by you. The pitch of the engine noise of a race car changes as the car passes an observer. When a source of sound is moving toward an observer, the pitch the observer hears is higher than when the source is at rest; and when the source is traveling away from the observer, the pitch is lower. This phenomenon is known as the **Doppler effect**[†] and occurs for all types of waves. Let us now see why it occurs, and calculate the difference between the perceived and source frequencies when there is relative motion between source and observer.



(a) At rest

(b) Fire truck moving

FIGURE 16–18 (a) Both observers on the sidewalk hear the same frequency from the fire truck at rest. (b) Doppler effect: the observer who sees the fire truck coming towards her hears a higher-frequency sound, and the observer behind the fire truck hears a lower-frequency sound.

Consider the siren of a fire truck at rest, which is emitting sound of a particular frequency in all directions as shown in Fig. 16–18a. The sound waves are moving at the speed of sound in air, v_{sound}, which is independent of the velocity of the source or observer. If our source, the fire truck, is moving, the siren emits sound at the same frequency as it does at rest. But the sound wavefronts it emits forward, in front of it, are closer together than when the fire truck is at rest, as shown in Fig. 16–18b. This is because the fire truck, as it moves, is partly "catching up" to the previously emitted wavefronts, and emits each crest closer to the previous one. Thus an observer on the sidewalk in front of the truck will detect more wave crests passing per second, so the frequency heard is higher. The wavefronts emitted behind the truck, on the other hand, are farther apart than when the truck is at rest because the truck is speeding away from them. Each new wavefront emitted is farther from the preceding one than when the truck is at rest. Hence, fewer wave crests per second pass by an observer behind the moving truck (Fig. 16–18b) and the perceived pitch is lower.

We can calculate the frequency shift by making use of Fig. 16–19. We assume the air (or other medium) is at rest in our reference frame. We consider first the stationary observer off to the right in Fig. 16–18. In Fig. 16–19a, the source of the sound is shown as a red dot, and is at rest. Two successive wave crests are shown, the first on the far right and the second, just emitted, is still near the source.

[†]After J. C. Doppler (1803–1853).

The distance between these crests in Fig. 16–19a is λ, the emitted wavelength. If the frequency of the source is f, then the time between emissions of wave crests is

$$T = \frac{1}{f} = \frac{\lambda}{v_{\text{sound}}}.$$

In Fig. 16–19b, the source is moving with a velocity v_{source} toward the observer. In the time T (just defined), the first wave crest has moved a distance $d = v_{\text{sound}}\,T = \lambda$, where v_{sound} is the velocity of the sound wave in air (which is the same whether the source is moving or not). In this same time, the source itself has moved a distance $d_{\text{source}} = v_{\text{source}}\,T$. Then the distance between successive wave crests, which is the wavelength λ', that the observer will perceive is

$$\begin{aligned}\lambda' &= d - d_{\text{source}} \\ &= \lambda - v_{\text{source}}\,T \\ &= \lambda - v_{\text{source}}\frac{\lambda}{v_{\text{sound}}} \\ &= \lambda\left(1 - \frac{v_{\text{source}}}{v_{\text{sound}}}\right).\end{aligned}$$

We subtract λ from both sides of this equation and find that the shift in wavelength, $\Delta\lambda$, is

$$\Delta\lambda = \lambda' - \lambda = -\lambda\frac{v_{\text{source}}}{v_{\text{sound}}}.$$

So the shift in wavelength is directly proportional to the source speed v_{source}. The frequency f' that will be perceived by our stationary observer on the ground is given by

$$f' = \frac{v_{\text{sound}}}{\lambda'} = \frac{v_{\text{sound}}}{\lambda\left(1 - \frac{v_{\text{source}}}{v_{\text{sound}}}\right)}.$$

Because $v_{\text{sound}}/\lambda = f$, then

$$f' = \frac{f}{\left(1 - \frac{v_{\text{source}}}{v_{\text{sound}}}\right)}. \qquad \left[\begin{array}{l}\text{source moving toward}\\ \text{stationary observer}\end{array}\right] \qquad \textbf{(16–9a)}$$

Because the denominator is less than 1, the observed frequency f' is higher than the source frequency f. That is, $f' > f$. For example, if a source emits a sound of frequency 400 Hz when at rest, then when the source moves toward a fixed observer with a speed of 30 m/s, the observer hears a frequency (at 20°C) of

$$f' = \frac{400\text{ Hz}}{1 - \frac{30\text{ m/s}}{343\text{ m/s}}} = 438\text{ Hz}.$$

Now consider a source moving *away* from the stationary observer at a speed v_{source} (observer on the left in Fig. 16–18). Using the same arguments as above, the wavelength λ' perceived by our observer will have the minus sign on d_{source} (second equation on this page) changed to plus:

$$\begin{aligned}\lambda' &= d + d_{\text{source}} \\ &= \lambda\left(1 + \frac{v_{\text{source}}}{v_{\text{sound}}}\right).\end{aligned}$$

The difference between the observed and emitted wavelengths will be $\Delta\lambda = \lambda' - \lambda = +\lambda\left(v_{\text{source}}/v_{\text{sound}}\right)$. The observed frequency of the wave is $f' = v_{\text{sound}}/\lambda' = (\lambda f)/\lambda'$, which equals

$$f' = \frac{f}{\left(1 + \frac{v_{\text{source}}}{v_{\text{sound}}}\right)}. \qquad \left[\begin{array}{l}\text{source moving away from}\\ \text{stationary observer}\end{array}\right] \qquad \textbf{(16–9b)}$$

If a source emitting at 400 Hz is moving away from a fixed observer at 30 m/s, the observer hears a frequency $f' = (400\text{ Hz})/\left[1 + (30\text{ m/s})/(343\text{ m/s})\right] = 368\text{ Hz}$.

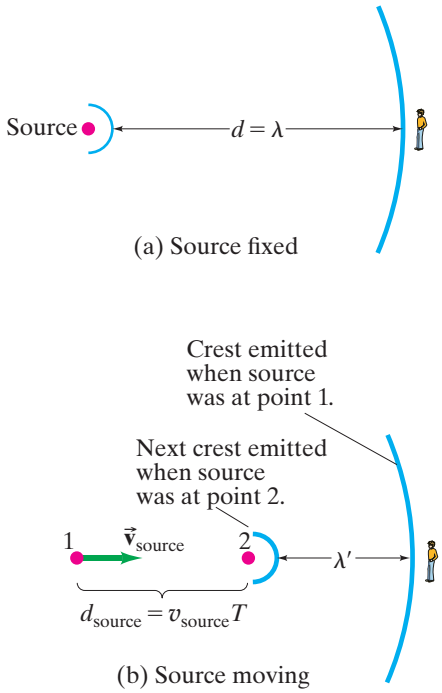


FIGURE 16–19 Determination of the frequency shift in the Doppler effect (see text). The red dot is the source. The observer is on the far right. (a) Source fixed. (b) Source moving toward observer.

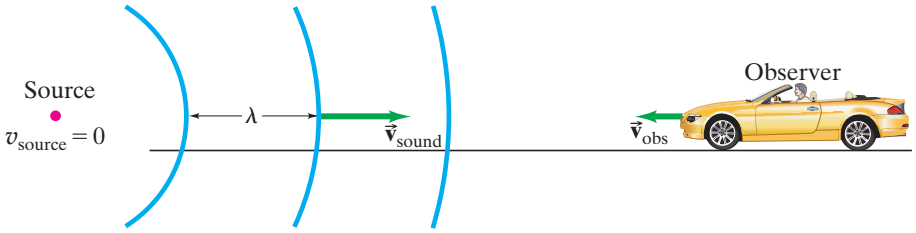


FIGURE 16–20 Observer moving with speed v_{obs} toward a stationary source detects wave crests passing at speed $v' = v_{sound} + v_{obs}$ where v_{sound} is the speed of the sound waves in air.

The Doppler effect also occurs when the source is at rest and the *observer* is in motion. If the observer is traveling *toward* the source, as in Fig. 16–20, the pitch heard is higher than that of the emitted source frequency. If the observer is traveling *away* from the source, the pitch heard is lower. Quantitatively the change in frequency is different than for the case of a moving source. With a fixed source and a moving observer, the distance between wave crests, the wavelength λ, is not changed. But the velocity of the crests with respect to the observer *is* changed. If the observer is moving toward the source, Fig. 16–20, the speed v' of the waves relative to the observer is a simple addition of velocities: $v' = v_{sound} + v_{obs}$, where v_{sound} is the velocity of sound in air (we assume the air is still) and v_{obs} is the velocity of the observer. Hence, the frequency f' heard is

$$f' = \frac{v'}{\lambda} = \frac{v_{sound} + v_{obs}}{\lambda}.$$

Because $\lambda = v_{sound}/f$, where f is the frequency of sound emitted by the source, then

$$f' = \frac{(v_{sound} + v_{obs})f}{v_{sound}},$$

or

$$f' = \left(1 + \frac{v_{obs}}{v_{sound}}\right)f. \qquad \begin{bmatrix}\text{observer moving toward}\\ \text{stationary source}\end{bmatrix} \quad \textbf{(16–10a)}$$

If the observer is moving *away* from the source, the relative velocity is $v' = v_{sound} - v_{obs}$, so

$$f' = \left(1 - \frac{v_{obs}}{v_{sound}}\right)f. \qquad \begin{bmatrix}\text{observer moving away}\\ \text{from stationary source}\end{bmatrix} \quad \textbf{(16–10b)}$$

EXAMPLE 16–14 **A police siren.** The siren of a police car at rest emits at a predominant frequency of 1600 Hz. (*a*) What frequency will you hear if you are at rest and the police car moves at 25.0 m/s toward you? (*b*) What frequency do you hear if you drive 25.0 m/s toward the police car which is at rest?

APPROACH The observer is fixed in (*a*), and the source moves, so we use Eqs. 16–9. The frequency you (the observer) hear is the emitted frequency f divided by the factor $\left(1 \pm v_{source}/v_{sound}\right)$ where v_{source} is the speed of the police car. Use the minus sign when the car moves toward you (giving a higher frequency); use the plus sign when the car moves away from you (lower frequency). In (*b*), you move but the source is at rest, so we use Eqs. 16–10.

SOLUTION (*a*) The car is moving toward you (at rest), so (Eq. 16–9a)

$$f' = \frac{f}{\left(1 - \frac{v_{source}}{v_{sound}}\right)} = \frac{1600\text{ Hz}}{\left(1 - \frac{25.0\text{ m/s}}{343\text{ m/s}}\right)} = 1726\text{ Hz} \approx 1730\text{ Hz}.$$

(*b*) You are moving toward a stationary source so we use Eq. 16–10a:

$$f' = \left(1 + \frac{v_{obs}}{v_{sound}}\right)f = \left(1 + \frac{25.0\text{ m/s}}{343\text{ m/s}}\right)(1600\text{ Hz}) = 1717\text{ Hz}.$$

EXERCISE F (*a*) What frequency would you hear if you were moving at 25.0 m/s away from a stationary police car whose siren emits at 1600 Hz? (*b*) What if you are at rest and the police car moves away from you at a speed of 25.0 m/s?

When a sound wave is reflected from a moving obstacle, the frequency of the reflected wave will be different from that of the incident wave because of the Doppler effect. This is illustrated in the following Example.

EXAMPLE 16–15 **Two Doppler shifts.** A pure 5000-Hz sound wave is emitted by a stationary source. This sound wave reflects from an object moving 3.50 m/s toward the source (Fig. 16–21). What is the frequency of the wave reflected by the moving object as detected by a detector at rest near the source?

APPROACH There are actually two Doppler shifts in this situation. First, the moving object acts like an observer moving toward the source with speed $v_{obs} = 3.50\,\text{m/s}$ (Fig. 16–21a) and so "detects" a sound wave of frequency (Eq. 16–10a) $f' = f\left[1 + (v_{obs}/v_{sound})\right]$. Second, reflection of the wave from the moving object is equivalent to the object reemitting the wave at the same frequency f', and thus acting effectively as a moving source with speed $v_{source} = 3.50\,\text{m/s}$ (Fig. 16–21b). The final frequency detected, f'', is given by $f'' = f'/\left[1 - (v_{source}/v_{sound})\right]$, Eq. 16–9a.

SOLUTION The frequency f' that is "detected" by the moving object is (Eq. 16–10a):

$$f' = \left(1 + \frac{v_{obs}}{v_{sound}}\right) f = \left(1 + \frac{3.50\,\text{m/s}}{343\,\text{m/s}}\right)(5000\,\text{Hz}) = 5051\,\text{Hz}.$$

The moving object now "emits" (reflects) a sound of frequency (Eq. 16–9a)

$$f'' = \frac{f'}{\left(1 - \dfrac{v_{source}}{v_{sound}}\right)} = \frac{5051\,\text{Hz}}{\left(1 - \dfrac{3.50\,\text{m/s}}{343\,\text{m/s}}\right)} = 5103\,\text{Hz}.$$

Thus the frequency shifts by 103 Hz.

NOTE Bats use this technique to be aware of their surroundings. This is also the principle behind Doppler radar as speed-measuring devices for vehicles, baseball pitches, tennis serves, storms such as tornadoes, and other objects.

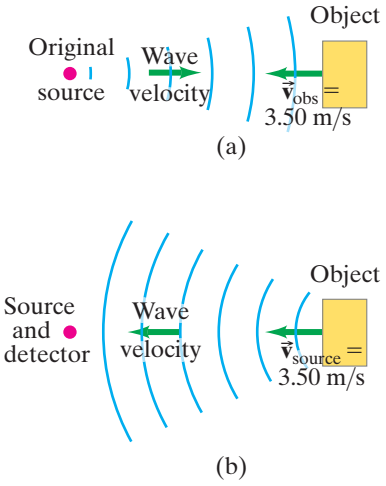


FIGURE 16–21 Example 16–15.

PHYSICS APPLIED
Bats use Doppler

The emitted wave and the reflected wave in Example 16–15, when mixed together (say, electronically), interfere with one another and beats are produced. The beat frequency is equal to the difference in the two frequencies, 103 Hz. This Doppler technique is used in a variety of medical applications, usually with ultrasonic waves in the megahertz frequency range. For example, ultrasonic waves reflected from red blood cells can be used to determine the velocity of blood flow. Similarly, the technique can be used to detect the movement of the chest of a young fetus and to monitor its heartbeat.

PHYSICS APPLIED
Doppler blood-flow meter and other medical uses

For convenience, we can write Eqs. 16–9 and 16–10 as a single equation that covers all cases of both source and observer in motion:

$$f' = f\left(\frac{v_{sound} \pm v_{obs}}{v_{sound} \mp v_{source}}\right). \qquad \left[\begin{array}{l}\text{source and}\\ \text{observer moving}\end{array}\right] \quad \textbf{(16–11)}$$

To get the signs right, recall from your own experience that the frequency is higher when observer and source approach each other, and lower when they move apart. Thus the upper signs in numerator and denominator apply if source and/or observer move toward each other; the lower signs apply if they are moving apart.

PROBLEM SOLVING
Getting the signs right

EXERCISE G How fast would a source have to approach an observer at rest for the observed frequency to be one octave above the produced frequency (frequency doubled)? (*a*) $\frac{1}{2}v_{sound}$, (*b*) v_{sound}, (*c*) $2v_{sound}$, (*d*) $4v_{sound}$.

Doppler Effect for Light

PHYSICS APPLIED
Doppler effect for EM waves, weather forecasting, speed traps

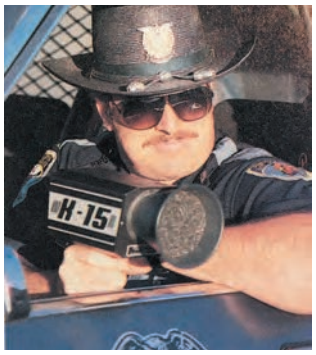


FIGURE 16–22 Gotcha. Your friendly sheriff checking your speed.

The Doppler effect occurs for waves other than sound. Light and other types of electromagnetic waves (such as for radar) exhibit the Doppler effect: although the formulas for the frequency shift are not identical to Eqs. 16–9 and 16–10, the effect is similar (see Chapter 44). One important application is for *weather forecasting* using radar. The time delay between the emission of a radar pulse and its reception after being reflected off raindrops gives the position of precipitation. Measuring the Doppler shift in frequency (as in Example 16–15) tells how fast the storm is moving and in which direction. *Radar guns* used by police (Fig. 16–22) work similarly, measuring a car's speed by the Doppler shift of electromagnetic waves. So too the speed of tennis balls and baseballs when shown on our TV screens.

Another important application is to astronomy, where the velocities of galaxies can be estimated from the Doppler shift. It is observed that light from all distant galaxies is shifted toward lower frequencies, indicating that the galaxies are moving away from us. This is called the **redshift** since red has the lowest frequency of visible light. The greater the frequency shift, the greater the velocity of recession. It is also observed that the farther the galaxies are from us, the faster they move away. These observations are the basis for the idea that the universe is expanding, and that the universe had a beginning called the "Big Bang," which would mean that the universe started very small and then expanded enormously over 14 billion years (Chapter 44).

PHYSICS APPLIED
Redshift in cosmology

*16–8 Shock Waves and the Sonic Boom

An object such as an airplane traveling faster than the speed of sound is said to have a **supersonic speed**. Such a speed is often given as a **Mach**[†] **number**, which is defined as the ratio of the speed of the object to the speed of sound in the surrounding medium. For example, a plane traveling 600 m/s high in the atmosphere, where the speed of sound is only 300 m/s, has a speed of Mach 2.

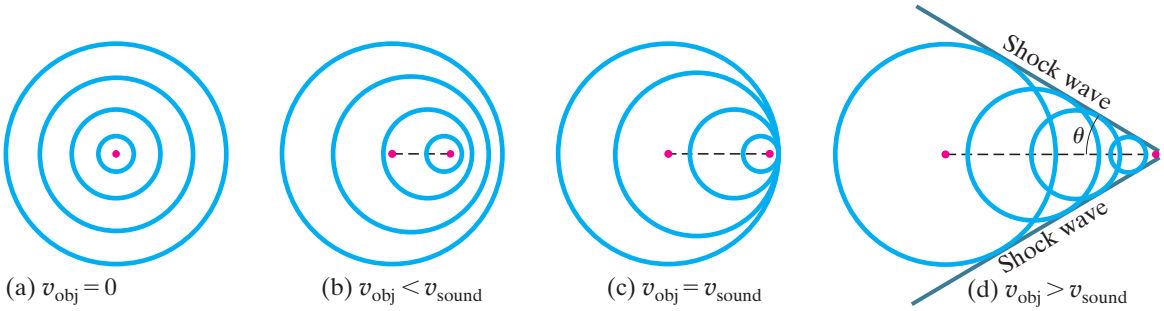
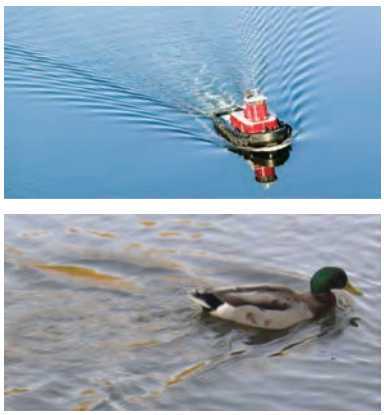


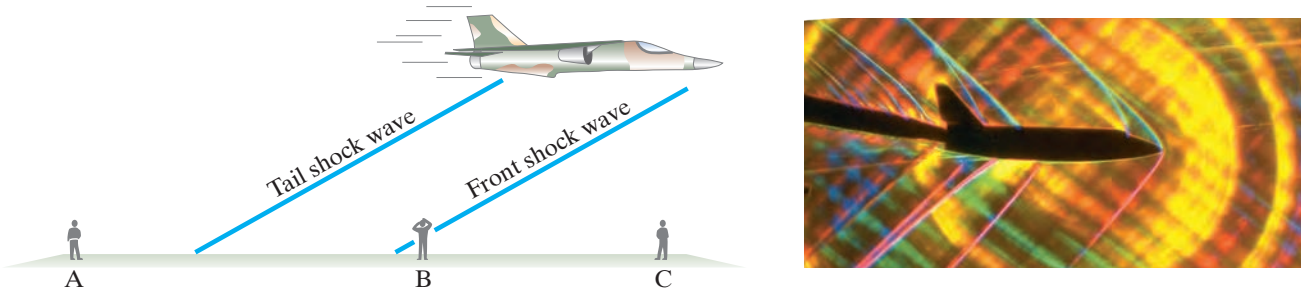
FIGURE 16–23 Sound waves emitted by an object, the red dot, (a) at rest or (b, c, and d) moving. (b) If the object's velocity is less than the velocity of sound, the Doppler effect occurs (note wavelength changes in front of and behind the moving object, as in Fig. 16–18b). (c) At $v_{obj} = v_{sound}$, the waves pile up in front, forming a "sound barrier." (d) If the object's velocity is greater than the velocity of sound, a shock wave is produced.

FIGURE 16–24 Bow waves produced by (a) a boat, (b) a duck.



When a source of sound moves at subsonic speeds (less than the speed of sound), the pitch of the sound is altered as we have seen (the Doppler effect); see also Figs. 16–23a and b. But if a source of sound moves faster than the speed of sound, a more dramatic effect known as a **shock wave** occurs. In this case, the source is actually "outrunning" the waves it produces. As shown in Fig. 16–23c, when the source is traveling *at* the speed of sound, the wave fronts it emits in the forward direction "pile up" directly in front of it, forming a *sound barrier* (see next page). When the object moves faster, at a supersonic speed, the wave fronts pile up on one another along the sides, as shown in Fig. 16–23d. The different wave crests overlap one another and form a single very large crest which is the shock wave. Behind this very large crest there is usually a very large trough. A shock wave is essentially the result of constructive interference of a large number of wave fronts. A shock wave in air is analogous to the *bow wave* of a boat (or duck) traveling faster than the speed of the water waves it produces, Fig. 16–24.

[†]After the Austrian physicist Ernst Mach (1838–1916).



(a) (b)

FIGURE 16–25 (a) The (double) sonic boom has already been heard by person A on the left. The front shock wave is just being heard by person B in the center. And both shock waves will shortly be heard by person C on the right. (b) Special photo of supersonic aircraft showing shock waves produced by different parts of the aircraft in the air.

When an airplane travels at supersonic speeds, greater than the speed of sound in the air, the noise it makes and its disturbance of the air form into a shock wave containing a tremendous amount of sound energy. When the shock wave passes a listener, it is heard as a loud **sonic boom**. A sonic boom lasts only a fraction of a second, but the energy it contains is sometimes sufficient to break windows and cause other damage. Actually, a sonic boom is made up of two or more booms since major shock waves can form at the front and the rear of the aircraft (Fig. 16–25), as well as at the wings and other parts. Bow waves of a boat are also multiple, as can be seen in Fig. 16–24a.

PHYSICS APPLIED
Sonic boom

When an aircraft accelerates toward the speed of sound, it encounters a barrier of sound waves in front of it (see Fig. 16–23c). To exceed the speed of sound, the aircraft needs extra thrust to pass through this **sound barrier**. This is called "breaking the sound barrier." Once a supersonic speed is attained, this barrier no longer impedes the motion. It is sometimes erroneously thought that a sonic boom is produced only at the moment an aircraft is breaking through the sound barrier. Actually, a shock wave follows the aircraft at all times it is traveling at supersonic speeds. A series of observers on the ground will each hear a loud "boom" as the shock wave passes, Fig. 16–25. The shock wave consists of a cone whose apex is at the aircraft. The angle of this cone, θ (see Fig. 16–23d), is given by

$$\sin\theta = \frac{v_{\text{sound}}}{v_{\text{obj}}}, \qquad \textbf{(16–12)}$$

where v_{obj} is the velocity of the object (the aircraft) and v_{sound} is the velocity of sound in the medium (air for an airplane) where the airplane is moving.

*16–9 Applications: Sonar, Ultrasound, and Medical Imaging

*Sonar

The reflection of sound is used in many applications to determine distance. The **sonar**[†] or **pulse-echo** technique is used to locate underwater objects. A transmitter sends out a sound pulse through the water, and a detector receives its reflection, or echo, a short time later. This time interval is carefully measured, and from it the distance to the reflecting object can be determined since the speed of sound in water is known. The depth of the sea and the location of reefs, sunken ships, submarines, or schools of fish can be determined in this way. The interior structure of the Earth is studied in a similar way by detecting reflections of waves traveling through the Earth whose source was a deliberate explosion (called "soundings"). An analysis of waves reflected from various structures and boundaries within the Earth reveals characteristic patterns that are also useful in the exploration for oil and minerals. (*Radar* used at airports to track aircraft involves a similar pulse-echo technique except that it uses electromagnetic (EM) waves, which, like visible light, travel with a speed of 3×10^8 m/s.)

PHYSICS APPLIED
Sonar: depth finding, Earth soundings

[†]Sonar stands for "*so*und *na*vigation *r*anging."

Sonar generally makes use of **ultrasonic** frequencies: that is, sound waves whose frequencies are above 20 kHz, beyond the range of human hearing. For sonar, the frequencies are typically in the range 20 kHz to 100 kHz. One reason for using ultrasound waves, other than the fact that they are inaudible, is that for these shorter wavelengths there is less diffraction (Section 15–11) so the beam spreads less, plus smaller objects can be detected.

*Ultrasound Medical Imaging

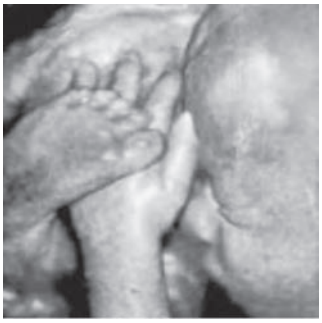


FIGURE 16–26 Ultrasound image of a human fetus within the uterus.

PHYSICS APPLIED
Ultrasound medical imaging

The diagnostic use of ultrasound in medicine, in the form of images (sometimes called *sonograms*), is an important and interesting application of physical principles. A **pulse-echo technique** is used, much like sonar, except that the frequencies used are in the range of 1 to 10 MHz $(1\text{ MHz} = 10^6\text{ Hz})$. A high-frequency sound pulse is directed into the body, and its reflections from boundaries or interfaces between organs and other structures and lesions in the body are then detected. *Tumors* and other abnormal growths, or pockets of fluid, can be distinguished; the action of heart valves and the development of a fetus (Fig. 16–26) can be examined; and information about various organs of the body, such as the brain, heart, liver, and kidneys, can be obtained. Although ultrasound does not replace X-rays, for certain kinds of diagnosis it is more helpful. Some kinds of tissue or fluid are not detected in X-ray photographs, but ultrasound waves are reflected from their boundaries. "Real-time" ultrasound images are like a movie of a section of the interior of the body.

The pulse-echo technique for medical imaging works like this. A brief pulse of ultrasound is emitted by a transducer that transforms an electrical pulse into a sound-wave pulse. Part of the pulse is reflected as echoes at each interface in the body, and most of the pulse (usually) continues on, Fig. 16–27a.

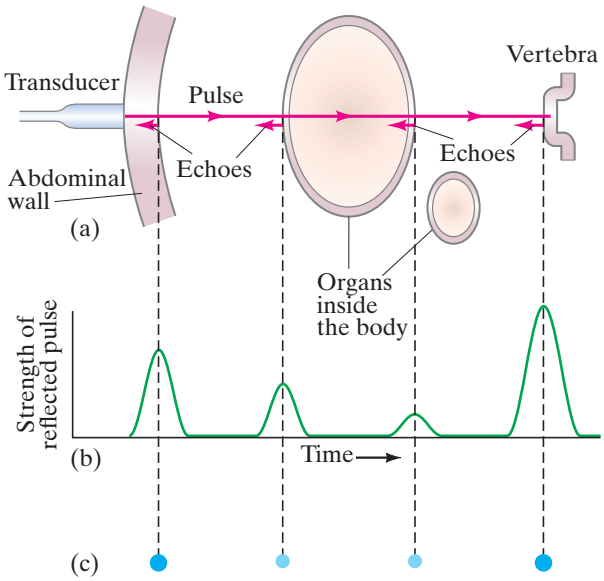


FIGURE 16–27 (a) Ultrasound pulse passes through the abdomen, reflecting from surfaces in its path. (b) Reflected pulses plotted as a function of time when received by transducer. The vertical dashed lines point out which reflected pulse goes with which surface. (c) "Dot display" for the same echoes: brightness of each dot is related to signal strength.

The detection of reflected pulses by the same transducer can then be displayed on the screen of a display monitor. The time elapsed from when the pulse is emitted to when each reflection (echo) is received is proportional to the distance to the reflecting surface. For example, if the distance from the transducer to the vertebra is 25 cm, the pulse travels a round-trip distance of $2 \times 25\text{ cm} = 0.50\text{ m}$. The speed of sound in human tissue is about 1540 m/s (close to that of sea water), so the time taken is

$$t = \frac{d}{v} = \frac{(0.50\text{ m})}{(1540\text{ m/s})} = 320\,\mu\text{s}.$$

The *strength* of a reflected pulse depends mainly on the difference in density of the two materials on either side of the interface and can be displayed as a pulse or as a dot (Figs. 16–27b and c). Each echo dot (Fig. 16–27c) can be represented as a point whose position is given by the time delay and whose brightness

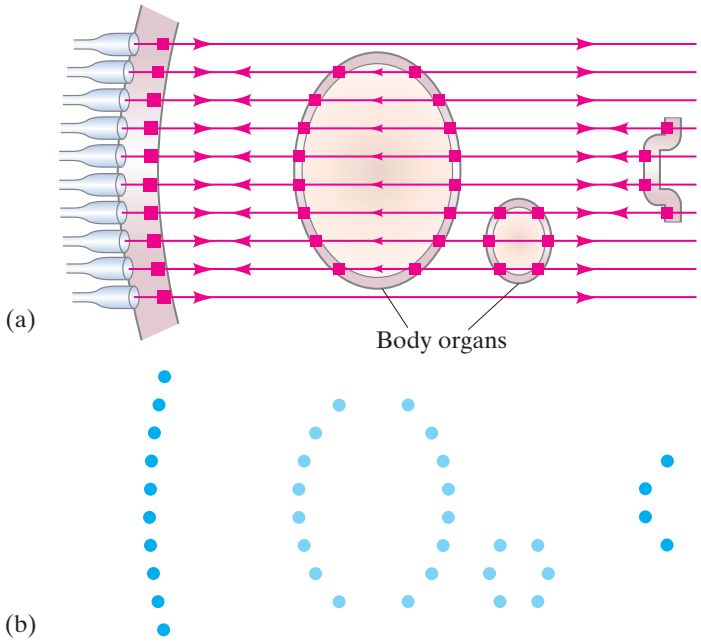


FIGURE 16–28 (a) Ten traces are made across the abdomen by moving the transducer, or by using an array of transducers. (b) The echoes from interfaces or boundaries (of organs) are plotted as dots to produce the image. More closely spaced traces would give a more detailed image.

depends on the strength of the echo. A two-dimensional image can then be formed out of these dots from a series of scans. The transducer is moved, or an array of transducers is used, each of which sends out a pulse at each position and receives echoes as shown in Fig. 16–28a. Each trace can be plotted, spaced appropriately one below the other, to form an image on a monitor screen as shown in Fig. 16–28b. Only 10 lines are shown in Fig. 16–28, so the image is very low resolution. More lines give a higher resolution image.

This type of ultrasound image shows the shape and position of objects such as internal organs. A **Doppler ultrasound image** uses the Doppler effect to measure movement of tissues (for a heart beating it is called, *Doppler echocardiography*) and the flow speed of fluids such as blood. Example 16–15, and the paragraph below it on page 479, discussed how this is done.

Summary

Sound travels as a longitudinal wave in air and other materials. In air, the speed of sound increases with temperature. At 20°C, it is about 343 m/s.

The **pitch** of a sound is determined by the frequency; the higher the frequency, the higher the pitch.

The **audible range** of frequencies for humans is at best about 20 Hz to 20,000 Hz (1 Hz = 1 cycle per second).

The **loudness** or **intensity** of a sound is related to the amplitude squared of the wave. Because the human ear can detect sound intensities from $10^{-12}\,\text{W/m}^2$ to over $1\,\text{W/m}^2$, sound levels are specified on a logarithmic scale. The **sound level** β, specified in decibels, is defined in terms of intensity I as

$$\beta(\text{in dB}) = 10\log\left(\frac{I}{I_0}\right), \qquad \textbf{(16–6)}$$

where the reference intensity I_0 is usually taken to be $10^{-12}\,\text{W/m}^2$.

Musical instruments are simple sources of sound in which *standing waves* are produced.

The strings of a stringed instrument may vibrate as a whole with nodes only at the ends; the frequency at which this standing wave occurs is called the **fundamental**. The fundamental frequency corresponds to a wavelength equal to twice the length of the string, $\lambda_1 = 2\ell$. The string can also vibrate at higher frequencies, called **overtones** or **harmonics**, in which there are one or more additional nodes on the string. The frequency of each harmonic is a whole-number multiple of the fundamental.

In wind instruments, standing waves are set up in the column of air within the tube.

The vibrating air in an **open tube** (open at both ends) has displacement antinodes at both ends. The fundamental frequency corresponds to a wavelength equal to twice the tube length: $\lambda_1 = 2\ell$. The harmonics have frequencies that are 1, 2, 3, 4, $\cdots$ times the fundamental frequency, just as for strings.

For a **closed tube** (closed at one end), the fundamental corresponds to a wavelength four times the length of the tube: $\lambda_1 = 4\ell$. Only the odd harmonics are present, equal to 1, 3, 5, 7, $\cdots$ times the fundamental frequency.

Sound waves from different sources can interfere with each other. If two sounds are at slightly different frequencies, **beats** can be heard at a frequency equal to the difference in frequency of the two sources.

The **Doppler effect** refers to the change in the pitch of a sound due to the motion of the source or of the observer. If source and observer are approaching each other, the perceived pitch is higher. If they are moving apart, the perceived pitch is lower.

[*Shock waves and a sonic boom occur when an object moves at a **supersonic** speed—faster than the speed of sound. **Ultrasonic**-frequency (higher than 20 kHz) sound waves are used in many applications, including sonar and medical imaging.]

Questions

1. What is the evidence that sound travels as a wave?
2. What is the evidence that sound is a form of energy?
3. Children sometimes play with a homemade "telephone" by attaching a string to the bottoms of two paper cups. When the string is stretched and a child speaks into one cup, the sound can be heard at the other cup (Fig. 16–29). Explain clearly how the sound wave travels from one cup to the other.



FIGURE 16–29 Question 3.

4. When a sound wave passes from air into water, do you expect the frequency or wavelength to change?
5. What evidence can you give that the speed of sound in air does not depend significantly on frequency?
6. The voice of a person who has inhaled helium sounds very high-pitched. Why?
7. Two tuning forks oscillate with the same amplitude, but one has twice the frequency. Which (if either) produces the more intense sound?
8. How will the air temperature in a room affect the pitch of organ pipes?
9. Explain how a tube might be used as a filter to reduce the amplitude of sounds in various frequency ranges. (An example is a car muffler.)
10. Why are the frets on a guitar (Fig. 16–30) spaced closer together as you move up the fingerboard toward the bridge?

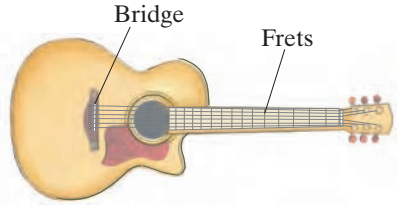


FIGURE 16–30 Question 10.

11. Why do the strings used for the lowest-frequency notes on a piano normally have wire wrapped around them?
12. A noisy truck approaches you from behind a building. Initially you hear it but cannot see it. When it emerges and you do see it, its sound is suddenly "brighter"—you hear more of the high-frequency noise. Explain. [*Hint*: See Section 15–11 on diffraction.]
13. Traditional methods of protecting the hearing of people who work in areas with very high noise levels have consisted mainly of efforts to block or reduce noise levels. *Noise-cancelling headphones* use a detector which detects the noise, inverts it electronically, then feeds it to the headphones *in addition to* the original loud noise. How could adding *more* noise reduce the sound levels reaching the ears?
14. In Fig. 16–16, if the frequency of the speakers is lowered, would the points D and C (where destructive and constructive interference occur) move farther apart or closer together?
15. Standing waves can be said to be due to "interference in space," whereas beats can be said to be due to "interference in time." Explain.
16. Consider the two waves shown in Fig. 16–31. Each wave can be thought of as a superposition of two sound waves with slightly different frequencies, as in Fig. 16–17. In which of the waves, (a) or (b), are the two component frequencies farther apart? Explain.

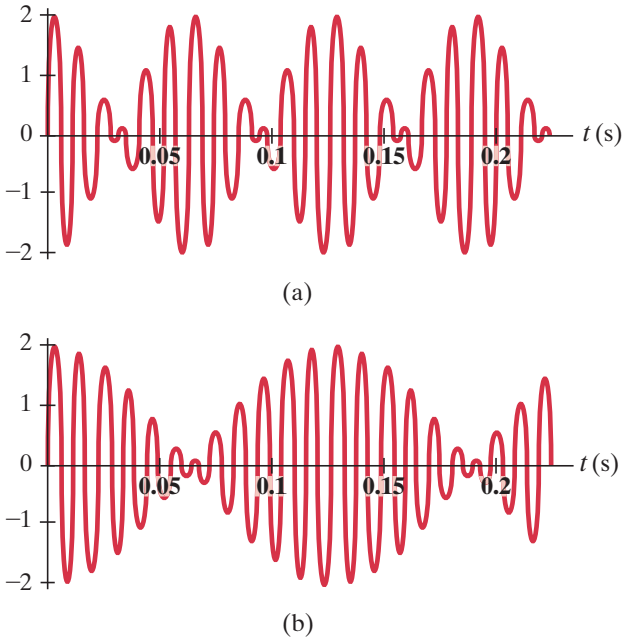


FIGURE 16–31 Question 16.

17. Is there a Doppler shift if the source and observer move in the same direction, with the same velocity? Explain.
18. If a wind is blowing, will this alter the frequency of the sound heard by a person at rest with respect to the source? Is the wavelength or velocity changed?
19. Figure 16–32 shows various positions of a child on a swing moving toward a person on the ground who is blowing a whistle. At which position, A through E, will the child hear the highest frequency for the sound of the whistle? Explain your reasoning.

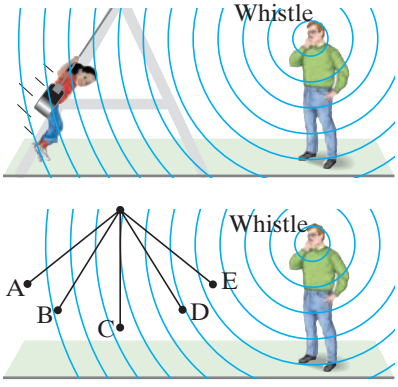


FIGURE 16–32 Question 19.

MisConceptual Questions

1. Do you expect an echo to return to you more quickly on a hot day or a cold day?
(*a*) Hot day.
(*b*) Cold day.
(*c*) Same on both days.

2. Sound waves are
(*a*) transverse waves characterized by the displacement of air molecules.
(*b*) longitudinal waves characterized by the displacement of air molecules.
(*c*) longitudinal waves characterized by pressure differences.
(*d*) Both (*b*) and (*c*).
(*e*) (*a*), (*b*), and (*c*).

3. The sound level near a noisy air conditioner is 70 dB. If two such units operate side by side, the sound level near them would be
(*a*) 70 dB.
(*b*) 73 dB.
(*c*) 105 dB.
(*d*) 140 dB.

4. To make a given sound seem twice as loud, how should a musician change the intensity of the sound?
(*a*) Double the intensity.
(*b*) Halve the intensity.
(*c*) Quadruple the intensity.
(*d*) Quarter the intensity.
(*e*) Increase the intensity by a factor of 10.

5. A musical note that is two octaves higher than a second note
(*a*) has twice the frequency of the second note.
(*b*) has four times the frequency of the second note.
(*c*) has twice the amplitude of the second note.
(*d*) is 3 dB louder than the second note.
(*e*) None of the above.

6. In which of the following is the wavelength of the fundamental vibration mode the same as the length of the string or tube?
(*a*) A string.
(*b*) An open tube.
(*c*) A tube closed at one end.
(*d*) All of the above.
(*e*) None of the above.

7. When a sound wave passes from air into water, what properties of the wave will change?
(*a*) Frequency.
(*b*) Wavelength.
(*c*) Wave speed.
(*d*) Both frequency and wavelength.
(*e*) Both wave speed and wavelength.

8. A guitar string vibrates at a frequency of 330 Hz with wavelength 1.40 m. The frequency and wavelength of this sound in air (20°C) as it reaches our ears are
(*a*) same frequency, same wavelength.
(*b*) higher frequency, same wavelength.
(*c*) lower frequency, same wavelength.
(*d*) same frequency, longer wavelength.
(*e*) same frequency, shorter wavelength.

9. A guitar player shortens the length of a guitar's vibrating string by pressing the string straight down onto a fret. The guitar then emits a higher-pitched note, because
(*a*) the string's tension has been increased.
(*b*) the string can vibrate with a much larger amplitude.
(*c*) the string vibrates at a higher frequency.

10. An organ pipe with a fundamental frequency f is open at both ends. If one end is closed off, the fundamental frequency will
(*a*) drop by half.
(*b*) not change.
(*c*) double.

11. "Hearing the ocean" by holding a sea shell close to your ear (Fig. 16–33) can be done also by using an empty paper towel tube. When you hold such a cardboard tube next to your ear, how would the sound of the "roaring ocean" change if you cover the other end of the tube with your hand?
(*a*) Pitch increases.
(*b*) Pitch decreases.
(*c*) Pitch doesn't change.



FIGURE 16–33
MisConceptual Question 11.

12. *Middle C* played on a piano sounds different from middle C played on a flute because
(*a*) the two instruments produce a different mix of harmonics.
(*b*) the frequency differs between the two instruments.
(*c*) the fundamental wave amplitudes are different.
(*d*) None of the above.

13. You are driving at 75 km/h. Your sister follows in the car behind at 75 km/h. When you honk your horn, your sister hears a frequency
(*a*) higher than the frequency you hear.
(*b*) lower than the frequency you hear.
(*c*) the same as the frequency you hear.
(*d*) You cannot tell without knowing the horn's frequency.

14. A guitar string is vibrating at its fundamental frequency f. Which of the following is *not* true?
(*a*) Each small section of the guitar string oscillates up and down at a frequency f.
(*b*) The wavelength of the standing wave on the guitar string is $\lambda = v/f$, where v is the velocity of the wave on the string.
(*c*) A sound wave created by this vibrating string propagates through the air with frequency f.
(*d*) A sound wave created by this vibrating string propagates through the air with wavelength $\lambda = v/f$, where v is the velocity of sound in air.
(*e*) The wavelength of the standing wave on the guitar string is $\lambda = \ell$, where ℓ is the length of the string.

Problems

[Unless stated otherwise, assume $T = 20°C$ and $v_{sound} = 343$ m/s in air.]

16–1 Characteristics of Sound

1. (I) A hiker determines the length of a lake by listening for the echo of her shout reflected by a cliff at the far end of the lake. She hears the echo 3 s after shouting. Estimate the length of the lake.
2. (I) (*a*) Calculate the wavelengths in air at 20°C for sounds in the maximum range of human hearing, 20 Hz to 20,000 Hz. (*b*) What is the wavelength of a 22-MHz ultrasonic wave?
3. (II) On a warm summer day (31°C), it takes 4.40 s for an echo to return from a cliff across a lake. On a winter day, it takes 4.80 s. What is the temperature on the winter day?
4. (II) An ocean fishing boat is drifting just above a school of tuna on a foggy day. Without warning, an engine backfire occurs on another boat 1.55 km away (Fig. 16–34). How much time elapses before the backfire is heard (*a*) by the fish, and (*b*) by the fishermen?

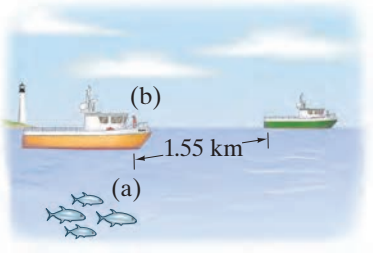


FIGURE 16–34 Problem 4.

5. (II) Calculate the percent error made over one mile of distance by the "5-second rule" for estimating the distance from a lightning strike if the temperature is (*a*) 30°C, and (*b*) 10°C.
6. (III) A stone is dropped from the top of a cliff. The splash it makes when striking the water below is heard 2.5 s later. How high is the cliff?

16–2 Mathematical Representation of Waves

7. (I) The pressure amplitude of a sound wave in air $(\rho = 1.29\ \text{kg/m}^3)$ at 0°C is 3.0×10^{-3} Pa. What is the displacement amplitude if the frequency is (*a*) 250 Hz and (*b*) 25 kHz?
8. (I) What must be the pressure amplitude in a sound wave in air (0°C) if the air molecules undergo a maximum displacement equal to the diameter of an oxygen molecule, about 3.0×10^{-10} m? Assume a sound-wave frequency of (*a*) 75 Hz and (*b*) 7.5 kHz.
9. (II) Write an expression that describes the pressure variation as a function of x and t for the waves described in Problem 8.
10. (II) The pressure variation in a sound wave is given by

$$\Delta P = 0.0035 \sin(0.38\pi x - 1220\pi t),$$

where ΔP is in pascals, x in meters, and t in seconds. Determine (*a*) the wavelength, (*b*) the frequency, (*c*) the speed, and (*d*) the displacement amplitude of the wave. Assume the density of the medium to be $\rho = 2.3 \times 10^3\ \text{kg/m}^3$.

16–3 Intensity of Sound; Decibels

11. (I) What is the sound level of a sound whose intensity is $3.0 \times 10^{-6}\ \text{W/m}^2$?
12. (I) A 5000-Hz tone must have what sound level to seem as loud as a 100-Hz tone that has a 50-dB sound level? (See Fig. 16–6.)
13. (I) What are the lowest and highest frequencies that an ear can detect when the sound level is 40 dB? (See Fig. 16–6.)
14. (II) (*a*) What is the intensity of a sound at the pain level of 120 dB? (*b*) By what factor is the intensity greater than that of a whisper at 20 dB?
15. (II) Your ears can accommodate a huge range of sound levels. What is the ratio of highest to lowest intensity at (*a*) 100 Hz, (*b*) 5000 Hz? (See Fig. 16–6.)
16. (II) One CD player is said to have a *signal-to-noise ratio* of 86 dB, whereas for a second CD player it is 98 dB. What is the ratio of intensities of the signal and the background noise for each device?
17. (II) A person standing a certain distance from an airplane with four equally noisy jet engines is experiencing a sound level of 140 dB. What sound level would this person experience if the captain shut down all but one engine? [*Hint*: Add intensities, not dBs.]
18. (II) A 45-dB sound wave strikes an eardrum whose area is $5.0 \times 10^{-5}\ \text{m}^2$. (*a*) How much energy is received by the eardrum per second? (*b*) At this rate, how long would it take your eardrum to receive a total energy of 1.0 J?
19. (II) A fireworks shell explodes 100 m above the ground, creating colorful sparks. How much greater is the sound level of the explosion for a person at a point directly below the explosion than for a person a horizontal distance of 200 m away (Fig. 16–35)?

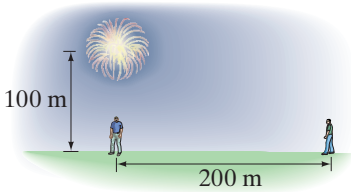


FIGURE 16–35 Problem 19.

20. (II) At a rock concert, a dB meter registered 130 dB when placed 2.8 m in front of a loudspeaker on stage. (*a*) What was the power output of the speaker, assuming uniform spherical spreading of the sound and neglecting absorption in the air? (*b*) How far away would the sound level be 85 dB?
21. (II) (*a*) Calculate the maximum displacement of air molecules when a 440-Hz sound wave passes whose intensity is at the threshold of pain (120 dB). (*b*) What is the pressure amplitude in this wave?
22. (II) If the amplitude of a sound wave is made 3.5 times greater, (*a*) by what factor will the intensity increase? (*b*) By how many dB will the sound level increase?
23. (II) Two sound waves have equal displacement amplitudes, but one has 3.5 times the frequency of the other. (*a*) Which has the greater pressure amplitude and by what factor is it greater? (*b*) What is the ratio of their intensities?

24. (II) A jet plane emits 5.0×10^5 J of sound energy per second. (*a*) What is the sound level 25 m away? Now consider that air absorbs sound at a rate of about 7.0 dB/km; calculate what the sound level will be (*b*) 1.00 km and (*c*) 5.00 km away from this jet plane, taking into account air absorption.

25. (II) (*a*) Estimate the power output of sound from a person speaking in normal conversation. Use Table 16–2. Assume the sound spreads roughly uniformly over a sphere centered on the mouth. (*b*) How many people would it take to produce a total sound output of 60 W of ordinary conversation? [*Hint*: Add intensities, not dBs.]

26. (II) What would be the sound level (in dB) of a sound wave in air that corresponds to a displacement amplitude of vibrating air molecules of 0.15 mm at 440 Hz?

27. (II) Expensive amplifier A is rated at 280 W, while the more modest amplifier B is rated at 45 W. (*a*) Estimate the sound level in decibels you would expect at a point 3.5 m from a loudspeaker connected in turn to each amp set to full output. (*b*) What output wattage amp would sound about twice as loud as the 45-W amp?

16–4 Sources of Sound: Strings and Air Columns

28. (I) What would you estimate for the length of a bass clarinet, assuming that it is modeled as a closed tube and that the lowest note that it can play is a $D^\flat$ whose frequency is 69 Hz?

29. (I) An organ pipe is 108 cm long. Determine the fundamental and first three audible overtones if the pipe is (*a*) closed at one end, and (*b*) open at both ends.

30. (I) If you were to build a pipe organ with open-tube pipes spanning the range of human hearing (20 Hz to 20 kHz), what would be the range of the lengths of pipes required?

31. (II) Estimate the number of octaves in the human audible range, 20 Hz to 20 kHz.

32. (II) Estimate the frequency of the "sound of the ocean" when you put your ear very near a 15-cm-diameter seashell.

33. (II) A tight guitar string has a frequency of 540 Hz as its third harmonic. What will be its fundamental frequency if it is fingered at a length of only 70% of its original length?

34. (II) How far from the mouthpiece of the flute in Example 16–11 should the hole be that must be uncovered to play F above middle C at 349 Hz?

35. (II) (*a*) Determine the length of an open organ pipe that emits middle C (262 Hz) when the temperature is 18°C. (*b*) What are the wavelength and frequency of the fundamental standing wave in the tube? (*c*) What are λ and f in the traveling sound wave produced in the outside air?

36. (II) An unfingered guitar string is 0.68 m long and is tuned to play E above middle C (330 Hz). (*a*) How far from the end of this string must a fret (and your finger, Fig. 16–8) be placed to play A above middle C (440 Hz)? (*b*) What is the wavelength on the string of this 440-Hz wave? (*c*) What are the frequency and wavelength of the sound wave produced in air at 22°C by this fingered string?

37. (II) An organ is in tune at 24°C. By what percent will the frequency be off at 11°C?

38. (II) A particular organ pipe can resonate at 264 Hz, 440 Hz, and 616 Hz, but not at any other frequencies in between. (*a*) Show why this is an open or a closed pipe. (*b*) What is the fundamental frequency of this pipe?

39. (II) A pipe in air at 21.5°C is to be designed to produce two successive harmonics at 308 Hz and 352 Hz. How long must the pipe be, and is it open or closed?

40. (II) A uniform narrow tube 1.50 m long is open at both ends. It resonates at two successive harmonics of frequencies 275 Hz and 330 Hz. What is (*a*) the fundamental frequency, and (*b*) the speed of sound in the gas in the tube?

41. (II) How many overtones are present within the audible range for a 2.18-m-long organ pipe at 20°C (*a*) if it is open, and (*b*) if it is closed?

42. (II) In a *quartz oscillator*, used as a stable "clock" in electronic devices, transverse (shear) standing sound waves are excited across the thickness d of a quartz disk and its frequency f is detected electronically. The parallel faces of the quartz are unsupported and so behave as "free ends" when the standing waves are excited (see Fig. 16–36). If the oscillator is designed to operate with the first overtone, f_3, determine the required disk thickness d if $f_3 = 24.0$ MHz. The density and shear modulus of quartz are $\rho = 2650\ \text{kg/m}^3$ and $G = 2.95 \times 10^{10}\ \text{N/m}^2$.

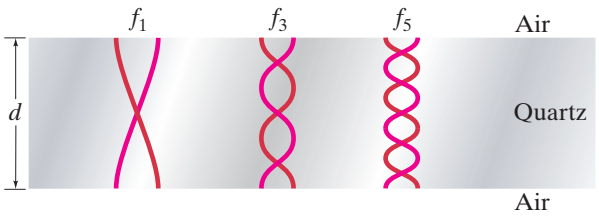


FIGURE 16–36 Problem 42.

43. (II) Determine the fundamental and first overtone frequencies when you are in a 9.0-m-long hallway with all doors closed. Model the hallway as a *tube closed at both ends*.

44. (II) The human ear canal is approximately 2.5 cm long. It is open to the outside and is closed at the other end by the eardrum. Estimate the frequencies (in the audible range) of the standing waves in the ear canal. What is the relationship of your answer to the information in the graph of Fig. 16–6?

45. (III) When a player's finger presses a guitar string down onto a fret, the length of the vibrating portion of the string is shortened, thereby increasing the string's fundamental frequency (see Fig. 16–37). The string's tension and mass per unit length remain unchanged. If the unfingered length of the string is $\ell = 75.0$ cm, determine the positions x of the first six frets, if each fret raises the pitch of the fundamental by one musical note in comparison to the neighboring fret. On the equally tempered *chromatic scale*, the ratio of frequencies of neighboring notes is $2^{1/12}$.

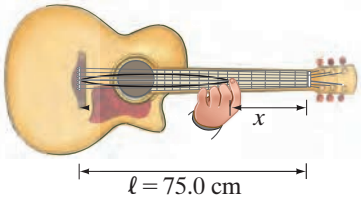


FIGURE 16–37 Problem 45.

*16–5 Quality of Sound; Superposition

*46. (II) Approximately what are the intensities of the first two overtones of a violin compared to the fundamental, Fig. 16–14? How many decibels softer than the fundamental are the first and second overtones?

16–6 Interference; Beats

47. (I) A piano tuner hears one beat every 2.0 s when trying to adjust two strings, one of which is sounding 340 Hz. How far off in frequency is the other string?

48. (I) A certain *dog whistle* operates at 23.5 kHz, while another (brand X) operates at an unknown frequency. If humans can hear neither whistle when played separately, but a shrill whine of frequency 3000 Hz occurs when they are played simultaneously, estimate the operating frequency of brand X.

49. (I) A guitar string produces 3 beats/s when sounded with a 350-Hz tuning fork and 8 beats/s when sounded with a 355-Hz tuning fork. What is the vibrational frequency of the string? Explain your reasoning.

50. (II) What is the beat frequency if middle C (262 Hz) and C♯ (277 Hz) are played together? What if each is played two octaves lower (each frequency reduced by a factor of 4)?

51. (II) Two violin strings are tuned to the same frequency, 294 Hz. The tension in one string is then decreased by 2.5%. What will be the beat frequency heard when the two strings are played together? [*Hint*: Recall Eq. 15–2.]

52. (II) The two sources of sound in Fig. 16–15 face each other and emit sounds of equal amplitude and equal frequency (305 Hz) but 180° out of phase. For what minimum separation of the two speakers will there be some point at which (*a*) complete constructive interference occurs and (*b*) complete destructive interference occurs? (Assume $T = 20°C$.)

53. (II) How many beats will be heard if two identical flutes, each 0.66 m long, try to play middle C (262 Hz), but one is at 5.0°C and the other at 25°C?

54. (II) Two loudspeakers are placed 3.00 m apart, as shown in Fig. 16–38. They emit 454-Hz sounds, in phase. A microphone is placed 3.20 m distant from a point midway between the two speakers, where an intensity maximum is recorded. (*a*) How far must the microphone be moved to the right to find the first intensity minimum? (*b*) Suppose the speakers are reconnected so that the 454-Hz sounds they emit are exactly out of phase. At what positions are the intensity maximum and minimum now?

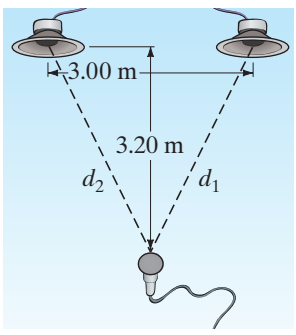


FIGURE 16–38 Problem 54.

55. (II) Two loudspeakers are 1.60 m apart. A person stands 3.00 m from one speaker and 3.50 m from the other. (*a*) What is the lowest frequency at which destructive interference will occur at this point if the speakers are in phase? (*b*) Calculate two other frequencies that also result in destructive interference at this point (give the next two highest). Let $T = 20°C$.

56. (III) A source emits sound of wavelengths 2.54 m and 2.72 m in air. (*a*) How many beats per second will be heard? (Assume $T = 20°C$.) (*b*) How far apart in space are the regions of maximum intensity?

16–7 Doppler Effect

57. (I) The predominant frequency of a certain fire truck's siren is 1650 Hz when at rest. What frequency do you detect if you move with a speed of 25.0 m/s (*a*) toward the fire truck, and (*b*) away from it?

58. (II) A bat at rest sends out ultrasonic sound waves at 50.0 kHz and receives them returned from an object moving directly away from it at 22.0 m/s. What is the received sound frequency?

59. (II) In one of the original Doppler experiments, a tuba was played at a frequency of 75 Hz on a moving flat train car, and a second identical tuba played the same tone while at rest in the railway station. What beat frequency was heard in the station if the train car approached the station at a speed of 14.0 m/s?

60. (II) Two trains emit 508-Hz whistles. One train is stationary. The conductor on the stationary train hears a 4.5-Hz beat frequency when the other train approaches. What is the speed of the moving train?

61. (II) A wave on the ocean surface with wavelength 44 m travels east at a speed of 11 m/s relative to the ocean floor. If, on this stretch of ocean, a ship is moving at 5.0 m/s (relative to the ocean floor), how often does the boat encounter a wave crest, if the boat is traveling (*a*) west, and (*b*) east?

62. (II) (*a*) Compare the shift in frequency if an 1800-Hz source is moving toward you at 18 m/s, versus you moving toward it at 18 m/s. Are the two frequencies exactly the same? Are they close? (*b*) Repeat the calculation for 160 m/s and then again (*c*) for 320 m/s. What can you conclude about the asymmetry of the Doppler formulas? (*d*) Show that at low speeds (relative to the speed of sound), the two formulas—source approaching and detector approaching—yield the same result.

63. (II) A police car sounding a siren with a frequency of 1580 Hz is traveling at 120.0 km/h. (*a*) What frequencies does an observer standing next to the road hear as the car approaches and as it recedes? (*b*) What siren frequencies are heard in a car traveling at 90.0 km/h in the opposite direction before and after passing the police car? (*c*) The police car passes a car traveling in the same direction at 80.0 km/h. What two frequencies are heard in this car?

64. (III) The Doppler effect using ultrasonic waves of frequency 2.25×10^6 Hz is used to monitor the heartbeat of a fetus. A (maximum) beat frequency of 220 Hz is observed. Assuming that the speed of sound in tissue is 1540 m/s, calculate the maximum velocity of the surface of the beating heart.

65. (III) A factory whistle emits sound of frequency 770 Hz. The wind velocity is 15.0 m/s from the north (heading south). What frequency will observers hear who are located, at rest, (*a*) due north, (*b*) due south, (*c*) due east, and (*d*) due west, of the whistle? What frequency is heard by a cyclist heading (*e*) north or (*f*) west, toward the whistle at 12.0 m/s? Assume $T = 20°C$.

*16–8 Shock Waves; Sonic Boom

***66.** (I) (*a*) How fast is an object moving on land if its speed at 24°C is Mach 0.33? (*b*) A high-flying jet cruising at 3000 km/h displays a Mach number of 3.1 on a screen. What is the speed of sound at that altitude?

***67.** (I) The wake of a speedboat is 12° in a lake where the speed of water waves is 2.2 km/h. What is the speed of the boat? The wake is the full width, 2θ in Fig. 16–24.

***68.** (II) A meteorite traveling 9400 m/s strikes the ocean. Determine the shock wave angle it produces (*a*) in the air just before entering the ocean, and (*b*) in the water just after entering. Assume $T = 20°C$.

***69.** (II) A space probe enters the thin atmosphere of a planet where the speed of sound is only about 55 m/s. (*a*) What is the probe's Mach number if its initial speed is 15,000 km/h? (*b*) What is the angle of the shock wave relative to the direction of motion?

***70.** (II) A supersonic jet traveling at Mach 2.0 at an altitude of 9500 m passes directly over an observer on the ground. Where will the plane be relative to the observer when the latter hears the sonic boom? (See Fig. 16–39.)

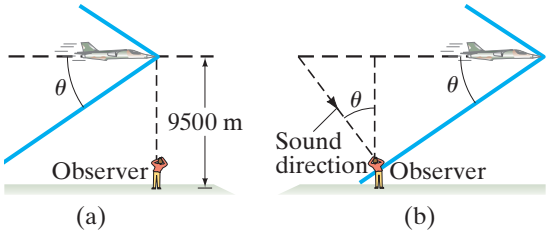


FIGURE 16–39 Problem 70.

***71.** (II) Show that the angle θ a sonic boom makes with the path of a supersonic object is given by Eq. 16–12.

***72.** (II) You look directly overhead and see a plane exactly 1.45 km above the ground flying faster than the speed of sound. By the time you hear the sonic boom, the plane has traveled a horizontal distance of 2.0 km. See Fig. 16–40. Determine (*a*) the angle of the shock cone, θ, and (*b*) the speed of the plane and its Mach number. Assume the speed of sound is 330 m/s.

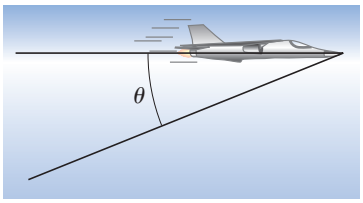


FIGURE 16–40 Problem 72.

General Problems

73. A *fish finder* uses a sonar device that sends 20,000-Hz sound pulses downward from the bottom of the boat, and then detects echoes. If the maximum depth for which it is designed to work is 85 m, what is the minimum time between pulses (in fresh water)?

74. A single mosquito 5.0 m from a person makes a sound close to the threshold of human hearing (0 dB). What will be the sound level of 200 such mosquitoes?

75. A **motion sensor** can accurately measure the distance d to an object repeatedly via the sonar technique used in Example 16–2. A short ultrasonic pulse is emitted and reflects from any objects it encounters, creating echo pulses upon their arrival back at the sensor. The sensor measures the time interval t between the emission of the original pulse and the arrival of the first echo. (*a*) The smallest time interval t that can be measured with high precision is 1.0 ms. What is the smallest distance (at 20°C) that can be measured with the motion sensor? (*b*) To measure an object's speed the motion sensor makes 15 distance measurements every second (that is, it emits 15 sound pulses per second at evenly spaced time intervals), the measurement of t must be completed within the time interval between the emissions of successive pulses. What is the largest distance (at 20°C) that can be measured with the motion sensor? (*c*) Assume that during a lab period the room's temperature increases from 20°C to 23°C. What percent error will this introduce into the motion sensor's distance measurements?

76. Two loudspeakers face each other at opposite ends of a long corridor. They are connected to the same source which produces a pure tone of 315 Hz. A person walks from one speaker toward the other at a speed of 1.6 m/s. What "beat" frequency does the person hear?

77. A stereo amplifier is rated at 225 W output at 1000 Hz. The power output drops by 12 dB at 15 kHz. What is the power output in watts at 15 kHz?

78. The sound level 9.75 m from a loudspeaker, placed in the open, is 115 dB. What is the acoustic power output (W) of the speaker, assuming it radiates equally in all directions?

79. A science museum has a display called a sewer pipe symphony. It consists of many plastic pipes of various lengths, which are open on both ends. (*a*) If the pipes have lengths of 3.0 m, 2.5 m, 2.0 m, 1.5 m and 1.0 m, what frequencies will be heard by a visitor's ear placed near the ends of the pipes? (*b*) Why does this display work better on a noisy day than on a quiet day?

80. Workers around jet aircraft typically wear protective devices over their ears. Assume that the sound level of a jet airplane engine, at a distance of 35 m, is 130 dB, and that the average human ear has an effective radius of 2.0 cm. What would be the power intercepted by an unprotected ear at a distance of 25 m from a jet airplane engine?

81. The radius of the hydrogen atom is 0.0529 nm. (*a*) What is the intensity (in dB) of a 100-Hz sound wave whose displacement amplitude is this size? (*b*) Repeat for a 2.0-kHz wave. Is either sound audible? Refer to Fig. 16–6.

82. In audio and communications systems, the **gain**, β, in decibels is defined as

$$\beta = 10 \log\left(\frac{P_{\text{out}}}{P_{\text{in}}}\right),$$

where P_{in} is the power input to the system and P_{out} is the power output. (*a*) A particular stereo amplifier puts out 135 W of power for an input of 1.0 mW. What is its gain in dB? (*b*) If a *signal-to-noise* ratio of 93 dB is specified, what is the noise power if the output signal is 10 W?

83. At a race track, you can estimate the speed of cars just by listening to the difference in pitch of the engine noise between approaching and receding cars. Suppose the sound of a certain car drops by a full octave (frequency halved) as it goes by on the straightaway. How fast is it going?

84. The diameter D of a tube does affect the position of the antinode at the open end of a tube. The end correction can be roughly approximated by adding $D/3$ to ℓ to give us an effective length for the tube in calculations. For a closed tube of length 55.0 cm and diameter 3.0 cm, what are the first four harmonics, taking the end correction into consideration?

85. A tuning fork is set into vibration above a vertical open tube filled with water (Fig. 16–41). The water level is allowed to drop slowly. As it does so, the air in the tube above the water level is heard to resonate with the tuning fork when the distance from the tube opening to the water level is 0.125 m and again at 0.395 m. What is the frequency of the tuning fork?

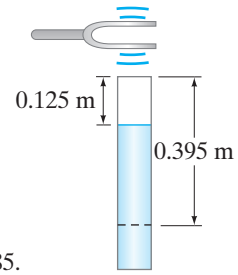


FIGURE 16–41 Problem 85.

86. A person hears the same pure tone in the 500 to 1000-Hz range coming from two sources. The sound is loudest at points equidistant from the two sources. To determine exactly what the frequency is, the person moves about and finds that the sound level is minimal at a point 0.25 m farther from one source than the other. What is the frequency of the sound?

87. Each string on a *violin* is tuned to a frequency $1\frac{1}{2}$ times that of its neighbor. The four equal-length strings are to be placed under the same tension; what must be the mass per unit length of each string relative to that of the lowest string?

88. A particular whistle produces sound by setting up the fundamental standing wave in an air column 8.40 cm long. The tube is closed at one end. The whistle blower is riding in a car moving away from you at 25 m/s. What frequency do you hear?

89. The A string of a violin is 32 cm long between fixed points with a fundamental frequency of 440 Hz and a mass per unit length of 7.2×10^{-4} kg/m. (*a*) What are the wave speed and tension in the string? (*b*) What is the length of the tube of a simple wind instrument (say, an organ pipe) closed at one end whose fundamental is also 440 Hz if the speed of sound is 343 m/s in air? (*c*) What is the frequency of the first overtone of each instrument?

90. Two loudspeakers are at opposite ends of a railroad car as it moves past a stationary observer at 12.0 m/s, as shown in Fig. 16–42. If the speakers emit identical sound frequencies of 348 Hz, what is the beat frequency heard by the observer when (*a*) he listens from position A, in front of the car, (*b*) he is between the speakers, at B, and (*c*) he hears the speakers after they have passed him, at C?

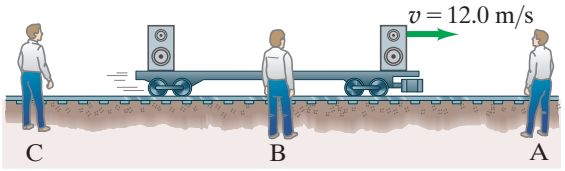


FIGURE 16–42 Problem 90.

91. At a painfully loud concert, a 120-dB sound wave travels away from a loudspeaker at 343 m/s. How much sound wave energy is contained in each 1.0-cm^3 volume of air in the region near this loudspeaker?

92. Two automobiles are equipped with the same single-frequency horn. When one is at rest and the other is moving toward the first at 6 m/s, the driver at rest hears a beat frequency of about 5 Hz. What is the frequency the horns emit? Assume $T = 20°\text{C}$.

93. Two open organ pipes, sounding together, produce a beat frequency of 6.0 Hz. The shorter one is 2.40 m long. How long is the other?

94. If the velocity of blood flow in the aorta is normally about 0.32 m/s, what beat frequency would you expect if 4.80-MHz ultrasound waves were directed along the flow and reflected from the red blood cells? Assume that the waves travel with a speed of 1.54×10^3 m/s.

95. A bat emits a series of high-frequency sound pulses as it approaches a moth. The pulses are approximately 65.0 ms apart, and each is about 3.0 ms long. How far away can the moth be detected by the bat so that the echo from one pulse returns before the next pulse is emitted?

96. The vibrating portion of an A-string on a particular *violin* measures 32 cm, and its fundamental frequency is precisely 440 Hz. (*a*) How far from the end of the string should the violinist place a finger so that the string plays a note at 588 Hz? (*b*) "Vibrato" in a violin is produced by sliding the finger back and forth along the vibrating string. If the violinist executes "vibrato" by moving the finger 0.5 cm to either side of the position found in part (*a*), what range of frequencies will result?

97. A *bugle* is a tube of fixed length that behaves as if it is open at both ends. A bugler, by adjusting his lips correctly and blowing with proper air pressure, can cause a harmonic (usually other than the fundamental) of the air column within the tube to sound loudly. Standard military tunes like *Taps* and *Reveille* require only four musical notes: G4 (392 Hz), C5 (523 Hz), E5 (659 Hz), and G5 (784 Hz). (*a*) For a certain length ℓ, a bugle will have a sequence of four consecutive harmonics whose frequencies very nearly equal those associated with the notes G4, C5, E5, and G5. Determine this ℓ. (*b*) Which harmonic is each of the (approximate) notes G4, C5, E5, and G5 for the bugle?

98. (*a*) Use the binomial expansion to show that Eqs. 16–9a and 16–10a become essentially the same for small relative velocity between source and observer. (*b*) What percent error would result if Eq. 16–10a were used instead of Eq. 16–9a for a relative velocity of 18.0 m/s?

99. The "alpenhorn" (Fig. 16–43) was once used to send signals from one Alpine village to another. Since lower frequency sounds are less susceptible to intensity loss, long horns were used to create deep sounds. When played as a musical instrument, the alpenhorn must be blown in such a way that only one of the overtones is resonating. The most popular alpenhorn is about 3.4 m long, and it is called the F♯ (or G♭) horn. What is the fundamental frequency of this horn, and which overtone is close to F♯? (See Table 16–3.) Model as a tube open at both ends.



FIGURE 16–43 Problem 99.

100. A source of sound waves (wavelength λ) is a distance ℓ from a detector. Sound reaches the detector directly, and also by reflecting off an obstacle, as shown in Fig. 16–44. The obstacle is equidistant from source and detector. When the obstacle is a distance d to the right of the line of sight between source and detector, as shown, the two waves arrive in phase. How much farther to the right must the obstacle be moved if the two waves are to be out of phase by $\frac{1}{2}$ wavelength, so destructive interference occurs? (Assume $\lambda \ll \ell, d$.)

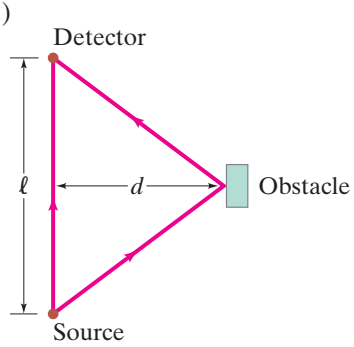


FIGURE 16–44
Problem 100.

101. A dramatic demonstration, called a "singing rod," involves a long, slender aluminum rod held in the hand near the rod's midpoint. The rod is stroked with the other hand. With a little practice, the rod can be made to "sing," or emit a clear, loud, ringing sound. For an 80-cm-long rod, (*a*) what is the fundamental frequency of the sound? (*b*) What is its wavelength in the rod, and (*c*) what is the traveling wavelength of the sound in air at 20°C?

102. A sound-insulating door reduces the sound level by 30 dB. What fraction of the initial sound intensity passes through this door?

103. A *Doppler flow meter* is used to measure the speed of blood flow. The transmitter and receiver are placed on the skin, as shown in Fig. 16–45. Typical sound-wave frequencies of about 5.0 MHz are used, which have a reasonable chance of being reflected from red blood cells. By measuring the frequency of the reflected waves, which are Doppler-shifted because the red blood cells are moving, the speed of the blood flow can be deduced. "Normal" blood flow speed is about 0.1 m/s. Suppose that an artery is partly constricted, so that the speed of the blood flow is increased, and the flow meter measures a Doppler shift of 780 Hz. What is the speed of blood flow in the constricted region? The effective angle between the sound waves (both transmitted and reflected) and the direction of blood flow is 45°. Assume the velocity of sound in tissue is 1540 m/s.

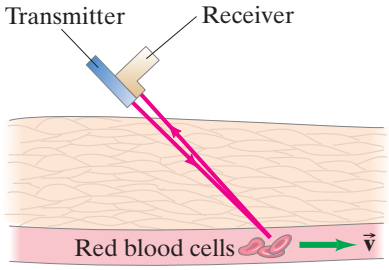


FIGURE 16–45
Problem 104.

104. Assuming that the maximum displacement of the air molecules in a sound wave is about the same as that of the speaker cone that produces the sound (Fig. 16–46), estimate by how much a loudspeaker cone moves for a fairly loud (105 dB) sound of (*a*) 8.0 kHz, and (*b*) 35 Hz.

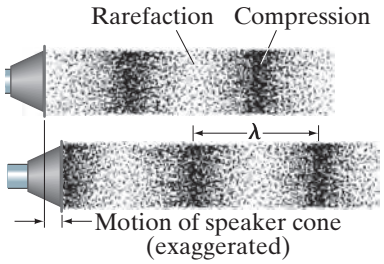


FIGURE 16–46
Problem 105.

ANSWERS TO EXERCISES

A: 4 times as intense.
B: (*b*).
C: 7 octaves, plus; each octave ends a factor of 2 higher in frequency than the previous one: 1 → 2, 4, 8, 16, 32, 64, 128, 256, … . [This is like counting in binary, $2^7 = 128$. See also Section 23–9.]
D: (*b*).
E: 257 Hz.
F: (*a*) 1483 Hz; (*b*) 1491 Hz.
G: (*a*).

Heating the air inside a "hot-air" balloon raises the air's temperature, causing it to expand, and forces air out the opening at the bottom. The reduced amount of air inside means its density is lower than the outside air, so there is a net buoyant force upward on the balloon. In this Chapter we study temperature and its effects on matter: thermal expansion and the gas laws.





CHAPTER 17

Temperature, Thermal Expansion, and the Ideal Gas Law

CONTENTS

17 – 1 Atomic Theory of Matter
17 – 2 Temperature and Thermometers
17 – 3 Thermal Equilibrium and the Zeroth Law of Thermodynamics
17 – 4 Thermal Expansion
*17 – 5 Thermal Stresses
17 – 6 The Gas Laws and Absolute Temperature
17 – 7 The Ideal Gas Law
17 – 8 Problem Solving with the Ideal Gas Law
17 – 9 Ideal Gas Law in Terms of Molecules: Avogadro's Number
*17–10 Ideal Gas Temperature Scale—a Standard

CHAPTER-OPENING QUESTION—Guess now!

A hot-air balloon, open at one end (see photos above), rises when the air inside is heated by a flame. For the following properties, is the air inside the balloon higher, lower, or the same as for the air outside the balloon?

(i) Temperature,
(ii) Pressure,
(iii) Density.

This Chapter is the first of four, 17 through 20, devoted to temperature, kinetic theory, heat, and the thermodynamics of processes of energy transfer as work and heat, including engines (Chapter 20). Much of this Chapter 17 deals with the concept of temperature, how it is measured, and how it affects materials. We also discuss the measured properties of gases such as pressure, volume, and temperature and how they affect each other (the gas laws).

The emphasis in this Chapter is on the concept of temperature. We begin, however, with a brief discussion of the theory that matter is made up of atoms and that these atoms are in continuous random motion. This theory is called *kinetic theory* ("kinetic," you may recall from Chapter 7, is Greek for "moving"), and we discuss it in more detail in Chapter 18.

17–1 Atomic Theory of Matter

The idea that matter is made up of atoms dates back to the ancient Greeks. According to the Greek philosopher Democritus, if a pure substance—say, a piece of iron—were cut into smaller and smaller bits, eventually a smallest piece of that substance would be obtained which could not be divided further. This smallest piece was called an **atom**, which in Greek means "indivisible." Today an atom is still the smallest piece of a substance, but we do not consider it indivisible. Rather it is viewed as consisting of a central nucleus (containing protons and neutrons) surrounded by electrons.

Today the atomic theory is universally accepted. The experimental evidence in its favor, however, came mainly in the eighteenth, nineteenth, and twentieth centuries, and much of it was obtained from the analysis of chemical reactions.

We will often speak of the relative masses of individual atoms and molecules—what we call the **atomic mass** or **molecular mass**, respectively. (The terms *atomic weight* and *molecular weight* are sometimes used.) These masses are based on arbitrarily assigning the most abundant form of carbon atom, ^{12}C, the atomic mass of exactly 12.0000 **unified atomic mass units** (u). In terms of kilograms,

$$1\,\text{u} = 1.6605 \times 10^{-27}\,\text{kg}.$$

The atomic mass of natural hydrogen is 1.0079 u, and the values for other atoms are included in the Periodic Table[†] inside the back cover of this book (and also in Appendix G). The molecular mass of a compound is the sum of atomic masses of the atoms making up the molecules of that compound.

An **element** is a substance, such as neon, gold, iron, or copper, that cannot be broken down into simpler substances by chemical means. **Compounds** are substances made up of elements, and can be broken down into them; examples are carbon dioxide and water. The smallest piece of an element is an **atom**; the smallest piece of a compound is a **molecule**. Molecules are made up of atoms; a molecule of water, for example, is made up of two atoms of hydrogen and one of oxygen; its chemical formula is H_2O.

An important piece of evidence for the atomic theory is called **Brownian motion**, named after the biologist Robert Brown, who is credited with its discovery in 1827. While he was observing tiny pollen grains suspended in water under his microscope, Brown noticed that the tiny grains moved about in erratic paths (Fig. 17–1), even though the water appeared to be perfectly still. The atomic theory easily explains Brownian motion if we assume that the atoms of any substance are continually in motion. Then Brown's tiny pollen grains are jostled about by the vigorous barrage of rapidly moving molecules of water.

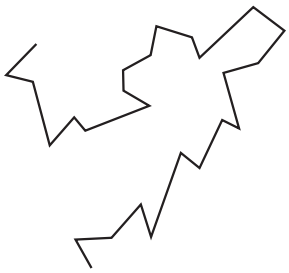
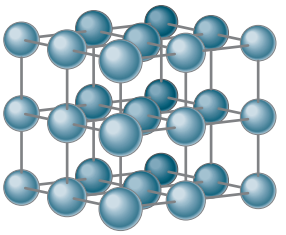
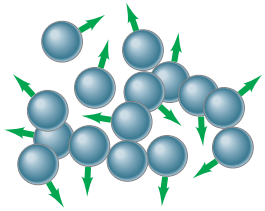


FIGURE 17–1 Path of a tiny particle (pollen grain, for example) suspended in water. The straight lines connect observed positions of the particle at equal time intervals.

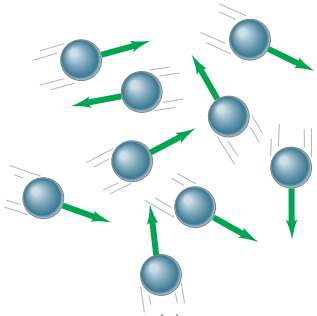
[†]The relative masses of different atoms came from analysis of chemical reactions, and the **law of definite proportions**. It states that when two or more elements combine to form a compound, they always do so in the same proportions by mass. For example, table salt is always formed from 23 parts sodium and 35 parts chlorine; and water from one part hydrogen and eight parts oxygen. A continuous theory of matter could not account for the law of definite proportions but atomic theory does: the proportions of each element that form a compound correspond to the relative masses of the combining atoms. One atom of sodium (Na) combines with one atom of chlorine (Cl) to form one molecule of salt (NaCl), and one atom of sodium has a mass 23/35 times as large as one of chlorine. Hydrogen, the lightest atom, was arbitrarily assigned the relative mass of 1. On this scale, carbon was about 12, oxygen 16, sodium 23, and so on. It was sometimes more complicated. For example, from the various compounds oxygen formed, its relative mass was judged to be 16; but this was inconsistent with the mass ratio in water of oxygen to hydrogen, only 8 to 1. This was explained by assuming two H atoms combine with one O atom to form a water molecule.



(a)



(b)



(c)

FIGURE 17–2 Atomic arrangements in (a) a crystalline solid, (b) a liquid, and (c) a gas.

In 1905, Albert Einstein examined Brownian motion from a theoretical point of view and was able to calculate from the experimental data the approximate size and mass of atoms and molecules. His calculations showed that the diameter of a typical atom is about 10^{-10} m.

At the start of Chapter 13, we distinguished the three common phases (or states) of matter—solid, liquid, gas—based on **macroscopic**, or "large-scale," properties. Now let us see how these three phases of matter differ, using the atomic or **microscopic** point of view. First of all, atoms and molecules must exert attractive forces on each other, because only this explains why a brick or a block of aluminum holds together in one piece. The attractive forces between molecules are of an electrical nature (more on this in later Chapters). When molecules come too close together, the force between them must become repulsive (electric repulsion between their outer electrons). We need this assumption to explain that matter takes up space. Thus molecules maintain a minimum distance from each other. In a solid material, the attractive forces are strong enough that the atoms or molecules move only slightly (oscillate) about relatively fixed positions, often in an array known as a crystal lattice, as shown in Fig. 17–2a. In a liquid, the atoms or molecules are moving more rapidly, or the forces between them are weaker, so that they are sufficiently free to pass around one another, as in Fig. 17–2b. In a gas, the forces are so weak, or the speeds so high, that the molecules do not even stay close together. They move rapidly every which way, Fig. 17–2c, filling any container and occasionally colliding with one another. On average, the speeds are sufficiently high in a gas that when two molecules collide, the force of attraction is not strong enough to keep them close together and they fly off in new directions.

EXAMPLE 17–1 ESTIMATE **Distance between atoms.** The density of copper is $8.9 \times 10^3\,\text{kg/m}^3$, and each copper atom has a mass of 63 u. Estimate the average distance between the centers of neighboring copper atoms.

APPROACH We consider a cube of copper 1 m on a side. From the given density ρ we can calculate the mass m of a cube of volume $V = 1\,\text{m}^3$ $(m = \rho V)$. We divide this mass m by the mass of one atom (63 u) to obtain the number of atoms in $1\,\text{m}^3$. We assume the atoms are in a uniform array, and we let N be the number of atoms in a 1-m length; then $(N)(N)(N) = N^3$ equals the total number of atoms in $1\,\text{m}^3$.

SOLUTION The mass of 1 copper atom is $63\,\text{u} = 63 \times 1.66 \times 10^{-27}\,\text{kg} = 1.05 \times 10^{-25}\,\text{kg}$. This means that in a cube of copper 1 m on a side $(\text{volume} = 1\,\text{m}^3)$, there are

$$\frac{8.9 \times 10^3\,\text{kg/m}^3}{1.05 \times 10^{-25}\,\text{kg/atom}} = 8.5 \times 10^{28}\,\text{atoms/m}^3.$$

The volume of a cube of side ℓ is $V = \ell^3$, so on one edge of the 1-m-long cube there are $(8.5 \times 10^{28})^{\frac{1}{3}}\,\text{atoms} = 4.4 \times 10^9\,\text{atoms}$. Hence the distance between neighboring atoms is

$$\frac{1\,\text{m}}{4.4 \times 10^9\,\text{atoms}} = 2.3 \times 10^{-10}\,\text{m}.$$

NOTE Watch out for units. Even though "atoms" is not a unit, it is helpful to include it to make sure you calculate correctly.

NOTE The distance between atoms in a solid is essentially what we mean when we speak of the size or diameter of an atom. So we have calculated the size of a copper atom.

PROBLEM SOLVING
"Size" of an atom

17–2 Temperature and Thermometers

In everyday life, **temperature** is a measure of how hot or cold something is. A hot oven is said to have a high temperature, whereas the ice of a frozen lake is said to have a low temperature.

Many properties of matter change with temperature. For example, most materials expand when their temperature is increased.[†] An iron beam is longer when hot than when cold. Concrete roads and sidewalks expand and contract slightly according to temperature, which is why compressible *spacers* or *expansion joints* (Fig. 17–3) are placed at regular intervals. The electrical resistance of matter changes with temperature (Chapter 25). So too does the color radiated by objects, at least at high temperatures: you may have noticed that the heating element of an electric stove glows with a red color when hot. At higher temperatures, solids such as iron glow orange or even white. The white light from an incandescent lightbulb comes from an extremely hot tungsten wire. The surface temperatures of the Sun and other stars can be measured by the predominant color (more precisely, wavelengths) of light they emit.

Instruments designed to measure temperature are called **thermometers**. There are many kinds of thermometers, but their operation always depends on some property of matter that changes with temperature. Many common thermometers rely on the expansion of a material with an increase in temperature. The first idea for a thermometer, by Galileo, made use of the expansion of a gas. Common thermometers today consist of a hollow glass tube filled with alcohol colored with a red dye, as were the earliest usable thermometers (Fig. 17–4).

Inside a common liquid-in-glass thermometer, the liquid expands more than the glass when the temperature is increased, so the liquid level rises in the tube (Fig. 17–5a). Although metals also expand with temperature, the change in length of a metal rod, say, is generally too small to measure accurately for ordinary changes in temperature. However, a useful thermometer can be made by bonding together two different metals with different rates of expansion (Fig. 17–5b). When the temperature is increased, the different amounts of expansion cause the bimetallic strip to bend. Often the bimetallic strip is in the form of a coil, one end of which is fixed while the other is attached to a pointer, Fig. 17–6. Such thermometers are used as ordinary air thermometers, oven thermometers, automatic off switches in electric coffeepots, and in room thermostats for determining when the heater or air conditioner should go on or off. Very precise thermometers make use of electrical properties (Chapter 25), such as resistance thermometers, thermocouples, and thermistors, often with a digital readout.

[†]Most materials expand when their temperature is raised, but not all. Water, for example, in the range 0°C to 4°C contracts with an increase in temperature (see Section 17–4).

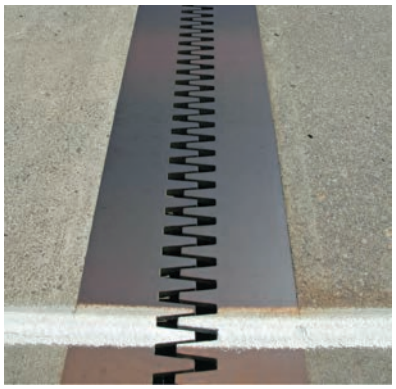


FIGURE 17–3 Expansion joint on a bridge. Note center white line of highway.

FIGURE 17–4 Thermometers built by the Accademia del Cimento (1657–1667) in Florence, Italy, are among the earliest known. These sensitive and exquisite instruments contained alcohol, sometimes colored, like many thermometers today.



FIGURE 17–5 (a) Mercury- or alcohol-in-glass thermometer; (b) bimetallic strip.

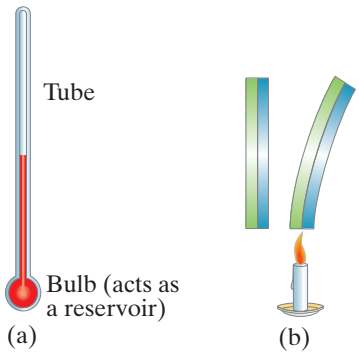
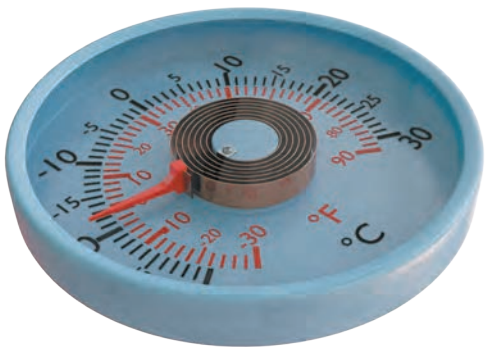
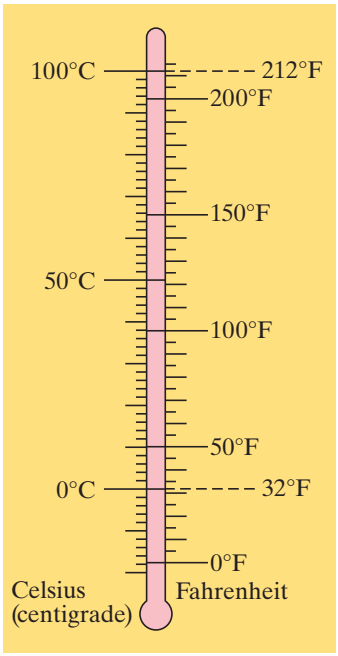
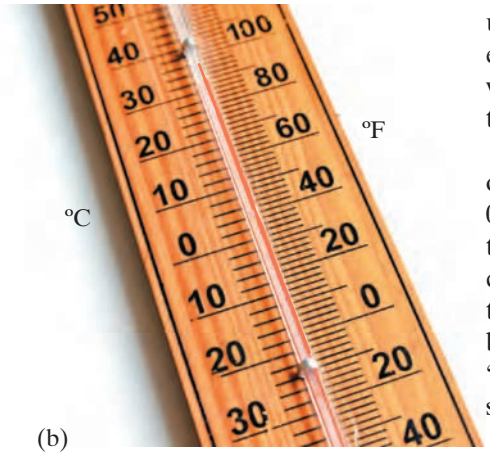


FIGURE 17–6 Photograph of a thermometer using a bimetallic coil.





(a)



(b)

FIGURE 17–7 Celsius and Fahrenheit scales compared.

PROBLEM SOLVING

Convert temperature by remembering $0°C = 32°F$ *and a change of* $5\,C° = 9\,F°$

Temperature Scales

In order to measure temperature quantitatively, some sort of numerical scale must be defined. The most common scale today is the **Celsius** or **centigrade** scale. In the United States, the **Fahrenheit** scale is also common. The most important scale in scientific work is the absolute, or Kelvin, scale, and it will be discussed later in this Chapter.

One way to define a temperature scale is to assign arbitrary values to two readily reproducible temperatures. For both the Celsius and Fahrenheit scales these two fixed points are chosen to be the freezing point and the boiling point of water, both taken at standard atmospheric pressure. On the Celsius scale, the freezing point† of water is chosen to be 0°C ("zero degrees Celsius") and the boiling point 100°C. On the Fahrenheit scale, the freezing point is defined as 32°F and the boiling point 212°F. A practical thermometer is calibrated by placing it in carefully prepared environments at each of the two temperatures and marking the position of the liquid or pointer. For a Celsius scale, the distance between the two marks is divided into one hundred equal intervals representing each degree between 0°C and 100°C (hence the name "centigrade scale" meaning "hundred steps"). For the Fahrenheit scale, the two points are labeled 32°F and 212°F and the distance between them is divided into 180 equal intervals. For temperatures below the freezing point of water and above the boiling point of water, the scales may be extended using the same equally spaced intervals. However, thermometers can be used only over a limited temperature range because of their own limitations—for example, an alcohol-in-glass thermometer is rendered useless above temperatures where the alcohol vaporizes. For very low or very high temperatures, specialized thermometers are required, some of which we will mention later.

Every temperature on the Celsius scale corresponds to a particular temperature on the Fahrenheit scale, Fig. 17–7. To convert from one to the other, remember that 0°C corresponds to 32°F and that a range of 100° on the Celsius scale corresponds to a range of 180° on the Fahrenheit scale. Thus, one Fahrenheit degree (1 F°) corresponds to $100/180 = \frac{5}{9}$ of a Celsius degree (1 C°). That is, $1\,F° = \frac{5}{9}\,C°$. (Notice that when we refer to a specific temperature, we say "degrees Celsius," as in 20°C; but when we refer to a *change* in temperature or a temperature interval, we say "Celsius degrees," as in "2 C°.") The conversion between the two temperature scales can be written

$$T(°C) = \tfrac{5}{9}[T(°F) - 32]$$

or

$$T(°F) = \tfrac{9}{5}T(°C) + 32.$$

Rather than memorizing these relations, it is simpler to remember that $0°C = 32°F$ and that a change of $5\,C° =$ a change of $9\,F°$.

EXAMPLE 17–2 **Taking your temperature.** "Normal" body temperature on average is considered to be 98.6°F. What is this on the Celsius scale?

APPROACH We recall that $0°C = 32°F$ and that a change of $5\,C° = 9\,F°$.

SOLUTION First we relate the given temperature to the freezing point of water (0°C). That is, 98.6°F is $98.6 - 32.0 = 66.6\,F°$ above the freezing point of water. Since each F° is equal to $\frac{5}{9}$ C°, this corresponds to $66.6 \times \frac{5}{9} = 37.0$ Celsius degrees above the freezing point. The freezing point of water is 0°C, so average normal body temperature is 37.0°C.

†The freezing point of a substance is defined as that temperature at which the solid and liquid phases coexist in equilibrium—that is, without any net liquid changing into the solid or vice versa. Experimentally, this is found to occur at only one definite temperature, for a given pressure. Similarly, the boiling point is defined as that temperature at which the liquid and gas coexist in equilibrium. Since these points vary with pressure, the pressure must be specified (usually it is 1 atm).

EXERCISE A Determine the temperature at which both scales give the same numerical reading $(T_C = T_F)$.

Different materials do not expand precisely linearly over a wide temperature range. Consequently, if we calibrate different kinds of thermometers exactly as just described, they may not agree precisely. We calibrate them so they agree at 0°C and at 100°C. But they may not agree precisely at intermediate temperatures; a carefully calibrated alcohol-in-glass thermometer might register 52.0°C, whereas a carefully calibrated thermometer of another type might read 52.6°C.

Because of such discrepancies, some standard kind of thermometer must be chosen so that all temperatures can be precisely defined. The chosen standard for this purpose is the **constant-volume gas thermometer**. As shown in the simplified diagram of Fig. 17–8, this thermometer consists of a hollow rigid bulb filled with a low-pressure gas connected by a thin tube to a mercury manometer (Section 13–6). The volume of the gas is kept constant by raising or lowering the right-hand tube of the manometer so that the mercury in the left-hand tube coincides with the reference mark. An increase in temperature causes a proportional increase in pressure in the bulb, so the right-hand tube must be lifted higher to keep the gas volume constant; the height difference of the mercury in the tube is then a measure of the temperature. This thermometer gives the same results for all gases in the limit of reducing the gas pressure in the bulb toward zero. The resulting scale serves as a basis for the *standard temperature scale* (Section 17–10).

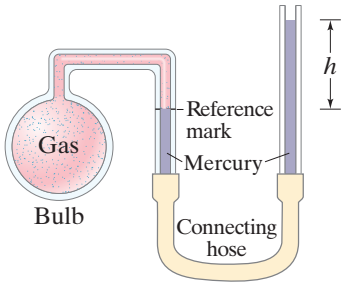


FIGURE 17–8 Constant-volume gas thermometer.

17–3 Thermal Equilibrium and the Zeroth Law of Thermodynamics

We are all familiar with the fact that if two objects at different temperatures are placed in thermal contact (meaning thermal energy can transfer from one to the other), the two objects will eventually reach the same temperature. They are then said to be in **thermal equilibrium**. For example, you leave a fever thermometer in your mouth until it comes into thermal equilibrium with that environment; only then do you read it. Two objects are defined to be in thermal equilibrium if, when placed in thermal contact, no net energy flows from one to the other, and their temperatures don't change. Experiments indicate that

> **if two systems are in thermal equilibrium with a third system, then they are in thermal equilibrium with each other.**

This postulate is called the **zeroth law of thermodynamics**. It has this unusual name because it was not until after the great first and second laws of thermodynamics (Chapters 19 and 20) were worked out that scientists realized that this postulate needed to be stated first.

Temperature is a property of a system that determines whether the system will be in thermal equilibrium with other systems. When two systems are in thermal equilibrium, their temperatures are (by definition) equal, and no net thermal energy will be exchanged between them. This is consistent with our everyday notion of temperature: when a hot object and a cold one are put into contact, they eventually come to the same intermediate temperature. Thus the importance of the zeroth law is that it allows a useful definition of temperature.

17–4 Thermal Expansion

Most substances expand when heated and contract when cooled. The amount of expansion or contraction depends on the material. From a microscopic point of view, if the temperature of an object increases, its length increases, which means its molecules are a bit farther apart on average.

Linear Expansion

Experiments indicate that the change in length $\Delta\ell$ of almost all solids is, to a good approximation, directly proportional to the change in temperature ΔT, as long as ΔT is not too large. The change in length is also proportional to the original length of the object, ℓ_0. That is, for the same temperature increase, a 4-m-long iron rod will increase in length twice as much as a 2-m-long iron rod. We can write this proportionality as an equation:

$$\Delta\ell = \alpha\ell_0\Delta T, \qquad \textbf{(17–1a)}$$

where α, the proportionality constant, is called the **coefficient of linear expansion** for the particular material and has units of $(\text{C}^\circ)^{-1}$. We write $\ell = \ell_0 + \Delta\ell$ (see Fig. 17–9), and rewrite Eq. 17–1a as

$$\ell = \ell_0(1 + \alpha\Delta T), \qquad \textbf{(17–1b)}$$

where ℓ_0 is the length initially, at temperature T_0, and ℓ is the length after heating or cooling to a temperature T. If the temperature change $\Delta T = T - T_0$ is negative, then $\Delta\ell = \ell - \ell_0$ is also negative; the length shortens as the temperature decreases.

The values of α for various materials at 20°C are listed in Table 17–1. Actually, α does vary slightly with temperature (which is why thermometers made of different materials do not agree precisely). However, if the temperature range is not too great, the variation can usually be ignored.

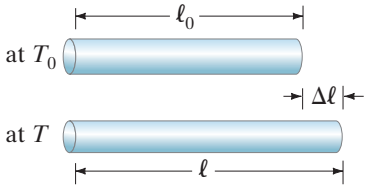


FIGURE 17–9 A thin rod of length ℓ_0 at temperature T_0 is heated to a new uniform temperature T and acquires length ℓ, where $\ell = \ell_0 + \Delta\ell$.

TABLE 17–1 Coefficients of Expansion, near 20°C

Material	Coefficient of Linear Expansion, $\alpha\,(\text{C}^\circ)^{-1}$	Coefficient of Volume Expansion, $\beta\,(\text{C}^\circ)^{-1}$
Solids		
Aluminum	25×10^{-6}	75×10^{-6}
Brass	19×10^{-6}	56×10^{-6}
Copper	17×10^{-6}	50×10^{-6}
Gold	14×10^{-6}	42×10^{-6}
Iron or steel	12×10^{-6}	35×10^{-6}
Lead	29×10^{-6}	87×10^{-6}
Glass (Pyrex®)	3×10^{-6}	9×10^{-6}
Glass (ordinary)	9×10^{-6}	27×10^{-6}
Quartz	0.6×10^{-6}	2×10^{-6}
Concrete and brick	$\approx 12\times10^{-6}$	$\approx 36\times10^{-6}$
Marble	$1.4\text{–}3.5\times10^{-6}$	$4\text{–}10\times10^{-6}$
Liquids		
Gasoline		950×10^{-6}
Mercury		180×10^{-6}
Ethyl alcohol		1100×10^{-6}
Glycerin		500×10^{-6}
Water		210×10^{-6}
Gases		
Air (and most other gases at atmospheric pressure)		3400×10^{-6}

PHYSICS APPLIED
Expansion in structures

EXAMPLE 17–3 **Bridge expansion.** The steel bed of a suspension bridge is 200 m long at 20°C. If the extremes of temperature to which it might be exposed are −30°C to +40°C, how much will it contract and expand?

APPROACH We assume the bridge bed will expand and contract linearly with temperature, as given by Eq. 17–1a.

SOLUTION From Table 17–1, we find that $\alpha = 12\times10^{-6}(\text{C}^\circ)^{-1}$ for steel. The increase in length when it is at 40°C will be

$$\Delta\ell = \alpha\ell_0\Delta T = (12\times10^{-6}/\text{C}^\circ)(200\,\text{m})(40^\circ\text{C} - 20^\circ\text{C}) = 4.8\times10^{-2}\,\text{m},$$

or 4.8 cm. When the temperature decreases to −30°C, $\Delta T = -50\,\text{C}^\circ$. Then

$$\Delta\ell = (12\times10^{-6}/\text{C}^\circ)(200\,\text{m})(-50\,\text{C}^\circ) = -12.0\times10^{-2}\,\text{m},$$

or a decrease in length of 12 cm. The total range the expansion joints must accommodate is 12 cm + 4.8 cm ≈ 17 cm (Fig. 17–3).

CONCEPTUAL EXAMPLE 17–4 **Do holes expand or contract?** If you heat a thin, circular ring (Fig. 17–10a) in the oven, does the ring's hole get larger or smaller?

RESPONSE You might guess that the metal expands into the hole, making the hole smaller, but it is not so. Imagine the ring is solid, like a coin (Fig. 17–10b). Draw a circle on it with a pen as shown. When the metal expands, the material inside the circle will expand along with the rest of the metal; so the dashed circle expands. Cutting the metal where the circle is shows that the hole in Fig. 17–10a increases in diameter. To see this in another way, consider the row or circle of atoms on the inner surface of the ring. With an increase in temperature, they get farther apart. So the circumference of the inside surface of the ring increases.

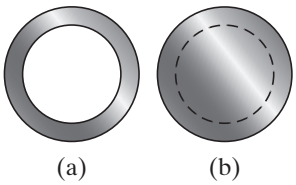


FIGURE 17–10 Example 17–4.

EXAMPLE 17–5 **Ring on a rod.** An iron ring is to fit snugly on a cylindrical iron rod (Fig. 17–11). At 20°C, the diameter of the rod is 6.445 cm and the inside diameter of the ring is 6.420 cm. To slip over the rod, the ring must be slightly larger than the rod diameter by about 0.008 cm. To what temperature must the ring be brought if its hole is to be large enough so it will slip over the rod?

APPROACH The hole in the ring must be increased from a diameter of 6.420 cm to 6.445 cm + 0.008 cm = 6.453 cm. The ring must be heated since the hole diameter will increase linearly with temperature (Example 17–4).

SOLUTION We solve for ΔT in Eq. 17–1a and find

$$\Delta T = \frac{\Delta \ell}{\alpha \ell_0} = \frac{6.453\ \text{cm} - 6.420\ \text{cm}}{(12 \times 10^{-6}/\text{C}^\circ)(6.420\ \text{cm})} = 430\ \text{C}^\circ.$$

So the ring must be raised at least to $T = (20°\text{C} + 430\ \text{C}^\circ) = 450°\text{C}$.

NOTE In doing Problems, do not forget the last step, adding in the initial temperature (20°C here).

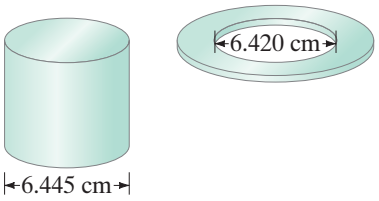


FIGURE 17–11 Example 17–5.

CONCEPTUAL EXAMPLE 17–6 **Opening a tight jar lid.** When the lid of a glass jar is tight, holding the lid under hot water for a short time will often make it easier to open (Fig. 17–12). Why?

RESPONSE The lid may be struck by the hot water more directly than the glass and so expand sooner. But even if not, metals generally expand more than glass for the same temperature change (α is greater—see Table 17–1).

PHYSICS APPLIED
Opening a tight lid

FIGURE 17–12 Example 17–6.



Volume Expansion

The change in *volume* of a material which undergoes a temperature change is given by a relation similar to Eq. 17–1a, namely,

$$\Delta V = \beta V_0 \Delta T, \qquad \textbf{(17–2)}$$

where V_0 is the original volume, ΔV is the change in volume when the temperature changes by ΔT, and β is the **coefficient of volume expansion**. The units of β are $(\text{C}^\circ)^{-1}$.

Values of β for various materials are given in Table 17–1. Notice that for solids, β is normally equal to approximately 3α. (See Problem 16.) Note also that linear expansion has no meaning for liquids and gases because they do not have fixed shapes.

EXERCISE B A long thin bar of aluminum at 0°C is 1.0 m long and has a volume of $1.0000 \times 10^{-3}\ \text{m}^3$. When heated to 100°C, the length of the bar becomes 1.0025 m. What is the approximate volume of the bar at 100°C? (*a*) $1.0000 \times 10^{-3}\ \text{m}^3$; (*b*) $1.0025 \times 10^{-3}\ \text{m}^3$; (*c*) $1.0050 \times 10^{-3}\ \text{m}^3$; (*d*) $1.0075 \times 10^{-3}\ \text{m}^3$; (*e*) $2.5625 \times 10^{-3}\ \text{m}^3$.

Equations 17–1 and 17–2 are accurate only if $\Delta\ell$ (or ΔV) is small compared to ℓ_0 (or V_0). This is of particular concern for liquids and even more so for gases because of the large values of β. Furthermore, β itself varies substantially with temperature for gases. Therefore, a more convenient way of dealing with gases is needed, and will be discussed starting in Section 17–6.

EXAMPLE 17–7 **Gas tank in the Sun.** The 70-liter (L) steel gas tank of a car is filled to the top with gasoline at 20°C. The car sits in the Sun and the tank reaches a temperature of 40°C (104°F). How much gasoline do you expect to overflow from the tank?

PHYSICS APPLIED
Gas tank overflow

APPROACH Both the gasoline and the tank expand as the temperature increases, and we assume they do so linearly as described by Eq. 17–2. The volume of overflowing gasoline equals the volume increase of the gasoline minus the increase in volume of the tank.

SOLUTION The gasoline expands by

$$\Delta V = \beta V_0 \Delta T = (950 \times 10^{-6}/\text{C}^\circ)(70\,\text{L})(40^\circ\text{C} - 20^\circ\text{C}) = 1.3\,\text{L}.$$

The tank also expands. It is a steel shell that undergoes volume expansion (Table 17–1: $\beta = 35 \times 10^{-6}/\text{C}^\circ \approx 3\alpha$). If you thought of the tank as being solid, the surface layer (the shell) would expand just the same. Thus the tank increases in volume by

$$\Delta V = (35 \times 10^{-6}/\text{C}^\circ)(70\,\text{L})(40^\circ\text{C} - 20^\circ\text{C}) = 0.049\,\text{L},$$

so the tank expansion has little effect. More than a liter of gas could spill out.

Anomalous Behavior of Water Below 4 °C

Most substances expand more or less uniformly with an increase in temperature, as long as no phase change occurs. Water, however, does not follow the usual pattern. If water at 0°C is heated, it actually *decreases* in volume until it reaches 4°C. Above 4°C water behaves normally and expands in volume as the temperature is increased, Fig. 17–13. Water thus has its greatest density at 4°C. This anomalous behavior of water is of great importance for the survival of aquatic life during cold winters. When the water in a lake (or river) is above 4°C and begins to cool by contact with cold air, the water at the surface sinks because it is denser. It is replaced by warmer water from below. This mixing continues until the temperature of the entire lake reaches 4°C. As the surface water cools further, it remains on the surface because it is less dense than the 4°C water below. Water then freezes first at the surface, and the ice remains on the surface since ice (specific gravity = 0.917) is less dense than water. The water at the bottom remains liquid unless it is so cold that the whole body of water freezes. If water were like most substances, becoming more dense as it cools, the water at the bottom of a lake would be frozen first. Lakes would freeze solid more easily because circulation would bring the warmer water to the surface to be efficiently cooled.

PHYSICS APPLIED
Life under ice

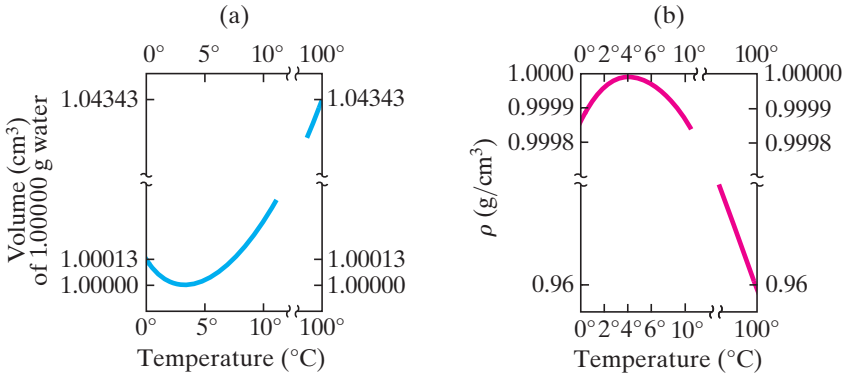


FIGURE 17–13 Behavior of water as a function of temperature near 4°C. (a) Volume of 1.00000 gram of water, as a function of temperature. (b) Density vs. temperature. [Note the break in each axis.]

The complete freezing of a lake would cause severe damage to its plant and animal life. Because of the unusual behavior of water below 4°C, it is rare for any large and deep body of water to freeze completely, and this is helped by the layer of ice on the surface which acts as an insulator to reduce the flow of heat out of the water into the cold air above. Without this peculiar but wonderful property of water, life on this planet as we know it might not have been possible.

Not only does water expand as it cools from 4°C to 0°C, it expands even more as it freezes to ice. This is why ice cubes float in water and pipes break when water inside them freezes.

*17–5 Thermal Stresses

In many situations, such as in buildings and roads, the ends of a beam or slab of material are rigidly fixed, which greatly limits expansion or contraction. If the temperature should change, large compressive or tensile stresses, called **thermal stresses**, will occur. The magnitude of such stresses can be calculated using the concept of elastic modulus developed in Chapter 12. To calculate the internal stress in a beam, we can think of this process as occurring in two steps: (1) the beam tries to expand (or contract) by an amount $\Delta\ell$ given by Eq. 17–1; (2) the rigid braces that hold the beam in place exert a force to compress the beam (or expand it), keeping it at its original length. The force F required is given by Eq. 12–4:

$$\Delta\ell = \frac{1}{E}\frac{F}{A}\ell_0,$$

where E is Young's modulus for the material. To calculate the internal stress, F/A, we then set $\Delta\ell$ in Eq. 17–1a equal to $\Delta\ell$ in the equation above and find

$$\alpha\ell_0\,\Delta T = \frac{1}{E}\frac{F}{A}\ell_0.$$

Hence, the stress is

$$\frac{F}{A} = \alpha E\,\Delta T.$$

EXAMPLE 17–8 **Stress in concrete on a hot day.** A highway is to be made of blocks of concrete 10 m long placed end to end with no space between them to allow for expansion. If the blocks were placed at a temperature of 10°C, what compressive stress would occur if the temperature reached 40°C? The contact area between each block is 0.20 m^2. Will fracture occur?

PHYSICS APPLIED
Highway buckling

APPROACH We use the expression for the stress F/A we just derived, and find the value of E from Table 12–1. To see if fracture occurs, we compare this stress to the ultimate strength of concrete in Table 12–2.

SOLUTION

$$\frac{F}{A} = \alpha E\,\Delta T = \left(12\times10^{-6}/\text{C}^\circ\right)\left(20\times10^{9}\,\text{N/m}^2\right)\left(30\,\text{C}^\circ\right) = 7.2\times10^{6}\,\text{N/m}^2.$$

This stress is not far from the ultimate strength of concrete under compression (Table 12–2) and exceeds it for tension and shear. If the concrete is not perfectly aligned, part of the force will act in shear, and fracture is likely. This is why soft spacers or expansion joints (Fig. 17–3) are used in concrete sidewalks, highways, and bridges.

EXERCISE C How much space would you allow between the 10-m-long concrete blocks in Example 17–8 if you expected a temperature range of 0°F to 110°F?

17–6 The Gas Laws and Absolute Temperature

Equation 17–2 for volume expansion is not often useful for describing the expansion of a gas, mainly because gases expand to fill the full volume of whatever container they are in. If the temperature is raised, the pressure in the gas will increase if the container is more or less rigid; if the container is very flexible like a toy balloon, the volume will increase instead. The pressure, temperature, and volume of a gas are intimately related. It is therefore valuable to determine a relation between the volume, the pressure, the temperature, and the quantity of a gas. Such a relation is called an **equation of state**.

We describe the **state** (or condition) of a system, such as a gas in a container, using quantities that are detectable directly by our senses and instruments. Such quantities are called **state variables**. For a gas in a container, they are pressure P, volume V, temperature T, and quantity of the gas which could be mass m, or number of "moles" (which are defined in the next Section).

If the state of a system is changed, we will always wait until the pressure and temperature have reached the same values throughout. We thus consider only **equilibrium states** of a system—when the variables that describe it (such as temperature and pressure) are the same throughout the system and are not changing in time. (This is an idealization: real systems cannot be perfectly insulated, so some heat exchange with the environment can occur and the edges of the container will have a slightly different temperature than the center.)

We also consider here gases that are not too dense (the pressure is not too high, on the order of an atmosphere or less), and not cold enough to be close to condensing and becoming a liquid.

Consider a fixed quantity of gas in a flexible container (its volume can change) such as the cylinder shown in Fig. 17–14a, which is fitted with a piston that can move, thereby changing the volume. It is found experimentally that, to a good approximation, *the volume of a gas is inversely proportional to the absolute pressure applied to it when the temperature is kept constant*. That is,

$$V \propto \frac{1}{P}, \qquad \text{[constant } T\text{]}$$

where P is the absolute pressure (*not* "gauge pressure"—see Section 13–4). For example, if the pressure on a fixed amount of gas in a flexible container is doubled, the volume is reduced to half the original volume. This relation is known as **Boyle's law**, after the Englishman Robert Boyle (1627–1691), who first stated it on the basis of his own experiments. A graph of P vs. V for a fixed temperature is shown in Fig. 17–14b. Boyle's law can also be written

$$PV = \text{constant} \qquad \text{[constant } T\text{]}$$

for a fixed quantity of gas kept at constant temperature. If either the pressure or volume of a fixed amount of gas is allowed to vary, the other variable also changes so that the product PV remains constant.

Temperature also affects the volume of a gas, but a quantitative relationship between V and T was not found until more than a century after Boyle's work. The Frenchman Jacques Charles (1746–1823) found that when the pressure is not too high and is kept constant, the volume of a gas increases with temperature at a nearly linear rate, as shown in Fig. 17–15a. However, all gases liquefy at low temperatures (for example, oxygen liquefies at −183°C), so the graph cannot be extended below the liquefaction point. Nonetheless, the graph is essentially a straight line and if projected to lower temperatures, as shown by the dashed line, it crosses the axis at about −273°C.

Such a graph can be drawn for any gas, and the straight line always projects back to −273°C at zero volume. This seems to imply that if a gas could be cooled to −273°C, it would have zero volume, and at lower temperatures a negative volume, which makes no sense. It could be argued that −273°C is the lowest temperature possible; indeed, many other more recent experiments indicate that this is so. This temperature is called the **absolute zero** of temperature. Its value has been determined to be −273.15°C.

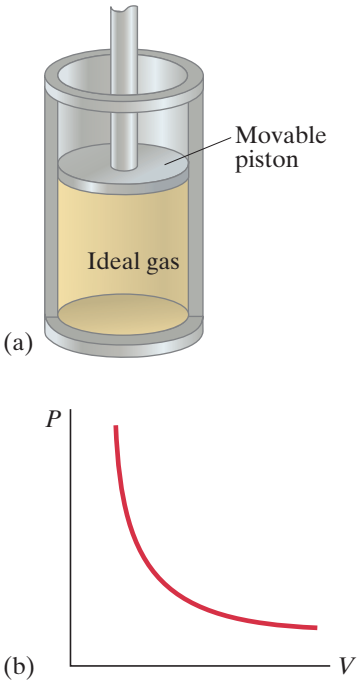
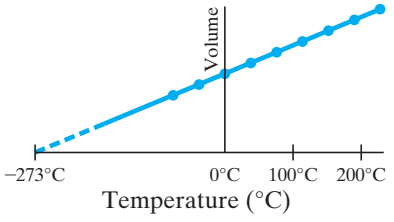
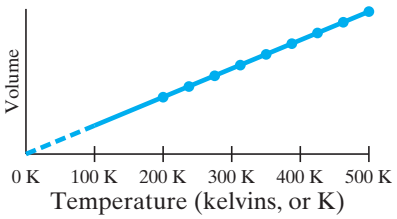


FIGURE 17–14 (a) An ideal gas in a cylinder fitted with a movable piston. (b) Pressure vs. volume of a fixed amount of gas at a constant temperature, showing the inverse relationship as given by Boyle's law: as the pressure decreases, the volume increases.

FIGURE 17–15 Volume of a fixed amount of gas as a function of (a) Celsius temperature, and (b) Kelvin temperature, when the pressure is kept constant.



(a)



(b)

Absolute zero forms the basis of a temperature scale known as the **absolute scale** or **Kelvin scale**, and it is used extensively in scientific work. On this scale the temperature is specified as degrees Kelvin or, preferably, simply as *kelvins* (K) without the degree sign. The intervals are the same as for the Celsius scale, but the zero on this scale (0 K) is chosen as absolute zero. Thus the freezing point of water (0°C) is 273.15 K, and the boiling point of water is 373.15 K. Indeed, any temperature on the Celsius scale can be changed to kelvins by adding 273.15 to it:

$$T(\text{K}) = T(^\circ\text{C}) + 273.15.$$

CAUTION
K for kelvins has no degree sign

Now let us look at Fig. 17–15b, where the graph of the volume of a gas versus absolute temperature is a straight line that passes through the origin. Thus, to a good approximation, *the volume of a fixed quantity of gas is directly proportional to the absolute temperature when the pressure is kept constant*. This is known as **Charles's law**, and is written

$$V \propto T. \qquad [\text{constant } P]$$

A third gas law, known as **Gay-Lussac's law**, after Joseph Gay-Lussac (1778–1850), states that *at constant volume, the absolute pressure of a gas is directly proportional to the absolute temperature*:

$$P \propto T. \qquad [\text{constant } V]$$

The laws of Boyle, Charles, and Gay-Lussac are not really laws in the sense that we use this term today (precise, deep, wide-ranging validity). They are really only approximations that are accurate for real gases only as long as the pressure and density of the gas are not too high, and the gas is not too close to condensing into a liquid (condensation). The term *law* applied to these three relationships has become traditional, however, so we have stuck with that usage.

PHYSICS APPLIED
Throwing a jar into a campfire

CONCEPTUAL EXAMPLE 17–9 **Why you should not put a closed glass jar into a campfire.** What could happen if you tossed an empty glass jar, with the lid on tight, into a fire, and why?

RESPONSE The inside of the jar is not empty. It is filled with air. As the fire heats the air inside, its temperature rises. The volume of the glass jar changes only slightly due to the heating. According to Gay-Lussac's law the pressure P of the air inside the jar increases, and that pressure can be large enough to cause the jar to explode, throwing glass pieces outward.

17–7 The Ideal Gas Law

The gas laws of Boyle, Charles, and Gay-Lussac were obtained by means of an important scientific technique: namely, *considering one quantity and how it is affected by changing only one other variable, keeping all other variables constant*. These laws can now be combined into a single more general relation among all three variables—absolute pressure, volume, and absolute temperature of a fixed amount of gas:

$$PV \propto T.$$

This relation indicates how any of the quantities P, V, or T will vary when the other two quantities change. This relation reduces to Boyle's, Charles's, or Gay-Lussac's law when either T, P, or V, respectively, is held constant.

Finally, we must incorporate the effect of the amount of gas present. For example, when more air is forced into a balloon, the balloon gets bigger (Fig. 17–16). Indeed, careful experiments show that at constant temperature and pressure, the volume V of an enclosed gas increases in direct proportion to the mass m of gas present. Hence we write

$$PV \propto mT.$$

This proportion can be made into an equation by inserting a constant of proportionality. Experiment shows that this constant has a different value for different gases. However, the constant of proportionality turns out to be the same for all gases if, instead of the mass m, we use the number of *moles*, which we now define.



FIGURE 17–16 Blowing up a balloon means putting more air (more air molecules) into the balloon, which increases its volume. The pressure is nearly constant, at atmospheric pressure, except for the small effect of the balloon's elasticity.

The "mole" is an official SI unit for the amount of any substance. One **mole** (abbreviated mol) has the precise definition of being the number of atoms in exactly 12 grams of carbon-12, whose atomic mass is exactly 12 u (Section 17–1).

An equivalent definition is that 1 mol is that amount of a substance whose mass in grams is numerically equal to the molecular mass of the substance (Section 17–1). For example, the mass of 1 mol of CO_2 is $[12 + (2 \times 16)] = 44$ g because carbon has atomic mass 12 and oxygen 16 (see Periodic Table inside the rear cover).

A modern definition of the mole is: an amount of a substance that contains 6.02×10^{23} objects (usually atoms, molecules, or ions, etc.). This number is called *Avogadro's number*, as we discuss in Section 17–9.

In general, the number of moles, n, in a given sample of a pure substance is equal to the mass of the sample in grams divided by the molecular mass specified as grams per mole:

$$n\,(\text{mole}) = \frac{\text{mass (grams)}}{\text{molecular mass (g/mol)}}.$$

For example, the number of moles in 132 g of CO_2 (molecular mass 44 u) is

$$n = \frac{132\text{ g}}{44\text{ g/mol}} = 3.0\text{ mol}.$$

We can now write the proportion above $(PV \propto mT)$ as an equation:

IDEAL GAS LAW

$$PV = nRT, \qquad \textbf{(17–3)}$$

where n represents the number of moles and R is the constant of proportionality. R is called the **universal gas constant** because its value is found experimentally to be the same for all gases. The value of R, in several sets of units (only the first is the proper SI unit), is

$$\begin{aligned} R &= 8.314\text{ J/(mol}\cdot\text{K)} && \text{[SI units]} \\ &= 0.0821\text{ (L}\cdot\text{atm)/(mol}\cdot\text{K)}^{\dagger} \\ &= 1.99\text{ calories/(mol}\cdot\text{K)}.^{\dagger} \end{aligned}$$

Equation 17–3 is called the **ideal gas law**, or the **equation of state for an ideal gas**. We use the term "ideal" because real gases do not follow Eq. 17–3 precisely, particularly at high pressure (and density) or when the gas is near the condensation point (= boiling point). However, at pressures less than an atmosphere or so, and when T is not close to the condensation point of the gas, Eq. 17–3 is quite accurate and useful for real gases.

CAUTION

Always give T in kelvins and P as absolute (not gauge) pressure

Always remember, when using the ideal gas law, that temperatures must be given in kelvins (K) and that the pressure P must always be *absolute* pressure, not gauge pressure (Section 13–4).

EXERCISE D Return to the Chapter-Opening Question, page 492, and answer it again now. Try to explain why you may have answered differently the first time.

EXERCISE E An ideal gas is contained in a steel sphere at 27.0°C and 1.00 atm absolute pressure. If no gas is allowed to escape and the temperature is raised to 127°C, what will be the new pressure? (*a*) 0.21 atm; (*b*) 0.75 atm; (*c*) 1.00 atm; (*d*) 1.33 atm; (*e*) 4.7 atm.

17–8 Problem Solving with the Ideal Gas Law

The ideal gas law is an extremely useful tool, and we now consider some Examples. We will often refer to "standard conditions" or **standard temperature and pressure** (abbreviated **STP**), which means:

STP (standard temperature and pressure)

$$T = 273.15\text{ K}\ \ (0°\text{C}) \quad \text{and} \quad P = 1.00\text{ atm} = 1.013 \times 10^5\text{ N/m}^2 = 101.3\text{ kPa}.$$

[†]Sometimes it is useful to use R as given in these non-SI units; calories will be defined in Section 19–1.

EXAMPLE 17–10 **Volume of one mole at STP.** Determine the volume of 1.00 mol of any gas, assuming it behaves like an ideal gas, at STP.

APPROACH We use the ideal gas law, solving for V with $n = 1.00\,\text{mol}$.

SOLUTION We solve for V in Eq. 17–3:

$$V = \frac{nRT}{P} = \frac{(1.00\,\text{mol})(8.314\,\text{J/mol}\cdot\text{K})(273.15\,\text{K})}{(1.013\times10^5\,\text{N/m}^2)} = 22.4\times10^{-3}\,\text{m}^3.$$

Since 1 liter (L) is $1000\,\text{cm}^3 = 1.00\times10^{-3}\,\text{m}^3$, 1.00 mol of any (ideal) gas has volume $V = 22.4\,\text{L}$ at STP.

PROBLEM SOLVING
1.00 mol of gas at STP has $V = 22.4$ L

The value of 22.4 L for the volume of 1 mol of an ideal gas at STP is worth remembering, for it sometimes makes calculation simpler.

EXERCISE F What is the volume of 1.00 mol of ideal gas at $546\,\text{K}\ (=2\times273\,\text{K})$ and 1.0 atm absolute pressure? (*a*) 11.2 L; (*b*) 22.4 L; (*c*) 44.8 L; (*d*) 67.2 L; (*e*) 89.6 L.

EXAMPLE 17–11 **Helium balloon.** A helium party balloon, assumed to be a perfect sphere, has a radius of 18.0 cm. At room temperature (20°C), its internal pressure is 1.05 atm. Find the number of moles of helium in the balloon and the mass of helium needed to inflate the balloon to these values.

APPROACH We can use the ideal gas law to find n, since we are given P and T, and can find V from the given radius.

SOLUTION We get the volume V from the formula for a sphere:

$$V = \tfrac{4}{3}\pi r^3 = \tfrac{4}{3}\pi(0.180\,\text{m})^3 = 0.0244\,\text{m}^3.$$

The pressure is given as $1.05\,\text{atm} = 1.064\times10^5\,\text{N/m}^2$. The temperature must be expressed in kelvins, so we change 20°C to $(20+273)\,\text{K} = 293\,\text{K}$. Finally, we use the value $R = 8.314\,\text{J/(mol}\cdot\text{K)}$ because we are using SI units. Thus

$$n = \frac{PV}{RT} = \frac{(1.064\times10^5\,\text{N/m}^2)(0.0244\,\text{m}^3)}{(8.314\,\text{J/mol}\cdot\text{K})(293\,\text{K})} = 1.066\,\text{mol}.$$

The mass of helium (atomic mass = 4.00 g/mol as given in the Periodic Table or Appendix G) can be obtained from

$$\text{mass} = n\times\text{molecular mass} = (1.066\,\text{mol})(4.00\,\text{g/mol}) = 4.26\,\text{g}$$

or $4.26\times10^{-3}\,\text{kg}$.

EXAMPLE 17–12 **ESTIMATE** **Mass of air in a room.** Estimate the mass of air in a room whose dimensions are $5\,\text{m}\times3\,\text{m}\times2.5\,\text{m}$ high, at STP.

APPROACH First we determine the number of moles n using the given volume. Then we can multiply by the mass of one mole to get the total mass.

SOLUTION Example 17–10 told us that 1 mol of a gas at 0°C has a volume of $22.4\,\text{L} = 22.4\times10^{-3}\,\text{m}^3$. The room's volume is $5\,\text{m}\times3\,\text{m}\times2.5\,\text{m}$, so

$$n = \frac{(5\,\text{m})(3\,\text{m})(2.5\,\text{m})}{22.4\times10^{-3}\,\text{m}^3/\text{mol}} \approx 1700\,\text{mol}.$$

Air is a mixture of about 20% oxygen (O_2) and 80% nitrogen (N_2). The molecular masses are $2\times16\,\text{u} = 32\,\text{u}$ and $2\times14\,\text{u} = 28\,\text{u}$, respectively, for an average of about 29 u. Thus, 1 mol of air has a mass of about $29\,\text{g} = 0.029\,\text{kg}$, so our room has a mass of air

$$m \approx (1700\,\text{mol})(0.029\,\text{kg/mol}) \approx 50\,\text{kg}.$$

NOTE That is roughly 100 lb of air!

PHYSICS APPLIED
The mass of the air in a room is significant

EXERCISE G At 20°C, would there be (*a*) more, (*b*) less, or (*c*) the same mass of air in a room than at 0°C?

Frequently, volume is specified in liters and pressure in atmospheres. Rather than convert these to SI units, we can instead use the value of R given in Section 17–7 as $0.0821\ \text{L}\cdot\text{atm/mol}\cdot\text{K}$.

PROBLEM SOLVING
Using the ideal gas law as a ratio

In many situations it is not necessary to use the value of R at all. For example, many problems involve a change in the pressure, temperature, and volume of a fixed amount of gas. In this case, $PV/T = nR = \text{constant}$, since n and R remain constant. If we now let P_1, V_1, and T_1 represent the appropriate variables initially, and P_2, V_2, T_2 represent the variables after the change is made, then we can write

$$\frac{P_1 V_1}{T_1} = \frac{P_2 V_2}{T_2}. \qquad \text{[fixed } n\text{]}$$

If we know any five of the quantities in this equation, we can solve for the sixth. Or, if one of the three variables is constant ($V_1 = V_2$, or $P_1 = P_2$, or $T_1 = T_2$) then we can use this equation to solve for one unknown when given the other three quantities.

PHYSICS APPLIED
Pressure in a hot tire



FIGURE 17–17 Example 17–13.

EXAMPLE 17–13 Check tires cold. An automobile tire is filled to a gauge pressure of 210 kPa (= 30 psi) at 10°C (Fig. 17–17). After a drive of 100 km, the temperature within the tire rises to 40°C due to friction with the road. What is the pressure within the tire now?

APPROACH We do not know the number of moles of gas, or the volume of the tire, but we assume they are constant. We use the ratio form of the ideal gas law.

SOLUTION Since $V_1 = V_2$, then

$$\frac{P_1}{T_1} = \frac{P_2}{T_2}.$$

This is, incidentally, a statement of Gay-Lussac's law. Since the pressure given is the gauge pressure (Section 13–4), we must add atmospheric pressure (= 101 kPa) to get the absolute pressure $P_1 = (210\ \text{kPa} + 101\ \text{kPa}) = 311\ \text{kPa}$. We convert temperatures to kelvins by adding 273 and solve for P_2:

$$P_2 = P_1\left(\frac{T_2}{T_1}\right) = (3.11\times 10^5\ \text{Pa})\left(\frac{313\ \text{K}}{283\ \text{K}}\right) = 344\ \text{kPa}.$$

Subtracting atmospheric pressure, we find the resulting gauge pressure to be 243 kPa, which is a 16% increase (= 35 psi).

NOTE This Example shows why car manuals emphasize checking tire pressure when the tires are cold.

17–9 Ideal Gas Law in Terms of Molecules: Avogadro's Number

The fact that the gas constant R in the ideal gas law has the same value for all gases demonstrates a remarkable simplicity in nature. It was first recognized, although in a slightly different form, by the Italian scientist Amedeo Avogadro (1776–1856). Avogadro proposed that *equal volumes of gas at the same pressure and temperature contain equal numbers of molecules*. This relationship is called **Avogadro's hypothesis** or **Avogadro's Principle**.

This is consistent with R being the same for all gases because, from the ideal gas law (Eq. 17–3) $PV/T = nR$, then n is the same for all gases if and only if R is the same for all gases. Thus Avogadro's hypothesis is equivalent to R being the same for all gases.

The number of molecules in one mole of any pure substance is known as **Avogadro's number**, N_A. Although Avogadro conceived the notion, he was not

able to actually determine the value of N_A. Indeed, precise measurements were not done until the twentieth century.

A number of methods have been devised to measure N_A, and the accepted value today is

$$N_A = 6.02 \times 10^{23}. \qquad \text{[molecules/mole]}$$

Avogadro's number

Since the total number of molecules, N, in a gas is equal to N_A times the number of moles $(N = nN_A)$, the ideal gas law, Eq. 17–3, can be written in terms of the number of molecules present:

$$PV = nRT = \frac{N}{N_A} RT,$$

or

$$PV = NkT, \qquad \textbf{(17–4)}$$

IDEAL GAS LAW (in terms of molecules)

where $k = R/N_A$ is called the **Boltzmann constant** and has the value

$$k = \frac{R}{N_A} = \frac{8.314\,\text{J/mol}\cdot\text{K}}{6.02 \times 10^{23}/\text{mol}} = 1.38 \times 10^{-23}\,\text{J/K}.$$

EXAMPLE 17–14 **Hydrogen atom mass.** Use Avogadro's number to determine the mass of a hydrogen atom.

APPROACH The mass of one atom equals the mass of 1 mol divided by the number of atoms in 1 mol, N_A.

SOLUTION One mole of hydrogen atoms (atomic mass = 1.008 u, Section 17–1 or Appendix G) has a mass of 1.008×10^{-3} kg and contains 6.02×10^{23} atoms. Thus one atom has a mass

$$m = \frac{1.008 \times 10^{-3}\,\text{kg}}{6.02 \times 10^{23}} = 1.67 \times 10^{-27}\,\text{kg}.$$

NOTE Historically, the reverse was done: a precise value of N_A was obtained from a precise measurement of the mass of the hydrogen atom.

EXAMPLE 17–15 **ESTIMATE** **How many molecules in one breath?** Estimate how many molecules you breathe in with a 1.0-L breath of air.

PHYSICS APPLIED *Molecules in a breath*

APPROACH We determine what fraction of a mole 1.0 L is by using the result of Example 17–10 that 1 mole has a volume of 22.4 L at STP, and then multiply that by N_A to get the number of molecules in this number of moles.

SOLUTION One mole corresponds to 22.4 L at STP, so 1.0 L of air is $(1.0\,\text{L})/(22.4\,\text{L/mol}) = 0.045$ mol. Then 1.0 L of air contains

$$(0.045\,\text{mol})\,(6.02 \times 10^{23}\,\text{molecules/mol}) \approx 3 \times 10^{22}\,\text{molecules}.$$

*17–10 Ideal Gas Temperature Scale—a Standard

It is important to have a very precisely defined temperature scale so that measurements of temperature made at different laboratories around the world can be accurately compared. Called the **ideal gas temperature scale**, it is based on the property of an ideal gas that the pressure is directly proportional to the absolute temperature (Gay-Lussac's law). A real gas approaches this ideal at very low density. To set up a scale we need two fixed points. One fixed point will be $P = 0$ at $T = 0$ K.

The second fixed point is chosen to be the **triple point** of water, which is that point where water in the solid, liquid, and gas states can coexist in equilibrium. This occurs only at a unique temperature and pressure, and can be reproduced at different laboratories with great precision.[†] The pressure on the water at its triple point can only be 4.58 torr, and its temperature can only be 0.01°C. In fact, the triple point is now *defined* to have a temperature of exactly 273.16 K (= 273.15 K + 0.01 K).

The absolute or Kelvin temperature T of a substance is determined by putting that substance in good thermal contact with a constant-volume gas thermometer (page 497, Fig. 17–8). The temperature T is then defined as

$$T = (273.16\,\text{K}) \lim_{P_{tp} \to 0} \left(\frac{P}{P_{tp}} \right). \qquad \text{[constant volume]} \quad \textbf{(17–5)}$$

In this relation, P_{tp} is the pressure of the gas in the rigid bulb of the constant-volume gas thermometer when it is placed in water at its triple point, and P is the pressure in that same thermometer when it is instead in contact with the substance where T is being determined. (Do not confuse the pressure of 4.58 torr in water at its triple point, with P_{tp} which is the pressure in the bulb of the thermometer (Fig. 17–8) when that bulb is immersed in water at its triple point. Both P and P_{tp} refer to the pressure inside the thermometer bulb.) Note that if we let $P = P_{tp}$ in Eq. 17–5, then $T = 273.16$ K, as it must.

The pressure inside the bulb (P and P_{tp} at the two different temperatures) can be reduced by allowing some gas to escape, and that is just what we must do to take the limit in Eq. 17–5. The reason we use the limit process is because we get different results for the temperature depending on the type and amount of gas in the bulb of the thermometer (see Fig. 17–18). But whatever real gas we use, when making a series of measurements in which the amount of gas in the thermometer bulb is reduced to smaller and smaller amounts (so that both P and P_{tp} become smaller and smaller), it is found experimentally that an extrapolation of such data to $P_{tp} = 0$ always gives the *same value* for the temperature (Fig. 17–18). That is, all gases approach an ideal gas in this limit, as suggested at the end of Section 17–2. One of the great advantages of this **ideal gas temperature scale** is that the value for T does not depend on the kind of gas used. Helium has the lowest condensation point of all gases and at very low pressures it liquefies at about 1 K; so temperatures below this cannot be defined on this scale.

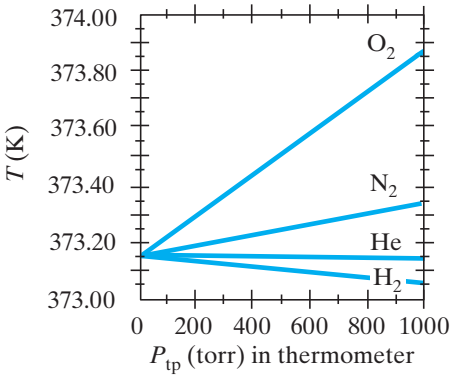


FIGURE 17–18 Temperature readings of a constant-volume gas thermometer immersed in boiling water at 1.00 atm are graphed, for different gases, as a function of the gas pressure in the thermometer at the triple point (P_{tp}). Note that as the amount of gas in the thermometer is reduced, so that $P_{tp} \to 0$, all gases give the same reading, 373.15 K. For pressure less than 0.10 atm (76 torr), the variation shown is less than 0.07 K.

[†]Liquid water and steam can coexist (the boiling point) at a range of temperatures depending on the pressure. Water boils at a lower temperature when the pressure is less, such as high in the mountains. The triple point represents a more precisely reproducible fixed point than does either the freezing point or boiling point of water at, say, 1 atm. See Section 18–3 for further discussion.

Summary

The atomic theory of matter postulates that all matter is made up of tiny entities called **atoms**, which are typically 10^{-10} m in diameter.

Atomic and **molecular masses** are specified on a scale where the most common form of carbon (^{12}C) is arbitrarily given the value 12.0000 u (atomic mass units), exactly.

The distinction between solids, liquids, and gases can be attributed to the strength of the attractive forces between the atoms or molecules and to their average speed.

Temperature is a measure of how hot or cold something is. **Thermometers** are used to measure temperature on the **Celsius** (°C), **Fahrenheit** (°F), and **Kelvin** (K) scales. Two standard points on each scale are the freezing point of water (0°C, 32°F, 273.15 K) and the boiling point of water (100°C, 212°F, 373.15 K). A one-kelvin change in temperature equals a change of one Celsius degree or $\frac{9}{5}$ Fahrenheit degrees. Kelvins are related to °C by

$$T(\text{K}) = T(^\circ\text{C}) + 273.15.$$

When two objects at different temperatures are placed in contact, they eventually reach the same temperature and are then said to be in **thermal equilibrium**.

The change in length, $\Delta\ell$, of a solid, when its temperature changes by an amount ΔT, is directly proportional to the temperature change and to its original length ℓ_0. That is,

$$\Delta\ell = \alpha\ell_0\Delta T, \qquad \textbf{(17–1a)}$$

where α is the *coefficient of linear expansion*.

The change in volume of most solids, liquids, and gases is proportional to the temperature change and to the original volume V_0:

$$\Delta V = \beta V_0\Delta T. \qquad \textbf{(17–2)}$$

The *coefficient of volume expansion*, β, is approximately equal to 3α for uniform solids.

Water is unusual because, unlike most materials whose volume increases with temperature, water's volume actually *decreases* as the temperature increases from 0°C to 4°C.

The **ideal gas law**, or **equation of state for an ideal gas**, relates the absolute pressure P, volume V, and temperature T (in kelvins) of n moles of gas by the equation

$$PV = nRT, \qquad \textbf{(17–3)}$$

where $R = 8.314\ \text{J/mol}\cdot\text{K}$ for all gases. Real gases obey the ideal gas law quite accurately if they are not at too high a pressure or near their liquefaction point.

One **mole** is that amount of a substance whose mass in grams is numerically equal to the atomic or molecular mass of that substance.

Avogadro's number, $N_A = 6.02\times10^{23}$, is the number of atoms or molecules in 1 mol of any pure substance.

The ideal gas law can be written in terms of the number of molecules N in the gas as

$$PV = NkT, \qquad \textbf{(17–4)}$$

where $k = R/N_A = 1.38\times10^{-23}\ \text{J/K}$ is **Boltzmann's constant**.

Questions

1. Which has more atoms: 1 kg of lead or 1 kg of copper? (See the Periodic Table or Appendix G.) Explain why.
2. Name several properties of materials that could be used to make a thermometer.
3. Which is larger, 1 C° or 1 F°? Explain why.
4. If system A is in thermal equilibrium with system B, but B is not in thermal equilibrium with system C, what can you say about the temperatures of A, B, and C?
5. Suppose system C is not in thermal equilibrium with system A nor in thermal equilibrium with system B. Does this imply that A and B are not in thermal equilibrium? What can you infer regarding the temperatures of A, B, and C?
6. In the relation $\Delta\ell = \alpha\ell_0\Delta T$, should ℓ_0 be the initial length, the final length, or does it matter?
7. A flat bimetallic strip consists of a strip of aluminum riveted to a strip of iron. When heated, the strip will bend. Which metal will be on the outside of the curve? Why?
8. Long steam pipes that are fixed at the ends often have a section in the shape of a U. Why?
9. Figure 17–19 shows a diagram of a simple bimetallic **thermostat** used to control a furnace (or other heating or cooling system). The electric switch (attached to the bimetallic strip) is a glass vessel containing liquid mercury that conducts electricity when it touches both contact wires. Explain how this device controls the furnace and how it can be set at different temperatures.

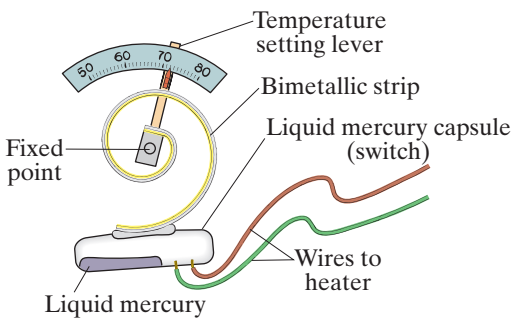


FIGURE 17–19 A thermostat (Question 9).

10. A lead cylinder floats in mercury at 0°C. Will the lead float higher or lower if the temperature is raised? Explain.
11. A glass container may break if one part of it is heated or cooled more rapidly than adjacent parts. Explain.
12. Explain why it is advisable to add water to an overheated automobile engine only slowly, and only with the engine running.
13. The units for the coefficients of linear expansion α are $(\text{C}^\circ)^{-1}$, and there is no mention of a length unit such as meters. Would the expansion coefficient change if we used feet or millimeters instead of meters? Explain.
14. When a cold alcohol-in-glass thermometer is first placed in a hot tub of water, the alcohol initially descends a bit and then rises. Explain.
15. The principal virtue of **Pyrex glass** is that its coefficient of linear expansion is much smaller than that for ordinary glass (Table 17–1). Explain why this gives rise to the higher heat resistance of Pyrex.
16. Will a clock using a pendulum supported on a long thin brass rod that is accurate at 20°C run fast or slow on a hot day (30°C)? Explain.
17. Freezing a can of soda will cause its bottom and top to bulge so badly the can will not stand up. What has happened?
18. Why might you expect an alcohol-in-glass thermometer to be more precise than a mercury-in-glass thermometer?
19. Will the buoyant force on an aluminum sphere submerged in water increase, decrease, or remain the same, if the temperature is increased from 20°C to 40°C? Explain.
20. If an atom is measured to have a mass of 6.7×10^{-27} kg, what atom do you think it is?

*21. From a practical point of view, does it really matter what gas is used in a constant-volume gas thermometer? If so, explain. [*Hint*: See Fig. 17–18.]

MisConceptual Questions

1. Rod A has twice the diameter of rod B, but both are made of iron and have the same initial length. Both rods are now subjected to the same change in temperature (but remain solid). How would the change in the rods' lengths compare?
(*a*) Rod A changes in length more than rod B.
(*b*) Rod B changes in length more than rod A.
(*c*) The same.
(*d*) Need to know whether the rods were cooled or heated.

2. Rod A is twice as long as rod B, but both are made of iron and have the same diameter. Both rods are now subjected to the same increase in temperature (but remain a solid). How would the change in the rods' lengths compare?
(*a*) Rod A increases in length more than rod B.
(*b*) Rod B increases in length more than rod A.
(*c*) The same.
(*d*) Not enough information.

3. The linear expansion of a rod depends on which of the following?
(*a*) The length of the rod.
(*b*) The change in temperature of the rod.
(*c*) The type of material.
(*d*) All of the above.
(*e*) Both (*b*) and (*c*).

4. A steel plate has a hole in it with a diameter of exactly 1.0 cm when the plate is at a temperature of 20°C. A steel ring has an inner diameter of exactly 1.0 cm at 20°C. Both the plate and the ring are heated to 100°C. Which statement is true?
(*a*) The hole in the plate gets smaller, and the opening in the ring gets larger.
(*b*) The opening in the ring gets larger, but we need the relative size of the plate and the hole to know what happens to the hole.
(*c*) The hole in the plate and the opening in the ring get larger.
(*d*) The hole in the plate and the opening in the ring get smaller.
(*e*) The hole in the plate gets larger, and the opening in the ring gets smaller.

5. A steel *tape measure* is marked such that it gives accurate length measurements at room temperature. If the tape measure is used outside on a very hot day, how will its length measurements be affected?
(*a*) Measured lengths will still be accurate.
(*b*) Measured lengths will be smaller than the real lengths.
(*c*) Measured lengths will be larger than the real lengths.

6. One mole of an ideal gas in a sealed rigid container is initially at a temperature of 100°C. The temperature is then increased to 200°C. The pressure in the gas
(*a*) remains constant.
(*b*) increases by about 25%.
(*c*) doubles.
(*d*) triples.
(*e*) is halved.

7. When an ideal gas is warmed from 20°C to 40°C, the gas's temperature T that appears in the ideal gas law increases by a factor
(*a*) of 2.
(*b*) of 1.07.
(*c*) that depends on the temperature scale you use.

8. When using the ideal gas law, which of the following rules must be obeyed?
(*a*) Always use temperature in kelvins and absolute pressure.
(*b*) Always use volume in m^3 and temperature in kelvins.
(*c*) Always use gauge pressure and temperature in degrees Celsius.
(*d*) Always use gauge pressure and temperature in kelvins.
(*e*) Always use volume in m^3 and gauge pressure.

9. The temperature of an ideal gas increases. Which of the following is true?
(*a*) The pressure must decrease.
(*b*) The pressure must increase.
(*c*) The pressure must increase while the volume decreases.
(*d*) The volume must increase while the pressure decreases.
(*e*) The pressure, the volume, or both, may increase.

10. When you blow up a balloon, which of the following increases significantly?
(*a*) The temperature inside the balloon.
(*b*) The pressure inside the balloon.
(*c*) The number of air molecules inside the balloon.
(*d*) The volume of the balloon.

11. You carry an airtight bag of potato chips onto an airplane. After the plane has ascended to cruising altitude, you notice that the unopened bag has puffed up. The most plausible explanation is that the volume of a gas (such as air) increases when
(*a*) temperature increases, if pressure is approximately constant.
(*b*) pressure decreases, if temperature is approximately constant.
(*c*) pressure increases, if temperature is approximately constant.
(*d*) temperature decreases, if pressure is approximately constant.

12. Two identical car tires are at the same temperature but contain different amounts of air. A pressure gauge indicates 1.0 atm for tire A and 2.0 atm for tire B. If the volume of the two tires is the same, how are the amounts of air in the two tires related?
(*a*) A contains twice as much air as B.
(*b*) B contains twice as much air as A.
(*c*) B contains more air than A, but not twice as much.
(*d*) B contains more than twice as much air as A.
(*e*) A contains more than B, but not twice as much.

Problems

17–1 Atomic Theory

1. (II) How many atoms are there in a 3.4-g copper coin?
2. (II) How does the number of atoms in a 31.5-g gold ring compare to the number in a silver ring of the same mass?

17–2 Temperature and Thermometers

3. (I) (*a*) "Room temperature" is often taken to be 68°F. What is this on the Celsius scale? (*b*) The temperature of the filament in a lightbulb is about 1900°C. What is this on the Fahrenheit scale?
4. (I) Among the highest and lowest natural air temperatures ever recorded are 136°F in the Libyan desert and −129°F in Antarctica. What are these temperatures on the Celsius scale?
5. (I) A thermometer tells you that you have a fever of 38.5°C. What is this in Fahrenheit?
6. (I) (*a*) 22° below zero on the Celsius scale is what Fahrenheit temperature? (*b*) 22° below zero on the Fahrenheit scale is what Celsius temperature?
7. (II) In an alcohol-in-glass thermometer, the alcohol column has length 12.61 cm at 0.0°C and length 22.79 cm at 100.0°C. What is the temperature if the column has length (*a*) 18.50 cm, and (*b*) 14.40 cm?

17–4 Thermal Expansion

8. (I) A concrete highway is built of slabs 14 m long at 15°C. How wide should the expansion cracks between the slabs be (at 15°C) to prevent buckling if the range of temperature is −30°C to +50°C?
9. (I) The Eiffel Tower (Fig. 17–20) is built of wrought iron approximately 300 m tall. Estimate how much its height changes between January (average temperature of 2°C) and July (average temperature of 25°C). Ignore the angles of the iron beams and treat the tower as a vertical beam.

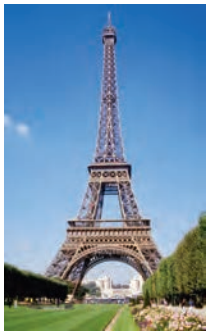


FIGURE 17–20 Problem 9. The Eiffel Tower in Paris.

10. (I) Super Invar™, an alloy of iron and nickel, is a strong material with a very low coefficient of thermal expansion $(0.20 \times 10^{-6}/\text{C}^\circ)$. A 1.8-m-long tabletop made of this alloy is used for sensitive laser measurements where extremely high tolerances are required. How much will this alloy table expand along its length if the temperature increases 7.5 C°? Compare to tabletops made of steel.
11. (II) The density of water at 4°C is $1.00 \times 10^3\ \text{kg/m}^3$. What is water's density at 94°C? Assume a constant coefficient of volume expansion.
12. (II) To what temperature would you have to heat a brass rod for it to be 1.5% longer than it is at 22°C?
13. (II) To make a secure fit, *rivets* that are larger than the rivet hole are often used and the rivet is cooled (usually in dry ice) before it is placed in the hole. A steel rivet 1.872 cm in diameter is to be placed in a hole 1.870 cm in diameter in a metal at 24°C. To what temperature must the rivet be cooled if it is to fit in the hole?
14. (II) At a given latitude, ocean water in the so-called *mixed layer* (from the surface to a depth of about 50 m) is at approximately the same temperature due to the mixing action of waves. Assume that because of *global warming*, the temperature of the mixed layer is everywhere increased by 0.5°C, while the temperature of the deeper portions of the ocean remains unchanged. Estimate the resulting rise in sea level. The ocean covers about 70% of the Earth's surface.
15. (II) An aluminum sphere is 8.75 cm in diameter. What will be its % change in volume if it is heated from 30°C to 140°C?
16. (II) A uniform rectangular plate of length ℓ and width w has a coefficient of linear expansion α. (*a*) Show that, if we neglect very small quantities, the change in area of the plate due to a temperature change ΔT is $\Delta A = 2\alpha\ell w\,\Delta T$. See Fig. 17–21. (*b*) The plate has thickness t. Show that the coefficient of volume expansion is $\beta \approx 3\alpha$.

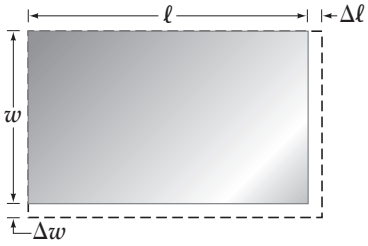


FIGURE 17–21 Problem 16(a). A rectangular plate is heated.

17. (II) A glass is filled to the brim with 450.0 mL of water, all at 100.0°C. If the temperature of glass and water is decreased to 20.0°C, how much water could be added to the glass?
18. (II) A certain car has 14.0 L of liquid coolant circulating at a temperature of 93°C through the engine's cooling system. Assume that, in this normal condition, the coolant completely fills the 3.5-L volume of the aluminum radiator and the 10.5-L internal cavities within the aluminum engine. When a car overheats, the radiator, engine, and coolant expand and a small reservoir connected to the radiator catches any resultant coolant overflow. Estimate how much coolant overflows to the reservoir if the system goes from 93°C to 105°C. Model the radiator and engine as hollow shells of aluminum. The coefficient of volume expansion for coolant is $\beta = 410 \times 10^{-6}/\text{C}^\circ$.
19. (II) It is observed that 55.50 mL of water at 20°C completely fills a container to the brim. When the container and the water are heated to 60°C, 0.35 g of water is lost. (*a*) What is the coefficient of volume expansion of the container? (*b*) What is the most likely material of the container? Density of water at 60°C is 0.98324 g/mL.
20. (II) A brass plug is to be placed in a ring made of iron. At 15°C, the diameter of the plug is 8.756 cm and that of the inside of the ring is 8.742 cm. They must both be brought to what common temperature in order to fit?
21. (II) If a fluid is contained in a long narrow vessel so it can expand in essentially one direction only, show that the effective coefficient of linear expansion α is approximately equal to the coefficient of volume expansion β.
22. (II) (*a*) Show that the change in the density ρ of a substance, when the temperature changes by ΔT, is given by $\Delta\rho = -\beta\rho\,\Delta T$. (*b*) What is the fractional change in density of a lead sphere whose temperature decreases from 35°C to −55°C?

23. (II) Wine bottles are never completely filled: a small volume of air is left in the glass bottle's cylindrically shaped neck (inner diameter $d = 18.5\,\text{mm}$) to allow for wine's fairly large coefficient of thermal expansion. The distance H between the surface of the liquid contents and the bottom of the cork is called the "headspace height" (Fig. 17–22), and is typically $H = 1.5\,\text{cm}$ for a 750-mL bottle filled at 20°C. Due to its alcoholic content, wine's coefficient of volume expansion is about double that of water; in comparison, the thermal expansion of glass can be neglected. Estimate H if the bottle is kept (*a*) at 10°C, (*b*) at 30°C.

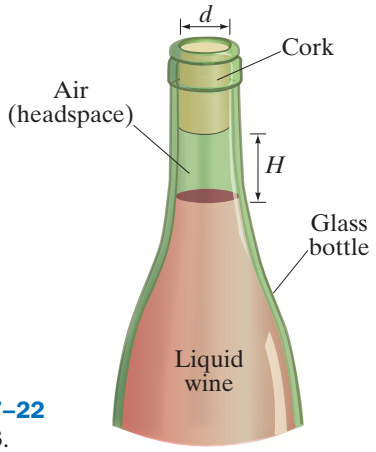


FIGURE 17–22
Problem 23.

24. (III) (*a*) Determine a formula for the change in surface area of a uniform solid sphere of radius r if its coefficient of linear expansion is α (assumed constant) and its temperature is changed by ΔT. (*b*) What is the increase in area of a solid iron sphere of radius 60.0 cm if its temperature is raised from 15°C to 225°C?

25. (III) The pendulum in a grandfather clock is made of brass and keeps perfect time at 17°C. How much time is gained or lost in a year if the clock is kept at 26°C? (Assume the frequency dependence on length for a simple pendulum applies; see Chapter 14.)

26. (III) Water's coefficient of volume expansion in the temperature range from 0°C to about 20°C is given approximately by $\beta = a + bT + cT^2$, with $a = -6.43 \times 10^{-5}(\text{C}^\circ)^{-1}$, $b = 1.70 \times 10^{-5}(\text{C}^\circ)^{-2}$, and $c = -2.02 \times 10^{-7}(\text{C}^\circ)^{-3}$. Using the formula for density from Problem 22, show that water has its greatest density at approximately 4.0°C.

*17–5 Thermal Stresses

*27. (II) An aluminum bar has the desired length when at 12°C. How much stress is required to keep it at this length if the temperature increases to 38°C? [See Table 12–1.]

*28. (II) (*a*) A horizontal steel I-beam of cross-sectional area $0.041\,\text{m}^2$ is rigidly connected to two fixed vertical supports. If the beam was installed when the temperature was 25°C, what stress is developed in the beam when the temperature drops to −25°C? (*b*) Is the ultimate strength of the steel exceeded? (*c*) What stress is developed if the beam is concrete and has a cross-sectional area of $0.15\,\text{m}^2$? Will it fracture?

*29. (III) A barrel of diameter 134.122 cm at 20°C is to be enclosed by an iron band. The circular band has an inside diameter of 134.110 cm at 20°C. It is 9.4 cm wide and 0.65 cm thick. (*a*) To what temperature must the band be heated so that it will fit over the barrel? (*b*) What will be the tension in the band when it cools to 20°C?

17–6 Gas Laws; Absolute Temperature

30. (I) What are the following temperatures on the Kelvin scale: (*a*) 58°C, (*b*) 86°F, (*c*) −55°C, (*d*) 5100°C?

31. (I) Absolute zero is what temperature on the Fahrenheit scale?

32. (II) A weather forecast states that tomorrow's temperature will be twice what it is today. Today's temperature is 10°C. Using the forecast, compute tomorrow's temperature on the Celsius, Fahrenheit, and Kelvin scales. Which value is most believable?

33. (II) Typical temperatures in the interior of the Earth and Sun are about 4000°C and 15×10^6 °C, respectively. (*a*) What are these temperatures in kelvins? (*b*) What percent error is made in each case if a person forgets to change °C to K?

17–7 and 17–8 Ideal Gas Law

34. (I) In an internal combustion engine, air at atmospheric pressure and a temperature of about 20°C is compressed in the cylinder by a piston to $\frac{1}{9}$ of its original volume (compression ratio = 9.0). Estimate the temperature of the compressed air, assuming the pressure reaches 40 atm.

35. (I) If $3.50\,\text{m}^3$ of a gas initially at STP is placed under a pressure of 2.80 atm, the temperature of the gas rises to 38.0°C. What is the volume?

36. (II) If 12.50 mol of helium gas is at 10.0°C and a gauge pressure of 0.350 atm, calculate (*a*) the volume of the helium gas under these conditions, and (*b*) the temperature if the gas is compressed to precisely half the volume at a gauge pressure of 1.00 atm.

37. (II) Calculate the density of nitrogen at STP using the ideal gas law.

38. (II) A storage tank contains 21.6 kg of nitrogen (N_2) at an absolute pressure of 3.45 atm. What will the pressure be if the nitrogen is replaced by an equal mass of CO_2 at the same temperature?

39. (II) A storage tank at STP contains 26.5 kg of nitrogen (N_2). (*a*) What is the volume of the tank? (*b*) What is the pressure if an additional 32.2 kg of nitrogen is added without changing the temperature?

40. (II) A *scuba* tank is filled with air to a gauge pressure of 202 atm when the air temperature is 29°C. A diver then jumps into the ocean and, after a short time on the ocean surface, checks the tank's pressure and finds that it is only 191 atm. Assuming the diver has inhaled a negligible amount of air from the tank, what is the temperature of the ocean water?

41. (II) What is the pressure inside a 41.0-L container holding 105.0 kg of argon gas at 21.6°C?

42. (II) A tank contains 30.0 kg of O_2 gas at a gauge pressure of 8.20 atm. If the oxygen is replaced by helium at the same temperature, how many kilograms of the latter will be needed to produce a gauge pressure of 8.00 atm?

43. (II) A sealed metal container contains a gas at 30.0°C and absolute pressure 1.00 atm. To what temperature must the gas be heated for the pressure to double to 2.00 atm? (Ignore expansion of the container.)

44. (II) A tire is filled with air at 15°C to a gauge pressure of 230 kPa. If the tire reaches a temperature of 38°C, what fraction of the original air must be removed if the original pressure of 230 kPa is to be maintained?

45. (II) If 59.2 L of oxygen at 18.0°C and an absolute pressure of 2.45 atm are compressed to 38.8 L and at the same time the temperature is raised to 56.0°C, what will the new pressure be?

46. (II) A helium-filled balloon escapes a child's hand at sea level and 20.0°C. When it reaches an altitude of 3600 m, where the temperature is 5.0°C and the pressure only 0.64 atm, how will its volume compare to that at sea level?

47. (II) A sealed metal container can withstand a pressure difference of 0.50 atm. The container initially is filled with an ideal gas at 18°C and 1.0 atm. To what temperature can you cool the container before it collapses? (Ignore any changes in the container's volume due to thermal expansion.)

48. (II) A helium balloon rises because of a buoyant force. By what percentage does the buoyant force on a helium balloon change if the temperature of the helium is increased from 15°C to 25°C while the temperature of the surrounding air is unchanged? Assume that the pressure of the helium remains constant.

49. (II) The atmospheric pressure on the surface of Mars has an average value of about 600 Pa. (*a*) What is the volume of 1.0 mole of the *Martian atmosphere* where the average temperature is about −50°C? (*b*) What is the density of the Martian atmosphere under these conditions, which is about 96% carbon dioxide? (*c*) What fraction is this density as compared to the Earth's atmospheric density of 1.29 kg/m^3, at 0°C and 1 atm pressure?

50. (II) You buy an "airtight" *potato chip bag* packaged at sea level, and take the chips on an airplane flight. When you take the potato chip bag out of your "carry-on" bag, you notice it has noticeably "puffed up." Airplane cabins are typically pressurized at 0.75 atm, and assuming the temperature inside an airplane is about the same as inside a potato chip processing plant, by what percentage has the bag "puffed up" in comparison to when it was packaged?

51. (II) A typical *scuba tank*, when fully charged, contains 12 L of air at 204 atm. Assume an "empty" tank contains air at 34 atm gauge pressure, and is connected to an air compressor at sea level. The air compressor intakes air from the atmosphere, compresses it to high pressure, and then inputs this high-pressure air into the scuba tank. If the (average) flow rate of air from the atmosphere into the intake port of the air compressor is 290 L/min, how long will it take to fully charge the scuba tank? Assume the tank remains at the same temperature as the surrounding air during the filling process.

52. (II) Compare the value for the density of water vapor at exactly 100°C and 1 atm (Table 13–1) with the value predicted from the ideal gas law. Why would you expect a difference?

53. (II) A sealed test tube traps 28.0 cm^3 of air at a pressure of 1.00 atm and temperature of 18°C. The test tube's stopper has a diameter of 1.50 cm and will "pop off" the test tube if a net upward force of 10.0 N is applied to it. To what temperature would you have to heat the trapped air in order to "pop off" the stopper? Assume the air surrounding the test tube is always at a pressure of 1.00 atm.

54. (III) A sealed container containing 3.5 mol of gas is squeezed, changing its volume from 0.020 m^3 to 0.018 m^3. During this process, the temperature decreases by 9.0 K while the pressure increases by 450 Pa. What were the original pressure and temperature of the gas in the container?

55. (III) An air bubble at the bottom of a lake 32.0 m deep has a volume of 1.00 cm^3. If the temperature at the bottom is 5.5°C and at the top 18.5°C, what is the radius of the bubble just before it reaches the surface?

17–9 Ideal Gas Law in Terms of Molecules: Avogadro's Number

56. (I) Calculate the number of molecules/m^3 in an ideal gas at STP.

57. (II) How many moles of water are there in 1.00 L at STP? How many molecules?

58. (II) Estimate the number of (*a*) moles and (*b*) molecules of water in all the Earth's oceans. Assume water covers 75% of the Earth to an average depth of 3 km.

59. (II) What is the pressure in a region of outer space where there is 1 molecule/cm^3 and the temperature is 3 K?

60. (II) Is a gas mostly empty space? Check by assuming that a typical width of a gas molecule in air is about $\ell_0 = 0.3$ nm so one gas molecule occupies an approximate volume equal to ℓ_0^3. Assume STP.

61. (II) The lowest pressure attainable using the best available vacuum techniques is about 10^{-12} N/m^2. At such a pressure, how many molecules are there per cm^3 at 0°C?

*17–10 Ideal Gas Temperature Scale

*62. (II) In a constant-volume gas thermometer, what is the limiting ratio of the pressure at the boiling point of water at 1 atm to that at the triple point? (Keep five significant figures.)

*63. (II) At the boiling point of sulfur (444.6°C) the pressure in a constant-volume gas thermometer is 187 torr. Estimate (*a*) the pressure at the triple point of water, (*b*) the temperature when the pressure in the thermometer is 128 torr.

*64. (II) Use Fig. 17–18 to determine the inaccuracy of a constant-volume gas thermometer using oxygen if it reads a pressure $P = 268$ torr at the boiling point of water at 1 atm. Express answer (*a*) in kelvins and (*b*) as a percentage.

*65. (III) A constant-volume gas thermometer is being used to determine the temperature of the melting point of a substance. The pressure in the thermometer at this temperature is 218 torr; at the triple point of water, the pressure is 286 torr. Some gas is now released from the thermometer bulb so that the pressure at the triple point of water becomes 163 torr. At the temperature of the melting substance, the pressure is 128 torr. Estimate, as accurately as possible, the melting-point temperature of the substance.

General Problems

66. A Pyrex measuring cup was calibrated at normal room temperature. How much error will be made in a recipe calling for 375 mL of cool water, if the water and the cup are hot, at 95°C, instead of at room temperature? Neglect the glass expansion.

67. A precise steel *tape measure* has been calibrated at 14°C. At 37°C, (*a*) will it read high or low, and (*b*) what will be the percentage error?

68. A cubic box of volume 8.25×10^{-2} m^3 is filled with air at atmospheric pressure at 15°C. The box is closed and heated to 165°C. What is the net force on each side of the box?

69. The gauge pressure in a helium gas cylinder is initially 28 atm. After many balloons have been blown up, the gauge pressure has decreased to 5 atm. What fraction of the original gas remains in the cylinder?

70. If a rod of original length ℓ_1 has its temperature changed from T_1 to T_2, determine a formula for its new length ℓ_2 in terms of T_1, T_2, and α. Assume (*a*) $\alpha =$ constant, (*b*) $\alpha = \alpha(T)$ is some function of temperature, and (*c*) $\alpha = \alpha_0 + bT$ where α_0 and b are constants.

71. If a *scuba diver* fills his lungs to full capacity of 4.8 L when 9.0 m below the surface of sea water, to what volume would his lungs expand if he quickly rose to the surface? Is this advisable?

72. (*a*) Use the ideal gas law to show that, for an ideal gas at constant pressure, the coefficient of volume expansion is equal to $\beta = 1/T$, where T is the kelvin temperature. Compare to Table 17–1 for gases at $T = 293$ K. (*b*) Show that the bulk modulus (Section 12–5) for an ideal gas held at constant temperature is $B = P$, where P is the pressure.

73. A house has a volume of 1200 m^3. (*a*) What is the total mass of air inside the house at 15°C? (*b*) If the temperature in the house drops to −12°C, what mass of air enters or leaves the house?

74. Estimate the number of air molecules in a room of length 6.0 m, width 3.0 m, and height 2.5 m. Assume the temperature is 22°C. How many moles does that correspond to?

75. Estimate how many *molecules of air* are in each 2.0-L *breath* you inhale that were also in the last breath Galileo took. Assume the atmosphere is about 10 km high and of constant density. What other assumptions did you make?

76. Assume that in an alternate universe, the laws of physics are very different from ours and that "ideal" gases behave as follows: (i) At constant temperature, pressure is inversely proportional to the square of the volume. (ii) At constant pressure, the volume varies directly with the $\frac{2}{3}$ power of the temperature. (iii) At 273.15 K and 1.00 atm pressure, 1.00 mole of an ideal gas is found to occupy 22.4 L. Obtain the form of the ideal gas law in this alternate universe, including the value of the gas constant R.

77. An iron cube floats in a bowl of liquid mercury at 0°C. (*a*) If the temperature is raised to 25°C, will the cube float higher or lower in the mercury? (*b*) By what percent will the fraction of volume submerged change? [*Hint*: See Chapter 13.]

78. (*a*) The tube of a mercury thermometer has an inside diameter of 0.140 mm. The bulb has a volume of 0.285 cm^3. How far will the thread of mercury move when the temperature changes from 10.5°C to 33.0°C? Take into account expansion of the Pyrex glass. (*b*) Determine a formula for the change in length of the mercury column in terms of relevant variables. Ignore tube volume compared to bulb volume.

79. Many cars warn the driver when the *tire pressure* is low. Suppose a warning light comes on when the pressure drops by 20% from the specified gauge pressure of 2.4 atm. Neglecting volume change and assuming constant temperature, find the fraction of air that has leaked out of the tire.

80. From the known value of atmospheric pressure at the surface of the Earth, estimate the total number of air molecules in the Earth's atmosphere.

81. What is the average distance between nitrogen molecules at STP?

82. Estimate the percent difference in the density of iron at STP, and when it is a solid deep in the Earth where the temperature is 2000°C and under 5000 atm of pressure. Assume the bulk modulus $(90 \times 10^9\ \text{N/m}^2)$ and the coefficient of volume expansion do not vary with temperature and are the same as at STP.

83. A standard *cylinder of oxygen* used in a hospital has gauge pressure = 2000 psi (13,800 kPa) and volume = 14 L $(0.014\ \text{m}^3)$ at $T = 295$ K. How long will the cylinder last if the flow rate, measured at atmospheric pressure, is constant at 2.0 L/min?

84. A helium balloon, assumed to be a perfect sphere, has a radius of 26.0 cm. At room temperature (20°C), its internal pressure is 1.08 atm. Determine the number of moles of helium in the balloon, and the mass of helium needed to inflate the balloon to these values.

85. The density of gasoline at 0°C is $0.68 \times 10^3\ \text{kg/m}^3$. (*a*) What is the density on a hot day, when the temperature is 33°C? (*b*) What is the percent change in density?

86. A brass lid screws tightly onto a glass jar at 15°C. To help open the jar, it can be placed into a bath of hot water. After this treatment, the temperatures of the lid and the jar are both 55°C. The inside diameter of the lid is 7.0 cm. Find the size of the gap (difference in radius) that develops by this procedure.

87. The first real length standard, adopted more than 200 years ago, was a platinum bar with two very fine marks separated by what was defined to be exactly one meter. If this standard bar was to be accurate to within $\pm 1.0\ \mu$m, how carefully would the trustees have needed to control the temperature? The coefficient of linear expansion is $9 \times 10^{-6}/\text{C}°$.

88. A helium balloon has volume V_0 and temperature T_0 at sea level where the pressure is P_0 and the air density is ρ_0. The balloon is allowed to float up in the air to altitude y where the temperature is T_1. (*a*) Show that the volume occupied by the balloon is then $V = V_0(T_1/T_0)e^{+cy}$ where $c = \rho_0 g/P_0 = 1.25 \times 10^{-4}\ \text{m}^{-1}$. (*b*) Show that the buoyant force does not depend on altitude y. Assume that the skin of the balloon maintains the helium pressure at a constant factor of 1.05 times greater than the outside pressure. [*Hint*: Assume that the pressure change with altitude is $P = P_0 e^{-cy}$, Eq. 13–6c in Chapter 13.]

89. If a steel band were to fit snugly around the Earth's equator at 25°C, but then was heated to 55°C, how high above the Earth would the band be (assume equal everywhere)?

90. A 28.4-kg solid aluminum cylindrical wheel of radius 0.41 m is rotating about its axle in frictionless bearings with angular velocity $\omega = 32.8$ rad/s. If its temperature is then raised from 15.0°C to 95.0°C, what is the fractional change in ω?

91. A copper wire sags 54.0 cm between two utility poles 30.0 m apart when the temperature is −15°C. Estimate the amount of sag when the temperature is +35°C. [*Hint*: An estimate can be made by assuming the shape of the wire is approximately an arc of a circle; hard equations can sometimes be solved by guessing values.]

92. A temperature controller, designed to work in a steam environment, involves a bimetallic strip constructed of brass and steel, connected at their ends by rivets. Each of the metals is 1.6 mm thick. At 20°C, the strip is 10.0 cm long and straight. Find the radius of curvature r of the assembly at 100°C. See Fig. 17–23.

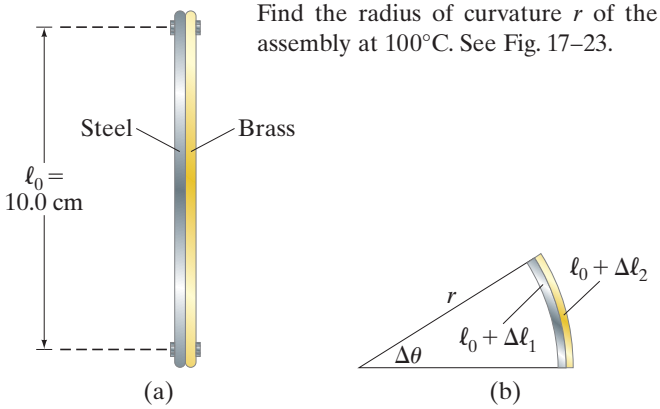


FIGURE 17–23 Problem 92.

93. You have a vial of an unknown liquid which might be octane (gasoline), water, glycerin, or ethyl alcohol. You are trying to determine its identity by studying how its volume changes with temperature changes. You fill a Pyrex graduated cylinder to 100.00 mL with the liquid when the liquid and the cylinder are at 0.000°C. You raise the temperature in five-degree increments, allowing the graduated cylinder and liquid to come to equilibrium at each temperature. You read the volumes listed below off the graduated cylinder at each temperature. Take into account the expansion of the Pyrex glass cylinder. Graph the data and determine the slope of the line to find the effective (combined) coefficient of volume expansion β. Then determine β for the liquid and which liquid is in the vial.

Temperature (°C)	Volume Reading (apparent mL)
0.000	100.00
5.000	100.24
10.000	100.50
15.000	100.72
20.000	100.96
25.000	101.26
30.000	101.48
35.000	101.71
40.000	101.97
45.000	102.20
50.000	102.46

94. A *scuba tank* when fully charged has an absolute pressure of 185 atm at 18°C. The volume of the tank is 11.3 L. (*a*) What would the volume of the air be at 1.00 atm and at the same temperature? (*b*) Before entering the water, a person consumes 2.0 L of air in each breath, and breathes 12 times a minute. At this rate, 2.0 L at 1 atm per breathe, how long would the tank last? (*c*) At a depth of 23.0 m in sea water at a temperature of 10°C, how long would the same tank last assuming the breathing rate does not change?

95. This Chapter gives two ways to calculate the thermal expansion of a gas at a constant pressure of 1.0 atm. Use both methods to calculate the volume change of 750 L of an ideal gas as it goes from −100°C to 0°C and from 0°C to 100°C. Why are the answers different?

96. A *hot-air balloon* achieves its buoyant lift by heating the air inside the balloon, which makes it less dense than the air outside. Suppose the volume of a balloon is 1800 m^3 and the required lift is 3300 N (rough estimate of the weight of the equipment and passenger). Calculate the temperature of the air inside the balloon which will produce the required lift. Assume the outside air is an ideal gas at 0°C. What factors limit the maximum altitude attainable by this method for a given load? [*Hint*: See Chapter 13, Example 13–5.]

97. **Why snorkels are not 4 feet long.** Snorkelers breathe through short tubular "snorkels" while swimming under water very near the surface (Fig. 17–24). One end of the snorkel is in the snorkeler's mouth and the other end protrudes just above the water's surface. Unfortunately, snorkels cannot support breathing to any great depth: it is said that a typical snorkeler below a water depth of only about 30 cm cannot draw a breath through a snorkel. Based on this observation, what is the approximate change in a typical person's lung pressure (in atm) when drawing a breath? (Note that your diaphragm muscles, which expand your lungs, must work also against the extra water pressure.)



FIGURE 17–24 Problem 97.

ANSWERS TO EXERCISES

A: −40°C = −40°F.
B: (*d*).
C: $7\frac{1}{2}$ mm [$\Delta T = 110(5/9)$].
D: (i) Higher, (ii) same, (iii) lower.
E: (*d*).
F: (*c*).
G: (*b*) Less.

In this winter scene in Yosemite National Park, we recognize the three states of matter for water: as a liquid (the river), as a solid (snow and ice), and as a gas (steam). In this Chapter we examine the microscopic theory of matter as atoms or molecules that are always in motion, which we call kinetic theory. The temperature of a gas is directly related to the average translational kinetic energy of its molecules. We will consider ideal gases, but will also look at real gases and how they change phase, including evaporation, vapor pressure, and humidity. We also discover how the temperature of boiling water decreases with altitude.



CHAPTER 18

Kinetic Theory of Gases

CONTENTS

18–1 The Ideal Gas Law and the Molecular Interpretation of Temperature
18–2 Distribution of Molecular Speeds
18–3 Real Gases and Changes of Phase
18–4 Vapor Pressure and Humidity
18–5 Temperature Decrease of Boiling Water with Altitude
18–6 Van der Waals Equation of State
18–7 Mean Free Path
18–8 Diffusion

CHAPTER-OPENING QUESTION—Guess now!

The typical speed of an air molecule at room temperature (20°C) is

(a) nearly at rest (<10 km/h).
(b) on the order of 10 km/h.
(c) on the order of 100 km/h.
(d) on the order of 1000 km/h.
(e) nearly the speed of light.

The analysis of matter in terms of atoms in continuous random motion (see Fig. 17–2) is called **kinetic theory**. We now investigate the properties of a gas from the point of view of kinetic theory, which is based on the laws of classical mechanics. But to apply Newton's laws to each one of the vast number of molecules in a gas ($>10^{25}/m^3$ at STP) is far beyond the capability of any present computer. Instead we take a statistical approach and determine averages of certain quantities, and connect these averages to macroscopic variables. We will demand that our microscopic description correspond to the macroscopic properties of gases; otherwise our theory would be of little value. Most importantly, we will arrive at an important relation between the average kinetic energy of molecules in a gas and the absolute temperature.

18–1 The Ideal Gas Law and the Molecular Interpretation of Temperature

We make the following assumptions about the molecules in a gas. These assumptions reflect a simple view of a gas, but nonetheless the results they predict correspond well to the essential features of real gases that are at low pressure and far from the liquefaction point. Under these conditions real gases follow the ideal gas law quite closely, and indeed the gas we now describe is referred to as an **ideal gas**.

The assumptions representing the basic postulates of the kinetic theory for an ideal gas are

1. There are a large number of molecules, N, each of mass m, moving in random directions with a variety of speeds. This assumption agrees with our observation that a gas fills its container and, in the case of air on Earth, is kept from escaping only by the force of gravity.
2. The molecules are, on average, far apart from one another. That is, their average separation is much greater than the diameter of each molecule.
3. The molecules are assumed to obey the laws of classical mechanics, and are assumed to interact with one another only when they collide. Although molecules exert weak attractive forces on each other between collisions, the potential energy associated with these forces is small compared to the kinetic energy.
4. Collisions with another molecule or the wall of the vessel are assumed to be perfectly elastic, like the collisions of perfectly elastic billiard balls (Chapter 9). We assume the collisions are of very short duration compared to the time between collisions. Then we can ignore the potential energy associated with collisions in comparison to the kinetic energy between collisions.

We can see how this kinetic view of a gas can explain Boyle's law (Section 17–6). The pressure exerted on a wall of a container of gas is due to the constant bombardment of molecules. If the volume is reduced by (say) half, the molecules are closer together and twice as many will be striking a given area of the wall per second. Hence we expect the pressure to be twice as great, in agreement with Boyle's law.

Now let us calculate quantitatively the pressure a gas exerts on its container in terms of microscopic quantities. We imagine that the molecules are inside a rectangular container (at rest) whose ends have area A and whose length is ℓ, as shown in Fig. 18–1a. The pressure exerted by the gas on the walls of its container is, according to our model, due to the collisions of the molecules with the walls. Let us focus our attention on the wall, of area A, at the left end of the container and examine what happens when one molecule strikes this wall, as shown in Fig. 18–1b. This molecule exerts a force on the wall, and according to Newton's third law the wall exerts an equal and opposite force back on the molecule. The magnitude of this force on the molecule, according to Newton's second law, is equal to the molecule's rate of change of momentum, $F = dp/dt$ (Eq. 9–2). Assuming the collision is elastic, only the x component of the molecule's momentum changes, and it changes from $-mv_x$ (it is moving in the negative x direction) to $+mv_x$. Thus the change in the molecule's momentum, $\Delta(mv)$, which is the final momentum minus the initial momentum, is

$$\Delta(mv) = mv_x - (-mv_x) = 2mv_x$$

for one collision. This molecule will make many collisions with the wall, each separated by a time Δt, which is the time it takes the molecule to travel across the container and back again, a distance (x component) equal to 2ℓ. Thus $2\ell = v_x\,\Delta t$, or

$$\Delta t = \frac{2\ell}{v_x}.$$

The time Δt between collisions of one molecule with the wall is very small, so the number of collisions per second is very large. Thus the average force due to one molecule—averaged over many collisions—will be equal to the momentum change during one collision with the wall divided by the time between collisions (Newton's second law):

$$F = \frac{\Delta(mv)}{\Delta t} = \frac{2mv_x}{2\ell/v_x} = \frac{mv_x^2}{\ell}. \qquad \text{[due to one molecule]}$$

During its passage back and forth across the container, the molecule may collide with the tops and sides of the container, but this does not alter its x component of momentum and thus does not alter our result. It may also collide with other molecules, which may change its v_x. However, any loss (or gain) of momentum is exchanged with the other molecule in the collision, and because we will eventually sum over all the molecules, this effect will be included. So our result above is not altered.

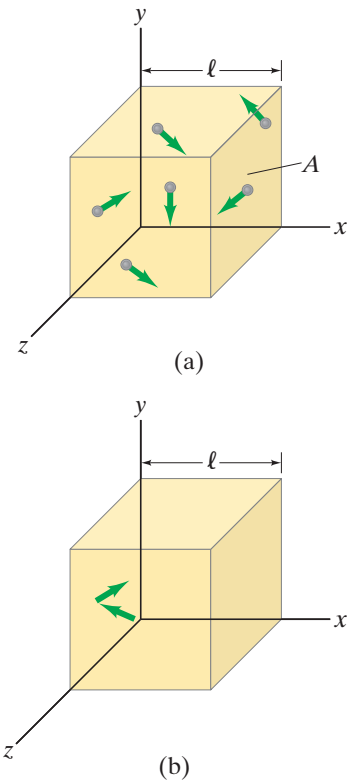


FIGURE 18–1 (a) Molecules of a gas moving about in a rectangular container. (b) Arrows indicate the momentum of one molecule as it rebounds from the end wall.

The actual force due to one molecule is intermittent, but because a huge number of molecules are striking the wall per second, the force is, on average, nearly constant. To calculate the force due to *all* the molecules in the container, we have to add the contributions of each. If all N molecules have the same mass m, the net force on the wall is

$$F = \frac{m}{\ell}(v_{x1}^2 + v_{x2}^2 + \cdots + v_{xN}^2),$$

where v_{x1} means v_x for molecule number 1 (we arbitrarily assign each molecule a number) and the sum extends over the total number of molecules N in the container. The average value of the square of the x component of velocity is

$$\overline{v_x^2} = \frac{v_{x1}^2 + v_{x2}^2 + \cdots + v_{xN}^2}{N}, \qquad \textbf{(18–1)}$$

CAUTION
Overbar ($\bar{}$) means average

where the overbar ($\bar{}$) means "average." Thus we can write the force as

$$F = \frac{m}{\ell} N\overline{v_x^2}. \qquad \textbf{(i)}$$

We know that the square of any vector is equal to the sum of the squares of its components (theorem of Pythagoras). That is, $v^2 = v_x^2 + v_y^2 + v_z^2$ for any velocity v. Taking averages, we obtain

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}. \qquad \textbf{(ii)}$$

Since the velocities of the molecules in our gas are assumed to be random, there is no preference to one direction or another. Hence

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}. \qquad \textbf{(iii)}$$

Combining Eqs. (iii) and (ii) we get:

$$\overline{v^2} = 3\overline{v_x^2}. \qquad \textbf{(iv)}$$

We substitute Eq. (iv) into Eq. (i) for the net force F, and find

$$F = \frac{m}{\ell} N \frac{\overline{v^2}}{3}.$$

The pressure on the wall is then

$$P = \frac{F}{A} = \frac{1}{3}\frac{Nm\overline{v^2}}{A\ell}$$

or

$$P = \frac{1}{3}\frac{Nm\overline{v^2}}{V}, \qquad \left[\begin{array}{c}\text{pressure in an}\\ \text{ideal gas}\end{array}\right] \qquad \textbf{(18–2)}$$

where $V = \ell A$ is the volume of the container. This is the result we were seeking, the pressure exerted by a gas on its container expressed in terms of molecular properties.

Equation 18–2 can be rewritten in a clearer form by multiplying both sides by V and rearranging the right-hand side:

$$PV = \tfrac{2}{3}N(\tfrac{1}{2}m\overline{v^2}). \qquad \textbf{(18–3)}$$

The quantity $\frac{1}{2}m\overline{v^2}$ is the average translational kinetic energy $\overline{K}$ of the molecules in the gas. If we compare Eq. 18–3 with Eq. 17–4, the ideal gas law $PV = NkT$, we see that the two agree if (recall k = Boltzmann constant = 1.38×10^{-23} J/K; Section 17–9)

TEMPERATURE RELATED TO AVERAGE KINETIC ENERGY OF MOLECULES

$$\tfrac{2}{3}(\tfrac{1}{2}m\overline{v^2}) = kT,$$

or

$$\overline{K} = \tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}kT. \qquad \text{[ideal gas]} \quad \textbf{(18–4)}$$

This equation tells us that

> **the average translational kinetic energy of molecules in random motion in an ideal gas is directly proportional to the absolute temperature of the gas.**

The higher the temperature, according to kinetic theory, the faster the molecules are moving on average. This relation is a fundamental one in physics: it relates

an everyday macroscopic quantity (T) to theoretical microscopic quantities. Because of this relationship (Eq. 18–4), the random motion of molecules in a gas is sometimes referred to as **thermal motion**.

EXAMPLE 18–1 Molecular kinetic energy. What is the average translational kinetic energy of molecules in an ideal gas at 37°C?

APPROACH We use the absolute temperature in Eq. 18–4.

SOLUTION We change 37°C to 310 K and insert into Eq. 18–4:

$$\overline{K} = \tfrac{3}{2}kT = \tfrac{3}{2}(1.38\times10^{-23}\,\text{J/K})(310\,\text{K}) = 6.42\times10^{-21}\,\text{J}.$$

NOTE A mole of molecules would have a total translational kinetic energy equal to $(6.42\times10^{-21}\,\text{J})(6.02\times10^{23}) = 3860\,\text{J}$, which equals the kinetic energy of a 1-kg stone traveling almost 90 m/s.

EXERCISE A If molecules of hydrogen gas and oxygen gas were placed in the same balloon at room temperature, how would the average kinetic energies of the molecules compare? (*a*) They would be the same. (*b*) The hydrogen molecules would have greater kinetic energy. (*c*) The oxygen molecules would have greater kinetic energy. (*d*) Need more information.

PHYSICS APPLIED
Molecules in human cells

Equation 18–4 holds not only for gases, but also applies reasonably accurately to liquids and solids. Thus the result of Example 18–1 would apply to molecules within living cells at body temperature (37°C).

We can use Eq. 18–4 to calculate how fast molecules are moving on the average. Notice that the average in Eqs. 18–1 through 18–4 is over the *square* of the speed. The square root of $\overline{v^2}$ is called the **root-mean-square** speed, v_{rms} (since we are taking the square *root* of the *mean* of the *square* of the speed):

$$v_{\text{rms}} = \sqrt{\overline{v^2}} = \sqrt{\frac{3kT}{m}}. \qquad \textbf{(18–5)}$$

EXAMPLE 18–2 Speeds of air molecules. What is the rms speed of air molecules (O_2 and N_2) at room temperature (20°C)?

APPROACH To obtain v_{rms}, we need the masses of O_2 and N_2 molecules and then apply Eq. 18–5 to oxygen and nitrogen separately, since they have different masses.

SOLUTION The masses of one molecule of O_2 (molecular mass = 32 u) and N_2 (molecular mass = 28 u) are (where $1\,\text{u} = 1.66\times10^{-27}\,\text{kg}$)

$$m(O_2) = (32)(1.66\times10^{-27}\,\text{kg}) = 5.3\times10^{-26}\,\text{kg},$$

$$m(N_2) = (28)(1.66\times10^{-27}\,\text{kg}) = 4.6\times10^{-26}\,\text{kg}.$$

Thus, for oxygen

$$v_{\text{rms}} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{(3)(1.38\times10^{-23}\,\text{J/K})(293\,\text{K})}{(5.3\times10^{-26}\,\text{kg})}} = 480\,\text{m/s},$$

and for nitrogen the result is $v_{\text{rms}} = 510\,\text{m/s}$.

NOTE These speeds are more than 1700 km/h or 1000 mi/h, and are greater than the speed of sound ≈ 340 m/s at 20°C (Chapter 16).

NOTE The speed v_{rms} is a magnitude only. The *velocity of molecules averages to zero*. Why? The velocity has direction, and as many molecules move to the right as to the left, as many up as down, as many inward as outward.

EXERCISE B What is the rms speed of O_2 molecules in your lungs (37°C)?

EXERCISE C Now you can return to the Chapter-Opening Question, page 516, and answer it correctly. Try to explain why you may have answered differently the first time.

EXERCISE D If you double the volume of a gas while keeping the pressure and number of moles constant, the average (rms) speed of the molecules (*a*) doubles, (*b*) quadruples, (*c*) increases by $\sqrt{2}$, (*d*) is half, (*e*) is $\frac{1}{4}$.

CONCEPTUAL EXAMPLE 18–3 **Less helium in the tank.** A tank of helium is used to fill balloons. As each balloon is filled, the number of helium atoms remaining in the tank decreases. How does this affect the rms speed of molecules remaining in the tank?

RESPONSE The rms speed is given by Eq. 18–5: $v_{\text{rms}} = \sqrt{3kT/m}$. So only the temperature matters, not pressure P or number of moles n. If the tank remains at a constant (ambient) temperature, then the rms speed remains constant even though the pressure of helium in the tank decreases.

In a collection of molecules, the **average speed**, $\bar{v}$, is the average of the magnitudes of the speeds themselves; $\bar{v}$ is generally not equal to v_{rms}. To see the difference between the average speed and the rms speed, consider the following Example.

EXAMPLE 18–4 **Average speed and rms speed.** Eight particles have the following speeds, given in m/s: 1.0, 6.0, 4.0, 2.0, 6.0, 3.0, 2.0, 5.0. Calculate (*a*) the average speed and (*b*) the rms speed.

APPROACH In (*a*) we sum the speeds and divide by $N = 8$. In (*b*) we square each speed, sum the squares, divide by $N = 8$, and take the square root.

SOLUTION (*a*) The average speed is

$$\bar{v} = \frac{1.0 + 6.0 + 4.0 + 2.0 + 6.0 + 3.0 + 2.0 + 5.0}{8} = 3.6\,\text{m/s}.$$

(*b*) The rms speed is (Eq. 18–1):

$$v_{\text{rms}} = \sqrt{\frac{(1.0)^2 + (6.0)^2 + (4.0)^2 + (2.0)^2 + (6.0)^2 + (3.0)^2 + (2.0)^2 + (5.0)^2}{8}}\,\text{m/s}$$
$$= 4.0\,\text{m/s}.$$

We see in this Example that $\bar{v}$ and v_{rms} are not necessarily equal. In fact, for an ideal gas they differ by about 8%. We will see in the next Section how to calculate $\bar{v}$ for an ideal gas. We already have the tool to calculate v_{rms} (Eq. 18–5).

Kinetic Energy Near Absolute Zero

Equation 18–4, $\overline{K} = \frac{3}{2}kT$, implies that as the temperature approaches absolute zero, the kinetic energy of molecules approaches zero. Modern quantum theory, however, tells us this is not quite so. Instead, as absolute zero is approached, the kinetic energy approaches a very small nonzero minimum value. Even though all real gases become liquid or solid near 0 K, molecular motion does not cease, even at absolute zero.

18–2 Distribution of Molecular Speeds

The Maxwell Distribution

The molecules in a gas are assumed to be in random motion, which means that many molecules have speeds less than the rms speed and others have greater speeds. In 1859, James Clerk Maxwell (1831–1879) worked out a formula for the

most probable distribution of speeds in a gas containing N molecules. We will not give a derivation here but merely quote his result:

$$f(v) = 4\pi N\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} v^2 e^{-\frac{1}{2}\frac{mv^2}{kT}} \qquad \textbf{(18–6)}$$

where $f(v)$ is called the **Maxwell distribution of speeds**, and is plotted in Fig. 18–2. The quantity $f(v)\,dv$ represents the *number of molecules* that have speed between v and $v + dv$. Notice that $f(v)$ does not give the number of molecules with speed v; $f(v)$ must be multiplied by dv to give the number of molecules (the number of molecules depends on the "width" or "range" of velocities included, dv). In the formula for $f(v)$, m is the mass of a single molecule, T is the absolute temperature, and k is the Boltzmann constant. Since N is the total number of molecules in the gas, when we sum over all the molecules in the gas we must get N; thus we must have

$$\int_0^\infty f(v)\,dv = N.$$

(Problem 23 is an exercise to show that this is true.)

Experiments to determine the distribution of molecular speeds in real gases, starting in the 1920s, confirmed with considerable accuracy the Maxwell distribution (for gases at not too high a pressure) and the direct proportion between average kinetic energy and absolute temperature, Eq. 18–4.

The Maxwell distribution for a pure gas depends only on its absolute temperature. Figure 18–3 shows the Maxwell distribution for two different temperatures. Just as v_{rms} increases with temperature, so the whole distribution curve shifts to the right at higher temperatures.

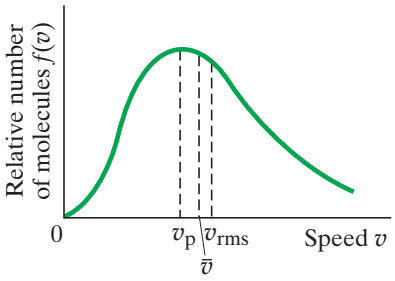


FIGURE 18–2 Distribution of speeds of molecules in an ideal gas. Note that $\bar{v}$ and v_{rms} are not at the peak of the curve. This is because the curve is not symmetrical. It is skewed to the right (where the curve can go on and on farther out). The speed at the peak of the curve is the "most probable speed," v_p.

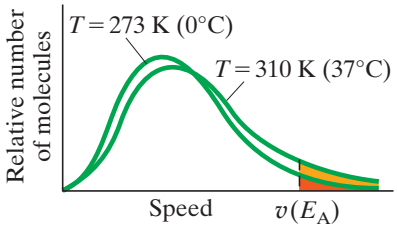


FIGURE 18–3 Distribution of molecular speeds for two different temperatures.

Figure 18–3 illustrates how kinetic theory can explain why many chemical reactions, including those in biological cells, take place more rapidly as the temperature increases. Most chemical reactions occur in a liquid solution, and the molecules have a speed distribution close to the Maxwell distribution. Two molecules may chemically react only if their kinetic energy is high enough so that when they collide, they partially penetrate into each other. The minimum energy required for a particular chemical reaction is called the *activation energy*, E_A. The molecular speed corresponding to a kinetic energy of E_A for a particular reaction is indicated in Fig. 18–3. The relative number of molecules with energy greater than this value is given by the area under the curve to the right of $v(E_A)$, shown in Fig. 18–3 by the two different shadings. We see that the number of molecules that have kinetic energies above E_A increases greatly for a small increase in temperature. The rate of a chemical reaction is proportional to the number of molecules with energy greater than E_A, and thus we see why reaction rates increase rapidly with increased temperature.

PHYSICS APPLIED
How chemical reactions depend on temperature

Calculations Using the Maxwell Distribution

Let us see how the Maxwell distribution can be used to obtain some interesting results.

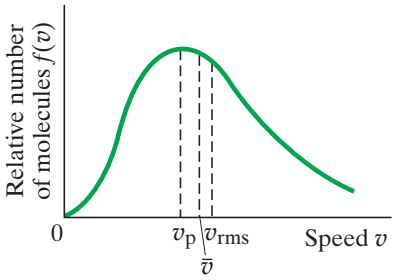


FIGURE 18–2 (Repeated.) Distribution of speeds of molecules in an ideal gas. Note that $\bar{v}$ and v_{rms} are not at the peak of the curve. This is because the curve is not symmetrical. It is skewed to the right (where the curve can go on and on farther out). The speed at the peak of the curve is the "most probable speed," v_p.

EXAMPLE 18–5 **Determining $\bar{v}$ and v_p.** Derive formulas for (*a*) the average speed, $\bar{v}$, and (*b*) the most probable speed, v_p, of molecules in an ideal gas at temperature T.

APPROACH (*a*) The average value of a quantity is found by multiplying each possible value of the quantity (here, speed) by the number of molecules having that value, then summing them and dividing by N (the total number). For (*b*), we want to find where the curve of Fig. 18–2 has a maximum, with zero slope; so we set $df/dv = 0$.

SOLUTION (*a*) To find the average speed of molecules, $\bar{v}$, we sum (integrate) over the product of v and the number $f(v)\,dv$ that have speed v using Eq. 18–6:

$$\bar{v} = \frac{\int_0^\infty v f(v)\,dv}{N} = 4\pi\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}}\int_0^\infty v^3 e^{-\frac{1}{2}\frac{mv^2}{kT}}\,dv.$$

We can integrate by parts or look up the definite integral in a Table (Appendix B–5), and obtain

$$\bar{v} = 4\pi\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}}\left(\frac{2k^2T^2}{m^2}\right) = \sqrt{\frac{8}{\pi}\frac{kT}{m}} \approx 1.60\sqrt{\frac{kT}{m}}.$$

(*b*) The **most probable speed**, v_p, occurs more than any other speed, where $f(v)$ has its maximum value. The slope there is zero: the derivative of Eq. 18–6 gives

$$\frac{df(v)}{dv} = 4\pi N\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}}\left(2ve^{-\frac{mv^2}{2kT}} - \frac{2mv^3}{2kT}e^{-\frac{mv^2}{2kT}}\right) = 0.$$

Solving for v, we find

$$v_p = \sqrt{\frac{2kT}{m}} \approx 1.41\sqrt{\frac{kT}{m}}.$$

In summary,

$$v_p = \sqrt{2\frac{kT}{m}} \approx 1.41\sqrt{\frac{kT}{m}} \quad \text{[most probable speed]} \quad \textbf{(18–7a)}$$

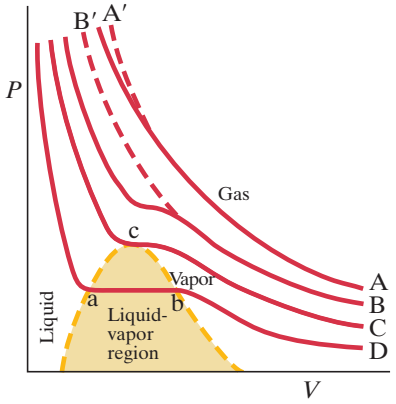
$$\bar{v} = \sqrt{\frac{8}{\pi}\frac{kT}{m}} \approx 1.60\sqrt{\frac{kT}{m}} \quad \text{[average speed]} \quad \textbf{(18–7b)}$$

and, repeating Eq. 18–5 for completeness,

$$v_{rms} = \sqrt{3\frac{kT}{m}} \approx 1.73\sqrt{\frac{kT}{m}}. \quad \text{[rms speed]} \quad \textbf{(18–5)}$$

All three are indicated in Fig. 18–2. The speeds of molecules in a gas vary from zero to many times the rms speed, but as Figs. 18–2 and 18–3 show, most molecules have speeds that are not far from the average. Less than 1% of the molecules exceed four times v_{rms}.

FIGURE 18–4 *PV* diagram for a real substance. Curves A, B, C, and D represent the same substance at different temperatures ($T_A > T_B > T_C > T_D$).



18–3 Real Gases and Changes of Phase

The ideal gas law $PV = NkT$ is an accurate description of the behavior of a real gas as long as the pressure is not too high and as long as the temperature is far from the liquefaction point. But what happens when these two criteria are not satisfied? First we discuss real gas behavior, and then we examine how kinetic theory can help us understand this behavior.

Let us look at a graph of pressure plotted against volume for a given amount of gas. On such a ***PV* diagram**, Fig. 18–4, each point represents the pressure and volume of an equilibrium state of the given substance. The four curves (labeled A, B, C, and D) show how the pressure varies, as the volume is changed at constant temperature, for four different values of constant temperature. The red dashed curve A′ represents the behavior of a gas as predicted by the ideal gas law; that is, PV = constant. The solid curve A represents the behavior of a *real* gas at the same temperature. Notice that at high pressure, the volume of a real gas is less than that predicted by the ideal gas law. The curves B and C in Fig. 18–4 represent the (real) gas at successively lower temperatures, and we see that the behavior deviates

even more from the curves predicted by the ideal gas law (for example, B′), and the difference is greater the closer the gas is to liquefying.

To explain this behavior, note that at higher pressure we expect the molecules to be closer together. And at lower temperatures, the potential energy associated with attractive forces between the molecules (which we ignored before) is no longer negligible compared to their kinetic energy. These attractive forces (of an electrical nature) tend to pull the molecules closer together so the volume of a real gas (solid curves in Fig. 18–4) is less than expected from the ideal gas law (dashed curves A′ and B′ in Fig. 18–4). At still lower temperatures, these forces cause **liquefaction** as the substance becomes a liquid (also called **condensation**), and the molecules become very close together. Section 18–6 discusses in more detail the effect of these attractive molecular forces, as well as the effect of the volume which the molecules themselves occupy.

Curve D represents the situation when liquefaction occurs. At low pressure on curve D (on the lower right in Fig. 18–4), the substance is a gas and occupies a large volume. As the pressure is increased, the volume decreases until point b is reached. From point b to point a, the volume of the substance decreases with no change in pressure, even though we try gently to compress it: the substance is gradually changing from the gas to the liquid phase. At point a, all of the substance has changed to liquid. Further increase in pressure reduces the volume only slightly—liquids are nearly incompressible—so on the left curve D is very steep as seen in Fig. 18–4. The shaded area under the gold dashed line represents the region where the gas and liquid phases exist together in equilibrium.

Curve C in Fig. 18–4 represents the behavior of the substance at its **critical temperature**; the point c (the one point where curve C is horizontal) is called the **critical point**. At temperatures less than the critical temperature, a gas will change to the liquid phase if sufficient pressure is applied. Above the critical temperature (and this is the definition of the term), no amount of pressure can cause a gas to change phase and become a liquid. (Thus curves A and B represent the substance at temperatures where it can only be a gas.) The critical temperatures for various gases are given in Table 18–1. Scientists tried for many years to liquefy oxygen without success, which led to the idea that there must be a critical point. Oxygen can be liquefied only if first cooled below its critical temperature of −118°C.

Often a distinction is made between the terms "gas" and "vapor": a substance below its critical temperature in the gaseous state is called a **vapor**; above the critical temperature, it is called a **gas**.

The behavior of a substance can be diagrammed not only on a *PV* diagram but also on a *PT* diagram. A ***PT* diagram**, often called a **phase diagram**, is particularly convenient for comparing the different phases of a substance. Figure 18–5 is the phase diagram for water. The curve labeled ℓ-v represents those points where the liquid and vapor phases are in equilibrium—it is thus a graph of the boiling point versus pressure. Note that the curve correctly shows that at a pressure of 1 atm the boiling point is 100°C, and that the boiling point is lowered for a decreased pressure. The curve s-ℓ represents points where solid and liquid exist in equilibrium and thus is a graph of the freezing point versus pressure. At 1 atm, the freezing point of water is 0°C, as shown. Notice also in Fig. 18–5 that at a pressure of 1 atm, the substance is in the liquid phase if the temperature is between 0°C and 100°C, but is in the solid or vapor phase if the temperature is below 0°C or above 100°C. The curve labeled s-v is the *sublimation point* versus pressure curve. **Sublimation** refers to the process whereby at low pressures a solid changes directly into the vapor phase without passing through the liquid phase. For water, sublimation occurs if the pressure of the water vapor is less than 0.006 atm. Carbon dioxide, which in the solid phase is called dry ice, sublimates even at atmospheric pressure (Fig. 18–6).

The intersection of the three curves (in Fig. 18–5) is the **triple point**. For water this occurs at $T = 273.16\ \text{K}$ $(= 0.01°\text{C})$ and $P = 6.03 \times 10^{-3}$ atm. It is only at the triple point that the three phases can exist together in equilibrium. Because the triple point corresponds to a unique value of temperature and pressure, it is precisely reproducible and is often used as a point of reference. For example, the standard of temperature is usually specified as exactly 273.16 K at the triple point of water, rather than 273.15 K at the freezing point of water at 1 atm.

⚠ CAUTION

Gases above a certain "critical temperature" cannot be liquefied

TABLE 18–1 Critical Temperatures and Pressures

Substance	Critical Temperature °C	Critical Temperature K	Critical Pressure (atm)
Water	374	647	218
CO_2	31	304	72.8
Oxygen	−118	155	50
Nitrogen	−147	126	33.5
Hydrogen	−239.9	33.3	12.8
Helium	−267.9	5.3	2.3

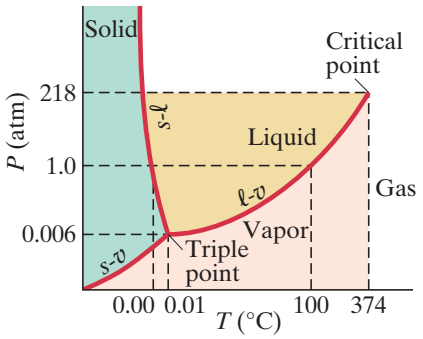


FIGURE 18–5 Phase diagram for water (note that the scales are not linear).

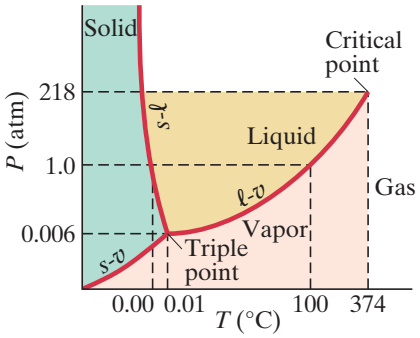
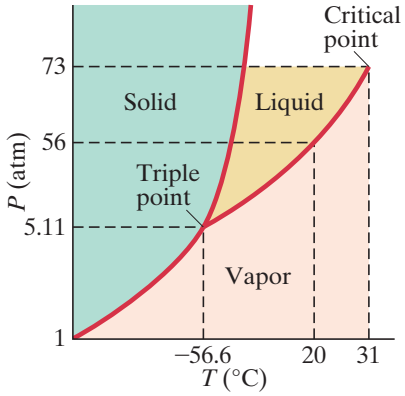


FIGURE 18–5 (Repeated.) Phase diagram for water (note that the scales are not linear).

FIGURE 18–6 Phase diagram for carbon dioxide.



The solid-liquid (s-ℓ) curve for water (Fig. 18–5) slopes upward to the left. This is true only of substances that *expand* upon freezing: at a higher temperature the pressure needs to be lower to allow that expansion. Most substances contract upon freezing and the s-ℓ curve slopes upward to the right, as shown for carbon dioxide (CO_2) in Fig. 18–6.

At temperatures above the critical temperature (374° for water), a substance is in the gas state. However, at temperatures below the critical temperature but at high pressure above the critical pressure (218 atm for water), the substance has some properties like a gas and some like a liquid. The lower the temperature, the more liquid-like it is, but there is no phase transition: this is the area directly above the tan-colored area labeled "liquid" in Figs. 18–5 and 18–6.

The phase transitions we have been discussing are the common ones. Some substances, however, can exist in several forms in the solid phase. A transition from one phase to another occurs at a particular temperature and pressure, just like ordinary phase changes. For example, ice has been observed in at least eight forms at very high pressure. Ordinary helium has two distinct liquid phases, called helium I and II. They exist only at temperatures within a few degrees of absolute zero. Helium II exhibits very unusual properties referred to as **superfluidity**. It has essentially zero viscosity and exhibits strange properties such as climbing up the sides of an open container. Also interesting are **liquid crystals** (used for computer and TV screens, Section 34–8) which can be considered to be in a phase between liquid and solid.

18–4 Vapor Pressure and Humidity

Evaporation

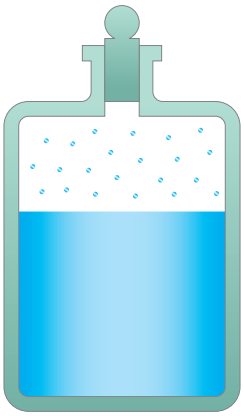
If a glass of water is left out overnight, by morning the water level drops. We say the water evaporated, meaning some of the water has changed to the vapor or gas phase.

This process of **evaporation** can be explained on the basis of kinetic theory. The molecules in a liquid move past one another with a variety of speeds that follow, approximately, the Maxwell distribution. There are strong attractive (electric) forces between these molecules, which keeps them close together in the liquid phase. A molecule near the surface of the liquid may, because of its speed, leave the liquid momentarily. But just as a rock thrown into the air returns to the Earth due to gravity, so the attractive electric forces due to the other molecules can pull the vagabond molecule back to the liquid surface—that is, if its velocity is not too large. A molecule with a high enough velocity, however, will escape the liquid entirely, like a rocket escaping the Earth, and become part of the gas phase. Only those molecules that have kinetic energy above a particular value can escape to the gas phase. We have already seen that kinetic theory predicts that the relative number of molecules with kinetic energy above a particular value (such as E_A in Fig. 18–3) increases with temperature. This is in accord with the well-known observation that the evaporation rate is greater at higher temperatures.

PHYSICS APPLIED
Evaporation cools

Because it is the fastest molecules that escape from the surface, the average speed of those remaining is less. When the average speed is less, the absolute temperature is less. Thus kinetic theory predicts that *evaporation is a cooling process*. You may have noticed this effect when you stepped out of a warm shower and felt cold as the water on your body began to evaporate; or after working up a sweat on a hot day, even a slight breeze makes you feel cool through evaporation. Try wetting your finger and then blow on it.

FIGURE 18–7 Vapor appears above a liquid in a closed container.



Vapor Pressure

Air normally contains water vapor (water in the gas phase), and it comes mainly from evaporation. To look at this process in a little more detail, consider a closed container that is partially filled with water (or another liquid) and from which the air has been removed (Fig. 18–7). The fastest moving molecules quickly evaporate into the empty space above the liquid's surface. As they move about, some of these molecules strike the liquid surface and again become part of the liquid phase: this is called **condensation**. The number of molecules in the vapor increases until the number of molecules returning to the liquid equals the number leaving in the same time interval. Equilibrium then exists, and the space above the liquid surface is said to be *saturated*.

The pressure of the vapor when it is saturated is called the **saturated vapor pressure**.

The saturated vapor pressure of any substance depends on the temperature, and not on the volume of the container. At higher temperatures, more molecules have sufficient kinetic energy to break from the liquid surface into the vapor phase. Hence equilibrium will be reached at a higher saturated vapor pressure. The saturated vapor pressure of water at various temperatures is given in Table 18–2. Notice that even solids—for example, ice—have a measurable vapor pressure.

In everyday situations, evaporation from a liquid takes place into the air above it rather than into a vacuum. This does not materially alter the discussion above relating to Fig. 18–7. Equilibrium will still be reached when there are sufficient molecules in the gas phase that the number reentering the liquid equals the number leaving. The concentration of particular molecules (such as water) in the gas phase is not affected by the presence of air, although collisions with air molecules may lengthen the time needed to reach equilibrium. Thus equilibrium occurs at the same value of the saturated vapor pressure as if air were not there.

If the container is large or is not closed, all the liquid may evaporate before saturation is reached. And if the container is not sealed—as, for example, a room in your house—it is not likely that the air will become saturated with water vapor (unless it is raining outside).

TABLE 18–2 Saturated Vapor Pressure of Water

Temperature (°C)	Vapor Pressure†	
	atm	Pa (= N/m^2)
−50	3.9×10^{-5}	4.0
−10	2.6×10^{-3}	2.60×10^2
0	6.1×10^{-3}	6.11×10^2
5	8.6×10^{-3}	8.72×10^2
10	0.012	1.23×10^3
15	0.017	1.71×10^3
20	0.023	2.33×10^3
25	0.031	3.17×10^3
30	0.042	4.24×10^3
40	0.073	7.37×10^3
50	0.12	1.23×10^4
60	0.20	1.99×10^4
70	0.31	3.12×10^4
80	0.47	4.73×10^4
90	0.70	7.01×10^4
100	1.00	1.01×10^5
120	2.00	1.99×10^5
150	4.70	4.76×10^5

†Other units (see Table 13–2) are 1 bar = 10^5 Pa, 1 mbar = 100 Pa, 1 torr = 1 mm-Hg = 133 Pa.

Boiling

The saturated vapor pressure of a liquid increases with temperature. When the temperature is raised to the point where the saturated vapor pressure at that temperature equals the external pressure, **boiling** occurs (Fig. 18–8). As the boiling point is approached, tiny bubbles tend to form in the liquid, which indicate a change from the liquid to the gas phase. However, if the vapor pressure inside the bubbles is less than the external pressure, the bubbles immediately are crushed. As the temperature is increased, the vapor pressure inside a bubble eventually becomes equal to or exceeds the external pressure. The bubble will then not collapse but can rise to the surface. Boiling has then begun. *A liquid boils when its saturated vapor pressure equals the external pressure*. This occurs for water at a pressure of 1 atm (760 torr) at 100°C, as confirmed in Table 18–2.



FIGURE 18–8 Boiling: bubbles of water vapor float upward from the bottom (where the temperature is highest).

Partial Pressure and Humidity

When we refer to the weather as being dry or humid, we are referring to the water vapor content of the air. In a gas such as air, which is a mixture of several types of gases, the total pressure is the sum of the *partial pressures* of each gas present.† By **partial pressure**, we mean the pressure each gas would exert if it alone were present. The partial pressure of water in the air can be as low as zero and can vary up to a maximum equal to the saturated vapor pressure of water at the given temperature. Thus, at 20°C, the partial pressure of water cannot exceed 0.023 atm (see Table 18–2). The **relative humidity** is defined as the ratio of the partial pressure of water vapor to the saturated vapor pressure at a given temperature. It is usually expressed as a percentage:

$$\text{Relative humidity} = \frac{\text{partial pressure of } H_2O}{\text{saturated vapor pressure of } H_2O} \times 100\%.$$

Thus, when the humidity is close to 100%, the air holds nearly all the water vapor it can.

EXAMPLE 18–6 **Relative humidity.** On a particular hot day, the temperature is 30°C and the partial pressure of water vapor in the air is 0.030 atm. What is the relative humidity?

APPROACH From Table 18–2, we see that the saturated vapor pressure of water at 30°C is 0.042 atm.

SOLUTION The relative humidity is thus

$$\frac{0.030 \text{ atm}}{0.042 \text{ atm}} \times 100\% = 71\%.$$

†For example, 78% (by volume) of air molecules are nitrogen and 21% oxygen, with much smaller amounts of water vapor, argon, carbon dioxide, and other gases. At an air pressure of 1 atm, oxygen exerts a partial pressure of 0.21 atm and nitrogen 0.78 atm.

PHYSICS APPLIED
Humidity and comfort

Humans are sensitive to humidity. A relative humidity of 40–50% is generally optimum for both health and comfort. High humidity, particularly on a hot day, reduces the evaporation of moisture from the skin, which is one of the body's vital mechanisms for regulating body temperature. Very low humidity, on the other hand, can dry the skin and mucous membranes.

PHYSICS APPLIED
Weather

Air is saturated with water vapor when the partial pressure of water in the air is equal to the saturated vapor pressure at that temperature. If the partial pressure of water exceeds the saturated vapor pressure, the air is said to be **supersaturated**. This situation can occur when a temperature decrease occurs. For example, suppose the temperature is 30°C and the partial pressure of water is 0.030 atm, which represents a humidity of 71% as we saw in Example 18–6. Suppose now that the temperature falls to, say, 20°C, as might happen at nightfall. From Table 18–2 we see that the saturated vapor pressure of water at 20°C is 0.023 atm. Hence the relative humidity would be greater than 100%, and the supersaturated air cannot hold this much water vapor. The excess water may condense and appear as dew, clouds, or as fog or rain (Fig. 18–9).

When air containing a given amount of water is cooled, a temperature is reached where the partial pressure of water equals the saturated vapor pressure. This is called the **dew point**. Measurement of the dew point is the most accurate means of determining the relative humidity. One method uses a polished metal surface while in contact with air, which is gradually cooled down. The temperature at which moisture begins to appear on the surface is the dew point, and the partial pressure of water can then be obtained from saturated vapor pressure Tables, such as Table 18–2. If, for example, on a given day the temperature is 20°C and the dew point is 5°C, then the partial pressure of water in the 20°C air is 8.6×10^{-3} atm (5°C in Table 18–2), whereas the saturated vapor pressure of the 20°C air is 0.023 atm; hence the relative humidity is $8.6 \times 10^{-3}/0.023 = 37\%$.

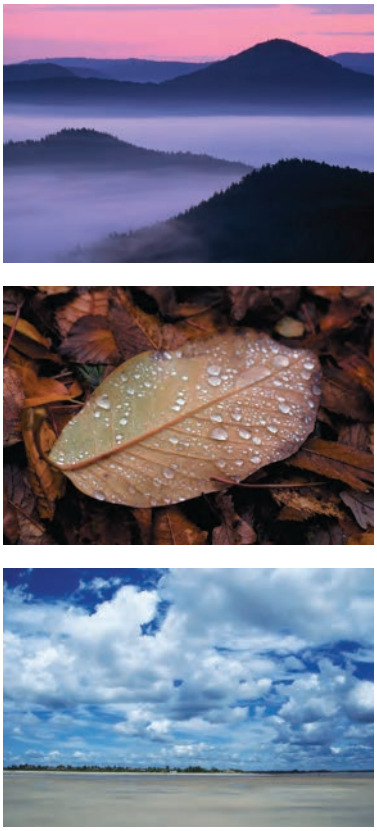


FIGURE 18–9 (a) Fog or mist settling in a valley where the temperature has dropped below the dew point. (b) Dew drops on a leaf. (c) Clouds form on a sunny day at the beach due to air, nearly saturated with water vapor, rising to an altitude where the cooler temperature is at the dew point.

EXERCISE E As the air warms up in the afternoon, how would the relative humidity change if there were no further evaporation? It would (*a*) increase, (*b*) decrease, (*c*) stay the same.

CONCEPTUAL EXAMPLE 18–7 **Dryness in winter.** Why does the air inside heated buildings seem very dry on a cold winter day?

RESPONSE Suppose the relative humidity outside on a −10°C day is 50%. Table 18–2 tells us the partial pressure of water in the air is about 1.3×10^{-3} atm. If this air is brought indoors and heated to +20°C, the relative humidity is $(1.3 \times 10^{-3}\ \text{atm})/(0.023\ \text{atm}) = 5.7\%$. Even if the outside air is fully saturated (partial pressure 2.6×10^{-3} atm at −10°C), the inside relative humidity would still be low at 11%.

18–5 Temperature Decrease of Boiling Water with Altitude

PHYSICS APPLIED
Food cooks slower at altitude

Backpackers, mountain climbers, campers, and others who travel or live at high altitudes are aware that cooking food by boiling takes longer at high elevations than at sea level (or won't cook at all†). Why? Because the temperature of the boiling water is lower.

We can estimate the temperature of boiling water at various altitudes by using Table 18–2 and Eq. 13–6c which gives the pressure P of the atmosphere (on average) as a function of altitude y above sea level:

$$P = P_0 e^{-(\rho_0 g/P_0)y} = P_0 e^{-(1.25 \times 10^{-4}\,\text{m}^{-1})y}.$$

P_0 is $1.00\ \text{atm} = 1.013 \times 10^5\ \text{N/m}^2$, atmospheric pressure at sea level, and in Section 13–3 (see page 370) we calculated $\rho_0 g/P_0 = 1.25 \times 10^{-4}\ \text{m}^{-1}$.

So let's do some calculations of atmospheric pressure at different altitudes.

(1) At sea level, water boils at 100°C = 212°F.

†For example, spaghetti usually cooks fine at 8,000 ft, but thick rigatoni may never quite become tender.

(2) At an altitude of $y = 1200\,\text{m} \approx 4{,}000\,\text{ft}$,

$$P = (1.00\,\text{atm})e^{-(1.25\times10^{-4}\,\text{m}^{-1})(1200\,\text{m})} = (1.00\,\text{atm})e^{-0.15} = 0.86\,\text{atm}.$$

Now we look at Table 18–2 and interpolate between 90°C (0.70 atm) and 100°C (1.00 atm) and write $[(0.86 - 0.70)/(1.00 - 0.70)] \times 10\,\text{C}^\circ = 5.3\,\text{C}^\circ$, so $T \approx 95°\text{C} = 203°\text{F}$ is the boiling point of water at 1200 m above sea level.

(3) At $y = 2400\,\text{m}$ (twice as high), about 8,000 ft elevation, $P = (1.00\,\text{atm})\,e^{-0.30} = 0.74\,\text{atm}$, so T is 91°C = 196°F, and you will notice a difference in cooking times.

(4) At the summit of Mt. Whitney, highest point in the lower 48 states at 4420 m = 14,500 ft, we have $(1.00\,\text{atm})e^{-(1.25\times10^{-4}\,\text{m}^{-1})(4420\,\text{m})} = 0.58\,\text{atm}$. Air pressure is reduced to almost half what it is at sea level. In Table 18–2 we interpolate between 80°C (0.47 atm) and 90°C (0.70 atm) and find $T = 85°\text{C} = 185°\text{F}$.

(5) The highest point on Earth is the summit of Mt. Everest, 8850 m = 29,040 ft above sea level. $P = (1.00\,\text{atm})e^{-(1.25\times10^{-4}\,\text{m}^{-1})(8850\,\text{m})} = 0.33\,\text{atm}$. Only $\frac{1}{3}$ as much air as at sea level—really hard to breathe. In Table 18–2, this corresponds to a temperature of about 71°C = 160°F.

(6) The lowest dry land on Earth is at the Dead Sea, 410 m below sea level, and $y = -410\,\text{m}$. Then $P = (1.00\,\text{atm})e^{+(1.25\times10^{-4}\,\text{m}^{-1})(410\,\text{m})} = 1.05\,\text{atm}$. Harder to interpolate: going up toward 120°C, this is a 5% change so $T \approx 101°\text{C}$.

Our interpolations are approximations because we assume linearity, which may not be true. See Table 18–3 for a summary of these results.

Pressure cookers reduce cooking time because they build up pressure as high as 2 atm, allowing higher cooking (boiling) temperature.

TABLE 18–3 Boiling of Water

Altitude	Boiling Temperature of Water
0 m	100°C = 212°F
1200 m = 4,000 ft	95°C = 203°F
2400 m = 8,000 ft	91°C = 196°F
Mt. Whitney	85°C = 185°F
Mt. Everest	71°C = 160°F
Dead Sea	101°C = 214°F

18–6 Van der Waals Equation of State

In Section 18–3, we discussed how real gases deviate from ideal gas behavior, particularly at high densities or when near condensing to a liquid. We would like to understand these deviations using a microscopic (molecular) point of view. J. D. van der Waals (1837–1923) analyzed this problem and in 1873 arrived at an equation of state which fits real gases more accurately than the ideal gas law. His analysis is based on kinetic theory but takes into account: (1) the finite size of molecules (we previously neglected the actual volume of the molecules themselves, compared to the total volume of the container, and this assumption becomes poorer as the density increases and molecules become closer together); (2) the range of the forces between molecules may be greater than the size of the molecules (we previously assumed that molecular forces act only during collisions, when the molecules are "in contact"). Let us now look at this analysis and derive the van der Waals equation of state.

Assume the molecules in a gas are spherical with radius r. If we assume these molecules behave like hard spheres, then two molecules collide and bounce off one another if the distance between their centers (Fig. 18–10) gets as small as $2r$. Thus the actual volume in which the molecules can move about is somewhat less than the volume V of the container holding the gas. The amount of "unavailable volume" depends on the number of molecules and on their size. Let b represent the "unavailable volume per mole" of gas. Then in the ideal gas law we replace V by $(V - nb)$, where n is the number of moles, and we obtain

$$P(V - nb) = nRT.$$

If we divide through by n, we have

$$P\left(\frac{V}{n} - b\right) = RT. \qquad \textbf{(18–8)}$$

This relation (sometimes called the **Clausius equation of state**) predicts that for a given temperature T and volume V, the pressure P will be greater than for an ideal gas. This makes sense since the reduced "available" volume means the number of collisions with the walls is increased.

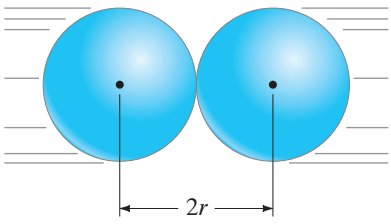


FIGURE 18–10 Molecules, of radius r, colliding.

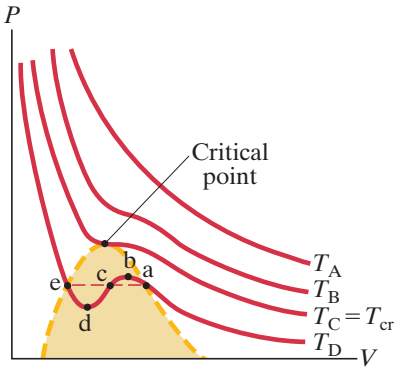


FIGURE 18–11 *PV* diagram for a van der Waals gas, shown for four different temperatures. For T_A, T_B, and T_C (T_C is chosen equal to the critical temperature), the curves fit experimental data quite well for most real gases. The curve labeled T_D, a temperature below the critical point, passes through the liquid–vapor region. The maximum (point b) and minimum (point d) would seem to be artifacts, since we usually see constant pressure, as indicated by the horizontal dashed line (and Fig. 18–4). However, for very pure supersaturated vapors or supercooled liquids, the sections ab and ed, respectively, have been observed. (The section bd would be unstable and has not been observed.)

Next we consider the effects of attractive forces between molecules, which are responsible for holding molecules in the liquid and solid states at lower temperatures. These forces are electrical in nature and although they act even when molecules are not touching, we assume their range is small—that is, they act mainly between nearest neighbors. Molecules at the edge of the gas, headed toward a wall of the container, are slowed down by a net force pulling them back into the gas. Thus these molecules will exert less force and less pressure on the wall than if there were no attractive forces. The reduction in pressure will be proportional to the density of molecules in the layer of gas at the surface, and also to the density in the next layer, which exerts the inward force.[†] Therefore we expect the pressure to be reduced by a factor proportional to the density squared $(n/V)^2$, here written as moles per volume. We solve for the pressure P in Eq. 18–8, and reduce this by an amount $a(n/V)^2$ where a is a proportionality constant. Thus we have

$$P = \frac{RT}{(V/n) - b} - \frac{a}{(V/n)^2}$$

or

$$\left(P + \frac{a}{(V/n)^2}\right)\left(\frac{V}{n} - b\right) = RT, \qquad \textbf{(18–9)}$$

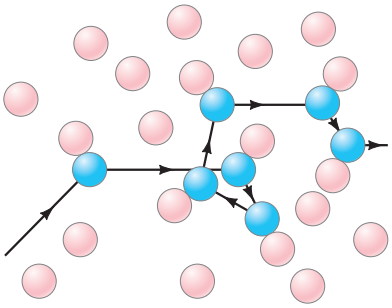
which is the **van der Waals equation of state**.

The constants a and b in the van der Waals equation are different for different gases and are determined by fitting to experimental data for each gas. For CO_2 gas, the best fit is obtained for $a = 0.36\,\text{N}\cdot\text{m}^4/\text{mol}^2$ and $b = 4.3 \times 10^{-5}\,\text{m}^3/\text{mol}$. Figure 18–11 shows a typical *PV* diagram for Eq. 18–9 (a "van der Waals gas") for four different temperatures, with detailed caption, and it should be compared to Fig. 18–4 for real gases.

Neither the van der Waals equation of state nor the many other equations of state that have been proposed are accurate for all gases under all conditions. Yet Eq. 18–9 is a very useful relation. And because it is quite accurate for many situations, its derivation gives us further insight into the nature of gases at the microscopic level. Note that at low densities, $a/(V/n)^2 \ll P$ and $b \ll V/n$, so that the van der Waals equation reduces to the equation of state for an ideal gas, $PV = nRT$.

18–7 Mean Free Path

FIGURE 18–12 Zigzag path of a molecule colliding with other molecules.



If gas molecules were truly point particles, they would have zero cross-section and never collide with one another. If you opened a perfume bottle, you would be able to smell it almost instantaneously across the room, since molecules travel hundreds of meters per second as we saw in Example 18–2. In reality, it takes time before you detect an odor and, according to kinetic theory, this must be due to collisions between molecules of nonzero size.

If we were to follow the path of a particular molecule, we would expect to see it follow a zigzag path as shown in Fig. 18–12. Between each collision the molecule would move in a straight-line path. (Not quite true if we take account of the small intermolecular forces that act between collisions.) An important parameter for a given situation is the **mean free path**, which is defined as the average distance a molecule travels between collisions. We would expect that the greater the gas density, and the larger the molecules, the shorter the mean free path would be. We now determine the nature of this relationship for an ideal gas.

Suppose our gas is made up of molecules which are hard spheres of radius r. A collision will occur whenever the centers of two molecules come within a distance $2r$ of one another. Let us follow a molecule as it traces a straight-line path.

[†]This is similar to the gravitational force in which the force on mass m_1 due to mass m_2 is proportional to the product of their masses (Newton's law of universal gravitation, Chapter 6).

In Fig. 18–13, the dashed line represents the path of our particle if it makes no collisions. Also shown is a cylinder of radius $2r$. If the center of another molecule lies within this cylinder, a collision will occur. (Of course, when a collision occurs the particle's path would change direction, as would our imagined cylinder, but our result won't be altered by unbending a zigzag cylinder into a straight one for purposes of calculation.) Assume our molecule is an average one, moving at the mean speed $\bar{v}$ in the gas. For the moment, let us assume that the other molecules are not moving, and that the concentration of molecules (number per unit volume) is N/V. Then the number of molecules whose centers lie within the cylinder of Fig. 18–13 is N/V times the volume of this cylinder, and this also represents the number of collisions that will occur. In a time Δt, our molecule travels a distance $\bar{v}\,\Delta t$, so the length of the cylinder is $\bar{v}\,\Delta t$ and its volume is $\pi(2r)^2\,\bar{v}\,\Delta t$. Hence the number of collisions that occur in a time Δt is $(N/V)\pi(2r)^2\,\bar{v}\,\Delta t$. We define the **mean free path**, ℓ_M, as the average distance between collisions. This distance is equal to the distance traveled $(\bar{v}\,\Delta t)$ in a time Δt divided by the number of collisions made in time Δt:

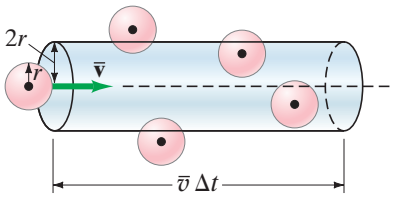


FIGURE 18–13 Molecule at left moves to the right with speed $\bar{v}$. It collides with any molecule whose center is within the cylinder of radius $2r$.

$$\ell_M = \frac{\bar{v}\,\Delta t}{(N/V)\pi(2r)^2\,\bar{v}\,\Delta t} = \frac{1}{4\pi r^2(N/V)}. \qquad \textbf{(18–10a)}$$

Thus we see that ℓ_M is inversely proportional to the cross-sectional area $(=\pi r^2)$ of the molecules and to their concentration (number/volume), N/V. However, Eq. 18–10a is not fully correct since we assumed the other molecules are all at rest. In fact, they are moving, and the number of collisions in a time Δt must depend on the *relative* speed of the colliding molecules, rather than on $\bar{v}$. Hence the number of collisions in time Δt is $(N/V)\pi(2r)^2 v_{rel}\,\Delta t$ (rather than $(N/V)\pi(2r)^2\,\bar{v}\,\Delta t$), where v_{rel} is the average relative speed of colliding molecules. A careful calculation shows that for a Maxwellian distribution of speeds $v_{rel} = \sqrt{2}\bar{v}$. Hence the mean free path is

$$\ell_M = \frac{1}{4\pi\sqrt{2}r^2(N/V)}. \qquad \textbf{(18–10b)}$$

Mean free path

EXAMPLE 18–8 ESTIMATE **Mean free path of air molecules at STP.** Estimate the mean free path of air molecules at STP, standard temperature and pressure (0°C, 1 atm). The diameter of O_2 and N_2 molecules is about 3×10^{-10} m.

APPROACH We saw in Example 17–10 that 1 mol of an ideal gas occupies a volume of $22.4 \times 10^{-3}\ m^3$ at STP. We can thus determine N/V and apply Eq. 18–10b.

SOLUTION

$$\frac{N}{V} = \frac{6.02 \times 10^{23}\ \text{molecules}}{22.4 \times 10^{-3}\ m^3} = 2.69 \times 10^{25}\ \text{molecules/m}^3.$$

Then

$$\ell_M = \frac{1}{4\pi\sqrt{2}\,(1.5 \times 10^{-10}\ m)^2(2.7 \times 10^{25}\ m^{-3})} \approx 9 \times 10^{-8}\ m.$$

NOTE This is about 300 times the diameter of an air molecule.

At very low densities, such as in an evacuated vessel, the concept of mean free path loses meaning since collisions with the container walls may occur more frequently than collisions with other molecules. For example, in a cubical box that is (say) 20 cm on a side containing air at $\approx 10^{-10}$ atm, the mean free path would be about 900 m, which means many more collisions are made with the walls than with other molecules. (Note, nonetheless, that the box contains over 10^{12} molecules.) If the concept of mean free path included also collision with the walls, it would be closer to 0.2 m than to the 900 m calculated from Eq. 18–10b.

18–8 Diffusion

If you carefully place a few drops of food coloring in a glass of water as in Fig. 18–14, you will find that the color spreads throughout the water. The process may take some time in a still liquid (you do not shake the glass), but eventually the color will become uniform. This mixing, known as **diffusion**, is explained by kinetic theory as due to the random movement of the molecules. Diffusion occurs in gases too. Common examples include perfume or smoke (or the odor of something cooking on a stove) diffusing in air, although convection (moving air currents) often plays a greater role in spreading odors than does diffusion. Diffusion depends on concentration, by which we mean the number of molecules or moles per unit volume. In general, *the diffusing substance moves from a region where its concentration is high to a region where its concentration is low.*



(a) (b) (c)

FIGURE 18–14 A few drops of food coloring (a) dropped into water, (b) spreads slowly throughout the water, eventually (c) becoming uniform.

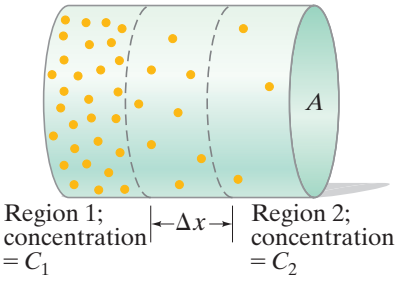


FIGURE 18–15 Diffusion occurs from a region of high concentration to one of lower concentration (only one type of molecule is shown).

Diffusion can be readily understood on the basis of kinetic theory and the random motion of molecules. Consider a tube of cross-sectional area A containing molecules in a higher concentration on the left than on the right, Fig. 18–15. We assume the molecules are in random motion. Yet there will be a net flow of molecules to the right. To see why, let us consider the small section of tube of length Δx as shown. Molecules from both regions 1 and 2 cross into this central section as a result of their random motion. The more molecules there are in a region, the more will strike a given area or cross a boundary. Since there is a greater concentration of molecules in region 1 than in region 2, more molecules cross into the central section from region 1 than from region 2. There is, then, a net flow of molecules from left to right, from high concentration toward low concentration. The net flow becomes zero only when the concentrations become equal.

We might expect that the greater the difference in concentration, the greater the flow rate. Indeed, the rate of diffusion, J (number of molecules or moles or kg per second), is directly proportional to the difference in concentration per unit distance, $(C_1 - C_2)/\Delta x$ (which is called the **concentration gradient**), and to the cross-sectional area A (see Fig. 18–15):

$$J = DA\frac{C_1 - C_2}{\Delta x},$$

or, in terms of derivatives,

$$J = DA\frac{dC}{dx}. \qquad \textbf{(18–11)}$$

D is a constant of proportionality called the **diffusion constant**. Equation 18–11 is known as the **diffusion equation**, or **Fick's law**. If the concentrations are given in mol/m^3, then J is the number of moles passing through a given area A per second. If the concentrations are given in kg/m^3, then J is the mass movement per second (kg/s). The length Δx is given in meters, and area A in m^2. The values of D for a variety of substances diffusing in a particular medium are given in Table 18–4.

TABLE 18–4 Diffusion Constants, D (20°C, 1 atm)

Diffusing Molecules	Medium	D (m^2/s)
H_2	Air	6.3×10^{-5}
O_2	Air	1.8×10^{-5}
O_2	Water	100×10^{-11}
Blood hemoglobin	Water	6.9×10^{-11}
Glycine (an amino acid)	Water	95×10^{-11}
DNA (mass 6×10^6 u)	Water	0.13×10^{-11}

*Applications of Diffusion

EXAMPLE 18–9 **ESTIMATE** **Diffusion of ammonia in air.** To get an idea of the time required for diffusion, estimate how long it might take for ammonia (NH_3) to be detected 10 cm from a bottle after it is opened, assuming only diffusion is occurring.

APPROACH This will be an order-of-magnitude calculation. The rate of diffusion J can be set equal to the number of molecules N diffusing across area A in a time t: $J = N/t$. Then the time $t = N/J$, where J is given by Eq. 18–11. We will have to make some assumptions and rough approximations about concentrations to use Eq. 18–11.

PHYSICS APPLIED
Diffusion time

SOLUTION Using Eq. 18–11, we find

$$t = \frac{N}{J} = \frac{N}{DA}\frac{\Delta x}{\Delta C}.$$

The average concentration (in molecules/m^3), midway between bottle and nose, can be approximated by $\overline{C} \approx N/V$, where V is the volume over which the molecules move and is roughly on the order of $V \approx A\,\Delta x$, where Δx is 10 cm = 0.10 m. We substitute $N = \overline{C}V = \overline{C}A\,\Delta x$ into the above equation:

$$t \approx \frac{(\overline{C}A\,\Delta x)\,\Delta x}{DA\,\Delta C} = \frac{\overline{C}}{\Delta C}\frac{(\Delta x)^2}{D}.$$

The concentration of ammonia is high near the bottle (C) and low near the detecting nose (≈ 0), so $\overline{C} \approx C/2 \approx \Delta C/2$, or $(\overline{C}/\Delta C) \approx \frac{1}{2}$. Since NH_3 molecules have a size somewhere between H_2 and O_2, from Table 18–4 we can estimate $D \approx 4 \times 10^{-5}\,m^2/s$. Then

$$t \approx \tfrac{1}{2}\frac{(0.10\,m)^2}{(4 \times 10^{-5}\,m^2/s)} \approx 100\,s,$$

or about a minute or two.

NOTE This result seems rather long from experience, suggesting that air currents (convection) are more important than diffusion for transmitting odors.

CONCEPTUAL EXAMPLE 18–10 **Colored rings on a paper towel.** A child colors a small spot on a wet paper towel with a brown marking pen. Later, she discovers that instead of a brown spot, there are concentric colored rings around the marked spot. What happened?

RESPONSE The ink in a brown marker is composed of several different inks that mix to make brown. These inks each diffuse at different rates (D is different for each) through the wet paper towel. After a period of time the inks have diffused far enough that the differences in distances traveled is sufficient to see the separation of the different colors. Biochemists and chemists use a similar technique, called **chromatography**, to separate large molecules or small cell components based on their diffusion rates through a medium.

PHYSICS APPLIED
Chromatography

Diffusion is extremely important for living organisms. For example, molecules produced in certain chemical reactions within cells diffuse to other areas where they take part in other reactions.

PHYSICS APPLIED
Diffusion in living organisms

Gas diffusion is important too. Plants require carbon dioxide for photosynthesis. The CO_2 diffuses into leaves from the outside air through tiny openings (stomata). As CO_2 is utilized by the cells, its concentration drops below that in the air outside, and more diffuses inward. Water vapor and oxygen produced by the cells diffuse outward into the air.

Animals also exchange oxygen and CO_2 with the environment. Oxygen required for energy-producing reactions in animals must diffuse into cells. CO_2 is produced as an end product of many metabolic reactions and must diffuse out of cells.

But diffusion is slow over longer distances, so only the smallest organisms in the animal world could survive without having developed complex respiratory and circulatory systems. In humans, oxygen is taken into the lungs, where it diffuses short distances across lung tissue and into the blood. The blood (not diffusion) circulates the oxygen to cells throughout the body. The blood also carries CO_2 produced by the cells back to the lungs, where it diffuses outward.

Summary

According to the **kinetic theory** of gases, which is based on the idea that a gas is made up of molecules that are moving rapidly and randomly, the average translational kinetic energy of the molecules is proportional to the Kelvin temperature T:

$$\overline{K} = \tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}kT \qquad \textbf{(18–4)}$$

where k is Boltzmann's constant.

At any moment, there exists a wide distribution of molecular speeds within a gas. The **Maxwell distribution of speeds** is derived from simple kinetic theory assumptions, and is in good accord with experiment for gases at not too high a pressure.

The behavior of real gases at high pressure, and/or when near their liquefaction point, deviates from the ideal gas law due to the finite size of molecules and to the attractive forces between molecules.

Below the **critical temperature**, a gas can change to a liquid if sufficient pressure is applied; but if the temperature is higher than the critical temperature, no amount of pressure will cause a liquid surface to form.

The **triple point** of a substance is that unique temperature and pressure at which all three phases—solid, liquid, and gas—can coexist in equilibrium. Because of its precise reproducibility, the triple point of water is often taken as a standard reference point.

Evaporation of a liquid is the result of the fastest moving molecules escaping from the surface. Because the average molecular velocity is less after the fastest molecules escape, the temperature decreases when evaporation takes place.

Saturated vapor pressure refers to the pressure of the vapor above a liquid when the two phases are in equilibrium. The vapor pressure of a substance (such as water) depends strongly on temperature and is equal to atmospheric pressure at the boiling point.

Relative humidity of air at a given place is the ratio of the partial pressure of water vapor in the air to the saturated vapor pressure at that temperature; it is usually expressed as a percentage.

The **van der Waals equation of state** takes into account the finite volume of molecules, and the attractive forces between molecules, to better approximate the behavior of real gases.

The **mean free path** is the average distance a molecule moves between collisions with other molecules.

Diffusion is the process whereby molecules of a substance move (on average) from one area to another because of a difference in that substance's concentration.

Questions

1. Why doesn't the size of different molecules enter into the ideal gas law?
2. When a gas is rapidly compressed (say, by pushing down a piston as in Fig. 17–14a), its temperature increases. When a gas expands against a piston, it cools. Explain these changes in temperature using the kinetic theory, in particular noting what happens to the momentum of molecules when they strike the moving piston.
3. In Section 18–1 we assumed the gas molecules made perfectly elastic collisions with the walls of the container. This assumption is not necessary as long as the walls are at the same temperature as the gas. Why?
4. Explain in words how Charles's law follows from kinetic theory and the relation between average kinetic energy and the absolute temperature.
5. Explain in words how Gay-Lussac's law follows from kinetic theory.
6. As you go higher in the Earth's atmosphere, the ratio of N_2 molecules to O_2 molecules increases. Why?
7. Can you determine the temperature of a vacuum?
8. Explain why the peak of the curve for 310 K in Fig. 18–3 is not as high as for 273 K. (Assume the total number of molecules is the same for both.)
9. Is temperature a macroscopic or microscopic variable?
10. Escape velocity from the Earth refers to the minimum speed an object must have to leave the Earth and never return. (*a*) The escape velocity from the Moon is about one-fifth what it is from the Earth due to the Moon's smaller mass. Explain why the Moon has practically no atmosphere. (*b*) If hydrogen was once in the Earth's atmosphere, why would it have probably escaped?
11. If a container of gas is at rest, the average velocity of molecules must be zero. Yet the average speed is not zero. Explain.
12. If the pressure in a gas is doubled while its volume is held constant, by what factor do (*a*) v_{rms} and (*b*) $\bar{v}$ change?
13. What everyday observation would tell you that not all molecules in a material have the same speed?
14. Alcohol evaporates more quickly than water at room temperature. What can you infer about the molecular properties of one relative to the other?
15. Explain why a hot humid day is far more uncomfortable than a hot dry day at the same temperature.
16. Is it possible to boil water at room temperature (20°C) without heating it? Explain.

17. What exactly does it mean when we say that oxygen boils at −183°C?
18. A length of thin wire is placed over a block of ice (or an ice cube) at 0°C. The wire hangs down both sides of the ice, and weights are hung from the ends of the wire. It is found that the wire cuts its way through the ice cube, but leaves solid ice above it. This process is called *regelation*. Explain how this happens by inferring how the freezing point of water depends on pressure.
19. Consider two days when the air temperature is the same but the humidity is different. Which is more dense, the dry air or the humid air at the same T? Explain.
20. (*a*) Why does food cook faster in a pressure cooker? (*b*) Why does pasta or rice need to boil longer at high altitudes? (*c*) Is it harder to boil water at high altitudes?
21. How do a gas and a vapor differ?
22. (*a*) At suitable temperatures and pressures, can ice be melted by applying pressure? (*b*) At suitable temperatures and pressures, can carbon dioxide be melted by applying pressure?
23. Why does dry ice not last long at room temperature?
24. Under what conditions can liquid CO_2 exist? Be specific. Can it exist as a liquid at normal room temperature?
25. Why does exhaled air appear as a little white cloud in the winter (Fig. 18–16)?



FIGURE 18–16
Question 25.

26. Discuss why sound waves can travel in a gas only if their wavelength is somewhat larger than the mean free path.
27. Name several ways to reduce the mean free path in a gas.

MisConceptual Questions

1. An ideal gas is in a sealed rigid container. The average kinetic energy of the gas molecules depends mainly on
   (*a*) the size of the container.
   (*b*) the number of molecules in the container.
   (*c*) the temperature of the gas.
   (*d*) the mass of the molecules.
2. Two ideal gases, A and B, are at the same temperature. If the molecular mass of the molecules in gas A is twice that of the molecules in gas B, the molecules' root-mean-square speed is
   (*a*) the same in both gases.
   (*b*) twice as great in A.
   (*c*) $\sqrt{2}$ times greater in A.
   (*d*) twice as great in B.
   (*e*) $\sqrt{2}$ times greater in B.
3. In a mixture of the gases oxygen and helium in equilibrium, which statement is valid?
   (*a*) The helium atoms will be moving faster than the oxygen molecules, on average.
   (*b*) They will be moving at the same average speed.
   (*c*) The oxygen molecules will be moving faster than the helium atoms, on average.
   (*d*) The kinetic energy of helium atoms will exceed that of oxygen molecules.
   (*e*) None of the above.
4. Which of the following is *not* true about an ideal gas?
   (*a*) The average kinetic energy of the gas molecules increases as the temperature increases.
   (*b*) The volume of an ideal gas increases with temperature if the pressure is held constant.
   (*c*) The pressure of an ideal gas increases with temperature if the volume is held constant.
   (*d*) All gas molecules have the same speed at a particular temperature.
   (*e*) The molecules are assumed to be far apart compared to their size.
5. By what factor must the absolute temperature change to double v_{rms}?
   (*a*) $\sqrt{2}$. (*b*) 2. (*c*) $2\sqrt{2}$. (*d*) 4. (*e*) 16.
6. The rms speed of the molecules of an ideal gas
   (*a*) is the same as the most probable speed of the molecules.
   (*b*) is always equal to $\sqrt{2}$ times the maximum molecular speed.
   (*c*) will increase as the temperature of a gas increases.
   (*d*) All of the above.
7. A sample of an ideal gas is slowly compressed to one-half its original volume with no change in temperature. What happens to the average speed of the molecules in the sample?
   (*a*) It doubles.
   (*b*) It does not change.
   (*c*) It halves.
   (*d*) It quadruples.
   (*e*) None of the above.
8. The air in this room is composed of mostly nitrogen and oxygen. Select all of the true statements below.
   (*a*) The nitrogen and oxygen molecules have the same rms speed.
   (*b*) The rms speed of the oxygen molecules is larger.
   (*c*) The rms speed of the nitrogen molecules is larger.
   (*d*) The average kinetic energies are the same.
   (*e*) The average kinetic energy of oxygen is larger.
   (*f*) The average kinetic energy of nitrogen is larger.
9. About how fast are molecules of nitrogen in air at room temperature moving? As fast as a
   (*a*) tortoise.
   (*b*) sprinter.
   (*c*) car at the highway speed limit.
   (*d*) passenger jet plane.
   (*e*) satellite in orbit around the Earth.
10. An ice cube could go directly from the solid phase to the gas phase (a process called sublimation)
    (*a*) never.
    (*b*) if the temperature and pressure are low enough.
    (*c*) if the temperature and pressure are high enough.
    (*d*) always.

Problems

18–1 Molecular Interpretation of Temperature

1. (I) Calculate the rms speed of helium atoms near the surface of the Sun at a temperature of about 6000 K.

2. (I) By what factor will the rms speed of gas molecules increase if the temperature is increased from 0°C to 160°C?

3. (I) Twelve molecules have the following speeds, given in arbitrary units: 6.0, 2.0, 4.0, 6.0, 0.0, 4.0, 1.0, 8.0, 5.0, 3.0, 7.0, and 8.0. Calculate (*a*) the mean speed, and (*b*) the rms speed.

4. (I) What speed would a 1.5-g paper clip have if it had the same kinetic energy as a molecule at 22°C?

5. (II) A gas is at 30°C. To what temperature must it be raised to triple the rms speed of its molecules?

6. (II) (*a*) What is the average translational kinetic energy of a nitrogen molecule at STP? (*b*) What is the total translational kinetic energy of 1.0 mol of N_2 molecules at 25°C?

7. (II) A 1.0-mol sample of helium gas has a temperature of 18°C. (*a*) What is the total kinetic energy of all the gas atoms in the sample? (*b*) How fast would a 65-kg person have to run to have the same kinetic energy?

8. (II) At what temperature do the atoms of helium gas have the same rms speed as molecules of water vapor at 20.0°C?

9. (II) If the pressure in a gas is tripled while its volume is held constant, by what factor does v_{rms} change?

10. (II) Show that the rms speed of molecules in a gas is given by $v_{\text{rms}} = \sqrt{3P/\rho}$, where P is the pressure in the gas, and ρ is the gas density.

11. (II) Show that for a mixture of two gases at the same temperature, the ratio of their rms speeds is equal to the inverse ratio of the square roots of their molecular masses, $v_1/v_2 = \sqrt{M_2/M_1}$.

12. (II) What is the rms speed of nitrogen molecules contained in a 7.2-m^3 volume at 2.9 atm if the total amount of nitrogen is 2100 mol?

13. (II) (*a*) For an ideal gas at temperature T show that

$$\frac{dv_{\text{rms}}}{dT} = \frac{1}{2}\frac{v_{\text{rms}}}{T},$$

and using the approximation $\Delta v_{\text{rms}} \approx \frac{dv_{\text{rms}}}{dT}\Delta T$, show that

$$\frac{\Delta v_{\text{rms}}}{v_{\text{rms}}} \approx \frac{1}{2}\frac{\Delta T}{T}.$$

(*b*) If the average air temperature changes from −5°C in winter to 25°C in summer, estimate the percent change in the rms speed of air molecules between these seasons.

14. (II) What is the average distance between oxygen molecules at STP?

15. (II) The rms speed of molecules in a gas at 20.0°C is to be increased by 5.0%. To what temperature must it be raised?

16. (II) Can pockets of vacuum persist in an ideal gas? Assume that a room is filled with air at 20°C and that somehow a small spherical region of radius 1 cm within the room becomes devoid of air molecules. Estimate how long it will take for air to refill this region of vacuum. Assume the atomic mass of air is 29 u.

17. (II) Calculate (*a*) the rms speed of an oxygen molecule at 0°C and (*b*) determine how many times per second it would move back and forth across a 6.0-m-long room on average, assuming it made very few collisions with other molecules.

18. (II) A friend calculates the room temperature rms speed of CO_2 molecules to be about 1.7×10^{-11} m/s. Looking back at Example 18–2, you conclude that there must be something wrong. What was the mistake and what is the correct rms speed?

19. (III) Estimate how many air molecules rebound from a wall in a typical room per second, assuming an ideal gas of N molecules contained in a cubic room with sides of length ℓ at temperature T and pressure P. (*a*) Show that the frequency f with which gas molecules strike a wall is

$$f = \frac{\bar{v}_x}{2}\frac{P}{kT}\ell^2$$

where $\bar{v}_x$ is the average x component of the molecule's velocity. (*b*) Show that the equation can then be written as

$$f \approx \frac{P\ell^2}{\sqrt{4mkT}}$$

where m is the mass of a gas molecule. (*c*) Determine f for a cubic air-filled room at sea level, at a temperature 20°C, with sides of length $\ell = 3$ m.

18–2 Distribution of Molecular Speeds

20. (II) A group of 25 particles have the following speeds: two have speed 10 m/s, seven have 15 m/s, four have 20 m/s, three have 25 m/s, six have 30 m/s, one has 35 m/s, and two have 40 m/s. Determine (*a*) the average speed, (*b*) the rms speed, and (*c*) the most probable speed.

21. (II) If you changed the type of molecules in a gas so that their mass is tripled, is it possible to change the temperature to keep the velocity distribution from changing? If so, what do you need to do to the temperature?

22. (II) A gas consisting of 14,500 molecules, each of mass 2.00×10^{-26} kg, has the following distribution of speeds, which crudely mimics the Maxwell distribution:

Number of Molecules	Speed (m/s)
1500	220
3900	440
4600	660
2900	880
1200	1100
400	1320

(*a*) Determine v_{rms} for this distribution of speeds. (*b*) Given your value for v_{rms}, what (effective) temperature would you assign to this gas? (*c*) Determine the mean speed $\bar{v}$ of this distribution and use this value to assign an (effective) temperature to the gas. Is the temperature you find here consistent with the one you determined in part (*b*)? If not, what is the difference between these two temperatures?

23. (III) Starting from the Maxwell distribution of speeds, Eq. 18–6, show (*a*)

$$\int_0^\infty f(v)\,dv = N,$$

which is the total number of molecules, and (*b*)

$$\overline{v^2} = v_{\text{rms}}^2 = \int_0^\infty v^2 f(v)\,dv/N = 3kT/m,$$

so $\frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$ (Eq. 18–4).

18–3 Real Gases; Phase Changes

24. (I) CO_2 exists in what phase when the pressure is 35 atm and the temperature is 35°C (Fig. 18–6)?

25. (I) (*a*) At atmospheric pressure, in what phases can CO_2 exist? (*b*) For what range of pressures and temperatures can CO_2 be a liquid? Refer to Fig. 18–6.

26. (I) Water is in which phase when the pressure is 0.01 atm and the temperature is (*a*) 90°C, (*b*) −20°C?

27. (II) You have a sample of water and are able to control temperature and pressure arbitrarily. (*a*) Using Fig. 18–5, describe the phase changes you would see if you started at a temperature of 85°C, a pressure of 180 atm, and decreased the pressure to 0.004 atm while keeping the temperature fixed. (*b*) Repeat part (*a*) with the temperature at 0.0°C. Assume that you held the system at the starting conditions long enough for the system to stabilize before making further changes.

18–4 Vapor Pressure and Humidity

28. (I) What is the partial pressure of water on a day when the temperature is 25°C and the relative humidity is 75%?

29. (I) What is the partial pressure of water vapor at 30°C if the humidity is 35%?

30. (I) What is the air pressure at a place where water boils at 80°C?

31. (II) If the air pressure at a particular place in the mountains is 0.80 atm, estimate the temperature at which water boils.

32. (II) What is the dew point if the humidity is 55% on a day when the temperature is 25°C?

33. (II) A **pressure cooker** is a sealed pot designed to cook food with the steam produced by boiling water somewhat above 100°C. The pressure cooker in Fig. 18–17 uses a weight of mass *m* to allow steam to escape at a certain pressure through a small hole (diameter *d*) in the cooker's lid. If $d = 3.0$ mm, what should *m* be in order to cook food at 120°C? Assume that atmospheric pressure outside the cooker is 1.01×10^5 Pa.

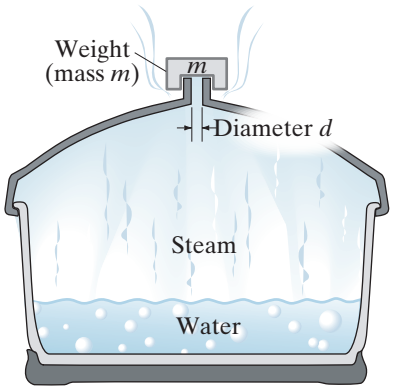


FIGURE 18–17
Problem 33.

34. (II) What is the mass of water vapor in a closed room 5.0 m × 6.0 m × 2.4 m when the temperature is 22°C and the relative humidity is 45%?

35. (II) What is the approximate pressure inside a pressure cooker if the water is boiling at a temperature of 130°C? Assume no air escaped during the heating process, which started at 18°C.

36. (II) If the humidity in a sealed room of volume 420 m^3 at 25°C is 65%, what mass of water can still evaporate from an open pan?

37. (III) When using a mercury barometer (Section 13–6), the vapor pressure of mercury is usually assumed to be zero. At room temperature mercury's vapor pressure is about 0.0015 mm-Hg. At sea level, the height *h* of mercury in a barometer is about 760 mm. (*a*) If the vapor pressure of mercury is neglected, is the true atmospheric pressure greater or less than the value read from the barometer? (*b*) What is the percent error? (*c*) What is the percent error if you use a water barometer and ignore water's saturated vapor pressure at STP?

38. (III) Air that is at its dew point of 5°C is drawn into a building where it is heated to 22°C. What will be the relative humidity at this temperature? Assume constant pressure of 1.0 atm. Take into account the expansion of the air.

18–5 Temperature of Boiling Water

39. (II) (*a*) At what temperature does water boil at 10,000 ft (3000 m) of elevation? (*b*) At what elevation would water boil at 80°C?

40. (II) (*a*) Estimate at what temperature the water boils in Denver, Colorado—called "the mile high city." (*b*) If water boils at 89°C, at what elevation are you?

18–6 Van der Waals Equation of State

41. (II) In the van der Waals equation of state, the constant *b* represents the amount of "unavailable volume" (per mole) occupied by the molecules themselves. Thus *V* is replaced by $(V - nb)$, where *n* is the number of moles. For oxygen, *b* is about 3.2×10^{-5} m^3/mol. Estimate the diameter of an oxygen molecule.

42. (II) For oxygen gas, the van der Waals equation of state achieves its best fit for $a = 0.14\ \text{N}\cdot\text{m}^4/\text{mol}^2$ and $b = 3.2 \times 10^{-5}\ \text{m}^3/\text{mol}$. Determine the pressure in 1.0 mol of the gas at 0°C if its volume is 0.70 L, calculated using (*a*) the van der Waals equation, (*b*) the ideal gas law.

43. (III) A 0.5-mol sample of O_2 gas is in a large cylinder with a movable piston on one end so it can be compressed (as in Fig. 17–14a). The initial volume is large enough that there is not a significant difference between the pressure given by the ideal gas law and that given by the van der Waals equation. As the gas is slowly compressed at constant temperature (use 300 K), at what volume does the van der Waals equation give a pressure that is 5% different than the ideal gas law pressure? Let $a = 0.14\ \text{N}\cdot\text{m}^4/\text{mol}^2$ and $b = 3.2 \times 10^{-5}\ \text{m}^3/\text{mol}$.

44. (III) How well does the ideal gas law describe the pressurized air in a scuba tank? (*a*) To fill a typical scuba tank, an air compressor intakes about 2300 L of air at 1.0 atm and compresses this gas into the tank's 12-L internal volume. If the filling process occurs at 20°C, show that a tank holds about 96 mol of air. (*b*) Assume the tank has 96 mol of air at 20°C. Use the ideal gas law to predict the air's pressure within the tank. (*c*) Use the van der Waals equation of state to predict the air's pressure within the tank. For air, the van der Waals constants are $a = 0.1373\ \text{N}\cdot\text{m}^4/\text{mol}^2$ and $b = 3.72 \times 10^{-5}\ \text{m}^3/\text{mol}$. (*d*) Assuming the van der Waals pressure is the true pressure inside the full tank, show that the ideal gas law predicts a pressure that is in error by only about 3%.

18–7 Mean Free Path

45. (II) Below a certain threshold pressure, the air molecules (0.3-nm diameter) within a research vacuum chamber are in the "collision-free regime," meaning that a particular air molecule is as likely to cross the container and collide with the opposite wall as it is to collide with another air molecule. Estimate the threshold pressure for a vacuum chamber of side 1.0 m at 20°C.

46. (II) At about what pressure would the mean free path of air molecules be (*a*) 0.30 m and (*b*) equal to the diameter of air molecules, $\approx 3 \times 10^{-10}$ m? Assume $T = 20°C$.

47. (II) (*a*) The mean free path of CO_2 molecules at STP is about 5.6×10^{-8} m. Estimate the diameter of a CO_2 molecule. (*b*) Do the same for He gas for which $\ell_M \approx 25 \times 10^{-8}$ m at STP.

48. (II) A very small amount of hydrogen gas is released into the air. If the air is at 1.0 atm and 15°C, estimate the mean free path for an H_2 molecule. What assumptions did you make?

49. (II) We saw in Example 18–8 that the mean free path ℓ_M of air molecules at STP is about 9×10^{-8} m. Estimate the collision frequency f, the number of collisions per unit time.

50. (II) A cubic box 1.40 m on a side is placed under one of the best possible vacuums, so the pressure of air inside is 10^{-13} atm. Estimate how many collisions molecules make with each other for each collision with a wall (0°C).

51. (III) Estimate the maximum allowable pressure in an electronic tube if 98% of all electrons leaving the hot negative electrode must hit the positive electrode 3.5 cm away without first striking an air molecule. Assume electrons are point particles (zero diameter).

52. (III) Sound waves in a gas can only propagate if the gas molecules collide with each other on the time scale of the sound wave's period. Thus the highest possible frequency f_{max} for a sound wave in a gas is approximately equal to $1/t_{coll}$, the inverse of the average collision time between molecules. Assume a gas, composed of molecules with mass m and radius r, is at pressure P and temperature T. (*a*) Show that the *highest sound frequency* in the gas is

$$f_{max} \approx 16Pr^2\sqrt{\frac{\pi}{mkT}}.$$

(*b*) Determine f_{max} for 20°C air at sea level. How many times larger is f_{max} than the highest frequency in the human audio range (20 kHz)?

18–8 Diffusion

***53.** (I) Approximately how long would it take for the ammonia of Example 18–9 to be detected 1.0 m from the bottle after it is opened? What does this suggest about the relative importance of diffusion and convection for carrying odors?

***54.** (II) Estimate the time needed for a glycine molecule (see Table 18–3) to diffuse a distance of 25 μm in water at 20°C if its concentration varies over that distance from 1.00 mol/m^3 to 0.50 mol/m^3. Compare this "speed" to its rms (thermal) speed. The molecular mass of glycine is about 75 u.

***55.** (II) *Oxygen diffuses* from the surface of *insects* to the interior through tiny tubes called tracheae. An average trachea is about 2 mm long and has cross-sectional area of 2×10^{-9} m^2. Assume the concentration of oxygen inside is half what it is outside in the atmosphere. (*a*) Show that the concentration of oxygen in the air (assume 21% is oxygen) at 20°C is about 8.7 mol/m^3. (*b*) Calculate the diffusion rate J. (*c*) Estimate the average time for a molecule to diffuse through a trachea. Assume the diffusion constant is 1×10^{-5} m^2/s.

General Problems

56. A sample of ideal gas must contain at least $N = 10^6$ molecules in order for the Maxwell distribution to be a valid description of the gas, and to assign it a meaningful temperature. For an ideal gas at STP, what is the smallest length scale ℓ (volume $V = \ell^3$) for which a valid temperature can be assigned?

57. In outer space the density of matter is about one atom per cm^3, mainly hydrogen atoms, and the temperature is about 2.7 K. Calculate the rms speed of these hydrogen atoms, and the pressure (in atmospheres).

58. Calculate approximately the total translational kinetic energy of all the molecules in an *E. coli* bacterium of mass 2.0×10^{-15} kg at 37°C. Assume 70% of the cell, by weight, is water, and the other molecules have an average molecular mass on the order of 10^5 u.

59. (*a*) Estimate the rms speed of an amino acid, whose molecular mass is 89 u, in a living cell at 37°C. (*b*) What would be the rms speed of a protein of molecular mass 85,000 u at 37°C?

60. A 0.40-kg trash-can lid is suspended against gravity by tennis balls thrown vertically upward at it. How many tennis balls per second must rebound from the lid elastically, assuming they have a mass of 0.060 kg and are thrown at 15 m/s?

61. The escape speed from the Earth is 1.12×10^4 m/s (Section 8–7). So a gas molecule traveling away from Earth near the outer boundary of the Earth's atmosphere would, at this speed, be able to escape from the Earth's gravitational field and be lost to the atmosphere. At what temperature is the rms speed of (*a*) oxygen molecules, and (*b*) helium atoms equal to 1.12×10^4 m/s? (*c*) Can you explain why our atmosphere contains oxygen but not helium?

62. Assuming a typical nitrogen or oxygen molecule is about 0.3 nm in diameter, what percent of the room you are sitting in is taken up by the volume of the molecules themselves?

63. A *scuba* tank has a volume of 3100 cm^3. For very deep dives, the tank is filled with 50% (by volume) pure oxygen and 50% pure helium. (*a*) How many molecules are there of each type in the tank if at 20°C the gauge pressure reads 14 atm? (*b*) What is the ratio of the average kinetic energies of the two types of molecule? (*c*) What is the ratio of the rms speeds of the two types of molecule?

64. A *sauna* has 7.8 m^3 of air volume, and the temperature is 85°C. The air is perfectly dry. How much water (in kg) should be evaporated if we want to increase the relative humidity from 0% to 10%? (See Table 18–2.)

65. (*a*) The second postulate of kinetic theory is that molecules in a gas are, on average, far apart from one another. That is, their average separation is much greater than the diameter of each molecule. Is this assumption reasonable? To check, calculate the average distance between molecules of a gas at STP, and compare it to the diameter of a typical gas molecule, about 0.3 nm. (*b*) If the molecules had the diameter of ping-pong balls, say 4 cm, how far away would the next ping-pong ball be on average? (*c*) Repeat part *a*, but now assume the gas has been compressed so that the pressure is now 3 atm but still at 273 K. (*d*) Estimate what % of the total volume of gas is taken up by molecules themselves in parts *a* and *c*. [Note that the volume of the molecules themselves can become a significant part of the total volume at lower temperatures and higher pressures. Hence the actual volume the molecules have to bounce around in is less than the total volume. This contributes to the effect shown in Fig. 18–4 at high pressures where real gases (solid red lines) deviate from ideal gas behavior (dashed lines A′ and B′).]

66. Two isotopes of uranium, ^{235}U and ^{238}U (the superscripts refer to their atomic masses), can be separated by a gas diffusion process by combining them with fluorine to make the gaseous compound UF_6. Calculate the ratio of the rms speeds of these molecules for the two isotopes, at constant T. Use Appendix G for masses.

67. Consider a container of oxygen gas at a temperature of 23°C that is 1.50 m tall. Compare the gravitational potential energy of a molecule at the top of the container (assuming the potential energy is zero at the bottom) with the average kinetic energy of the molecules. Is it reasonable to neglect the potential energy?

68. In humid climates, people constantly *dehumidify* their cellars to prevent rot and mildew. If the cellar in a house (kept at 20°C) has 105 m^2 of floor space and a ceiling height of 2.4 m, what is the mass of water that must be removed from it in order to drop the humidity from 95% to a more reasonable 45%?

69. A space vehicle returning from the Moon enters the Earth's atmosphere at a speed of about 42,000 km/h. Molecules (assume nitrogen) striking the nose of the vehicle with this speed correspond to what temperature? (Because of this high temperature, the nose of a space vehicle must be made of special materials; indeed, part of it does vaporize, and this is seen as a bright blaze upon reentry.)

70. Calculate the total water vapor pressure in the air on the following two days: (*a*) a hot summer day, with the temperature 30°C and the relative humidity at 75%; (*b*) a cold winter day, with the temperature 5°C and the relative humidity at 75%.

71. At room temperature, it takes approximately 2.45×10^3 J to evaporate 1.00 g of water. Estimate the average speed of evaporating molecules. What multiple of v_{rms} (at 20°C) for water molecules is this? (Assume Eq. 18–4 holds.)

72. If the humidity is 65% at 30.0°C, what is the dew point? Use linear interpolation to find the temperature of the dew point to the nearest degree.

73. A sample of cesium vapor is in an oven at 400°C. The volume of the oven is 75 cm^3, the vapor pressure of Cs at 400°C is 17 mm-Hg, and the diameter of cesium atoms in the vapor is 0.33 nm. (*a*) Calculate the mean speed of cesium atoms in the vapor. (*b*) Determine the number of collisions a single Cs atom undergoes with other cesium atoms per second. (*c*) Determine the total number of collisions per second between all of the cesium atoms in the vapor. Note that a collision involves two Cs atoms and assume the ideal gas law holds.

74. Using the ideal gas law, find an expression for the mean free path ℓ_M that involves pressure and temperature instead of (N/V). Use this expression to find the mean free path for nitrogen molecules at a pressure of 7.5 atm and 300 K.

75. The density of atoms, mostly hydrogen, in interstellar space is about one per cubic centimeter. Estimate the mean free path of the hydrogen atoms, assuming an atomic diameter of 1.0×10^{-10} m.

76. At 300 K, an 8.50-mol sample of carbon dioxide occupies a volume of 0.220 m^3. Calculate the gas pressure, first by the ideal gas law, and then by using the van der Waals equation of state. (The values for a and b are given in Section 18–6.) In this range of pressure and volume, the van der Waals equation is very accurate. What percent error did you make in assuming ideal-gas-law behavior?

77. (*a*) From the van der Waals equation of state, show that the critical temperature and pressure are given by

$$T_{cr} = \frac{8a}{27bR}, \qquad P_{cr} = \frac{a}{27b^2}.$$

[*Hint*: Use the fact that the P versus V curve has an inflection point at the critical point so that the first and second derivatives are zero.] (*b*) Determine a and b for CO_2 from the measured values of $T_{cr} = 304$ K and $P_{cr} = 72.8$ atm.

78. For oxygen gas the van der Waals constants are $a = 0.14\ N \cdot m^4/mol^2$ and $b = 3.2 \times 10^{-5}\ m^3/mol$. Using these values, graph six curves of pressure vs. volume between $V = 2 \times 10^{-5}\ m^3$ and $2.0 \times 10^{-4}\ m^3$, for 1 mol of oxygen gas at temperatures of 80 K, 100 K, 120 K, 130 K, 150 K, and 170 K. From the graphs determine approximately the critical temperature for oxygen.

ANSWERS TO EXERCISES

A: (*a*).
B: Almost 500 m/s.
C: (*d*).
D: (*c*).
E: (*b*).

When it is cold, warm clothes act as insulators to reduce heat loss from the body to the environment by conduction and convection. Heat radiation from a campfire can warm you and your clothes. The fire can also transfer energy directly by heat convection and conduction to what you are cooking. Heat, like work, represents a transfer of energy. Heat is defined as a transfer of energy due to a difference of temperature. Work is a transfer of energy by mechanical means, not due to a temperature difference. The first law of thermodynamics links the two in a general statement of energy conservation: the heat Q added to a system minus the net work W done by the system equals the change in internal energy ΔE_{int} of the system: $\Delta E_{\text{int}} = Q - W$. Internal energy E_{int} is the sum total of all the energy of the molecules of the system.



CHAPTER 19

Heat and the First Law of Thermodynamics

CONTENTS

19 – 1 Heat as Energy Transfer
19 – 2 Internal Energy
19 – 3 Specific Heat
19 – 4 Calorimetry—Solving Problems
19 – 5 Latent Heat
19 – 6 The First Law of Thermodynamics
19 – 7 Thermodynamic Processes and the First Law
19 – 8 Molar Specific Heats for Gases, and the Equipartition of Energy
19 – 9 Adiabatic Expansion of a Gas
19–10 Heat Transfer: Conduction, Convection, Radiation

CHAPTER-OPENING QUESTION—Guess now!

A 5-kg cube of warm iron (60°C) is put in thermal contact with a 10-kg cube of cold iron (15°C). Which statement is valid:

(a) Heat flows spontaneously from the warm cube to the cold cube until both cubes have the same heat content.
(b) Heat flows spontaneously from the warm cube to the cold cube until both cubes have the same temperature.
(c) Heat can flow spontaneously from the warm cube to the cold cube, but can also flow spontaneously from the cold cube to the warm cube.
(d) Heat flows from the larger cube to the smaller one because the larger one has more internal energy.

When a pot of cold water is placed on a hot burner of a stove, the temperature of the water increases. We say that heat "flows" from the hot burner to the cold water. When two objects at different temperatures are put in contact, heat spontaneously flows from the hotter one to the colder one. The spontaneous flow of heat is in the direction tending to equalize the temperature. If the two objects are kept in contact long enough for their temperatures to become equal, the objects are said to be in thermal equilibrium, and there is no further heat flow between them. For example, when a fever thermometer is first placed in your mouth, heat flows from your mouth to the thermometer.

When the thermometer reaches the same temperature as the inside of your mouth, the thermometer and your mouth are then in equilibrium, and no more heat flows.

Heat and temperature are often confused. They are very different concepts and we will make the clear distinction between them. We start this Chapter by defining and using the concept of heat. We also begin our discussion of *thermodynamics*, which is the name we give to the study of processes in which energy is transferred as heat and as work.

19–1 Heat as Energy Transfer

We use the term "heat" in everyday life as if we knew what we meant. But the term is often used inconsistently, so it is important for us to define heat clearly, and to clarify the phenomena and concepts related to heat.

We commonly speak of the flow of heat—heat flows from a stove burner to a pot of soup, from the Sun to the Earth, from a person's mouth into a fever thermometer. Heat flows spontaneously from an object at higher temperature to one at lower temperature. Indeed, an eighteenth-century model of heat pictured heat flow as movement of a fluid substance called *caloric*. However, the caloric fluid could never be detected. In the nineteenth century, it was found that the various phenomena associated with heat could be described consistently using a new model that views heat as akin to work, as we will discuss in a moment. First we note that a common unit for heat, still in use today, is named after caloric. It is called the **calorie** (cal) and is defined as *the amount of heat necessary to raise the temperature of 1 gram of water by 1 Celsius degree*. [To be precise, the particular temperature range from 14.5°C to 15.5°C is specified because the heat required is very slightly different at different temperatures. The difference is less than 1% over the range 0 to 100°C, and we will ignore it for most purposes.] More often used than the calorie is the **kilocalorie** (kcal), which is 1000 calories. Thus *1 kcal is the heat needed to raise the temperature of 1 kg of water by 1 C°*. Often a kilocalorie is called a **Calorie** (with a capital C), and this Calorie (or the kJ) is used to specify the energy value of food. In the British system of units, heat is measured in British thermal units (Btu). One Btu is defined as the heat needed to raise the temperature of 1 lb of water by 1 F°. It can be shown (Problem 4) that 1 Btu = 0.252 kcal = 1056 J. [Gas companies may use the **therm**, which is 10^5 Btu.]

CAUTION
Heat is not a fluid

The idea that heat is related to energy transfer was pursued by a number of scientists in the 1800s, particularly by an English brewer, James Prescott Joule (1818–1889). Joule and others performed a number of experiments that were crucial for establishing our present-day view that heat, like work, represents a transfer of energy. One of Joule's experiments is shown (simplified) in Fig. 19–1. The falling weight causes the paddle wheel to turn. The friction between the water and the paddle wheel causes the temperature of the water to rise slightly (barely measurable, in fact, by Joule). The same temperature rise could also be obtained by heating the water on a hot stove. In this and many other experiments (some involving electrical energy), Joule determined that a given amount of work done was always equivalent to a particular amount of heat input.

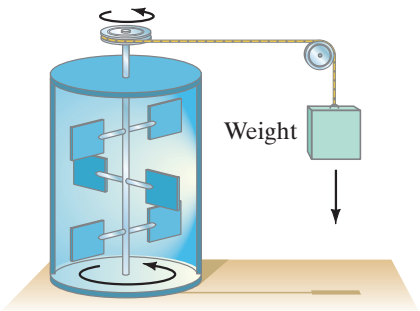


FIGURE 19–1 Joule's experiment on the mechanical equivalent of heat. [The work is transformed into internal energy (Section 19–2).]

Quantitatively, 4.186 joules (J) of work was found to be equivalent to 1 calorie (cal) of heat. This is known as the **mechanical equivalent of heat**:

$$4.186\,\text{J} = 1\,\text{cal};$$
$$4.186\,\text{kJ} = 1\,\text{kcal}.$$

CAUTION
Heat is energy transferred because of a ΔT

As a result of these and other experiments, scientists came to interpret heat not as a substance, and not exactly as a form of energy. Rather, heat refers to a *transfer of energy*: when heat flows from a hot object to a cooler one, it is energy that is being transferred from the hot to the cold object. Thus, **heat** is *energy transferred from one object to another because of a difference in temperature*. In SI units, the unit for heat, as for any form of energy, is the joule. Nonetheless, calories and kcal are still sometimes used. Today the calorie is *defined* in terms of the joule (via the mechanical equivalent of heat, above), rather than in terms of the properties of water, as given previously. The latter is still handy to remember: 1 cal raises 1 g of water by 1 C°, or 1 kcal raises 1 kg of water by 1 C°.

PHYSICS APPLIED
Working off Calories

EXAMPLE 19–1 **ESTIMATE** **Working off the extra Calories.** Suppose you throw caution to the wind and eat 500 Calories of ice cream and cake. To compensate, you want to do an equivalent amount of work climbing stairs or a mountain. How much total height must you climb?

APPROACH The work W you need to do in climbing stairs equals the change in gravitational potential energy: $W = mgh$, where h is the vertical height climbed. For this estimate, approximate your mass as $m \approx 60\,\text{kg}$.

SOLUTION 500 food Calories are 500 kcal, which in joules is

$$(500\,\text{kcal})\left(4.186 \times 10^3\,\text{J/kcal}\right) = 2.1 \times 10^6\,\text{J}.$$

The work done to climb a vertical height h is $W = mgh$. We solve for h:

$$h = \frac{W}{mg} = \frac{2.1 \times 10^6\,\text{J}}{(60\,\text{kg})\left(9.80\,\text{m/s}^2\right)} = 3600\,\text{m}.$$

This is a huge elevation change (over 11,000 ft).

NOTE The human body does not transform food energy with 100% efficiency—it is more like 20% efficient. As we shall discuss in the next Chapter, some energy is always "wasted," so you would actually have to climb only about $(0.2)(3600\,\text{m}) \approx 700\,\text{m}$, which is more reasonable (about 2300 ft of elevation gain).

19–2 Internal Energy

The sum total of all the energy of all the molecules in an object is called its **internal energy**. (Sometimes **thermal energy** is used to mean the same thing.) We introduce the concept of internal energy now since it will help clarify ideas about heat.

Distinguishing Temperature, Heat, and Internal Energy

CAUTION
Distinguish heat from internal energy and from temperature

Using the kinetic theory, we can make a clear distinction between temperature, heat, and internal energy. Temperature (in kelvins) is a measure of the average kinetic energy of individual molecules (Eq. 18–4). Internal energy refers to the *total* energy of all the molecules within the object. (Thus two equal-mass hot ingots of iron may have the same temperature, but two of them have twice as much internal energy as one does.) Heat, finally, refers to a *transfer* of energy from one object to another because of a difference in temperature.

CAUTION
Direction of heat flow depends on temperature (not on amount of internal energy)

Notice that the direction of heat flow between two objects depends on their temperatures, not on how much internal energy each has. Thus, if 50 g of water at 30°C is placed in contact (or mixed) with 200 g of water at 25°C, heat flows *from* the water at 30°C *to* the water at 25°C even though the internal energy of the 25°C water is much greater because there is so much more of it.

Internal Energy of an Ideal Gas

Let us calculate the internal energy of n moles of an ideal monatomic (one atom per molecule) gas. The internal energy, E_{int}, is the sum of the translational kinetic energies of all the atoms.[†] This sum is equal to the average kinetic energy per molecule times the total number of molecules, N:

$$E_{\text{int}} = N\left(\tfrac{1}{2}m\overline{v^2}\right).$$

Using Eq. 18–4, $\overline{K} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$, we can write this as

$$E_{\text{int}} = \tfrac{3}{2}NkT \qquad \textbf{(19–1a)}$$

or (recall Section 17–9)

$$E_{\text{int}} = \tfrac{3}{2}nRT, \qquad \left[\begin{array}{c}\text{internal energy of}\\ \text{ideal monatomic gas}\end{array}\right] \qquad \textbf{(19–1b)}$$

where n is the number of moles. Thus, the internal energy of an ideal gas depends only on the Kelvin temperature and the number of moles (or molecules) of gas.

If the gas molecules contain more than one atom, then the rotational and vibrational energy of the molecules (Fig. 19–2) must also be taken into account. The internal energy will then be greater at a given temperature than for a monatomic gas, but it will still be a function only of temperature for an ideal gas.

The internal energy of real gases also depends mainly on temperature, but where real gases deviate from ideal gas behavior, their internal energy depends also somewhat on pressure and volume (due to atomic potential energy).

The internal energy of liquids and solids is quite complicated, for it includes electrical potential energy associated with the forces (or "chemical" bonds) between atoms and molecules.

EXERCISE A Return to the Chapter-Opening Question, page 538, and answer it again now. Try to explain why you may have answered differently the first time.

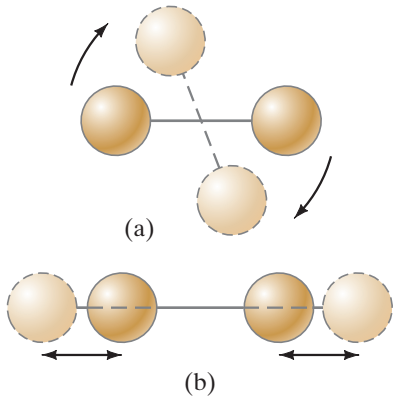


FIGURE 19–2 Besides translational kinetic energy, molecules can have (a) rotational kinetic energy, and (b) vibrational energy (both kinetic and potential).

19–3 Specific Heat

If heat flows into an object, the object's temperature rises (assuming no phase change). But how much does the temperature rise? That depends. As early as the eighteenth century, experimenters had recognized that the amount of heat Q required to change the temperature of a given material is proportional to the mass m of the material present and to the temperature change ΔT. This remarkable simplicity in nature can be expressed in the equation

$$Q = mc\,\Delta T, \qquad \textbf{(19–2)}$$

where c is a quantity characteristic of the material called its **specific heat**. Because $c = Q/(m\,\Delta T)$, specific heat is specified in units[‡] of $\text{J/kg}\cdot\text{C}^\circ$ (the proper SI unit) or $\text{kcal/kg}\cdot\text{C}^\circ$. For water at 15°C and a constant pressure of 1 atm, $c = 4.186\times10^3\,\text{J/kg}\cdot\text{C}^\circ$ or $1.00\,\text{kcal/kg}\cdot\text{C}^\circ$, since, by definition of the cal and the joule, it takes 1 kcal of heat to raise the temperature of 1 kg of water by 1 C°. Table 19–1 gives the values of specific heat for other substances at 20°C. The values of c depend to some extent on temperature (as well as slightly on pressure), but for temperature changes that are not too great, c can often be considered constant.[§] [Gases are complicated and are treated in Section 19–8.] Values for c are the same using kelvins: $1\,\text{J/kg}\cdot\text{C}^\circ = 1\,\text{J/kg}\cdot\text{K}$.

TABLE 19–1 Specific Heats
(at 1 atm constant pressure and 20°C unless otherwise stated)

Substance	Specific Heat, c	
	$\frac{\text{J}}{\text{kg}\cdot\text{C}^\circ}$	$\frac{\text{kcal}}{\text{kg}\cdot\text{C}^\circ}$ (= cal/g·C°)
Aluminum	900	0.22
Alcohol (ethyl)	2400	0.58
Copper	390	0.093
Glass	840	0.20
Iron or steel	450	0.11
Lead	130	0.031
Marble	860	0.21
Mercury	140	0.033
Silver	230	0.056
Wood	1700	0.4
Water		
Ice (−5°C)	2100	0.50
Liquid (15°C)	4186	1.00
Steam (110°C)	2010	0.48
Human body (average)	3470	0.83
Protein	1700	0.4

[†]The symbol U is used in some books for internal energy. The use of E_{int} avoids confusion with U which stands for potential energy (Chapter 8).

[‡]Note that $\text{J/kg}\cdot\text{C}^\circ$ means $\dfrac{\text{J}}{\text{kg}\cdot\text{C}^\circ}$ and *not* $(\text{J/kg})\cdot\text{C}^\circ = \text{J}\cdot\text{C}^\circ/\text{kg}$ (otherwise we would have written it that way).

[§]To take into account the dependence of c on T, we can write Eq. 19–2 in differential form: $dQ = mc(T)\,dT$, where $c(T)$ means c is a function of temperature T. Then the heat Q required to change the temperature from T_1 to T_2 is

$$Q = \int_{T_1}^{T_2} mc(T)\,dT.$$

EXAMPLE 19–2 **How heat transferred depends on specific heat.** (*a*) How much heat input is needed to raise the temperature of an empty 20-kg vat made of iron from 10°C to 90°C? (*b*) What if the vat is filled with 20 kg of water?

APPROACH We apply Eq. 19–2 to the different materials involved.

SOLUTION (*a*) Our system is the iron vat alone. From Table 19–1, the specific heat of iron is 450 J/kg · C°. The change in temperature is (90°C − 10°C) = 80 C°. Thus,

$$Q_{\text{vat}} = mc\,\Delta T = (20\,\text{kg})(450\,\text{J/kg}\cdot\text{C}^\circ)(80\,\text{C}^\circ) = 7.2\times10^5\,\text{J} = 720\,\text{kJ}.$$

(*b*) Our system is the vat plus the water. The water alone would require

$$Q_{\text{w}} = mc\,\Delta T = (20\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}^\circ)(80\,\text{C}^\circ) = 6.7\times10^6\,\text{J} = 6700\,\text{kJ},$$

or almost 10 times what an equal mass of iron requires. The total, for the vat plus the water, is 720 kJ + 6700 kJ = 7400 kJ.

NOTE In (*b*), the iron vat and the water underwent the same temperature change, $\Delta T = 80\,\text{C}^\circ$, but their specific heats are different.

PHYSICS APPLIED
Practical effects of water's high specific heat

If the iron vat in part (*a*) of Example 19–2 had been *cooled* from 90°C to 10°C, 720 kJ of heat would have flowed *out* of the iron. In other words, Eq. 19–2 is valid for heat flow either in or out, with a corresponding increase or decrease in temperature. We saw in part (*b*) that water requires almost 10 times as much heat as an equal mass of iron to make the same temperature change. Water has one of the highest specific heats of all substances, which makes it an ideal substance for hot-water space-heating systems and other uses that require a minimal drop in temperature for a given amount of heat transfer. It is the water content, too, that causes the apples rather than the crust in hot apple pie to burn our tongues through heat transfer.

CONCEPTUAL EXAMPLE 19–3 **A very hot frying pan.** You accidentally make a bad mistake by letting an empty iron frying pan get very hot on the stove (200°C or even more). What happens when you dunk it into a few inches of cool water in the bottom of the sink? Will the final temperature be midway between the initial temperatures of the water and pan? Will the water start boiling? Assume the mass of water is roughly the same as the mass of the frying pan.

RESPONSE Experience may tell you that the water warms up—perhaps by as much as 10 or 20 degrees. The water doesn't come close to boiling. The water's temperature increase is a lot less than the frying pan's temperature decrease. Why? Because the mass of water is roughly equal to that of the pan, and iron has a specific heat nearly 10 times smaller than that of water (Table 19–1). As heat leaves the frying pan and enters the water, the iron pan's temperature change will be about 10 times greater than that of the water. If, instead, you let a few drops of water fall onto the hot pan, that very small mass of water will sizzle and boil away (the pan's mass may be hundreds of times larger than that of the water drops).

19–4 Calorimetry—Solving Problems

CAUTION
Types of SYSTEM

In discussing heat and thermodynamics, we often consider a particular **system**, which is any object or set of objects we choose to consider. Everything else in the universe is its "environment" (or the "surroundings"). There are several categories of systems. A **closed system** is one for which no mass enters or leaves (but energy may be exchanged with the environment). In an **open system**, mass may enter or leave (as may energy). Many (idealized) systems we study in physics are closed systems. But many systems, including plants and animals, are open systems since they exchange materials (food, oxygen, waste products) with the environment. A closed system is said to be **isolated** if no energy in any form (as well as no mass) passes across its boundaries; otherwise it is not isolated.

A perfectly isolated system is an ideal, but we often try to set up a system that can be closely approximated as an isolated system (preferably one we can deal with fairly easily). When different parts of an isolated system are at different temperatures, heat will flow (energy is transferred) from the part at higher temperature to the part at lower temperature—that is, within the system—until **thermal equilibrium** is reached, meaning the entire system is at the same temperature. For an isolated system, no energy is transferred into or out of it. So we can apply *conservation of energy* to such an isolated system. A simple intuitive way to set up a conservation of energy equation is to write that the heat lost by one part of the system is equal to the heat gained by the other part:

$$\text{heat lost} = \text{heat gained} \qquad \text{[isolated system]}$$

or

$$\text{energy } \textit{out} \text{ of one part} = \text{energy } \textit{into} \text{ another part.} \qquad \text{[isolated system]}$$

EXAMPLE 19–4 **The cup cools the tea.** If 200 cm^3 of tea at 95°C is poured into a 150-g glass cup initially at 25°C (Fig. 19–3), what will be the common final temperature T of the tea and cup when equilibrium is reached, assuming no heat flows to the surroundings?

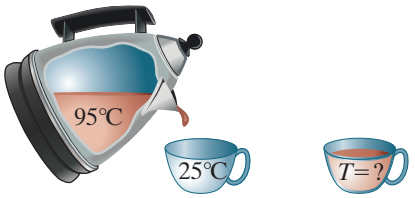


FIGURE 19–3 Example 19–4.

APPROACH We apply conservation of energy to our system of tea plus cup, which we are assuming is isolated: all of the heat that leaves the tea flows into the cup. We can use the specific heat equation, Eq. 19–2, to determine how the heat flow is related to the temperature changes.

SOLUTION Because tea is mainly water, we can take its specific heat as 4186 $\text{J/kg} \cdot \text{C}^\circ$ (Table 19–1), and its mass m is its density times its volume $(V = 200\ \text{cm}^3 = 200 \times 10^{-6}\ \text{m}^3)$: $m = \rho V = (1.0 \times 10^3\ \text{kg/m}^3)(200 \times 10^{-6}\ \text{m}^3) = 0.20\ \text{kg}$. We use Eq. 19–2, apply conservation of energy, and let T be the as yet unknown final temperature:

$$\begin{aligned} \text{heat lost by tea} &= \text{heat gained by cup} \\ m_{\text{tea}} c_{\text{tea}} (95^\circ\text{C} - T) &= m_{\text{cup}} c_{\text{cup}} (T - 25^\circ\text{C}). \end{aligned}$$

Putting in numbers and using Table 19–1 ($c_{\text{cup}} = 840\ \text{J/kg} \cdot \text{C}^\circ$ for glass), we solve for T, and find

$$\begin{aligned} (0.20\ \text{kg})(4186\ \text{J/kg} \cdot \text{C}^\circ)(95^\circ\text{C} - T) &= (0.15\ \text{kg})(840\ \text{J/kg} \cdot \text{C}^\circ)(T - 25^\circ\text{C}) \\ 79{,}500\ \text{J} - (837\ \text{J/C}^\circ)T &= (126\ \text{J/C}^\circ)T - 3150\ \text{J} \\ T &= 86^\circ\text{C}. \end{aligned}$$

The tea drops in temperature by 9 C° by coming into equilibrium with the cup.

NOTE The cup increases in temperature by $86^\circ\text{C} - 25^\circ\text{C} = 61\ \text{C}^\circ$. Its much greater change in temperature (compared with that of the tea water) is mainly due to its much smaller specific heat compared to that of water.

NOTE In this calculation, the ΔT (of Eq. 19–2, $Q = mc\,\Delta T$) is a positive quantity on both sides of our conservation of energy equation. On the left is "heat lost" and ΔT is the initial minus the final temperature $(95^\circ\text{C} - T)$, whereas on the right is "heat gained" and ΔT is the final minus the initial temperature.

CAUTION
When using heat lost = heat gained, ΔT is positive on both sides

Another way to set up the energy conservation equation for heat transfer within an isolated system, which is perhaps more general, is to write that the sum of all internal heat transfers within the system adds up to zero:

$$\Sigma Q = 0. \qquad \text{[isolated system]} \quad \textbf{(19–3)}$$

Each Q represents the heat entering or leaving one part of the system. Each term is written as $Q = mc(T_{\text{f}} - T_{\text{i}})$, and $\Delta T = T_{\text{f}} - T_{\text{i}}$ is always the final minus the initial temperature. ΔT can be either positive or negative, depending on whether heat flows into or out of that part. Let us redo Example 19–4 using $\Sigma Q = 0$.

PROBLEM SOLVING
Alternate approach: $\Sigma Q = 0$

EXAMPLE 19–4′ Alternate Solution, $\Sigma Q = 0$, for "cup cools tea."

APPROACH We use Eq. 19–3, $\Sigma Q = 0$.

SOLUTION We redo Example 19–4, using $m_{\text{cup}} = 0.15\,\text{kg}$, $m_{\text{tea}} = 0.20\,\text{kg}$, $c_{\text{cup}} = 840\,\text{J/kg}\cdot\text{C}^\circ$ and $c_{\text{tea}} = 4186\,\text{J/kg}\cdot\text{C}^\circ$. For each term in Eq. 19–3, $\Delta T = T_{\text{f}} - T_{\text{i}}$. If $T_{\text{f}} < T_{\text{i}}$, then $\Delta T < 0$. Equation 19–3, $\Sigma Q = 0$, becomes

$$m_{\text{cup}}\, c_{\text{cup}}(T - 25°\text{C}) + m_{\text{tea}}\, c_{\text{tea}}(T - 95°\text{C}) = 0.$$

The second term is negative because T will be less than 95°C. Solving the algebra gives the same result, $T = 86°\text{C}$.

You are free to use either approach. They are entirely equivalent, algebraically. For example, if you move the first term in the displayed equation of the alternate Example 19–4′ over to the other side of the equals sign, you obtain the "heat lost = heat gained" equation in the first version of Example 19–4.

The exchange of energy, as exemplified in Example 19–4, is the basis for a technique known as **calorimetry**, which is the quantitative measurement of heat exchange. To make such measurements, a **calorimeter** is used; a simple water calorimeter is shown in Fig. 19–4. It is very important that the calorimeter be well insulated so that almost no heat is exchanged with the surroundings. One important use of the calorimeter is in the determination of specific heats of substances. In the technique known as the "method of mixtures," a sample of a substance is heated to a high temperature, which is accurately measured, and then quickly placed in the cool water of the calorimeter. The heat lost by the sample will be gained by the water and the calorimeter cup. By measuring the final temperature of the mixture, the specific heat can be calculated, as illustrated in the following Example.

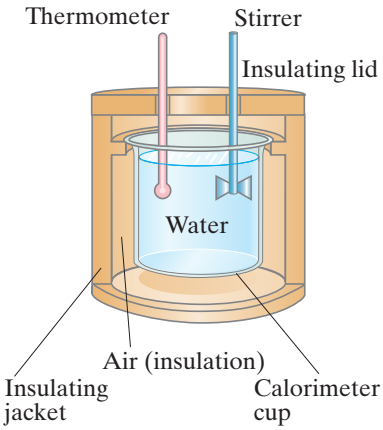


FIGURE 19–4 Simple water calorimeter. In Example 19–5, we assume the temperature of the insulating jacket doesn't change significantly because of the insulating air space between it and the calorimeter cup.

EXAMPLE 19–5 Unknown specific heat determined by calorimetry. An engineer wishes to determine the specific heat of a new metal alloy. A 0.150-kg sample of the alloy is heated to 540°C. It is then quickly placed in 0.400 kg of water at 10.0°C, which is contained in a 0.200-kg aluminum calorimeter cup. The final temperature of the system is 30.5°C. Calculate the specific heat of the alloy.

APPROACH We apply conservation of energy to our system, which we take to be the alloy sample, the water, and the calorimeter cup. We assume this system is isolated, and apply Eq. 19–3, $\Sigma Q = 0$.

SOLUTION Each term is of the form $Q = mc(T_{\text{f}} - T_{\text{i}})$. Thus $\Sigma Q = 0$ gives

$$m_{\text{a}}\, c_{\text{a}}\, \Delta T_{\text{a}} + m_{\text{w}}\, c_{\text{w}}\, \Delta T_{\text{w}} + m_{\text{cal}}\, c_{\text{cal}}\, \Delta T_{\text{cal}} = 0$$

where the subscripts a, w, and cal refer to the alloy, water, and calorimeter, respectively, and each ΔT is the final temperature (30.5°C) minus the initial temperature for each object. When we put in values and use Table 19–1 ($c_{\text{cal}} = 900\,\text{J/kg}\cdot\text{C}^\circ$ for aluminum), this equation becomes

$$(0.150\,\text{kg})(c_{\text{a}})(30.5°\text{C} - 540°\text{C}) + (0.400\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}^\circ)(30.5°\text{C} - 10.0°\text{C}) + (0.200\,\text{kg})(900\,\text{J/kg}\cdot\text{C}^\circ)(30.5°\text{C} - 10.0°\text{C}) = 0,$$

or

$$-(76.4\,\text{kg}\cdot\text{C}^\circ)c_{\text{a}} + 34{,}300\,\text{J} + 3690\,\text{J} = 0.$$

Solving for c_{a} we obtain:

$$c_{\text{a}} = 497\,\text{J/kg}\cdot\text{C}^\circ \approx 500\,\text{J/kg}\cdot\text{C}^\circ.$$

NOTE We rounded off because we have ignored any heat transferred to the thermometer and the stirrer (which is used to quicken the heat transfer process and thus reduce heat loss to the outside). To take them into account we would have to add (small) additional terms to the equation.

In all Examples and Problems of this sort, be sure to include *all* objects that gain or lose heat (within reason). For simplicity, we have ignored very small masses, such as the thermometer and the stirrer, which will affect the energy balance only very slightly.

PROBLEM SOLVING
Be sure to consider all possible sources of energy transfer

Bomb Calorimeter

A **bomb calorimeter** is used to measure the thermal energy released when a substance burns (including foods) to determine their Calorie content. A carefully weighed sample of the substance, with an excess amount of oxygen, is placed in a sealed container (the "bomb"). The bomb is placed in the water of the calorimeter and a fine wire passing into the bomb is then heated to ignite the mixture. The Calorie content of foods determined in this way can be unreliable because our human bodies may not metabolize all the available energy. Careful measurements and calculations need to take this into account.

PHYSICS APPLIED
Measuring Calorie content

19–5 Latent Heat

When a material changes phase from solid to liquid, or from liquid to gas (see also Section 18–3), a certain amount of energy is involved in this **change of phase**. For example, let us trace what happens when a 1.0-kg block of ice at $-40°C$ is heated at a slow steady rate until all the ice has changed to water, then the (liquid) water is heated to 100°C and changed to steam, and heated further above 100°C, all at 1 atm pressure. As shown in the graph of Fig. 19–5, as the ice is heated starting at $-40°C$, its temperature rises at a rate of about 2 C°/kcal of heat added (since for ice, $c \approx 0.50\,\text{kcal/kg}\cdot\text{C}°$). However, when 0°C is reached, the temperature stops increasing even though heat is still being added. The ice gradually changes to water in the liquid state, with no change in temperature. After about 40 kcal has been added at 0°C, half the ice remains and half has changed to water. After about 80 kcal, or 330 kJ, has been added, all the ice has changed to water, still at 0°C. Continued addition of heat causes the water's temperature to again increase, now at a rate of 1 C°/kcal ($c_{\text{water}} = 1.00\,\text{kcal/kg}\cdot\text{C}°$). When 100°C is reached, the temperature again remains constant as the heat added changes the liquid water to vapor (steam). About 540 kcal (2260 kJ) is required to change the 1.0 kg of water completely to steam, after which the graph rises again, indicating that the temperature of the steam rises as heat is added.

The heat required to change 1.0 kg of a substance from the solid to the liquid state is called the **heat of fusion**; it is denoted by L_F. The heat of fusion of water is 79.7 kcal/kg or, in proper SI units, 333 kJ/kg $(= 3.33 \times 10^5\,\text{J/kg})$. The heat required to change a substance from the liquid to the vapor phase is called the **heat of vaporization**, L_V. For water it is 539 kcal/kg or 2260 kJ/kg. Other substances follow graphs similar to Fig. 19–5, although the melting-point and boiling-point temperatures are different, as are the specific heats and heats of fusion and vaporization.

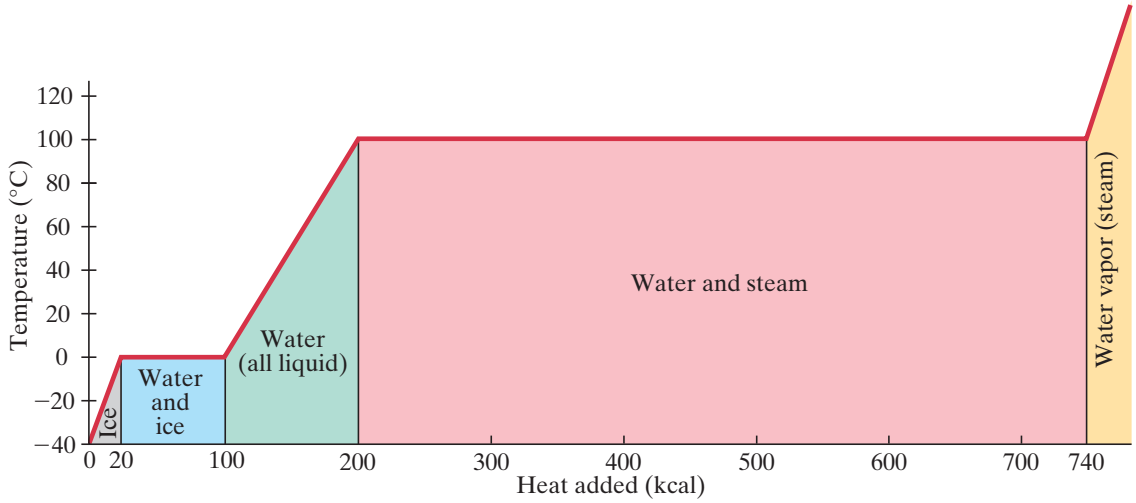


FIGURE 19–5 Temperature as a function of the heat added to bring 1.0 kg of ice at −40°C to steam above 100°C.

TABLE 19–2 Latent Heats (at 1 atm)

Substance	Melting Point (°C)	Heat of Fusion kJ/kg	Heat of Fusion kcal/kg†	Boiling Point (°C)	Heat of Vaporization kJ/kg	Heat of Vaporization kcal/kg†
Oxygen	−218.8	14	3.3	−183	210	51
Nitrogen	−210.0	26	6.1	−195.8	200	48
Ethyl alcohol	−114	104	25	78	850	204
Ammonia	−77.8	332	80	−33.4	1369	330
Water	0	333	79.7	100	2260	539
Lead	327	25	5.9	1750	870	208
Silver	961	88	21	2193	2300	558
Iron	1538	289	69.1	3023	6340	1520
Tungsten	3410	184	44	5900	4800	1150

†Numerical values in kcal/kg are the same in cal/g.

Values for the heats of fusion and vaporization, which are also called the **latent heats**, are given in Table 19–2 for a number of substances.

The heats of vaporization and fusion also refer to the amount of heat *released* by a substance when it changes from a gas to a liquid, or from a liquid to a solid. Thus, steam releases 2260 kJ/kg when it changes to water, and water releases 333 kJ/kg when it becomes ice. [In these cases of heat release, $Q < 0$ when using the $\Sigma Q = 0$ approach, Eq. 19–3.]

The heat involved in a change of phase depends not only on the latent heat but also on the total mass of the substance. That is,

$$Q = mL, \tag{19–4}$$

where L is the latent heat of the particular process and substance, m is the mass of the substance, and Q is the heat added or released during the phase change. For example, when 5.00 kg of water freezes at 0°C, $(5.00\,\text{kg})(3.33 \times 10^5\,\text{J/kg}) = 1.67 \times 10^6\,\text{J}$ of energy is released.

EXERCISE B A pot of water is boiling on a gas stove, and then you turn up the heat. What happens? (*a*) The temperature of the water starts increasing. (*b*) There is no change in the rate of water loss by evaporation. (*c*) The rate of water loss by evaporation increases. (*d*) There is an appreciable increase in both the rate of boiling and the temperature of the water. (*e*) None of these.

Calorimetry sometimes involves a change of state, as the following Examples show. Indeed, latent heats are often measured using calorimetry.

EXAMPLE 19–6 Making ice. How much energy does a freezer have to remove from 1.5 kg of water at 20°C to make ice at −12°C?

APPROACH We need to calculate the total energy removed by adding the heat outflow (1) to reduce the water temperature from 20°C to 0°C, (2) to change the liquid water to solid ice at 0°C, and (3) to lower the ice temperature from 0°C to −12°C.

SOLUTION The heat Q that needs to be removed from the 1.5 kg of water is

$$\begin{aligned} Q &= mc_w(20°\text{C} - 0°\text{C}) + mL_F + mc_{ice}[0° - (-12°\text{C})] \\ &= (1.5\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}°)(20\,\text{C}°) + (1.5\,\text{kg})(3.33 \times 10^5\,\text{J/kg}) \\ &\quad + (1.5\,\text{kg})(2100\,\text{J/kg}\cdot\text{C}°)(12\,\text{C}°) \\ &= 6.6 \times 10^5\,\text{J} = 660\,\text{kJ}. \end{aligned}$$

EXERCISE C Which process in Example 19–6 required the greatest heat loss?

EXAMPLE 19–7 **ESTIMATE** **Will all the ice melt?** At a party, a 0.50-kg chunk of ice at $-10°C$ is placed in 3.0 kg of "iced" tea at 20°C. At what temperature and in what phase will the final mixture be? The tea can be considered as water. Ignore any heat flow to the surroundings, including the container.

PROBLEM SOLVING
First determine (or estimate) the final state

APPROACH Before we can write down an equation applying conservation of energy, we must first check to see if the final state will be all ice, a mixture of ice and water at 0°C, or all water. To bring the 3.0 kg of water at 20°C down to 0°C would require an energy release of

$$m_w c_w (20°C - 0°C) = (3.0\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}°)(20\,\text{C}°) = 250{,}000\,\text{J}.$$

On the other hand, to raise the ice from $-10°C$ to 0°C would require

$$m_{ice} c_{ice}[0°C - (-10°C)] = (0.50\,\text{kg})(2100\,\text{J/kg}\cdot\text{C}°)(10\,\text{C}°) = 10{,}500\,\text{J},$$

and to change the ice to water at 0°C would require

$$m_{ice} L_F = (0.50\,\text{kg})(333\,\text{kJ/kg}) = 167{,}000\,\text{J}.$$

The sum of the last two quantities is $10.5\,\text{kJ} + 167\,\text{kJ} = 177\,\text{kJ}$. This is not enough energy to bring the 3.0 kg of water at 20°C down to 0°C, so we see that the mixture must end up all water, somewhere between 0°C and 20°C.

PROBLEM SOLVING
Then determine the final temperature

SOLUTION To determine the final temperature T, we apply conservation of energy. We present both of the techniques discussed in Section 19–4.

Method 1: "$\Sigma Q = 0$" gives

$$\begin{pmatrix}\text{heat to raise}\\ \text{0.50 kg of ice}\\ \text{from } -10°C\\ \text{to } 0°C\end{pmatrix} + \begin{pmatrix}\text{heat to change}\\ \text{0.50 kg}\\ \text{of ice}\\ \text{to water}\end{pmatrix} + \begin{pmatrix}\text{heat to raise}\\ \text{0.50 kg of water}\\ \text{from } 0°C\\ \text{to } T\end{pmatrix} + \begin{pmatrix}\text{heat lost by}\\ \text{3.0 kg of}\\ \text{water cooling}\\ \text{from } 20°C \text{ to } T\end{pmatrix} = 0.$$

Using some of the results from the "Approach" above, we obtain

$$10{,}500\,\text{J} + 167{,}000\,\text{J} + (0.50\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}°)(T - 0°C) + (3.0\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}°)(T - 20°C) = 0.$$

Solving for T we obtain

$$T = 5.0°C.$$

Method 2: "heat gained = heat lost" produces a word equation like the one above (the last plus sign becomes an equals sign and we lose the "$= 0$"). That is,

$$10{,}500\,\text{J} + 167{,}000\,\text{J} + (0.50\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}°)(T - 0°C) = (3.0\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}°)(20°C - T).$$

The term on the right is heat lost by 3.0 kg of water cooling from 20°C to T; here $\Delta T = T_i - T_f = 20°C - T$ for this approach. Algebraically, this equation is identical to the one above in the first method.

EXERCISE D How much more ice at $-10°C$ would be needed in Example 19–7 to bring the tea down to 0°C, while just melting all the ice?

PROBLEM SOLVING

Calorimetry

1. Be sure you have sufficient information to apply energy conservation. Define what your system is, and ask yourself: **is the system isolated** (or nearly so, enough to get a good estimate)? Do we know or can we calculate all significant sources of energy transfer?
2. Apply **conservation of energy**. Either write
   $$\text{heat gained} = \text{heat lost},$$
   or use
   $$\Sigma Q = 0.$$
3. If **no phase changes** occur, each term in the energy conservation equation will have the form
   $$Q = mc\,\Delta T.$$
4. If **phase changes** do or might occur, there may be terms in the energy conservation equation of the form $Q = mL$, where L is the latent heat. But *before* applying energy conservation, determine (or estimate) in which phase the final state will be, as we did in Example 19–7 by calculating the different contributing values for heat Q.
5. Note that when the system reaches thermal **equilibrium**, the final **temperature** of each substance will have the *same* value. There is only one T_f.
6. **Solve** your energy equation for the unknown.

EXAMPLE 19–8 Determining a latent heat. The specific heat of liquid mercury is $140\,\text{J/kg}\cdot\text{C}^\circ$. When 1.0 kg of solid mercury at its melting point of -39°C is placed in a 0.50-kg aluminum calorimeter filled with 1.2 kg of water at 20.0°C, all the mercury melts and the final temperature of the combination is found to be 16.5°C. What is the heat of fusion of mercury in J/kg?

APPROACH We follow the Problem Solving Strategy we just read, previous page.

SOLUTION

1. **Is the system isolated**? The mercury is placed in a calorimeter, which we assume is well insulated. Our isolated system is the calorimeter, the water, and the mercury.
2. **Conservation of energy**. The heat gained by the mercury = the heat lost by the water and calorimeter.
3. and **4. Phase changes**. There is a phase change (of mercury), plus we use specific heat equations. The heat gained by the mercury (Hg) includes a term representing the melting of the Hg,

$$Q(\text{melt solid Hg}) = m_{\text{Hg}} L_{\text{Hg}},$$

plus a term representing the heating of the liquid Hg from -39°C to $+16.5^\circ\text{C}$:

$$\begin{aligned} Q(\text{heat liquid Hg}) &= m_{\text{Hg}} c_{\text{Hg}}[16.5^\circ\text{C} - (-39^\circ\text{C})] \\ &= (1.0\,\text{kg})(140\,\text{J/kg}\cdot\text{C}^\circ)(55.5\,\text{C}^\circ) = 7770\,\text{J}. \end{aligned}$$

All of this heat gained by the mercury is obtained from the water and calorimeter (aluminum, $c_{\text{cal}} = 900\,\text{J/kg}\cdot\text{C}^\circ$), which cool down:

$$\begin{aligned} Q_{\text{cal}} + Q_{\text{w}} &= m_{\text{cal}} c_{\text{cal}}(20.0^\circ\text{C} - 16.5^\circ\text{C}) + m_{\text{w}} c_{\text{w}}(20.0^\circ\text{C} - 16.5^\circ\text{C}) \\ &= (0.50\,\text{kg})(900\,\text{J/kg}\cdot\text{C}^\circ)(3.5\,\text{C}^\circ) + (1.2\,\text{kg})(4186\,\text{J/kg}\cdot\text{C}^\circ)(3.5\,\text{C}^\circ) \\ &= 19{,}200\,\text{J}. \end{aligned}$$

5. **Energy equation**. The conservation of energy tells us the heat lost by the water and calorimeter cup must equal the heat gained by the mercury:

$$Q_{\text{cal}} + Q_{\text{w}} = Q(\text{melt solid Hg}) + Q(\text{heat liquid Hg})$$

or

$$19{,}200\,\text{J} = m_{\text{Hg}} L_{\text{Hg}} + 7770\,\text{J}.$$

6. **Equilibrium temperature**. It is given as 16.5°C, and we already used it.
7. **Solve**. The only unknown in our energy equation (point 5) is L_{Hg}, the latent heat of fusion (or melting) of mercury. We solve for it, putting in $m_{\text{Hg}} = 1.0\,\text{kg}$:

$$L_{\text{Hg}} = \frac{19{,}200\,\text{J} - 7770\,\text{J}}{1.0\,\text{kg}} = 11{,}400\,\text{J/kg} \approx 11\,\text{kJ/kg},$$

where we rounded off to 2 significant figures.

Evaporation

The latent heat to change a liquid to a gas is needed not only at the boiling point. Water can change from the liquid to the gas phase even at room temperature. This process is called **evaporation** (see also Section 18–4). The value of the heat of vaporization of water increases slightly with a decrease in temperature: at 20°C, for example, it is 2450 kJ/kg (585 kcal/kg) compared to 2260 kJ/kg (= 539 kcal/kg) at 100°C. When water evaporates, the remaining liquid cools, because the energy required (the latent heat of vaporization) comes from the water itself; so its internal energy, and therefore its temperature, must drop.[†]

PHYSICS APPLIED
Evaporation cools

[†]Also from kinetic theory, evaporation is a cooling process because it is the fastest-moving molecules that escape from the surface (Section 18–4). Hence the average speed of the remaining molecules is less, so by Eq. 18–4 the temperature is less.

Evaporation of water from the skin is one of the most important ways the body uses to control its temperature. When the temperature of the blood rises slightly above normal, the hypothalamus region of the brain detects this temperature increase and sends a signal to the sweat glands to increase their production. The energy (latent heat) required to vaporize this water comes from the body, so the body cools.

PHYSICS APPLIED
Body temperature and evaporation

Kinetic Theory of Latent Heats

We can make use of kinetic theory to see why energy is needed to melt or vaporize a substance. At the melting point, the latent heat of fusion does not act to increase the average kinetic energy (and the temperature) of the molecules in the solid, but instead is used to overcome the potential energy associated with the forces between the molecules. That is, work must be done against these attractive forces to break the molecules loose from their relatively fixed positions in the solid so they can freely roll over one another in the liquid phase. Similarly, energy is required for molecules held close together in the liquid phase to escape into the gaseous phase where they are far apart. The vaporization process is a more energetic reorganization of the molecules than is melting (the average distance between the molecules is greatly increased), and hence the heat of vaporization is generally much greater than the heat of fusion for a given substance.

19–6 The First Law of Thermodynamics

Thermodynamics is the study of processes in which energy is transferred as heat and work. In this Chapter we have discussed internal energy and heat, and now we bring in work.

In Chapter 8 we saw that work is done when energy is transferred from one object to another by mechanical means. In Section 19–1 we saw that heat is a transfer of energy from one object to a second one at a lower temperature. Thus, heat is much like work. To distinguish them, *heat* is defined as a *transfer of energy due to a difference in temperature*, whereas work is a transfer of energy that is not due to a temperature difference.

In Section 19–2, we defined the internal energy of a system as the sum total of all the energy of the molecules within the system. Then the internal energy of a system should increase if work is done on the system, or if heat is added to it. Similarly the internal energy should decrease if heat flows out of the system or if work is done by the system on something in the surroundings.

Thus it is reasonable to extend conservation of energy and propose an important law: the change in internal energy of a closed system, ΔE_{int}, will be equal to the energy added to the system by heating minus the work done by the system on the surroundings. In equation form we write

$$\Delta E_{\text{int}} = Q - W \qquad \textbf{(19–5)}$$

FIRST LAW OF THERMODYNAMICS

where Q is the net heat *added* to the system and W is the net work done *by* the system.[†] We must be careful and consistent in following the **sign conventions** for Q and W: Because W in Eq. 19–5 is the work done *by* the system, then if work is done *on* the system, W will be negative and E_{int} will increase. Similarly, Q is positive for heat added to the system, so if heat leaves the system, Q is negative.

CAUTION
Sign Conventions
Heat added is +
Heat lost is −
Work on *system is −*
Work by *system is +*

Equation 19–5 is known as the **first law of thermodynamics**. It is one of the great laws of physics, and its validity rests on experiments (such as Joule's) to which no exceptions have been seen. Since Q and W represent energy transferred into or out of the system, the internal energy changes accordingly. Thus, the first law of thermodynamics is a general statement of the **law of conservation of energy**.

Note that the conservation of energy law was not formulated until the 1800s, because it depended on the interpretation of heat as a transfer of energy.

[†]This convention relates historically to steam engines: the interest was in the heat *input* and the work *output*, both regarded as positive. *CAUTION*: You may see the first law of thermodynamics written in other books as $\Delta E_{\text{int}} = Q + W$, in which case W refers to the work done *on* the system.

Equation 19–5, $\Delta E_{\text{int}} = Q - W$, applies to a closed system. It also applies to an open system (page 542) if we take into account the change in internal energy due to the increase or decrease in the amount of matter. For an isolated system (page 542), no work is done and no heat enters or leaves the system, so $W = Q = 0$, and hence $\Delta E_{\text{int}} = 0$.

CAUTION

P, V, T, E_{int}, m, n are state variables. W and Q are not: a system does not have *an amount of heat or work*

A given system at any moment is in a particular state and can be said to have a certain amount of internal energy, E_{int}. But a system does not "have" a certain amount of heat or work. Rather, when work is done on a system (such as compressing a gas), or when heat is added or removed from a system, the state of the system *changes*. Thus, work and heat are involved in *thermodynamic processes* that can change the system from one state to another; they are not characteristic of the state itself. Quantities which describe the state of a system, such as internal energy E_{int}, pressure P, volume V, temperature T, and mass m or number of moles n, are referred to as **state variables**. Q and W are *not* state variables.

Because E_{int} is a *state variable*, which depends only on the state of the system and not on how the system arrived in that state, we can write

$$\Delta E_{\text{int}} = E_{\text{int},2} - E_{\text{int},1} = Q - W$$

where $E_{\text{int},1}$ and $E_{\text{int},2}$ represent the internal energy of the system in states 1 and 2, and Q and W are the heat added to the system and work done by the system in going from state 1 to state 2.

It is sometimes useful to write the first law of thermodynamics in differential form:

$$dE_{\text{int}} = đQ - đW.$$

Here, dE_{int} represents an infinitesimal change in internal energy when an infinitesimal amount of heat $đQ$ is added to the system, and the system does an infinitesimal amount of work $đW$ The horizontal bars on the $đ$ (for differential) are a reminder that Q and W can *not* be written as general mathematical functions of state variables such as P, V, T, n, because Q and W are *not* properties of the system. That is, $đQ$ and $đW$ mean only infinitesimal amounts of heat and work. Sometimes we don't bother to write the bars. (Internal energy E_{int} *is* a property of the system and *is* a function of state variables.)

EXAMPLE 19–9 **Using the first law.** 2500 J of heat is added to a system, and 1800 J of work is done on the system. What is the change in internal energy of the system?

APPROACH We apply the first law of thermodynamics, Eq. 19–5, to our system.

SOLUTION The heat added to the system is $Q = 2500\,\text{J}$. The work W done *by* the system is $-1800\,\text{J}$. Why the minus sign? Because 1800 J done *on* the system (as given) equals $-1800\,\text{J}$ done *by* the system, and it is the latter we need for the sign conventions we used in Eq. 19–5. Hence

$$\Delta E_{\text{int}} = 2500\,\text{J} - (-1800\,\text{J}) = 2500\,\text{J} + 1800\,\text{J} = 4300\,\text{J}.$$

NOTE We did this calculation in detail to emphasize the importance of keeping careful track of signs. Both the heat and the work are inputs to the system, so we expect ΔE_{int} to be increased by both.

EXERCISE E What would be the internal energy change in Example 19–9 if 2500 J of heat is added to the system and 1800 J of work is done *by* the system (i.e., as output)?

The First Law of Thermodynamics Extended

To write the first law of thermodynamics in a more complete form, consider a system that is moving so it has kinetic energy K, and suppose there is also potential energy U. Then the first law of thermodynamics would have to include these terms and would be written as

$$\Delta K + \Delta U + \Delta E_{\text{int}} = Q - W. \qquad \textbf{(19–6)}$$

This is a full summary of energy conservation.

EXAMPLE 19–10 **Kinetic energy transformed to thermal energy.** A 3.0-g bullet traveling at a speed of 400 m/s enters a tree and exits the other side with a speed of 200 m/s. Where did the bullet's lost kinetic energy go, and how much energy was transferred?

APPROACH Take the bullet and tree as our system. No potential energy is involved. No work is done on (or by) the system by outside forces, nor is any heat added because no energy was transferred to or from the system due to a temperature difference. Thus the kinetic energy gets transformed into internal energy of the bullet and tree. This answers the first question.

PROBLEM SOLVING
Always define your system

SOLUTION In the first law of thermodynamics as written in Eq. 19–6, we are given $Q = W = \Delta U = 0$, so we have

$$\Delta K + \Delta E_{\text{int}} = 0$$

or, using subscripts i and f for initial and final velocities

$$\Delta E_{\text{int}} = -\Delta K = -(K_{\text{f}} - K_{\text{i}}) = \tfrac{1}{2}m(v_{\text{i}}^2 - v_{\text{f}}^2)$$
$$= \tfrac{1}{2}(3.0 \times 10^{-3}\,\text{kg})\left[(400\,\text{m/s})^2 - (200\,\text{m/s})^2\right] = 180\,\text{J}.$$

NOTE The internal energy of the bullet and tree both increase, as both experience a rise in temperature. If we had chosen the bullet alone as our system, work would be done on it and heat transfer would occur.

19–7 Thermodynamic Processes and the First Law

Let us analyze some thermodynamic processes in light of the first law of thermodynamics.

Isothermal Processes ($\Delta T = 0$)

To begin, we choose a very simple system: a fixed mass of a monatomic ideal gas enclosed in a container fitted with a movable piston as shown in Fig. 19–6.

First we consider an idealized process, such as adding heat or doing work, that is carried out at constant temperature. Such a process is called an **isothermal** process (from the Greek meaning "same temperature"). If the system is an ideal gas, then $PV = nRT$ (Eq. 17–3), so for a fixed amount of gas kept at constant temperature, $PV =$ constant. Thus a graph of pressure P vs. volume V, a ***PV* diagram**, would follow a curve like AB in Fig. 19–7 for an isothermal process. Each point on the curve, such as point A, represents the state of the system at a given moment—that is, its pressure P and volume V. At a lower temperature, another isothermal process would be represented by a curve like A′B′ in Fig. 19–7 (the product $PV = nRT =$ constant is less when T is less). The curves shown in Fig. 19–7 are referred to as *isotherms*. We assume the gas at any instant is in contact with a **heat reservoir** (a body whose mass is so large that, ideally, its temperature does not change significantly when heat is exchanged with our system). We also assume that a process of compression (volume decrease) or expansion (volume increase) is done **quasistatically** (= almost statically), meaning very slowly, so that the process can be considered a *series of equilibrium states* all at the same constant temperature. Thus a PV diagram is an idealized approximation to a real process. We cannot plot P vs. V for a real process precisely, because a real process is not quasistatic, which would take forever. (For each change in P or V, the pressure P at any instant will not be the same at all points within the gas.)

If the gas is initially in a state represented by point A in Fig. 19–7, and an amount of heat Q is added to the system, the pressure and volume will change and the state of the system will be represented by another point, B, on the diagram.

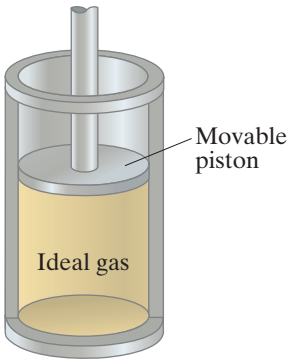
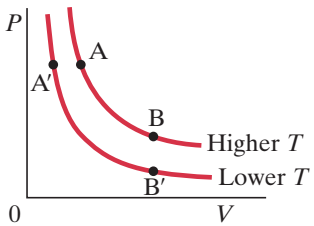


FIGURE 19–6 An ideal gas in a cylinder fitted with a movable piston.

FIGURE 19–7 *PV* diagram for ideal gas undergoing isothermal quasistatic processes at two different temperatures.



If the temperature is to remain constant, the gas will expand and do an amount of work W on the environment (it exerts a force on the piston in Fig. 19–6 and moves it through a distance). The temperature and mass are kept constant so, from Eq. 19–1, the internal energy does not change: $\Delta E_{\text{int}} = \frac{3}{2}nR\,\Delta T = 0$. Hence, by the first law of thermodynamics, Eq. 19–5, $\Delta E_{\text{int}} = Q - W = 0$, so $W = Q$: the work done by the gas in an isothermal process equals the heat added to the gas.

Adiabatic Processes ($Q = 0$)

An **adiabatic** process is one in which no heat is allowed to flow into or out of the system: $Q = 0$. This situation can occur if the system is extremely well insulated, or the process happens so quickly that heat—which flows slowly—has no time to flow in or out. The very rapid expansion of gases in an internal combustion engine is one example of a process that is very nearly adiabatic. Such a rapid expansion could not be graphed on a PV diagram because the system would be tumultuous and turbulent, without a fixed pressure P throughout the gas at any moment, and thus not a sequence of (quasistatic) equilibrium states needed for such a plot. If an ideal gas undergoes an adiabatic expansion very slowly, it can be represented by a curve like that labeled AC in Fig. 19–8. Since $Q = 0$, we have from Eq. 19–5 that $\Delta E_{\text{int}} = -W$. When a gas expands, it does work and W is positive, so the internal energy decreases; hence the temperature decreases as well (because $\Delta E_{\text{int}} = \frac{3}{2}nR\,\Delta T$). This is seen in Fig. 19–8 where the product $PV\ (= nRT)$ is less at point C than at point B. (Compare to curve AB for an isothermal process, in which $\Delta E_{\text{int}} = 0$ and $\Delta T = 0$.) In the reverse operation, an adiabatic compression (going from C to A, for example), work is done *on* the gas, and hence the internal energy increases and the temperature rises.

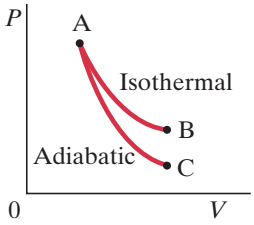


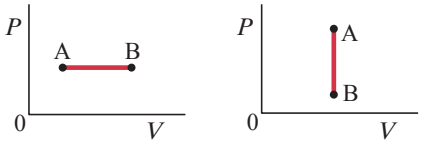
FIGURE 19–8 PV diagram for adiabatic (AC) and isothermal (AB) processes on an ideal gas.

In a diesel engine, the fuel–air mixture is rapidly compressed adiabatically by a factor of 15 or more; the temperature rise is so great that the mixture ignites spontaneously, without spark plugs.

Isobaric and Isovolumetric Processes

Isothermal and adiabatic processes are just two possible processes that can occur. Two other simple thermodynamic processes are illustrated on the PV diagrams of Fig. 19–9: (a) an **isobaric** process is one in which the pressure is kept constant, so the process is represented by a straight horizontal line on the PV diagram (Fig. 19–9a); (b) an **isovolumetric** (or *isochoric*) process is one in which the volume does not change (Fig. 19–9b). In these, and in all other processes, the first law of thermodynamics holds.

FIGURE 19–9 (a) Isobaric ("same pressure") process. (b) Isovolumetric ("same volume") process.



(a) Isobaric (b) Isovolumetric

Work Done in Volume Changes

We often want to calculate the work done in a process. Suppose we have a gas confined to a cylindrical container fitted with a movable piston (Fig. 19–10). We must always be careful to define exactly what our system is. In this case we choose our system to be the gas; so the container's walls and the piston are parts of the environment. Now let us calculate the work done by the gas when it expands quasistatically, so that P and T are defined for the system at all instants. The gas expands against the piston, whose area is A. The gas exerts a force $F = PA$ on the piston, where P is the pressure in the gas. The work done by the gas to move the piston an infinitesimal displacement $d\vec{\boldsymbol{\ell}}$ is

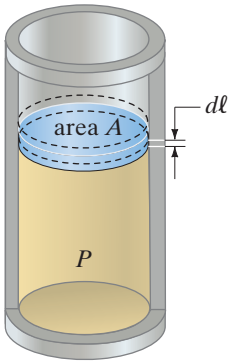
$$dW = \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}} = PA\,d\ell = P\,dV \qquad \textbf{(19–7)}$$

since the infinitesimal increase in volume is $dV = A\,d\ell$. If the gas was *compressed* so that $d\vec{\boldsymbol{\ell}}$ pointed into the gas, the volume would decrease and $dV < 0$. The work done by the gas in this case would then be negative, which is equivalent to saying that positive work was done *on* the gas, not by it. For a finite change in volume from V_{A} to V_{B}, the work W done by the gas will be

$$W = \int dW = \int_{V_{\text{A}}}^{V_{\text{B}}} P\,dV. \qquad \textbf{(19–8)}$$

Equations 19–7 and 19–8 are valid for the work done in any volume change—by a gas, a liquid, or a solid—as long as it is done quasistatically.

FIGURE 19–10 The work done by a gas when its volume increases by $dV = A\,d\ell$ is $dW = P\,dV$.



In order to integrate Eq. 19–8, we need to know how the pressure varies during the process, and this depends on the type of process. Let us first consider a quasistatic isothermal expansion of an ideal gas. This idealized process is represented by the curve between points A and B on the PV diagram of Fig. 19–11. The work done by the gas in this process, according to Eq. 19–8, is just the area between the PV curve and the V axis, and is shown shaded in Fig. 19–11. We can do the integral in Eq. 19–8 for an ideal gas by using the ideal gas law, $P = nRT/V$. The work done at constant T is

$$W = \int_{V_A}^{V_B} P\,dV = nRT\int_{V_A}^{V_B}\frac{dV}{V} = nRT\ln\frac{V_B}{V_A}. \quad \left[\begin{array}{c}\text{isothermal process;}\\ \text{ideal gas}\end{array}\right] \quad \textbf{(19–9)}$$

Let us next consider a different way of taking an ideal gas between the same states A and B. This time, let us lower the pressure in the gas from P_A to P_B, as indicated by the line AD in Fig. 19–12. (In this *isovolumetric* process, heat must be allowed to flow out of the gas so its temperature drops.) Then let the gas expand from V_A to V_B at constant pressure $(= P_B)$, which is indicated by the line DB in Fig. 19–12. (In this *isobaric* process, heat is added to the gas to raise its temperature.) No work is done in the isovolumetric process AD, since $dV = 0$:

$$W = 0. \qquad \text{[isovolumetric process]}$$

In the isobaric process DB the pressure remains constant, so

$$W = \int_{V_A}^{V_B} P\,dV = P_B(V_B - V_A) = P\,\Delta V. \quad \text{[isobaric process]} \quad \textbf{(19–10a)}$$

The work done is again represented on the PV diagram by the area between the curve (ADB) and the V axis, as indicated by the shading in Fig. 19–12. Using the ideal gas law, we can also write

$$W = P_B(V_B - V_A) = nRT_B\left(1 - \frac{V_A}{V_B}\right). \quad \left[\begin{array}{c}\text{isobaric process;}\\ \text{ideal gas}\end{array}\right] \quad \textbf{(19–10b)}$$

As can be seen from the shaded areas in Figs. 19–11 and 19–12, or by putting in numbers in Eqs. 19–9 and 19–10 (try it for $V_B = 2V_A$), the work done in these two processes is different. This is a general result. *The work done in taking a system from one state to another depends not only on the initial and final states but also on the type of process* (*or "path"*).

This result reemphasizes the fact that work cannot be considered a property of a system. The same is true of heat. The heat input required to change the gas from state A to state B depends on the process; for the isothermal process of Fig. 19–11, the heat input turns out to be greater than for the process ADB of Fig. 19–12. In general, *the amount of heat added or removed in taking a system from one state to another depends not only on the initial and final states but also on the path or process*.

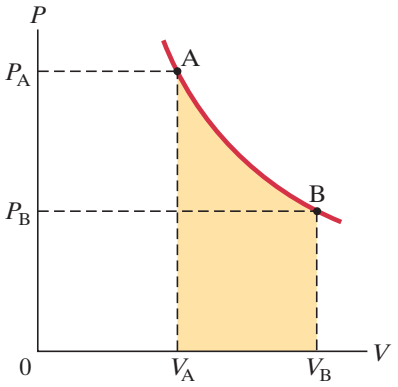
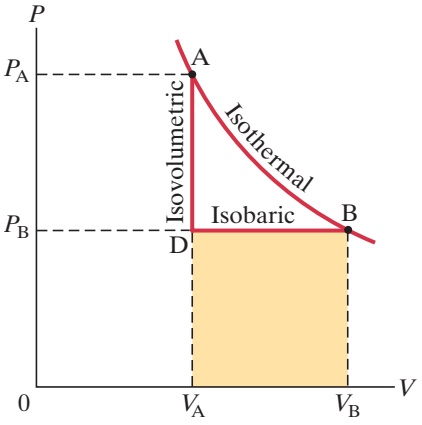


FIGURE 19–11 Work done by an ideal gas in an isothermal process equals the area under the PV curve. Shaded area equals the work done by the gas when it expands from V_A to V_B.

FIGURE 19–12 Process ADB consists of an isovolumetric (AD) and an isobaric (DB) process.



CONCEPTUAL EXAMPLE 19–11 **Work in isothermal and adiabatic processes.** In Fig. 19–8 we saw the PV diagrams for a gas expanding in two ways, isothermally and adiabatically. The initial volume V_A was the same in each case, and the final volumes were the same $(V_B = V_C)$. In which process was more work done by the gas?

RESPONSE Our system is the gas. More work was done by the gas in the isothermal process, which we can see in two simple ways by looking at Fig. 19–8. First, the "average" pressure was higher during the isothermal process AB, so $W = \bar{P}\,\Delta V$ was greater (ΔV is the same for both processes). Second, we can look at the area under each curve: the area under curve AB in Fig. 19–11, which represents the work done in an isothermal process, is greater (because curve AB is higher) than the area under AC in Fig. 19–8.

EXERCISE F Is the work done by the gas in process ADB of Fig. 19–12 greater than, less than, or equal to the work done in the isothermal process AB?

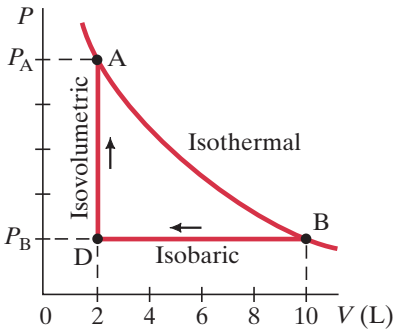


FIGURE 19–13 Example 19–12.

EXAMPLE 19–12 First law in isobaric and isovolumetric processes. An ideal gas is slowly compressed at a constant pressure of 2.0 atm from 10.0 L to 2.0 L. This process is represented in Fig. 19–13 as the path B to D. (In this process, some heat flows out of the gas and the temperature drops.) Heat is then added to the gas, holding the volume constant, and the pressure and temperature are allowed to rise (line DA) until the temperature reaches its original value $(T_A = T_B)$. Calculate (*a*) the total work done by the gas in the process BDA, and (*b*) the total heat flow into the gas.

APPROACH (*a*) Work is done only in the compression process BD. In process DA, the volume is constant so $\Delta V = 0$ and no work is done. (*b*) We use the first law of thermodynamics, Eq. 19–5.

SOLUTION (*a*) During the compression BD, the pressure is $2.0\,\text{atm} = 2(1.01 \times 10^5\,\text{N/m}^2)$ and the work done is $(\text{since } 1\,\text{L} = 10^3\,\text{cm}^3 = 10^{-3}\,\text{m}^3)$

$$W = P\,\Delta V = (2.02 \times 10^5\,\text{N/m}^2)(2.0 \times 10^{-3}\,\text{m}^3 - 10.0 \times 10^{-3}\,\text{m}^3)$$
$$= -1.6 \times 10^3\,\text{J}.$$

The total work done *by* the gas is $-1.6 \times 10^3\,\text{J}$, where the minus sign means that $+1.6 \times 10^3\,\text{J}$ of work is done *on* the gas.

(*b*) The temperature at the beginning and at the end of process BDA is the same, so even though the internal energy changes along the path BDA, it is the same at the end (A) as at the start (B): $\Delta E_{\text{int}} = 0$. From the first law of thermodynamics we have

$$0 = \Delta E_{\text{int}} = Q - W$$

so

$$Q = W = -1.6 \times 10^3\,\text{J}.$$

Because Q is negative, 1600 J of heat flows out of the gas for the whole process, BDA.

EXERCISE G In Example 19–12, if the heat lost from the gas in the process BD is $8.4 \times 10^3\,\text{J}$, what is the change in internal energy of the gas during process BD?

EXAMPLE 19–13 Work done in an engine. In an engine, 0.25 mol of an ideal monatomic gas in a cylinder expands rapidly and adiabatically against the piston. In the process, the temperature of the gas drops from 1150 K to 400 K. How much work does the gas do?

APPROACH We take the gas as our system (the piston is part of the surroundings). The pressure is not constant, and its varying value is not given. Instead, we can use the first law of thermodynamics to find W because $Q = 0$ (the process is adiabatic), and we can determine ΔE_{int} (from ΔT).

SOLUTION We determine ΔE_{int} from Eq. 19–1 for the internal energy of an ideal monatomic gas, using subscripts f and i for final and initial states:

$$\Delta E_{\text{int}} = E_{\text{int,f}} - E_{\text{int,i}} = \tfrac{3}{2}nR(T_f - T_i)$$
$$= \tfrac{3}{2}(0.25\,\text{mol})(8.314\,\text{J/mol}\cdot\text{K})(400\,\text{K} - 1150\,\text{K})$$
$$= -2300\,\text{J}.$$

Then, from the first law of thermodynamics, Eq. 19–5,

$$W = Q - \Delta E_{\text{int}} = 0 - (-2300\,\text{J}) = 2300\,\text{J}.$$

Table 19–3 (next page) shows a brief summary of a few processes we have discussed.

TABLE 19–3 Simple Thermodynamic Processes and the First Law

Process	What is constant:	The first law, $\Delta E_{\text{int}} = Q - W$, predicts
Isothermal	T = constant	$\Delta T = 0$ makes $\Delta E_{\text{int}} = 0$, so $Q = W$
Isobaric	P = constant	$Q = \Delta E_{\text{int}} + W = \Delta E_{\text{int}} + P\,\Delta V$
Isovolumetric	V = constant	$\Delta V = 0$ makes $W = 0$, so $Q = \Delta E_{\text{int}}$
Adiabatic	$Q = 0$	$\Delta E_{\text{int}} = -W$

EXAMPLE 19–14 **ΔE_{int} for boiling water to steam.** Determine the change in internal energy of 1.00 liter of water (mass 1.00 kg) at 100°C when it is fully boiled from liquid to gas, which results in 1671 liters of steam at 100°C. Assume the process is done at atmospheric pressure.

APPROACH Our system is the water. The heat required here does not result in a temperature change; rather, a change in phase occurs. We can determine the heat Q required using the latent heat of water, as in Section 19–5. Work too will be done because the volume changes. The first law of thermodynamics will then give us ΔE_{int}.

SOLUTION The latent heat of vaporization of water (Table 19–2) is $L_V = 22.6 \times 10^5\,\text{J/kg}$. So the heat input required for this process is (Eq. 19–4)

$$Q = mL = (1.00\,\text{kg})(22.6 \times 10^5\,\text{J/kg}) = 22.6 \times 10^5\,\text{J}.$$

The work done by the water (P = constant = 1 atm) is (Eq. 19–10a)

$$W = P\,\Delta V = (1.01 \times 10^5\,\text{N/m}^2)[(1671 \times 10^{-3}\,\text{m}^3) - (1 \times 10^{-3}\,\text{m}^3)]$$
$$= 1.69 \times 10^5\,\text{J},$$

where we used $1\,\text{atm} = 1.01 \times 10^5\,\text{N/m}^2$ and $1\,\text{L} = 10^3\,\text{cm}^3 = 10^{-3}\,\text{m}^3$. Then

$$\Delta E_{\text{int}} = Q - W = (22.6 \times 10^5\,\text{J}) - (1.7 \times 10^5\,\text{J}) = 20.9 \times 10^5\,\text{J}.$$

NOTE Most of the heat added goes to increasing the internal energy of the water (increasing molecular energy to overcome the attraction that held the molecules close together in the liquid state). Only a small part ($< 10\%$) goes into doing work.

NOTE Equation 19–1, $E_{\text{int}} = \frac{3}{2}nRT$, would tell us that $\Delta E_{\text{int}} = 0$ here because $\Delta T = 0$. Yet we determined that $\Delta E_{\text{int}} \approx 21 \times 10^5\,\text{J}$. What is wrong? Equation 19–1 applies only to an ideal monatomic gas, not to liquid water.

Free Expansion

One type of adiabatic process is a **free expansion** in which a gas is allowed to expand in volume adiabatically without doing any work. The apparatus to accomplish a free expansion is shown in Fig. 19–14. It consists of two well-insulated compartments (to ensure no heat flow in or out) connected by a valve or stopcock. One compartment is filled with gas, the other is empty. When the valve is opened, the gas expands to fill both containers. No heat flows in or out ($Q = 0$), and no work is done because the gas does not move any other object. Thus $Q = W = 0$ and by the first law of thermodynamics, $\Delta E_{\text{int}} = 0$. *The internal energy of a gas does not change in a free expansion*. For an ideal gas, $\Delta T = 0$ also, since E_{int} depends only on T (Section 19–2). Experimentally, the free expansion has been used to determine if the internal energy of *real gases* depends only on T. The experiments are very difficult to do accurately, but it has been found that the temperature of a real gas drops very slightly in a free expansion. Thus the internal energy of real gases does depend, a little, on pressure or volume as well as on temperature.

A free expansion can not be plotted on a PV diagram, because the process is rapid, not quasistatic. The intermediate states are not equilibrium states, and hence the pressure (and even the volume at some instants) is not clearly defined.

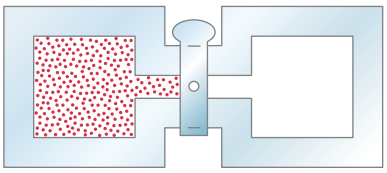


FIGURE 19–14 Free expansion.

19–8 Molar Specific Heats for Gases, and the Equipartition of Energy

In Section 19–3 we discussed the concept of specific heat and applied it to solids and liquids. The values of the specific heat for gases depend much more on how the thermodynamic process is carried out. Two important processes are those in which either the volume or the pressure is kept constant. Although for solids and liquids it matters little, Table 19–4 shows that for gases the specific heats at constant volume (c_V) and at constant pressure (c_P) are quite different.

TABLE 19–4 Specific Heats of Gases at 15°C

Gas	Specific heats (kcal/kg · K)		Molar specific heats (cal/mol · K)		$C_P - C_V$ (cal/mol · K)	$\gamma = \frac{C_P}{C_V}$
	c_V	c_P	C_V	C_P		
Monatomic						
He	0.75	1.15	2.98	4.97	1.99	1.67
Ne	0.148	0.246	2.98	4.97	1.99	1.67
Diatomic						
N_2	0.177	0.248	4.96	6.95	1.99	1.40
O_2	0.155	0.218	5.03	7.03	2.00	1.40
Triatomic						
CO_2	0.153	0.199	6.80	8.83	2.03	1.30
H_2O (100°C)	0.350	0.482	6.20	8.20	2.00	1.32
Polyatomic						
C_2H_6	0.343	0.412	10.30	12.35	2.05	1.20

Molar Specific Heats for Gases

The difference in specific heats for gases is nicely explained in terms of the first law of thermodynamics and kinetic theory. For gases, we usually use **molar specific heats**, C_V and C_P, which are defined as the heat required to raise 1 mol of the gas by 1 C° at constant volume and at constant pressure, respectively. We use moles n in place of m in Eq. 19–2: the heat Q needed to raise the temperature of n moles of gas by ΔT is

$$Q = nC_V \Delta T \qquad \text{[volume constant]} \qquad \textbf{(19–11a)}$$

$$Q = nC_P \Delta T. \qquad \text{[pressure constant]} \qquad \textbf{(19–11b)}$$

We can see from the definition of molar specific heat (or by comparing Eqs. 19–2 and 19–11) that

$$C_V = Mc_V$$
$$C_P = Mc_P,$$

where M is the molecular mass of the gas $(M = m/n$ in grams/mol$)$.[†] The values for molar specific heats are included in Table 19–4. These values are nearly the same for different gases that have the same number of atoms per molecule.

Now we use kinetic theory and imagine that an ideal gas is slowly heated via two different processes—first at constant volume, and then at constant pressure. In both processes, we let the temperature increase by the same amount, ΔT. In the constant-volume process, no work is done since $\Delta V = 0$. Thus, according to the first law of thermodynamics, the heat added (which we denote by Q_V) all goes into increasing the internal energy of the gas:

$$Q_V = \Delta E_{int}.$$

In the constant-pressure process, work *is* done. Hence the heat added, Q_P, must not only increase the internal energy but also is used to do the work $W = P\,\Delta V$.

[†]For example, $M = 4$ g/mol for He, and $M = 32$ g/mol for O_2 (see Periodic Table inside rear cover).

Thus, more heat will be needed in the process at constant pressure than in the process at constant volume. For the process at constant pressure, the first law of thermodynamics gives

$$Q_P = \Delta E_{int} + P\,\Delta V.$$

Since ΔE_{int} is the same in the two processes (we chose ΔT to be the same), we can combine the two preceding equations:

$$Q_P - Q_V = P\,\Delta V.$$

From the ideal gas law, $V = nRT/P$, so for a process at constant pressure $\Delta V = nR\,\Delta T/P$. Putting this into the last equation and using Eqs. 19–11, we find

$$nC_P\,\Delta T - nC_V\,\Delta T = P\left(\frac{nR\,\Delta T}{P}\right)$$

or, after cancellations,

$$C_P - C_V = R. \qquad \textbf{(19–12)}$$

The gas constant $R = 8.314\,\text{J/mol}\cdot\text{K} = 1.99\,\text{cal/mol}\cdot\text{K}$, so our prediction is that C_P will be larger than C_V by about $1.99\,\text{cal/mol}\cdot\text{K}$. Indeed, this is very close to what is obtained experimentally, as shown in the next to last column of Table 19–4.

Now we calculate the value of the molar specific heat of a monatomic gas using kinetic theory. In a process carried out at constant volume, no work is done; so the first law of thermodynamics tells us that

$$\Delta E_{int} = Q_V.$$

For an ideal monatomic gas, the internal energy E_{int} is the total kinetic energy of all the N molecules,

$$E_{int} = N\left(\tfrac{1}{2}m\overline{v^2}\right) = \tfrac{3}{2}nRT$$

as we saw in Section 19–2. Then, using Eq. 19–11a, we can write $\Delta E_{int} = Q$ as

$$\Delta E_{int} = \tfrac{3}{2}nR\,\Delta T = nC_V\,\Delta T \qquad \textbf{(19–13)}$$

or

$$C_V = \tfrac{3}{2}R. \qquad \textbf{(19–14)}$$

Since $R = 8.314\,\text{J/mol}\cdot\text{K} = 1.99\,\text{cal/mol}\cdot\text{K}$, kinetic theory predicts that $C_V = 2.98\,\text{cal/mol}\cdot\text{K}$ for an ideal monatomic gas. This is very close to the experimental values for monatomic gases such as helium and neon (Table 19–4). From Eq. 19–12, C_P is predicted to be $R + C_V = (1.99 + 2.98)\,\text{cal/mol}\cdot\text{K} = 4.97\,\text{cal/mol}\cdot\text{K}$, also in agreement with experiment (Table 19–4). So we are quite happy with kinetic theory's predictions.

Equipartition of Energy

The measured molar specific heats for more complex gases (Table 19–4), such as diatomic (two atoms) and triatomic (three atoms) gases, increase with the increased number of atoms per molecule. We can explain this by assuming that the internal energy includes not only translational kinetic energy but other forms of energy as well. For example, in a diatomic gas (Fig. 19–15), the two atoms can rotate about two different axes (but rotation about a third axis passing through the two atoms would give rise to very little energy since the moment of inertia is so small). The molecules can have rotational as well as translational kinetic energy. It is useful to introduce the idea of **degrees of freedom**, by which we mean the number of independent ways molecules can possess energy. For example, a monatomic gas has three degrees of freedom, because an atom can have velocity along the x, y, and z axes. These are considered to be three independent motions because a change in any one of the components would not affect the others.

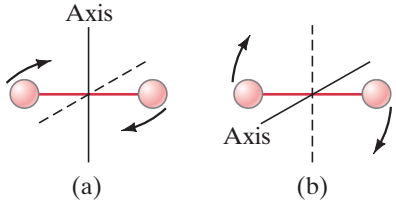


FIGURE 19–15 A diatomic molecule can rotate about two different axes.

A diatomic molecule has the same three degrees of freedom associated with translational kinetic energy plus two more degrees of freedom associated with rotational kinetic energy (Fig. 19–15), for a total of five degrees of freedom.

Table 19–4 indicates that the C_V for diatomic gases is about $\frac{5}{3}$ times as great as for a monatomic gas—that is, in the same ratio as their degrees of freedom. This result led nineteenth-century physicists to the **principle of equipartition of energy**. This principle states that energy is shared equally among the active degrees of freedom, and each active degree of freedom of a molecule has on average an energy equal to $\frac{1}{2}kT$. Thus, the average energy for a molecule of a monatomic gas, with its 3 degrees of freedom, would be $\frac{3}{2}kT$ (which we already knew) and for a diatomic gas, with its 5 degrees of freedom, $\frac{5}{2}kT$. Hence the internal energy of a diatomic gas would be $E_{\text{int}} = N\left(\frac{5}{2}kT\right) = \frac{5}{2}nRT$, where N is the number of molecules and n is the number of moles. Using the same argument we did for monatomic gases, we see that for diatomic gases the molar specific heat at constant volume would be $C_V = \frac{5}{2}R = 4.97\ \text{cal/mol} \cdot \text{K}$, close to measured values (Table 19–4). More complex molecules have even more degrees of freedom and thus greater molar specific heats.

However, measurements show that for diatomic gases at very low temperatures, C_V has a value of only $\frac{3}{2}R$, as if it had only three degrees of freedom. And at very high temperatures, C_V was about $\frac{7}{2}R$, as if there were seven degrees of freedom. The explanation is that at low temperatures, nearly all molecules have only translational kinetic energy. That is, no energy goes into rotational energy and only three degrees of freedom are "active." At very high temperatures, all five degrees of freedom are active plus two additional ones. We can interpret the two new degrees of freedom as being associated with the two atoms vibrating, as if they were connected by a spring (Fig. 19–16). One degree of freedom comes from the kinetic energy of the vibrational motion, and the second comes from the potential energy of vibrational motion $\left(\frac{1}{2}kx^2\right)$. At room temperature, these two degrees of freedom are apparently not active. See Fig. 19–17.

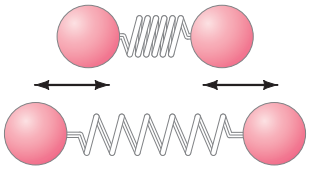


FIGURE 19–16 A diatomic molecule can vibrate, as if the two atoms were connected by a spring. Of course they are not connected by a spring; rather they exert forces on each other that are electrical in nature, but of a form that resembles a spring force.

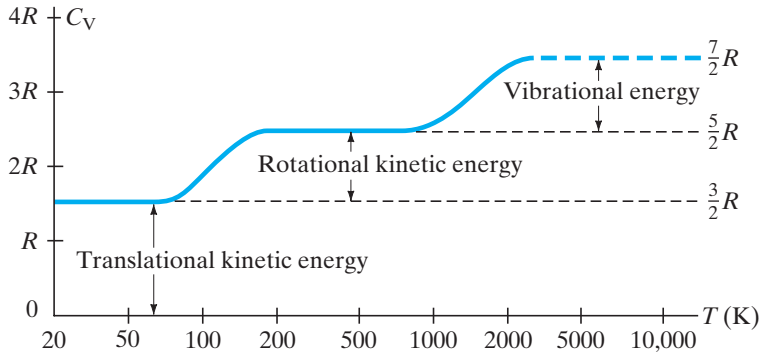
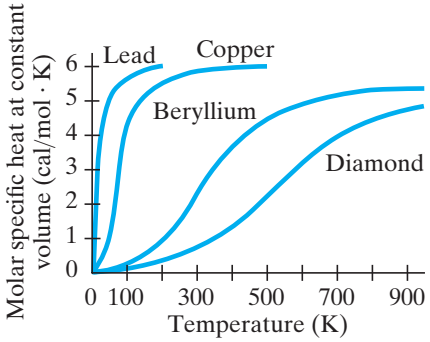


FIGURE 19–17 Molar specific heat C_V as a function of temperature for hydrogen molecules (H_2). As the temperature is increased, some of the translational kinetic energy can be transferred in collisions into rotational kinetic energy and, at still higher temperature, into vibrational energy. [Note: H_2 dissociates into two atoms at about 3200 K, so the last part of the curve is shown dashed. Also note the temperature axis is logarithmic.]

Just why fewer degrees of freedom are "active" at lower temperatures was eventually explained by Einstein using the quantum theory. [According to quantum theory, energy does not take on continuous values but is quantized—it can have only certain values, and there is a certain minimum energy. The minimum rotational and vibrational energies are higher than for simple translational kinetic energy, so at lower temperatures and lower translational kinetic energy, there is not enough energy to excite the rotational or vibrational kinetic energy.] Calculations based on kinetic theory and the principle of equipartition of energy (as modified by the quantum theory) give numerical results in accord with experiment.

FIGURE 19–18 Molar specific heats of solids as a function of temperature.



*Solids

The principle of equipartition of energy can be applied to solids as well. The molar specific heat of any solid, at high temperature, is close to $3R$ $(6.0\ \text{cal/mol} \cdot \text{K})$, Fig. 19–18. This is called the *Dulong and Petit* value after the scientists who first measured it in 1819. (We are using *molar* here, not per kg as in Table 19–1.) At high temperatures, each atom apparently has six degrees of freedom, although some are not active at low temperatures. Each atom in a crystalline solid can vibrate about

its equilibrium position as if it were connected by springs to each of its neighbors (Fig. 19–19). Thus it can have three degrees of freedom for kinetic energy and three more associated with potential energy of vibration in each of the x, y, and z directions, which is in accord with the measured values.

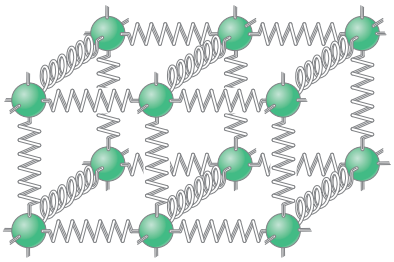


FIGURE 19–19 The atoms in a crystalline solid can vibrate about their equilibrium positions as if they were connected to their neighbors by springs. (The forces between atoms are actually electrical in nature.)

19–9 Adiabatic Expansion of a Gas

The PV curve for the quasistatic (slow) adiabatic expansion $(Q = 0)$ of an ideal gas was shown in Fig. 19–8 (curve AC). It is somewhat steeper than for an isothermal process $(\Delta T = 0)$, which indicates that for the same change in volume the change in pressure will be greater. Hence the temperature of the gas must drop during an adiabatic expansion. Conversely, the temperature rises during an adiabatic compression.

We can derive the relation between the pressure P and the volume V of an ideal gas that is allowed to slowly expand adiabatically. We begin with the first law of thermodynamics, written in differential form:

$$dE_{\text{int}} = dQ - dW = -dW = -P\,dV,$$

since $dQ = 0$ for an adiabatic process. Equation 19–13 gives us a relation between ΔE_{int} and C_V, which is valid for any ideal gas process since E_{int} is a function only of T for an ideal gas. We write this in differential form:

$$dE_{\text{int}} = nC_V\,dT.$$

When we combine these last two equations, we obtain

$$nC_V\,dT + P\,dV = 0.$$

We next take the differential of the ideal gas law, $PV = nRT$, allowing P, V, and T to vary:

$$P\,dV + V\,dP = nR\,dT.$$

We solve for dT in this relation and substitute it into the previous relation and get

$$nC_V\left(\frac{P\,dV + V\,dP}{nR}\right) + P\,dV = 0$$

or, multiplying through by R and rearranging,

$$(C_V + R)P\,dV + C_V V\,dP = 0.$$

We note from Eq. 19–12 that $C_V + R = C_P$, so we have

$$C_P P\,dV + C_V V\,dP = 0,$$

or

$$\frac{C_P}{C_V}P\,dV + V\,dP = 0.$$

We define

$$\gamma = \frac{C_P}{C_V} \qquad \textbf{(19–15)}$$

so that our last equation becomes

$$\frac{dP}{P} + \gamma\frac{dV}{V} = 0.$$

This is integrated to become

$$\ln P + \gamma \ln V = \text{constant}.$$

This simplifies (using the rules for addition and multiplication of logarithms) to

$$PV^{\gamma} = \text{constant}. \qquad \left[\begin{array}{c}\text{quasistatic adiabatic}\\ \text{process; ideal gas}\end{array}\right] \qquad \textbf{(19–16)}$$

This equation relates P and V for a quasistatic adiabatic expansion or contraction.

We will find Eq. 19–16, $PV^\gamma = \text{constant}$, very useful when we discuss heat engines in the next Chapter. Table 19–4 (page 556) gives values of γ for some real gases. Figure 19–8 compares an adiabatic expansion (Eq. 19–16) in curve AC to an isothermal expansion ($PV = \text{constant}$) in curve AB. It is important to remember that the ideal gas law, $PV = nRT$, continues to hold even for an adiabatic expansion ($PV^\gamma = \text{constant}$); clearly PV is not constant, meaning T is not constant.

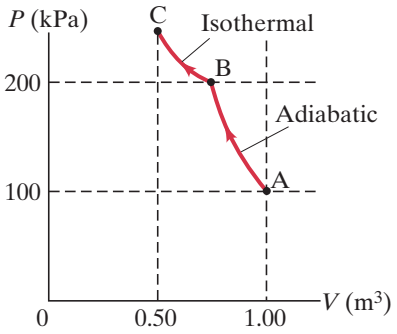


FIGURE 19–20 Example 19–15.

EXAMPLE 19–15 Compressing an ideal gas. An ideal monatomic gas is compressed starting at point A in the PV diagram of Fig. 19–20, where $P_A = 100\,\text{kPa}$, $V_A = 1.00\,\text{m}^3$, and $T_A = 300\,\text{K}$. The gas is first compressed adiabatically to state B ($P_B = 200\,\text{kPa}$). The gas is then further compressed from point B to point C ($V_C = 0.50\,\text{m}^3$) in an isothermal process. (*a*) Determine V_B. (*b*) Calculate the work done *on* the gas for the whole process.

APPROACH Volume V_B is obtained using Eq. 19–16. The work done *by* a gas is given by Eq. 19–8, $W = \int P\,dV$. The work *on* the gas is the negative of this: $W_{on} = -\int P\,dV$.

SOLUTION In the adiabatic process, Eq. 19–16 tells us $PV^\gamma = \text{constant}$. Therefore, $PV^\gamma = P_A V_A^\gamma = P_B V_B^\gamma$ where for a monatomic gas $\gamma = C_P/C_V = (5/2)/(3/2) = \frac{5}{3}$.

(*a*) Equation 19–16 gives $V_B = V_A(P_A/P_B)^{\frac{1}{\gamma}} = (1.00\,\text{m}^3)(100\,\text{kPa}/200\,\text{kPa})^{\frac{3}{5}} = 0.66\,\text{m}^3$.

(*b*) The pressure P at any instant during the adiabatic process is given by $P = P_A V_A^\gamma V^{-\gamma}$. The work done on the gas in going from V_A to V_B is

$$W_{AB} = -\int_A^B P\,dV = -P_A V_A^\gamma \int_{V_A}^{V_B} V^{-\gamma}\,dV$$
$$= -P_A V_A^\gamma \left(\frac{1}{-\gamma + 1}\right)(V_B^{1-\gamma} - V_A^{1-\gamma}).$$

Since $\gamma = \frac{5}{3}$, then $-\gamma + 1 = 1 - \gamma = -\frac{2}{3}$, so

$$W_{AB} = -\left(P_A V_A^{\frac{5}{3}}\right)\left(-\frac{3}{2}\right)\left(V_A^{-\frac{2}{3}}\right)\left[\left(\frac{V_B}{V_A}\right)^{-\frac{2}{3}} - 1\right]$$
$$= +\frac{3}{2}P_A V_A\left[\left(\frac{V_B}{V_A}\right)^{-\frac{2}{3}} - 1\right]$$
$$= +\frac{3}{2}(100\,\text{kPa})(1.00\,\text{m}^3)\left[(0.66)^{-\frac{2}{3}} - 1\right] = +48\,\text{kJ}.$$

For the isothermal process from B to C, the work is done at constant temperature, so the pressure at any instant during the process is $P = nRT_B/V$ and

$$W_{BC} = -\int_B^C P\,dV = -nRT_B\int_{V_B}^{V_C}\frac{dV}{V} = -nRT_B \ln\frac{V_C}{V_B}$$
$$= -P_B V_B \ln\frac{V_C}{V_B} = +37\,\text{kJ}.$$

The total work done on the gas is $48\,\text{kJ} + 37\,\text{kJ} = 85\,\text{kJ}$.

19–10 Heat Transfer: Conduction, Convection, Radiation

Heat is transferred from one place or body to another in three different ways: by *conduction*, by *convection*, and by *radiation*. We now discuss each of these in turn; but in practical situations, any two or all three may be operating at the same time. We start with conduction.

Conduction

When a metal poker is put in a hot fire, or a silver spoon is placed in a hot bowl of soup, the exposed end of the poker or spoon soon becomes hot as well, even though it is not in the source of heat. We say that heat has been conducted from the hot end to the cold end.

Heat **conduction** in many materials can be visualized as being carried out via molecular collisions. As one end of an object is heated, the molecules there move faster and faster (= higher temperature). As these faster molecules collide with slower-moving neighbors, they transfer some of their kinetic energy to them, which in turn transfer some energy by collision with molecules still farther along the object. Thus the kinetic energy of thermal motion is transferred by molecular collision along the object. In metals, collisions of free electrons are mainly responsible for conduction. Conduction between objects in physical contact occurs similarly.

Heat conduction from one point to another takes place only if there is a difference in temperature between the two points. Indeed, it is found experimentally that the rate of heat flow through a substance is proportional to the difference in temperature between its ends. The rate of heat flow also depends on the size and shape of the object. To investigate this quantitatively, let us consider the heat flow through a uniform cylinder, as illustrated in Fig. 19–21. It is found experimentally that the heat flow Q over a time interval t is given by the relation

$$\frac{Q}{t} = kA\frac{T_1 - T_2}{\ell} \qquad \textbf{(19–17a)}$$

where A is the cross-sectional area of the object, ℓ is the distance between the two ends, which are at temperatures T_1 and T_2, and k is a proportionality constant called the **thermal conductivity** which is characteristic of the material. From Eq. 19–17a, we see that the rate of heat flow (units of J/s) is directly proportional to the cross-sectional area and to the temperature gradient $(T_1 - T_2)/\ell$.

In some cases (such as when k or A cannot be considered constant) we need to consider the limit of an infinitesimally thin slab of thickness dx. Then Eq. 19–17a becomes

$$\frac{dQ}{dt} = -kA\frac{dT}{dx}, \qquad \textbf{(19–17b)}$$

where dT/dx is the **temperature gradient** and the negative sign is included since the heat flow is in the direction opposite to the temperature gradient.

[Equations 19–17 are quite similar to the relations describing diffusion (Section 18–8) and the flow of fluids through a pipe (Section 13–12). In those cases, the flow of matter was found to be proportional to the *concentration gradient* dC/dx, or to the *pressure gradient* $(P_1 - P_2)/\ell$. This close similarity is one reason we speak of the "flow" of heat. Yet we must keep in mind that no substance is flowing in the case of heat—it is energy that is being transferred.]

The thermal conductivities, k, for a variety of substances are given in Table 19–5. Substances for which k is large conduct heat rapidly and are said to be good thermal **conductors**. Most metals fall in this category, although there is a wide range even among them, as you may observe by holding the ends of a silver spoon and a stainless-steel spoon immersed in the same hot cup of soup. Substances for which k is small, such as wool, fiberglass, polyurethane, and goose down, are poor conductors of heat and are therefore good thermal **insulators**.

The relative magnitudes of k can explain simple phenomena such as why a tile floor is much colder on the feet than a rug-covered floor at the same temperature. Tile is a better conductor of heat than the rug. Heat that flows from your foot to the rug is not conducted away rapidly, so the rug's surface quickly warms up to the temperature of your foot and feels good. But the tile conducts the heat away rapidly and thus can take more heat from your foot quickly, so your foot's surface temperature drops.

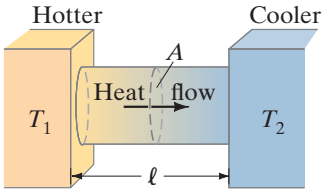


FIGURE 19–21 Heat conduction between areas at temperatures T_1 and T_2. If T_1 is greater than T_2, the heat flows to the right; the rate is given by Eq. 19–17a.

TABLE 19–5
Thermal Conductivities

Substance	Thermal conductivity, k	
	J (s · m · C°)	kcal (s · m · C°)
Silver	420	10×10^{-2}
Copper	380	9.2×10^{-2}
Aluminum	200	5.0×10^{-2}
Steel	40	1.1×10^{-2}
Ice	2	5.0×10^{-4}
Glass	0.84	2.0×10^{-4}
Brick	0.84	2.0×10^{-4}
Concrete	0.84	2.0×10^{-4}
Water	0.56	1.4×10^{-4}
Human tissue	0.2–2	$0.5–5 \times 10^{-4}$
Wood	0.1	0.3×10^{-4}
Fiberglass	0.048	0.12×10^{-4}
Cork	0.042	0.10×10^{-4}
Wool	0.040	0.10×10^{-4}
Goose down	0.025	0.060×10^{-4}
Polyurethane	0.024	0.057×10^{-4}
Air	0.023	0.055×10^{-4}

PHYSICS APPLIED
Heat loss through windows

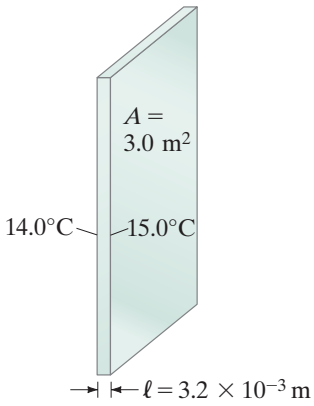


FIGURE 19–22 Example 19–16.

EXAMPLE 19–16 Heat loss through windows. A major source of heat loss from a house in cold weather is through the windows. Calculate the rate of heat flow through a glass window $2.0\,\text{m} \times 1.5\,\text{m}$ in area and 3.2 mm thick, if the temperatures at the inner and outer surfaces are 15.0°C and 14.0°C, respectively (Fig. 19–22).

APPROACH Heat flows by conduction through the 3.2-mm thickness of glass from the higher inside temperature to the lower outside temperature. We use the heat conduction equation, Eq. 19–17a.

SOLUTION Here $A = (2.0\,\text{m})(1.5\,\text{m}) = 3.0\,\text{m}^2$ and $\ell = 3.2 \times 10^{-3}\,\text{m}$. Using Table 19–5 to get k, we have

$$\frac{Q}{t} = kA\frac{T_1 - T_2}{\ell}$$

$$= \frac{(0.84\,\text{J/s}\cdot\text{m}\cdot\text{C}^\circ)(3.0\,\text{m}^2)(15.0^\circ\text{C} - 14.0^\circ\text{C})}{(3.2 \times 10^{-3}\,\text{m})} = 790\,\text{J/s}.$$

NOTE This rate of heat flow is equivalent to $(790\,\text{J/s})/(4.186 \times 10^3\,\text{J/kcal}) = 0.19\,\text{kcal/s}$, or $(0.19\,\text{kcal/s}) \times (3600\,\text{s/h}) = 680\,\text{kcal/h}$.

You might notice in Example 19–16 that 15°C is not very warm for the living room of a house. The room itself may indeed be much warmer, and the outside might be colder than 14°C. But the temperatures of 15°C and 14°C were specified as those at the window surfaces, and there is usually a considerable drop in temperature of the air in the vicinity of the window both on the inside and the outside. That is, the layer of air on either side of the window acts as an insulator, and normally the major part of the temperature drop between the inside and outside of the house takes place across the air layer. If there is a heavy wind, the air outside a window will constantly be replaced with cold air; the temperature gradient across the glass will be greater and there will be a much greater rate of heat loss. Increasing the width of the air layer, such as using two panes of glass separated by an air gap, will reduce the heat loss more than simply increasing the glass thickness, since the thermal conductivity of air is much less than that for glass. Such "double-pane windows" are often called *thermal windows*.

PHYSICS APPLIED
Thermal windows

PHYSICS APPLIED
Clothes insulate by trapping an air layer

The insulating properties of clothing come from the insulating properties of air. Without clothes, our bodies in still air would heat the air in contact with the skin and would soon become reasonably comfortable because air is a very good insulator. But since air moves—there are breezes and drafts, and people move about—the warm air would (in cooler climates) be replaced by cold air, thus increasing the temperature difference and the heat loss from the body. Clothes keep us warm by trapping air so it cannot move readily. It is not the cloth that insulates us, but the air that the cloth traps. Goose down is a very good insulator because even a small amount of it fluffs up and traps a great amount of air.

EXERCISE H Explain why drapes in front of a window reduce heat loss from a house.

TABLE 19–6 *R*-values

Material	Thickness	*R*-value ($\text{ft}^2\cdot\text{h}\cdot\text{F}^\circ/\text{Btu}$)
Glass	$\frac{1}{8}$ inch	1
Brick	$3\frac{1}{2}$ inches	0.6–1
Plywood	$\frac{1}{2}$ inch	0.6
Fiberglass insulation	4 inches	12

PHYSICS APPLIED
R-values of thermal insulation

*R-Values for Building Materials

The insulating properties of building materials are often specified by *R*-values (or "thermal resistance"), defined for a given thickness ℓ of material as:

$$R = \frac{\ell}{k}.$$

The *R*-value of a given piece of material combines the thickness ℓ and the thermal conductivity k in one number. Larger R means better insulation from heat or cold. In the United States, *R*-values are given in British units as $\text{ft}^2\cdot\text{h}\cdot\text{F}^\circ/\text{Btu}$ (for example, *R*-19 means $R = 19\,\text{ft}^2\cdot\text{h}\cdot\text{F}^\circ/\text{Btu}$). Table 19–6 gives *R*-values for some common building materials. *R*-values increase directly with material thickness: for example, 2 inches of fiberglass is *R*-6, whereas 4 inches is *R*-12.

Convection

Although liquids and gases are generally not very good conductors of heat, they can transfer heat rapidly by convection. **Convection** is the process whereby heat flows by the bulk movement of molecules from one place to another. Whereas conduction involves molecules (and/or electrons) moving only over small distances and colliding, convection involves the movement of large numbers of molecules over large distances.

A forced-air furnace, in which air is heated and then blown by a fan into a room, is an example of **forced convection**. **Natural convection** occurs as well, and one familiar example is that hot air rises. For instance, the air above a radiator (or other type of heater) expands as it is heated (Chapter 17), and hence its density decreases. Because its density is less than that of the surrounding cooler air, it rises, via buoyancy, just as a log submerged in water floats upward because its density is less than that of water. Warm or cold ocean currents, such as the balmy Gulf Stream, represent natural convection on a global scale. Wind is another example of convection, and weather in general is strongly influenced by convective air currents.

When a pot of water is heated (Fig. 19–23), convection currents are set up as the heated water at the bottom of the pot rises because of its reduced density. That heated water is replaced by cooler water from above. This principle is used in many heating systems, such as the hot-water radiator system shown in Fig. 19–24. Water is heated in the furnace, and as its temperature increases, it expands and rises up along the pipes as shown. This causes the water to circulate in the heating system. Hot water then enters the radiators, heat is transferred by conduction to the air, and the cooled water returns to the furnace. Thus, the water circulates because of convection; pumps are sometimes used to improve circulation. The air throughout the room also becomes heated as a result of convection. The air heated by the radiators rises and is replaced by cooler air, resulting in convective air currents, as shown by the green arrows in Fig. 19–24.

Other types of furnaces also depend on convection. Hot-air furnaces with registers (openings) near the floor often do not have fans but depend on natural convection, which can be appreciable. In other systems, a fan is used. In either case, it is important that cool air can return to the furnace so that convective currents circulate throughout the room to keep it uniformly heated. Be sure not to block the cool air return opening.

The human body produces a great deal of thermal energy. Of the food energy transformed within the body, at best 20% is used to do work, so over 80% appears as thermal energy. During light activity, for example, if this thermal energy is not dissipated, the body temperature would rise about 3 C° per hour. Clearly, the heat generated by the body must be transferred to the outside. Is the heat transferred by conduction? The temperature of the skin in a comfortable environment is 33 to 35°C, whereas the interior of the body is at 37°C. Calculation shows (Problem 82) that, because of this small temperature difference, plus the low thermal conductivity of tissue, direct conduction from the body's interior is not responsible for much of the heat that must leave the body. Instead, the heat is carried to the skin by the blood. In addition to all its other important responsibilities, blood acts as a convective fluid to transfer heat to just beneath the surface of the skin. It is then conducted (over a very short distance) to the surface. Once at the surface, the heat is transferred to the environment by convection, evaporation, and radiation.

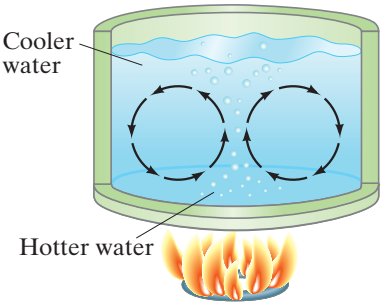
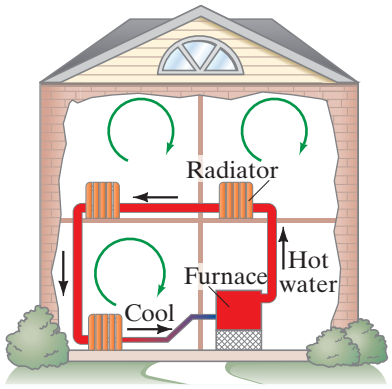


FIGURE 19–23 Convection currents in a pot of water being heated on a stove.

PHYSICS APPLIED
Ocean currents and wind

PHYSICS APPLIED
Convective home heating

FIGURE 19–24 Convection plays a role in heating a house. The circular arrows show the convective air currents in the rooms.



PHYSICS APPLIED
Body heat: convection by blood

Radiation

Convection and conduction require the presence of matter as a medium to carry the heat from the hotter to the colder region. But a third type of heat transfer occurs without any medium at all. All life on Earth depends on the transfer of energy from the Sun, and this energy is transferred to the Earth over empty (or nearly empty) space. This form of energy transfer is heat—since the Sun's surface temperature (6000 K) is much higher than Earth's ($\approx$ 300 K)—and is referred to as **radiation**. The warmth we receive from a fire is mainly radiant energy.

We will see in Chapter 31 that radiation consists of electromagnetic waves. Infrared (IR) wavelengths are mainly responsible for heating the Earth.

The rate at which an object radiates energy has been found to be proportional to the fourth power of the Kelvin temperature, T. That is, an object at 2000 K, as compared to one at 1000 K, radiates energy at a rate $2^4 = 16$ times as much. The rate of radiation is also proportional to the area A of the emitting object, so the rate at which energy leaves the radiating object, Q/t, is

$$\frac{Q}{t} = \epsilon\sigma A T^4. \qquad \textbf{(19–18)}$$

This is called the **Stefan-Boltzmann equation**, and σ is a universal constant called the **Stefan-Boltzmann constant** which has the value

$$\sigma = 5.67 \times 10^{-8}\,\mathrm{W/m^2 \cdot K^4}.$$

The factor ϵ, called the **emissivity**, is a number[†] between 0 and 1 that is characteristic of the surface of the radiating material. Very black surfaces, such as charcoal, have emissivity close to 1, whereas shiny metal surfaces have ϵ close to zero and thus emit correspondingly less radiation. The value of ϵ depends somewhat on the temperature of the material.

PHYSICS APPLIED
Dark vs. light clothing

Not only do shiny surfaces emit less radiation, but they absorb little of the radiation that falls upon them (most is reflected). Black and very dark objects are good emitters ($\epsilon \approx 1$); they also absorb nearly all the radiation that falls on them—which is why light-colored clothing is usually preferable to dark clothing on a hot day. Thus, **a good absorber is also a good emitter**.

Any object not only emits energy by radiation but also absorbs energy radiated by other objects. If an object of emissivity ϵ and area A is at a temperature T_1, it radiates energy at a rate $\epsilon\sigma A T_1^4$. If the object is surrounded by an environment at temperature T_2, the rate at which the surroundings radiate energy is proportional to T_2^4; then the rate that energy is *absorbed* by the object is proportional to T_2^4. The *net* rate of radiant heat flow from the object is given by the equation

$$\frac{Q}{t} = \epsilon\sigma A\left(T_1^4 - T_2^4\right), \qquad \textbf{(19–19)}$$

where A is the surface area of the object, T_1 its temperature and ϵ its emissivity (at temperature T_1), and T_2 is the temperature of the surroundings. This equation is consistent with the experimental fact that equilibrium between the object and its surroundings is reached when they come to the same temperature. That is, Q/t must equal zero when $T_1 = T_2$, so ϵ must be the same for emission and absorption. This confirms the idea that a good emitter is a good absorber. Because both the object and its surroundings radiate energy, there is a net transfer of energy from one to the other unless everything is at the same temperature.

From Eq. 19–19, we see that if $T_1 > T_2$, the net flow of heat is from the object to the surroundings, so the object cools. But if $T_1 < T_2$, the net heat flow is from the surroundings into the object, and the object's temperature rises. If different parts of the surroundings are at different temperatures, Eq. 19–19 becomes more complicated.

PHYSICS APPLIED
The body's radiative heat loss

EXAMPLE 19–17 **ESTIMATE** **Cooling by radiation.** An athlete is sitting unclothed in a locker room whose dark walls are at a temperature of 15°C. Estimate the body's rate of heat loss by radiation, assuming a skin temperature of 34°C and $\epsilon = 0.70$. Take the surface area of the body not in contact with the chair to be $1.5\,\mathrm{m^2}$.

APPROACH We use Eq. 19–19, which requires Kelvin temperatures.

SOLUTION We have

$$\frac{Q}{t} = \epsilon\sigma A\left(T_1^4 - T_2^4\right)$$
$$= (0.70)\left(5.67 \times 10^{-8}\,\mathrm{W/m^2 \cdot K^4}\right)(1.5\,\mathrm{m^2})\left[(307\,\mathrm{K})^4 - (288\,\mathrm{K})^4\right] = 120\,\mathrm{W}.$$

PROBLEM SOLVING
Must use the Kelvin temperature

NOTE This resting person's "output" is a bit more than what a 100-W bulb uses.

PROBLEM SOLVING
$T_1^4 - T_2^4 \neq (T_1 - T_2)^4$

NOTE Avoid a common error: $(T_1^4 - T_2^4) \neq (T_1 - T_2)^4$.

[†] ϵ is the Greek letter epsilon.

A resting person naturally produces heat internally at a rate of about 100 W, which is less than the heat loss by radiation as calculated in Example 19–17. Hence, our person's temperature would drop, causing considerable discomfort. The body responds to excessive heat loss by increasing its metabolic rate, and shivering is one method by which the body increases its metabolism. Naturally, clothes help a lot. Example 19–17 illustrates that a person may be uncomfortable even if the temperature of the air is, say, 25°C, which is quite a warm room. If the walls or floor are cold, radiation to them occurs no matter how warm the air is. Indeed, it is estimated that radiation accounts for about 50% of the heat loss from a sedentary person in a normal room. Rooms are most comfortable when the walls and floor are warm and the air is not so warm. Floors and walls can be heated by means of hot-water conduits or electric heating elements. Such first-rate heating systems are becoming more common today, and it is interesting to note that 2000 years ago the Romans, even in houses in the remote province of Great Britain, made use of hot-water and steam conduits in the floor to heat their houses.

PHYSICS APPLIED
Room comfort and metabolism

PHYSICS APPLIED
Warm walls, cool air are more comfortable

EXAMPLE 19–18 **ESTIMATE** **Two teapots.** A ceramic teapot ($\epsilon = 0.70$) and a shiny metal one ($\epsilon = 0.10$) each hold 0.75 L of tea at 95°C. (*a*) Estimate the rate of heat loss from each, and (*b*) estimate the temperature drop after 30 min for each. Consider only radiation, and assume the surroundings are at 20°C.

APPROACH We are given all the information necessary to calculate the heat loss due to radiation, except for the area. Each teapot holds 0.75 L, and we can approximate it as a cube 10 cm on a side (volume = 1.0 L), with five sides exposed. To estimate the temperature drop in (*b*), we use the concept of specific heat and ignore the contribution of the pots compared to that of the water.

SOLUTION (*a*) Each teapot, approximated by a cube 10 cm on a side with five sides exposed, has a surface area of about $5 \times (0.1\ \text{m})^2 = 5 \times 10^{-2}\ \text{m}^2$. The rate of heat loss would be about

$$\begin{aligned}\frac{Q}{t} &= \epsilon\sigma A\left(T_1^4 - T_2^4\right) \\ &= \epsilon\left(5.67 \times 10^{-8}\ \text{W/m}^2\cdot\text{K}^4\right)\left(5 \times 10^{-2}\ \text{m}^2\right)\left[(368\ \text{K})^4 - (293\ \text{K})^4\right] \\ &\approx \epsilon\,(30)\ \text{W},\end{aligned}$$

or about 20 W for the ceramic pot ($\epsilon = 0.70$) and 3 W for the shiny one ($\epsilon = 0.10$).

(*b*) To estimate the temperature drop, we use the specific heat of water and ignore the contribution of the pots. The mass of 0.75 L of water is 0.75 kg. (Recall that $1.0\ \text{L} = 1000\ \text{cm}^3 = 1 \times 10^{-3}\ \text{m}^3$ and $\rho = 1000\ \text{kg/m}^3$.) Using Eq. 19–2 and Table 19–1, we get

$$\frac{Q}{t} = mc\frac{\Delta T}{t}.$$

Then

$$\frac{\Delta T}{t} = \frac{Q/t}{mc} \approx \frac{\epsilon\,(30)\ \text{J/s}}{(0.75\ \text{kg})\left(4186\ \text{J/kg}\cdot\text{C}^\circ\right)} \approx \epsilon\,(0.01)\ \text{C}^\circ/\text{s}.$$

After 30 min (= 1800 s), $\Delta T = \epsilon\,(0.01\ \text{C}^\circ/\text{s})t = \epsilon\,(0.01\ \text{C}^\circ/\text{s})\,(1800\ \text{s}) \approx 18\epsilon\ \text{C}^\circ$, or about 12 C° for the ceramic pot ($\epsilon = 0.70$) and about 2 C° for the shiny one ($\epsilon = 0.10$). The shiny one clearly has an advantage, at least in terms of radiation.

NOTE Convection and conduction could play a greater role than radiation.

PHYSICS APPLIED
Radiation from the Sun

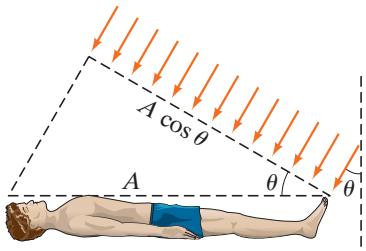


FIGURE 19–25 Radiant energy striking a body at an angle θ.

Heating of an object by radiation from the Sun cannot be calculated using Eq. 19–19 because this equation assumes a uniform temperature, T_2, of the environment surrounding the object, whereas the Sun is essentially a point source. Hence the Sun must be treated as a separate source of energy. About 1350 J of energy from the Sun strikes Earth's atmosphere per second per square meter of area at right angles to the Sun's rays. This number, 1350 W/m^2, is called the **solar constant**. The atmosphere may absorb as much as 70% of this energy before it reaches the ground, depending on the cloud cover. On a clear day, about 1000 W/m^2 reaches the Earth's surface. An object of emissivity ϵ with area A facing the Sun absorbs energy from the Sun on a sunny day at a rate, in watts, of about

$$\frac{Q}{t} = (1000\ \text{W/m}^2)\epsilon A \cos\theta, \qquad \textbf{(19–20)}$$

where θ is the angle between the Sun's rays and a line perpendicular to the area A (Fig. 19–25). That is, $A\cos\theta$ is the "effective" area, at right angles to the Sun's rays.

The explanation for the **seasons** and the polar ice caps (see Fig. 19–26) depends on this $\cos\theta$ factor in Eq. 19–20. The seasons are *not* a result of how close the Earth is to the Sun—in fact, in the Northern Hemisphere, summer occurs when the Earth is farthest from the Sun. It is the angle (i.e., $\cos\theta$) that really matters. Furthermore, the reason the Sun heats the Earth more at midday than at sunrise or sunset is also related to this $\cos\theta$ factor.

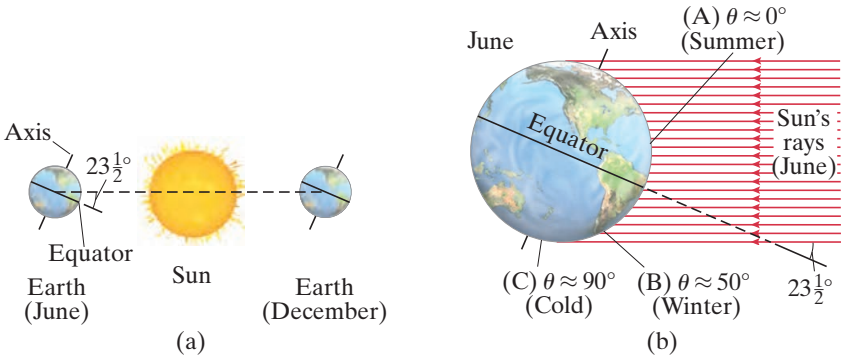


FIGURE 19–26 (a) Earth's seasons arise from the tilt of Earth's axis relative to its orbit around the Sun. (b) June sunlight makes an angle of $23\frac{1}{2}°$ with the equator. Thus θ in the southern U.S. (label A in part b) is near 0° (direct summer sunlight), but in the Southern Hemisphere (B), θ is 50° or 60°, and less heat can be absorbed—hence it is winter. Near the poles (C), there is never strong direct sunlight; $\cos\theta$ varies from about $\frac{1}{2}$ in summer to 0 in winter; so with little heating, ice can form.

PHYSICS APPLIED
Astronomy—size of a star

EXAMPLE 19–19 **ESTIMATE** **Star radius.** The giant star Betelgeuse emits radiant energy at a rate 10^4 times greater than our Sun, whereas its surface temperature is only half (2900 K) that of our Sun. Estimate the radius of Betelgeuse (B), assuming $\epsilon = 1$. The radius of the Sun (S) is $r_S = 7 \times 10^8$ m.

APPROACH We assume both stars are spherical, with surface area $4\pi r^2$.

SOLUTION We solve Eq. 19–18 for A:

$$4\pi r^2 = A = \frac{(Q/t)}{\epsilon\sigma T^4}.$$

Then

$$\frac{r_B^2}{r_S^2} = \frac{(Q/t)_B}{(Q/t)_S}\cdot\frac{T_S^4}{T_B^4} = (10^4)(2^4) = 16 \times 10^4.$$

Hence $r_B = \sqrt{16 \times 10^4}\, r_S = (400)(7 \times 10^8\ \text{m}) \approx 3 \times 10^{11}$ m.

NOTE If Betelgeuse were our Sun, it would envelop us (Earth is 1.5×10^{11} m from the Sun).

EXERCISE I Fanning yourself on a hot day cools you by (*a*) increasing the radiation rate of the skin; (*b*) increasing conductivity; (*c*) decreasing the mean free path of air; (*d*) increasing the evaporation of perspiration; (*e*) none of these.

(a)  (b) 

FIGURE 19–27 Thermograms of a healthy person's arms and hands (a) before and (b) after smoking a cigarette, showing a temperature decrease due to impaired blood circulation associated with smoking. The thermograms have been color-coded according to temperature; the scale on the right goes from blue (cold) to white (hot).

PHYSICS APPLIED
Thermography

An interesting application of thermal radiation to diagnostic medicine is **thermography**. A special instrument, the thermograph, scans the body, measuring the intensity of infrared[†] radiation from many points and forming a picture that resembles an X-ray (Fig. 19–27). Areas where metabolic activity is high, such as in tumors, can often be detected on a thermogram as a result of their higher temperature and consequent increased radiation.

Summary

Internal energy, E_{int}, refers to the total energy of all the molecules in an object. For an ideal monatomic gas,

$$E_{\text{int}} = \tfrac{3}{2}NkT = \tfrac{3}{2}nRT \qquad \textbf{(19–1)}$$

where N is the number of molecules and n is the number of moles.

Heat refers to the transfer of energy from one object to another because of a difference of temperature. Heat is thus measured in energy units, such as joules.

Heat and internal energy are also sometimes specified in calories or kilocalories (kcal), where

$$1\,\text{kcal} = 4.186\,\text{kJ}$$

is the amount of heat needed to raise the temperature of 1 kg of water by 1 C°.

The **specific heat**, c, of a substance is defined as the energy (or heat) required to change the temperature of unit mass of the substance by 1 degree; as an equation,

$$Q = mc\,\Delta T, \qquad \textbf{(19–2)}$$

where Q is the heat absorbed or given off, ΔT is the temperature increase or decrease, and m is the mass of the substance.

When heat flows between parts of an isolated system, conservation of energy tells us that the heat gained by one part of the system is equal to the heat lost by the other part of the system. This is the basis of **calorimetry**, which is the quantitative measurement of heat exchange.

Exchange of energy occurs, without a change in temperature, whenever a substance changes phase. The **heat of fusion** is the heat required to melt 1 kg of a solid into the liquid phase; it is also equal to the heat given off when the substance changes from liquid to solid. The **heat of vaporization** is the energy required to change 1 kg of a substance from the liquid to the vapor phase; it is also the energy given off when the substance changes from vapor to liquid.

The **first law of thermodynamics** states that the change in internal energy ΔE_{int} of a system is equal to the heat *added to* the system, Q, minus the work, W, done *by* the system:

$$\Delta E_{\text{int}} = Q - W. \qquad \textbf{(19–5)}$$

This important law is a broad restatement of the conservation of energy and is found to hold for all processes.

An **isothermal** process is a process carried out at constant temperature $(\Delta T = 0)$.

In an **adiabatic** process, no heat is exchanged $(Q = 0)$.

An **isobaric** process is one carried out at constant pressure. An **isovolumetric** process is one carried out at constant volume. (See Table 19–3, page 555, for a summary.)

The work done by (or on) a gas to change its volume by dV is $dW = P\,dV$, where P is the pressure.

Work and heat are not functions of the state of a system (as are P, V, T, n, and E_{int}) but depend on the type of process that takes a system from one state to another.

The **molar specific heat** of an ideal gas at constant volume, C_{V}, and at constant pressure, C_{P}, are related by

$$C_{\text{P}} - C_{\text{V}} = R, \qquad \textbf{(19–12)}$$

where R is the gas constant. For a monatomic ideal gas, $C_{\text{V}} = \frac{3}{2}R$.

For ideal gases made up of diatomic or more complex molecules, C_{V} is equal to $\frac{1}{2}R$ times the number of **degrees of freedom** of the molecule. Unless the temperature is very high, some of the degrees of freedom may not be active and so do not contribute. According to the **principle of equipartition of energy**, energy is shared equally among the active degrees of freedom in an amount $\frac{1}{2}kT$ per molecule on average.

When an ideal gas expands (or contracts) adiabatically $(Q = 0)$, the relation $PV^{\gamma} = \text{constant}$ holds, where

$$\gamma = \frac{C_{\text{P}}}{C_{\text{V}}}. \qquad \textbf{(19–15)}$$

Heat is transferred from one place (or object) to another in three different ways: conduction, convection, and radiation.

In **conduction**, energy is transferred through a substance by means of collisions between hotter (faster) molecules or electrons with their slower moving neighbors. The rate of heat flow through area A is

$$\frac{dQ}{dt} = -kA\frac{dT}{dx} \qquad \textbf{(19–7b)}$$

where dT/dx is the *temperature gradient*.

Convection is the transfer of energy by the mass movement of molecules over considerable distances.

Radiation, which does not require the presence of matter, is energy transfer by electromagnetic waves, such as from the Sun. All objects radiate energy in an amount that is proportional to the fourth power of their Kelvin temperature (T^4) and to their surface area. The energy radiated (or absorbed) also depends on the nature of the surface, which is characterized by the emissivity, ϵ (dark surfaces absorb and radiate more than do bright shiny ones).

Radiation from the Sun arrives at the surface of the Earth on a clear day at a rate of about $1000\,\text{W/m}^2$.

[†]Infrared radiation is light whose wavelengths are longer than visible light (see Fig. 31–12).

Questions

1. What happens to the work done on a bottle of orange juice when it is vigorously shaken?
2. When a hot object warms a cooler object, does temperature flow between them? Are the temperature changes of the two objects equal? Explain.
3. (*a*) If two objects of different temperature are placed in contact, will heat naturally flow from the object with higher internal energy to the object with lower internal energy? (*b*) Is it possible for heat to flow even if the internal energies of the two objects are the same? Explain.
4. The specific heat of water is quite large. Explain why this fact makes water particularly good for heating systems (that is, hot-water radiators).
5. In warm regions where tropical plants grow but the temperature may drop below freezing a few times in winter evenings, the destruction of sensitive plants due to freezing can be reduced by watering them in the evening. Explain.
6. Why does water in a metal canteen stay cooler if the cloth jacket surrounding the canteen is kept moist?
7. Explain why burns caused by steam at 100°C on the skin are often more severe than burns caused by water at 100°C.
8. Explain why water cools (its temperature drops) when it evaporates, using the concepts of latent heat and internal energy.
9. Will potatoes cook faster if the water is boiling more vigorously?
10. Very high in the Earth's atmosphere the temperature can be 700°C. Yet an animal there would freeze to death rather than roast. Explain.
11. Use the conservation of energy to explain why the temperature of a well-insulated gas increases when it is compressed—say, by pushing down on a piston—whereas the temperature decreases when the gas expands.
12. Explorers on failed Arctic expeditions have survived by covering themselves with snow. Why would they do that?
13. Why is wet sand at the beach cooler to walk on than dry sand?
14. When hot-air furnaces are used to heat a house, why is it important that there be a vent for air to return to the furnace? What happens if this vent is blocked by a bookcase?
15. Explain why the temperature of a gas increases when it is compressed adiabatically.
16. An ideal monatomic gas expands slowly to twice its volume (1) isothermally; (2) adiabatically; (3) isobarically. Plot each on a PV diagram. In which process is ΔE_{int} the greatest, and in which is ΔE_{int} the least? In which is W the greatest and the least? In which is Q the greatest and the least?
17. Ceiling fans are sometimes reversible, so that they drive the air down in one season and pull it up in another season. Explain which way you should set the fan (*a*) for summer, (*b*) for winter.
18. Goose down sleeping bags and parkas are often specified as so many inches or centimeters of *loft*, the actual thickness of the garment when it is fluffed up. Explain.
19. Microprocessor chips have a "heat sink" glued on top that looks like a series of fins. Why are they shaped like that?
20. Sea breezes are often encountered on sunny days at the shore of a large body of water. Explain, noting that the temperature of the land rises more rapidly than that of the nearby water.
21. The Earth cools off at night much more quickly when the weather is clear than when cloudy. Why?
22. Explain why air temperature readings are always taken with the thermometer in the shade.
23. A premature baby in an incubator can be dangerously cooled even when the air temperature in the incubator is warm. Explain.
24. A 22°C day is warm, while a swimming pool at 22°C feels cool. Why?
25. In the Northern Hemisphere the amount of heat required to heat a room where the windows face north is much higher than that required where the windows face south. Explain.
26. The floor of a house on a foundation under which the air can flow is often cooler than a floor that rests directly on the ground (such as a concrete slab foundation). Explain.
27. Why is the liner of a **thermos bottle** silvered (Fig. 19–28), and why does it have a vacuum between its two walls?

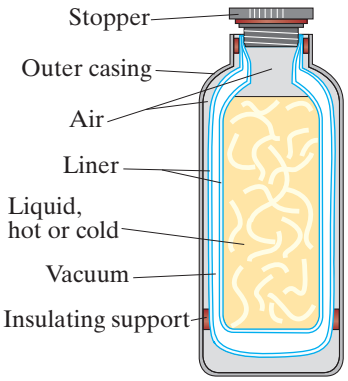


FIGURE 19–28
Question 27.

28. Heat loss occurs through windows by the following processes: (1) through the glass panes; (2) through the frame, particularly if it is metal; (3) ventilation leaks around edges; and (4) radiation. (*a*) For the first three, what is (are) the mechanism(s): conduction, convection, or radiation? (*b*) Heavy curtains reduce which of these heat losses? Explain in detail.
29. A paper cup placed among hot coals will burn if empty (note burn spots at top of cup in Fig. 19–29), but won't burn if filled with water. Explain. Forget the marshmallows.



FIGURE 19–29
Question 29.

30. On a cold windy day, a window will feel colder than on an equally cold day with no wind. This is true even if no air leaks in near the window. Why?
31. Early in the day, after the Sun has reached the slope of a mountain, there tends to be a gentle upward movement of air. Later, after a slope goes into shadow, there is a gentle downdraft. Explain.
32. A piece of wood lying in the Sun absorbs more heat than a piece of shiny metal. Yet the metal feels hotter than the wood when you pick it up. Explain.
33. An *emergency blanket* is a thin shiny (metal-coated) plastic foil. Explain how it can help to keep an immobile person warm.
34. Explain why cities situated by the ocean tend to have less extreme temperatures than inland cities at the same latitude.

MisConceptual Questions

1. When you put an ice cube in a glass of warm tea, which of the following happens?
(*a*) Cold flows from the ice cube into the tea.
(*b*) Cold flows from the ice cube into the tea and heat flows from the tea into the ice cube.
(*c*) Heat flows from the tea into the ice cube.
(*d*) Neither heat nor cold flows. Only temperature flows between the ice and the tea.

2. Both beakers A and B in Fig. 19–30 contain a mixture of ice and water at equilibrium. Which beaker is the coldest, or are they equal in temperature?
(*a*) Beaker A.
(*b*) Beaker B.
(*c*) Equal.

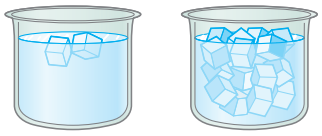


FIGURE 19–30
MisConceptual Question 2.

3. For objects at thermal equilibrium, which of the following is true?
(*a*) Each is at the same temperature.
(*b*) Each has the same internal energy.
(*c*) Each has the same heat.
(*d*) All of the above.
(*e*) None of the above.

4. As heat is slowly added to water, is it possible for the temperature measured by a thermometer in the water to remain constant?
(*a*) Yes, the water could be changing phase.
(*b*) No, adding heat will always change the temperature.
(*c*) Maybe; it depends on the rate at which the heat is added.
(*d*) Maybe; it depends on the initial water temperature.

5. Radiation is emitted
(*a*) only by glowing objects such as the Sun.
(*b*) only by objects whose temperature is greater than the temperature of the surroundings.
(*c*) only by objects with more caloric than their surroundings.
(*d*) by any object not at 0 K.[†]
(*e*) only by objects that have a large specific heat.

6. Ten grams of water is added to ten grams of ice in an insulated container. Will all of the ice melt?
(*a*) Yes.
(*b*) No.
(*c*) More information is needed.

7. Two objects are made of the same material, but they have different masses and temperatures. If the objects are brought into thermal contact, which one will have the greater temperature change?
(*a*) The one with the higher initial temperature.
(*b*) The one with the lower initial temperature.
(*c*) The one with the larger mass.
(*d*) The one with the smaller mass.
(*e*) The one with the higher specific heat.
(*f*) Not enough information.

8. It has been a hot summer, so when you arrive at a lake, you decide to go for a swim even though it is nighttime. The water is cold! The next day, you go swimming again during the hottest part of the day, and even though the air is warmer the water is still almost as cold. Why?
(*a*) Water is fairly dense compared with many other liquids.
(*b*) Water remains in a liquid state for a wide range of temperatures.
(*c*) Water has a high bulk modulus.
(*d*) Water has a high specific heat.

9. Two equal-mass liquids, initially at the same temperature, are heated for the same time over the same stove. You measure the temperatures and find that one liquid has a higher temperature than the other. Which liquid has the higher specific heat?
(*a*) The cooler one.
(*b*) The hotter one.
(*c*) Both are the same.

10. An ideal gas undergoes an isothermal expansion from state A to state B. In this process (use sign conventions, page 549),
(*a*) $Q = 0, \quad \Delta E_{\text{int}} = 0, \quad W > 0.$
(*b*) $Q > 0, \quad \Delta E_{\text{int}} = 0, \quad W < 0.$
(*c*) $Q = 0, \quad \Delta E_{\text{int}} > 0, \quad W > 0.$
(*d*) $Q > 0, \quad \Delta E_{\text{int}} = 0, \quad W > 0.$
(*e*) $Q = 0, \quad \Delta E_{\text{int}} < 0, \quad W < 0.$

11. An ideal gas undergoes an isothermal process. Which of the following statements are true? (i) No heat is added to or removed from the gas. (ii) The internal energy of the gas does not change. (iii) The average kinetic energy of the molecules does not change.
(*a*) (i) only.
(*b*) (i) and (ii) only.
(*c*) (i) and (iii) only.
(*d*) (ii) and (iii) only.
(*e*) (i), (ii), and (iii).
(*f*) None of the above.

12. An ideal gas is in an insulated cylinder with a movable piston on one end. The piston is slowly pushed down, compressing the gas. The piston is then released, allowing the gas to expand adiabatically. The piston accelerates, gaining kinetic energy. Where does the kinetic energy come from?
(*a*) From the heat developed by the expanding gas.
(*b*) From the decrease in internal energy of the expanding gas.
(*c*) From the increase in internal energy of the expanding gas.
(*d*) Both (*a*) and (*b*).
(*e*) Both (*a*) and (*c*).

13. An ideal gas undergoes an isobaric compression and then an isovolumetric process that brings it back to its initial temperature. Had the gas undergone *one isothermal* process instead of those two,
(*a*) the work done on the gas would be the same.
(*b*) the work done on the gas would be less.
(*c*) the work done on the gas would be greater.
(*d*) Need to know the temperature of the isothermal process.

[†]Which means *all* objects because 0 K is unattainable, as discussed in Section 20–10.

Problems

19–1 Heat as Energy Transfer

1. (I) To what temperature will 6800 J of heat raise 3.0 kg of water that is initially at 10.0°C?
2. (II) When a diver jumps into the ocean, water leaks into the gap region between the diver's skin and her wetsuit, forming a water layer about 0.5 mm thick. Assuming the total surface area of the wetsuit covering the diver is about 1.0 m^2, and that ocean water enters the suit at 10°C and is warmed by the diver to skin temperature of 35°C, estimate how much energy (in units of candy bars = 300 kcal) is required by this heating process.
3. (II) An average active person consumes about 2500 Cal a day. (*a*) What is this in joules? (*b*) What is this in kilowatt-hours? (*c*) If your power company charges about 12 ¢ per kilowatt-hour, how much would your energy cost per day if you bought it from the power company? Could you feed yourself on this much money per day?
4. (II) A British thermal unit (Btu) is a unit of heat in the British system of units. One Btu is defined as the heat needed to raise 1 lb of water by 1 F°. Show that

$$1\,\text{Btu} = 0.252\,\text{kcal} = 1056\,\text{J}.$$

5. (II) How many joules and kilocalories are generated when the brakes are used to bring a 1200-kg car to rest from a speed of 95 km/h?
6. (II) A water heater can generate 32,000 kJ/h. How much water can it heat from 12°C to 42°C per hour?
7. (II) A small immersion heater is rated at 375 W. Estimate how long it will take to heat a cup of soup (assume this is 250 mL of water) from 15°C to 75°C.

19–3 and 19–4 Specific Heat; Calorimetry

8. (I) An automobile cooling system holds 18 L of water. How much heat does the water absorb if its temperature rises from 15°C to 95°C?
9. (I) What is the specific heat of a metal substance if 165 kJ of heat is needed to raise 4.1 kg of the metal from 18.0°C to 37.2°C?
10. (II) Samples of copper, aluminum, and water experience the same temperature rise when they absorb the same amount of heat. What is the ratio of their masses? [*Hint*: See Table 19–1.]
11. (II) (*a*) How long does it take a 750-W coffeepot to bring to a boil 0.75 L of water at sea level initially at 11°C? Assume that the part of the pot which is heated with the water is made of 250 g of aluminum, and that no water boils away. (*b*) For how long could this amount of energy run a 60-W lightbulb?
12. (II) When a 290-g piece of iron at 180°C is placed in a 95-g aluminum calorimeter cup containing 250 g of glycerin at 10°C, the final temperature is observed to be 38°C. Estimate the specific heat of glycerin.
13. (II) What will be the equilibrium temperature when a 265-g block of copper at 245°C is placed in a 145-g aluminum calorimeter cup containing 825 g of water at 12.0°C?
14. (II) A 0.40-kg iron horseshoe just forged and very hot (Fig. 19–31), is dropped into 1.35 L of water in a 0.30-kg iron pot initially at 20.0°C. If the final equilibrium temperature is 25.0°C, estimate the initial temperature of the hot horseshoe.



FIGURE 19–31
Problem 14.

15. (II) Estimate the *Calorie content* of 65 g of candy from the following measurements. A 15-g sample of the candy is placed in a small aluminum container of mass 0.325 kg filled with oxygen. This container is placed in 1.75 kg of water in an aluminum calorimeter cup of mass 0.624 kg at an initial temperature of 15.0°C. The oxygen-candy mixture in the small container (a *bomb calorimeter*, page 545) is ignited, and the final temperature of the whole system is 53.5°C.
16. (II) The *heat capacity, C*, of an object is defined as the amount of heat needed to raise its temperature by 1 C°. Thus, to raise the temperature by ΔT requires heat Q given by

$$Q = C\,\Delta T.$$

(*a*) Write the heat capacity C in terms of the specific heat, c, of the material. (*b*) What is the heat capacity of 1.0 kg of water? (*c*) Of 38 kg of water?

17. (II) The 1.20-kg head of a hammer has a speed of 7.5 m/s just before it strikes a nail (Fig. 19–32) and is brought to rest. Estimate the temperature rise of a 14-g iron nail generated by eight such hammer blows done in quick succession. Assume the nail absorbs all the energy.

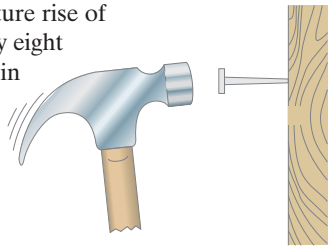


FIGURE 19–32
Problem 17.

18. (II) A 215-g sample of a substance is heated to 330°C and then plunged into a 105-g aluminum calorimeter cup containing 185 g of water and a 17-g glass thermometer at 10.5°C. The final temperature is 35.0°C. What is the specific heat of the substance? (Assume no water boils away.)
19. (II) A 0.095-kg aluminum sphere is dropped from the roof of a 55-m-high building. If 65% of the thermal energy produced when it hits the ground is absorbed by the sphere, what is its temperature increase?

19–5 Latent Heat

20. (I) If 3.40×10^5 J of energy is supplied to a container of liquid oxygen at −183°C, how much oxygen can evaporate?
21. (I) How much heat is needed to melt 26.50 kg of silver that is initially at 25°C?
22. (II) What mass of steam at 100°C must be added to 1.00 kg of ice at 0°C to yield liquid water at 30°C?

23. (II) During exercise, a person may give off 185 kcal of heat in 25 min by evaporation of water (at 20°C) from the skin. How much water has been lost? [*Hint*: See page 548.]

24. (II) A 28-g ice cube at its melting point is dropped into an insulated container of liquid nitrogen. How much nitrogen evaporates if it is at its boiling point of 77 K and has a latent heat of vaporization of 200 kJ/kg? Assume for simplicity that the specific heat of ice is a constant and is equal to its value near its melting point.

25. (II) High-altitude mountain climbers do not eat snow, but always melt it first with a stove. To see why, calculate the energy absorbed from your body if you (*a*) eat 1.0 kg of −15°C snow which your body warms to body temperature of 37°C; (*b*) melt 1.0 kg of −15°C snow using a stove and drink the resulting 1.0 kg of water at 2°C, which your body has to warm to 37°C.

26. (II) An iron boiler of mass 180 kg contains 710 kg of water at 18°C. A heater supplies energy at the rate of 58,000 kJ/h. How long does it take for the water (*a*) to reach the boiling point, and (*b*) to all have changed to steam?

27. (II) A bicyclist consumes 9.0 L of water over the span of 3.5 hours during a race. Making the approximation that 80% of the cyclist's energy goes into evaporating this water (at 20°C) as sweat, how much energy in kcal did the rider use during the ride? [*Hint*: See page 548].

28. (II) Determine the latent heat of fusion of mercury using the following calorimeter data: 1.00 kg of solid Hg at its melting point of −39.0°C is placed in a 0.620-kg aluminum calorimeter with 0.400 kg of water at 12.80°C; the resulting equilibrium temperature is 5.06°C.

29. (II) At a crime scene, the forensic investigator notes that the 6.2-g lead bullet that was stopped in a doorframe apparently melted completely on impact. Assuming the bullet was shot at room temperature (20°C), what does the investigator calculate as the minimum muzzle velocity of the gun?

30. (II) A 55-g bullet traveling at 250 m/s penetrates a block of ice at 0°C and comes to rest within the ice. Assuming that the temperature of the bullet doesn't change appreciably, how much ice is melted as a result of the collision?

31. (II) A cube of ice is taken from the freezer at −8.5°C and placed in an 85-g aluminum calorimeter filled with 310 g of water at room temperature of 20.0°C. The final situation is all water at 17.0°C. What was the mass of the ice cube?

19–6 and 19–7 First Law of Thermodynamics

[Recall sign conventions, page 549.]

32. (I) An ideal gas expands isothermally, performing 4.30×10^3 J of work in the process. Calculate (*a*) the change in internal energy of the gas, and (*b*) the heat absorbed during this expansion.

33. (I) One liter of air is cooled at constant pressure until its volume is halved, and then it is allowed to expand isothermally back to its original volume. Draw the process on a *PV* diagram.

34. (II) Sketch a *PV* diagram of the following process: 2.5 L of ideal gas at atmospheric pressure is cooled at constant pressure to a volume of 1.0 L, and then expanded isothermally back to 2.5 L, after which the pressure is increased at constant volume until the original pressure is reached.

35. (II) When 6.80×10^5 J of heat is added to a gas enclosed in a cylinder fitted with a light frictionless piston maintained at atmospheric pressure, the volume is observed to increase from 2.2 m^3 to 4.1 m^3. Calculate (*a*) the work done by the gas, and (*b*) the change in internal energy of the gas. (*c*) Graph this process on a *PV* diagram.

36. (II) The pressure in an ideal gas is cut in half slowly, while being kept in a container with rigid walls. In the process, 425 kJ of heat left the gas. (*a*) How much work was done during this process? (*b*) What was the change in internal energy of the gas during this process?

37. (II) A 64-kg ice skater moving at 8.5 m/s glides to a stop. Assuming the ice is at 0°C and that 50% of the heat generated by friction is absorbed by the ice, how much ice melts?

38. (II) A 1.0-L volume of air initially at 3.5 atm of (gauge) pressure is allowed to expand isothermally until the (gauge) pressure is 1.0 atm. It is then compressed at constant pressure to its initial volume, and lastly is brought back to its original pressure by heating at constant volume. (*a*) Draw the process on a *PV* diagram, including numbers and labels for the axes. (*b*) How much work does the 1.0 L of air do in this process?

39. (II) Consider the following two-step process. Heat is allowed to flow out of an ideal gas at constant volume so that its pressure drops from 2.2 atm to 1.4 atm. Then the gas expands at constant pressure, from a volume of 5.9 L to 9.3 L, where the temperature reaches its original value. Calculate (*a*) the total work done by the gas in the process, (*b*) the change in internal energy of the gas in the process, and (*c*) the total heat flow into or out of the gas.

40. (II) Suppose 3.20 mol of an ideal gas of volume $V_1 = 3.50\ m^3$ at $T_1 = 295$ K is allowed to expand isothermally to $V_2 = 7.00\ m^3$ at $T_2 = 295$ K. Determine (*a*) the work done by the gas, (*b*) the heat added to the gas, and (*c*) the change in internal energy of the gas.

41. (II) In an engine, an almost ideal gas is compressed adiabatically to half its volume. In doing so, 2630 J of work is done on the gas. (*a*) How much heat flows into or out of the gas? (*b*) What is the change in internal energy of the gas? (*c*) Does its temperature rise or fall?

42. (II) An ideal gas expands at a constant total pressure of 2.5 atm from 410 mL to 690 mL. Heat then flows out of the gas at constant volume, and the pressure and temperature are allowed to drop until the temperature reaches its original value. Calculate (*a*) the total work done by the gas in the process, and (*b*) the total heat flow into the gas.

43. (II) Determine (*a*) the work done and (*b*) the change in internal energy of 1.00 kg of water when it is all boiled to steam at 100°C. Assume a constant pressure of 1.00 atm.

44. (II) How much work is done by a pump to slowly compress, isothermally, 3.20 L of nitrogen at 0°C and 1.00 atm to 1.80 L at 0°C?

45. (III) Determine the work done by 1.00 mol of a van der Waals gas (Section 18–6) when it expands from volume V_1 to V_2 isothermally.

46. (III) When a gas is taken from a to c along the curved path in Fig. 19–33, the work done by the gas is $W = -35\,\text{J}$ and the heat added to the gas is $Q = -175\,\text{J}$. Along path abc, the work done by the gas is $W = -56\,\text{J}$. (That is, 56 J of work is done *on* the gas.) (*a*) What is Q for path abc? (*b*) If $P_c = \frac{1}{2}P_b$, what is W for path cda? (*c*) What is Q for path cda? (*d*) What is $E_{\text{int,a}} - E_{\text{int,c}}$? (*e*) If $E_{\text{int,d}} - E_{\text{int,c}} = 42\,\text{J}$, what is Q for path da?

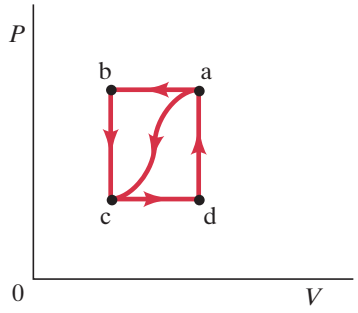


FIGURE 19–33
Problems 46, 47, 48.

47. (III) In the process of taking a gas from state a to state c along the curved path shown in Fig. 19–33, 85 J of heat leaves the system and 55 J of work is done *on* the system. (*a*) Determine the change in internal energy, $E_{\text{int,a}} - E_{\text{int,c}}$. (*b*) When the gas is taken along the path cda, the work done by the gas is $W = 38\,\text{J}$. How much heat Q is added to the gas in the process cda? (*c*) If $P_a = 2.2P_d$, how much work is done by the gas in the process abc? (*d*) What is Q for path abc? (*e*) If $E_{\text{int,a}} - E_{\text{int,b}} = 15\,\text{J}$, what is Q for the process bc? Here is a summary of what is given:

$$\begin{aligned} Q_{a\to c} &= -85\,\text{J} \\ W_{a\to c} &= -55\,\text{J} \\ W_{cda} &= 38\,\text{J} \\ E_{\text{int,a}} - E_{\text{int,b}} &= 15\,\text{J} \\ P_a &= 2.2P_d. \end{aligned}$$

48. (III) Suppose a gas is taken clockwise around the rectangular cycle shown in Fig. 19–33, starting at b, then to a, to d, to c, and returning to b. Using the values given in Problem 47, (*a*) describe each leg of the process, and then calculate (*b*) the net work done during the cycle, (*c*) the total internal energy change during the cycle, and (*d*) the net heat flow during the cycle. (*e*) What percentage of the *intake* heat was turned into usable work: i.e., how efficient is this "rectangular" cycle (give as a percentage)?

19–8 Molecular Specific Heat for Gases; Equipartition of Energy

49. (I) What is the internal energy of 5.40 mol of an ideal diatomic gas at 2450 K, assuming all possible degrees of freedom are active?

50. (I) If a heater supplies $1.8 \times 10^6\,\text{J/h}$ to a room $3.5\,\text{m} \times 4.6\,\text{m} \times 3.0\,\text{m}$ containing air at 20°C and 1.0 atm, by how much will the temperature rise in one hour, assuming no losses of heat or air mass to the outside? Assume air is an ideal diatomic gas with molecular mass 29.

51. (II) Show that if the molecules of a gas have f degrees of freedom, then theory predicts $C_V = \frac{1}{2}fR$ and $C_P = \frac{1}{2}(f+2)R$.

52. (II) A certain monatomic gas has specific heat $c_V = 0.0356\,\text{kcal/kg}\cdot\text{C}^\circ$, which changes little over a wide temperature range. What is the atomic mass of this gas? What gas is it?

53. (II) Show that the work done by n moles of an ideal gas when it expands adiabatically is $W = nC_V(T_1 - T_2)$, where T_1 and T_2 are the initial and final temperatures, and C_V is the molar specific heat at constant volume.

54. (II) An audience of 1600 fills a concert hall of volume $22{,}000\,\text{m}^3$. If there were no ventilation, by how much would the temperature of the air rise over a period of 2.0 h due to the metabolism of the people (70 W/person)?

55. (II) The specific heat at constant volume of a particular gas is $0.182\,\text{kcal/kg}\cdot\text{K}$ at room temperature, and its molecular mass is 34. (*a*) What is its specific heat at constant pressure? (*b*) What do you think is the molecular structure of this gas?

56. (II) A 2.00-mole sample of N_2 gas at 0°C is heated to 150°C at constant pressure (1.00 atm). Determine (*a*) the change in internal energy, (*b*) the work the gas does, and (*c*) the heat added to it.

57. (III) A 1.00-mol sample of an ideal diatomic gas at a pressure of 1.00 atm and temperature of 460 K undergoes a process in which its pressure increases linearly with temperature ($P = P_0 + aT$). The final temperature and pressure are 720 K and 1.60 atm. (*a*) First graph the process on a PV diagram. Then determine (*b*) the change in internal energy, (*c*) the work done by the gas, and (*d*) the heat added to the gas. (Assume five active degrees of freedom.)

19–9 Adiabatic Expansion of a Gas

58. (II) A 1.00-mol sample of an ideal diatomic gas, originally at 1.00 atm and 20°C, expands adiabatically to 1.75 times its initial volume. What are the final pressure and temperature for the gas? (Assume no molecular vibration.)

59. (II) Show, using Eqs. 19–7 and 19–16, that the work done by a gas that slowly expands adiabatically from pressure P_1 and volume V_1, to P_2 and V_2, is given by $W = (P_1V_1 - P_2V_2)/(\gamma - 1)$.

60. (II) A 3.65-mol sample of an ideal diatomic gas expands adiabatically from a volume of $0.1210\,\text{m}^3$ to $0.750\,\text{m}^3$. Initially the pressure was 1.00 atm. Determine: (*a*) the initial and final temperatures; (*b*) the change in internal energy; (*c*) the heat lost by the gas; (*d*) the work done *on* the gas. (Assume no molecular vibration.)

61. (II) An ideal monatomic gas, consisting of 2.5 mol of volume $0.086\,\text{m}^3$, expands adiabatically. The initial and final temperatures are 25°C and −68°C. What is the final volume of the gas?

62. (III) A 1.00-mol sample of an ideal monatomic gas, originally at a pressure of 1.00 atm, undergoes a three-step process: (1) it is expanded adiabatically from $T_1 = 588\,\text{K}$ to $T_2 = 389\,\text{K}$; (2) it is compressed at constant pressure until its temperature reaches T_3; (3) it then returns to its original pressure and temperature by a constant-volume process. (*a*) Plot these processes on a PV diagram. (*b*) Determine T_3. (*c*) Calculate the change in internal energy, the work done by the gas, and the heat added to the gas for each process, and (*d*) for the complete cycle.

63. (III) Consider a **parcel of air** moving to a different altitude y in the Earth's atmosphere (Fig. 19–34). As the parcel changes altitude it acquires the pressure P of the surrounding air. From Eq. 13–4 we have

$$\frac{dP}{dy} = -\rho g$$

where ρ is the parcel's altitude-dependent mass density.

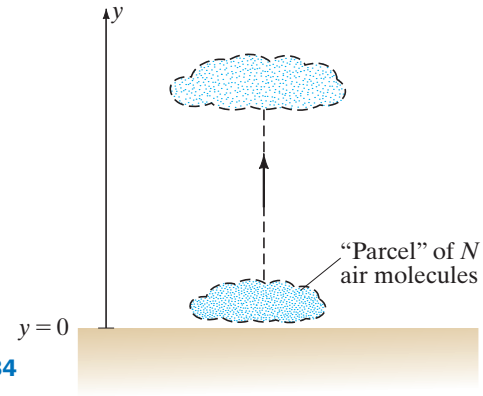


FIGURE 19–34
Problem 63.

During this motion, the parcel's volume will change and, because air is a poor heat conductor, we assume this expansion or contraction will take place adiabatically. (*a*) Starting with Eq. 19–16, $PV^\gamma = \text{constant}$, show that for an ideal gas undergoing an adiabatic process, $P^{1-\gamma}T^\gamma = \text{constant}$. Then show that the parcel's pressure and temperature are related by

$$(1-\gamma)\frac{dP}{dy} + \gamma\frac{P}{T}\frac{dT}{dy} = 0$$

and thus

$$(1-\gamma)(-\rho g) + \gamma\frac{P}{T}\frac{dT}{dy} = 0.$$

(*b*) Use the ideal gas law with the result from part (*a*) to show that the change in the parcel's temperature with change in altitude is given by

$$\frac{dT}{dy} = \frac{1-\gamma}{\gamma}\frac{mg}{k}$$

where m is the average mass of an air molecule and k is the Boltzmann constant. (*c*) Given that air is a diatomic gas with an average molecular mass of 29, show that $dT/dy = -9.8\,\text{C°/km}$. This value is called the **adiabatic lapse rate** for dry air. (*d*) In California, the prevailing westerly winds descend from one of the highest elevations (the 4000-m Sierra Nevada mountains) to one of the lowest elevations (Death Valley, -100 m) in the continental United States. If a dry wind has a temperature of -5°C at the top of the Sierra Nevada, what is the wind's temperature after it has descended to Death Valley?

19–10 Conduction, Convection, Radiation

64. (I) One end of a 64-cm-long copper rod with a diameter of 2.0 cm is kept at 460°C, and the other is immersed in water at 22°C. Calculate the heat conduction rate along the rod.

65. (I) Calculate the rate of heat flow by conduction through the windows of Example 19–16, assuming that there are strong gusty winds and the external temperature is -5°C.

66. (I) (*a*) How much power is radiated by a tungsten sphere (emissivity $\epsilon = 0.35$) of radius 19 cm at a temperature of 25°C? (*b*) If the sphere is enclosed in a room whose walls are kept at -5°C, what is the *net* flow rate of energy out of the sphere?

67. (II) How long does it take the Sun to melt a block of ice at 0°C with a flat horizontal area 1.0 m^2 and thickness 1.0 cm? Assume that the Sun's rays make an angle of 35° with the vertical and that the emissivity of ice is 0.050.

68. (II) **Heat conduction to skin**. Suppose 150 W of heat flows by conduction from the blood capillaries beneath the skin to the body's surface area of 1.5 m^2. If the temperature difference is 0.50 C°, estimate the average distance of capillaries below the skin surface.

69. (II) Two rooms, each a cube 4.0 m per side, share a 14-cm-thick brick wall. Because of a number of 100-W lightbulbs in one room, the air is at 30°C, while in the other room it is at 10°C. How many of the 100-W bulbs are needed to maintain the temperature difference across the wall?

70. (II) A ceramic teapot ($\epsilon = 0.70$) and a shiny metal one ($\epsilon = 0.10$) each hold 0.55 L of tea at 85°C. (*a*) Estimate the rate of heat loss from each, and (*b*) estimate the temperature drop after 30 min for each. Consider only radiation, and assume the surroundings are at 20°C.

71. (II) A copper rod and an aluminum rod of the same length and cross-sectional area are attached end to end (Fig. 19–35). The copper end is placed in a furnace maintained at a constant temperature of 205°C. The aluminum end is placed in an ice bath held at a constant temperature of 0.0°C. Calculate the temperature at the point where the two rods are joined.

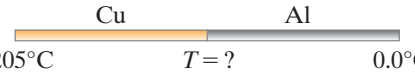


FIGURE 19–35
Problem 71.

72. (III) A cylindrical pipe has inner radius R_1 and outer radius R_2. The interior of the pipe carries hot water at temperature T_1. The temperature outside is T_2 $(< T_1)$. (*a*) Show that the rate of heat loss for a length ℓ of pipe is

$$\frac{dQ}{dt} = \frac{2\pi k(T_1 - T_2)\ell}{\ln(R_2/R_1)},$$

where k is the thermal conductivity of the pipe. (*b*) Suppose the pipe is steel with $R_1 = 3.3$ cm, $R_2 = 4.0$ cm, and $T_2 = 18$°C. If the pipe holds still water at $T_1 = 71$°C, what will be the initial rate of change of its temperature? (*c*) Suppose water at 71°C enters the pipe and moves at a speed of 8.0 cm/s. What will be its temperature drop per centimeter of travel?

***73.** (III) Suppose the insulating qualities of the wall of a house come mainly from a 4.0-in. layer of brick and an R-19 layer of insulation, as shown in Fig. 19–36. What is the total rate of heat loss through such a wall, if its total area is 195 ft^2 and the temperature difference across it is 35 F°? Give result both in British and in SI units.

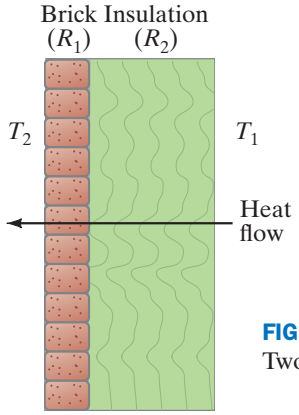


FIGURE 19–36 Problem 73.
Two layers insulating a wall.

General Problems

74. A soft-drink can contains about 0.35 kg of liquid at 5°C. Drinking this liquid can actually consume some of the fat in the body, since energy is needed to warm the liquid to body temperature (37°C). How many food Calories should the drink have so that it is in perfect balance with the heat needed to warm the liquid (essentially water)?

75. (*a*) Estimate the total power radiated into space by the Sun, assuming it to be a perfect emitter at $T = 5500$ K. The Sun's radius is 7.0×10^8 m. (*b*) From this, determine the power per unit area arriving at the Earth, 1.5×10^{11} m away (Fig. 19–37).

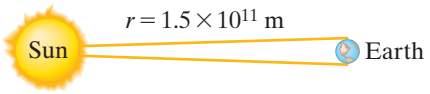


FIGURE 19–37
Problem 75.

76. To get an idea of how much thermal energy is contained in the world's oceans, estimate the heat liberated when a cube of ocean water, 1 km on each side, is cooled by 1 K. (Approximate the ocean water as pure water for this estimate.)

77. What will be the final result when equal masses of ice at 0°C and steam at 100°C are mixed together?

78. During light activity, a 70-kg person may generate 200 kcal/h. Assuming that 20% of this goes into useful work and the other 80% is converted to heat, estimate the temperature rise of the body after 45 min if none of this heat is transferred to the environment.

79. A mountain climber wears a goose-down jacket 3.8 cm thick with total surface area 0.95 m^2. The temperature at the surface of the clothing is −18°C and at the skin is 34°C. Determine the rate of heat flow by conduction through the jacket assuming (*a*) it is dry and the thermal conductivity k is that of goose down, and (*b*) the jacket is wet, so k is that of water and the jacket has matted to 0.50 cm thickness.

80. A marathon runner has an average metabolism rate of about 950 kcal/h during a race. If the runner has a mass of 62 kg, estimate how much water she would lose to evaporation from the skin for a race that lasts 2.2 h.

81. A house has well-insulated walls 19.5 cm thick (assume conductivity of air) and area 410 m^2, a roof of wood 5.5 cm thick and area 250 m^2, and uncovered windows 0.65 cm thick and total area 33 m^2. (*a*) Assuming that heat is lost only by conduction, calculate the rate at which heat must be supplied to this house to maintain its inside temperature at 23°C if the outside temperature is −15°C. (*b*) If the house is initially at 15°C, estimate how much heat must be supplied to raise the temperature to 23°C within 30 min. Assume that only the air needs to be heated and that its volume is 750 m^3. (*c*) If natural gas costs \$0.080/kg and its heat of combustion is 5.4×10^7 J/kg, what is the monthly cost to maintain the house as in part (*a*) for 24 h each day, assuming 90% of the heat produced is used to heat the house? Take the specific heat of air to be 0.24 kcal/kg · C°.

82. Estimate the rate at which heat can be conducted from the interior of the body to the surface. As a model, assume that the thickness of tissue is 4.0 cm, that the skin is at 34°C and the interior at 37°C, and that the surface area is 1.5 m^2. Compare this to the measured value of about 230 W that must be dissipated by a person working lightly. This clearly shows the necessity of *convective cooling* by the blood.

83. In a typical squash game (Fig. 19–38), two people hit a soft rubber ball at a wall. Assume that the ball hits the wall at a velocity of 22 m/s and bounces back at a velocity of 11 m/s, and that the kinetic energy lost in the process heats the ball. What will be the temperature increase of the ball after one bounce? (The specific heat of rubber is about 1200 J/kg · C°.)



FIGURE 19–38
Problem 83.

84. A bicycle pump is a cylinder 22 cm long and 3.0 cm in diameter. The pump contains air at 20.0°C and 1.0 atm. If the outlet at the base of the pump is blocked and the handle is pushed in very quickly, compressing the air to half its original volume, how hot does the air in the pump become?

85. A microwave oven is used to heat 250 g of water. On its maximum setting, the oven can raise the temperature of the liquid water from 20°C to 100°C in 1 min 45 s (= 105 s). (*a*) At what rate does the oven put energy into the liquid water? (*b*) If the power input from the oven to the water remains constant, determine how many grams of water will boil away if the oven is operated for 2 min (rather than just 1 min 45 s).

86. The temperature within the Earth's crust increases about 1.0 C° for each 30 m of depth. The thermal conductivity of the crust is 0.80 J/s · C° · m. (*a*) Determine the heat transferred from the interior to the surface for the entire Earth in 1.0 h. (*b*) Compare this heat to the 1000 W/m^2 that reaches the Earth's surface in 1.0 h from the Sun.

87. An ice sheet forms on a lake. The air above the sheet is at −18°C, whereas the water is at 0°C. Assume that the heat of fusion of the water freezing on the lower surface is conducted through the sheet to the air above. How much time will it take to form a sheet of ice 18 cm thick?

88. An iron meteorite melts when it enters the Earth's atmosphere. If its initial temperature was −105°C outside of Earth's atmosphere, calculate the minimum velocity the meteorite must have had before it entered Earth's atmosphere.

89. A 0.15-mole sample of ideal monatomic gas is in a balloon at an absolute pressure of 1.10 atm. Initially, the volume of the balloon is 3.0 liters. The balloon is gently heated until the rms speed of the atoms has increased by 16%. The pressure does not change appreciably during the expansion. How much heat flows into the gas?

90. A reciprocating compressor is a device that compresses air by a back-and-forth straight-line motion, like a piston in a cylinder. Consider a reciprocating compressor running at 120 rpm. During a compression stroke, 1.00 mol of air is compressed. The initial temperature of the air is 390 K, the engine of the compressor is supplying 7.5 kW of power to compress the air, and heat is being removed at the rate of 1.5 kW. Calculate the temperature change per compression stroke.

91. A scuba diver releases a 3.60-cm-diameter (spherical) bubble of air from a depth of 14.0 m. Assume the temperature is constant at 298 K, and that the air behaves as an ideal gas. (*a*) How large is the bubble when it reaches the surface? (*b*) Sketch a *PV* diagram for the process. (*c*) Apply the first law of thermodynamics to the bubble, and find the work done by the air in rising to the surface, the change in its internal energy, and the heat added or removed from the air in the bubble as it rises. Take the density of water to be $1000\ \text{kg/m}^3$.

92. Suppose 3.0 mol of neon (a monatomic gas, assume ideal) at STP are compressed slowly and isothermally to 0.22 the original volume. The gas is then allowed to expand quickly and adiabatically back to its original volume. Find the highest and lowest temperatures and pressures attained by the gas, and show on a *PV* diagram where these values occur.

93. The temperature of the glass surface of a 75-W lightbulb is 75°C when the room temperature is 18°C. Estimate the temperature of a 150-W lightbulb with a glass bulb the same size. Consider only radiation, and assume that 90% of the energy is emitted as heat.

94. A *diesel engine* accomplishes ignition without a spark plug by an adiabatic compression of air to a temperature above the ignition temperature of the diesel fuel, which is injected into the cylinder at the peak of the compression. Suppose air is taken into the cylinder at 280 K and volume V_1 and is compressed adiabatically to $560°\text{C} (\approx 1000°\text{F})$ and volume V_2. Assuming that the air behaves as an ideal gas whose ratio of C_P to C_V is 1.4, calculate the compression ratio V_1/V_2 of the engine.

95. At very low temperatures, the molar specific heat of many substances varies as the cube of the absolute temperature:

$$C = k\frac{T^3}{T_0^3},$$

which is sometimes called Debye's law. For rock salt, $T_0 = 281\ \text{K}$ and $k = 1940\ \text{J/mol} \cdot \text{K}$. Determine the heat needed to raise 2.75 mol of salt from 22.0 K to 46.0 K.

96. In a cold environment, a person can lose heat by conduction and radiation at a rate of about 200 W. Estimate how long it would take for the body temperature to drop from 36.6°C to 35.6°C if metabolism were nearly to stop. Assume a mass of 65 kg. (See Table 19–1.)

97. (*a*) Using the solar constant, estimate the rate at which the whole Earth receives energy from the Sun. (*b*) Assume the Earth radiates an equal amount back into space (that is, the Earth is in equilibrium). Then, assuming the Earth is a perfect emitter $(\epsilon = 1.0)$, estimate its average surface temperature. [*Hint*: Discuss why you use area $A = \pi r_E^2$ or $A = 4\pi r_E^2$ in each part.]

98. Approximately how long should it take 8.2 kg of ice at 0°C to melt when it is placed in a carefully sealed Styrofoam ice chest of dimensions $25\ \text{cm} \times 35\ \text{cm} \times 55\ \text{cm}$ whose walls are 1.5 cm thick? Assume that the conductivity of Styrofoam is double that of air and that the outside temperature is 34°C.

99. Calculate what will happen when 1000 J of heat is added to 100 grams of (*a*) ice at −20°C, (*b*) ice at 0°C, (*c*) water at 10°C, (*d*) water at 100°C, and (*e*) steam at 110°C.

100. A 12-g lead bullet traveling at 220 m/s passes through a thin wall and emerges at a speed of 160 m/s. If the bullet absorbs 50% of the heat generated, (*a*) what will be the temperature rise of the bullet? (*b*) If the bullet's initial temperature was 20°C, will any of the bullet melt, and if so, how much?

101. A leaf of area $40\ \text{cm}^2$ and mass $4.5 \times 10^{-4}\ \text{kg}$ directly faces the Sun on a clear day. The leaf has an emissivity of 0.85 and a specific heat of $0.80\ \text{kcal/kg} \cdot \text{K}$. (*a*) Estimate the energy absorbed per second by the leaf from the Sun, and then (*b*) estimate the rate of rise of the leaf's temperature. (*c*) Will the temperature rise continue for hours? Why or why not? (*d*) Calculate the temperature the leaf would reach if it lost all its heat by radiation to the surroundings at 24°C. (*e*) In what other ways can the heat be dissipated by the leaf?

102. Using the result of part (*a*) in Problem 101, take into account radiation from the leaf to calculate how much water must be transpired (evaporated) by the leaf per hour to maintain a temperature of 35°C.

103. A house thermostat is normally set to 22°C, but at night it is turned down to 16°C for 9.0 h. Estimate how much more heat would be needed (state as a percentage of daily usage) if the thermostat were not turned down at night. Assume that the outside temperature averages 0°C for the 9.0 h at night and 8°C for the remainder of the day, and that the heat loss from the house is proportional to the temperature difference inside and out. To obtain an estimate from the data, you must make other simplifying assumptions; state what these are.

104. *Metabolizing* 1.0 kg of fat results in about $3.7 \times 10^7\ \text{J}$ of internal energy in the body. (*a*) In one day, how much fat does the body burn to maintain the body temperature of a person staying in bed and metabolizing at an average rate of 95 W? (*b*) How long would it take to burn 1.0 kg of fat this way assuming there is no food intake?

ANSWERS TO EXERCISES

A: (*b*).
B: (*c*).
C: The phase change, liquid to ice (second process). The latent heat is large.
D: 0.21 kg.
E: 700 J.
F: Less.
G: $-6.8 \times 10^3\ \text{J}$.
H: The drapes trap a layer of air between the inside of the window and the room, which acts as an excellent insulator.
I: (*d*).

There are many uses for a heat engine such as a coal or gas burning power plant, or a steam locomotive. The photograph shows an old locomotive producing steam that does work on a piston which moves linkage to turn the locomotive wheels. The efficiency of any engine is limited by nature as described in the second law of thermodynamics. This great law is best stated in terms of a quantity called entropy, which is *not* conserved, but instead is constrained always to increase in any real process. Entropy is a measure of disorder. The second law of thermodynamics tells us that as time moves forward, the disorder in the universe increases.

We also discuss practical matters such as heat engines, heat pumps, refrigerators, and air conditioners.



CHAPTER 20

Second Law of Thermodynamics

CONTENTS

20 – 1 The Second Law of Thermodynamics—Introduction
20 – 2 Heat Engines
20 – 3 The Carnot Engine; Reversible and Irreversible Processes
20 – 4 Refrigerators, Air Conditioners, and Heat Pumps
20 – 5 Entropy
20 – 6 Entropy and the Second Law of Thermodynamics
20 – 7 Order to Disorder
20 – 8 Unavailability of Energy; Heat Death
20 – 9 Statistical Interpretation of Entropy and the Second Law
*20–10 Thermodynamic Temperature; Third Law of Thermodynamics
20–11 Thermal Pollution, Global Warming, and Energy Resources

CHAPTER-OPENING QUESTION—Guess now!

Fossil-fuel electric generating plants produce "thermal pollution." Part of the heat produced by the burning fuel is not converted to electric energy. The reason for this waste is

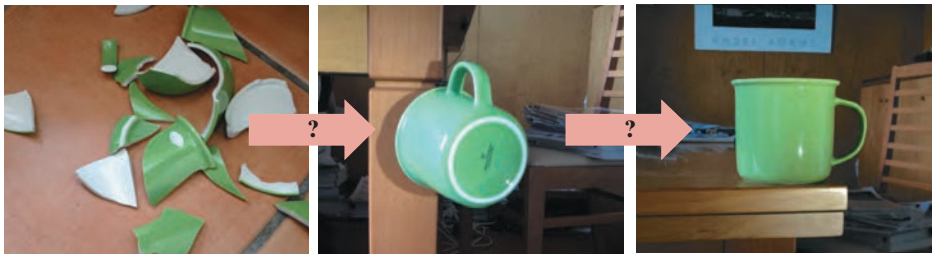
(a) the efficiency is higher if some heat is allowed to escape.
(b) engineering technology has not yet reached the point where 100% waste heat recovery is possible.
(c) when converting heat to useful work, you cannot avoid producing waste heat. This is a fundamental property of nature.
(d) the plants rely on fossil fuels, not nuclear fuel.
(e) None of the above.

In this final Chapter on heat and thermodynamics, we discuss the famous second law of thermodynamics, and the quantity "entropy" that arose from this fundamental law and is its quintessential expression. We also discuss heat engines—the useful engines that transform heat into work in power plants, trains, and motor vehicles, which first showed us that a new law was needed. Finally, we briefly discuss the third law of thermodynamics.

20–1 The Second Law of Thermodynamics—Introduction

The first law of thermodynamics states that energy is conserved. There are, however, many processes we can imagine that conserve energy but are not observed to occur in nature. For example, when a hot object is placed in contact with a cold object, heat flows from the hotter one to the colder one, never spontaneously from colder to hotter. If heat were to leave the colder object and pass to the hotter one, energy could still be conserved. Yet it does not happen spontaneously.[†] As a second example, consider what happens when you drop a rock and it hits the ground. The initial potential energy of the rock changes to kinetic energy as the rock falls. When the rock hits the ground, this energy in turn is transformed into internal energy of the rock and the ground in the vicinity of the impact; the molecules move faster and the temperature rises slightly. But have you seen the reverse happen—a rock at rest on the ground suddenly rise up in the air because the thermal energy of molecules is transformed into kinetic energy of the rock as a whole? Energy could be conserved in this process, yet we never see it happen.

There are many other examples of processes that occur in nature but whose reverse does not. Here are two more. (1) If you put a layer of salt in a jar and cover it with a layer of similar-sized grains of pepper, when you shake it you get a thorough mixture. But no matter how long you shake it, the mixture does not separate into two layers again. (2) Coffee cups and glasses break if you drop them. But they do not go back together spontaneously (Fig. 20–1).



(a) Initial state. (b) Later: cup reassembles and rises up. (c) Later still: cup lands on table.

FIGURE 20–1 Have you ever observed this process, a broken cup spontaneously reassembling and rising up onto a table? This process could conserve energy and other laws of mechanics.

The first law of thermodynamics (conservation of energy) would not be violated if any of these processes occurred in reverse. To explain this lack of reversibility, scientists in the latter half of the nineteenth century formulated a new principle known as the second law of thermodynamics.

The **second law of thermodynamics** is a statement about which processes occur in nature and which do not. It can be stated in a variety of ways, all of which are equivalent. One statement, due to R. J. E. Clausius (1822–1888), is that

> **heat can flow spontaneously from a hot object to a cold object; heat will not flow spontaneously from a cold object to a hot object.**

SECOND LAW OF THERMODYNAMICS (Clausius statement)

Since this statement applies to one particular process, it is not obvious how it applies to other processes. A more general statement is needed that will include other possible processes in a more obvious way.

The development of a general statement of the second law of thermodynamics was based partly on the study of heat engines. A **heat engine** is any device that changes thermal energy into mechanical work, such as a steam engine or an automobile engine. We now examine heat engines, both from a practical point of view and to show their importance in developing the second law of thermodynamics.

[†]By spontaneously, we mean by itself without input of work of some sort. (A refrigerator does move heat from a cold environment to a warmer one, but only because its motor does work—Section 20–4.)

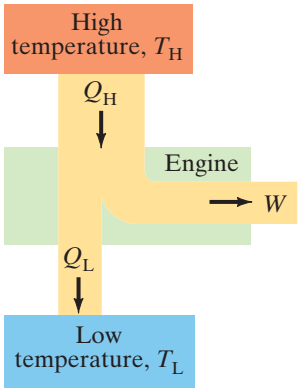


FIGURE 20–2 Schematic diagram of energy transfers for a heat engine.

CAUTION

Sign convention only for heat engines: $Q_H > 0,\ Q_L > 0,\ W > 0$

20–2 Heat Engines

It is easy to produce thermal energy by doing work—for example, by simply rubbing your hands together briskly, or indeed by any frictional process. But to get work from thermal energy is more difficult, and a practical device was invented only about 1700, the steam engine.

The basic idea behind any heat engine is that mechanical energy can be obtained from thermal energy only when heat is allowed to flow from a high temperature to a low temperature. In the process, some of the heat can then be transformed to mechanical work, as diagrammed schematically in Fig. 20–2. Useful heat engines run in a *repeating cycle*: that is, the system returns repeatedly to its starting point and thus can run continuously. In each cycle the change in internal energy of the system is $\Delta E_{int} = 0$ because it returns to the starting state. Thus a heat input Q_H at a high temperature T_H is partly transformed into work W and partly exhausted as heat Q_L at a lower temperature T_L (Fig. 20–2). By conservation of energy, $Q_H = W + Q_L$. The high and low temperatures, T_H and T_L, are called the **operating temperatures** of the engine. Note that we are now using a new (and intuitive) sign convention, but only for heat engines: we take Q_H, Q_L, and W as always positive. The direction of each energy transfer is shown by the arrow on the applicable diagram, such as Fig. 20–2.

PHYSICS APPLIED

Engines

Steam Engine and Internal Combustion Engine

The operation of a steam engine is illustrated in Fig. 20–3. Steam engines are of two main types, each using steam heated by combustion of coal, oil or gas, or by nuclear energy. In a reciprocating engine, Fig. 20–3a, the heated steam passes through the intake valve and expands against a piston, forcing it to move. As the piston returns to its original position, it forces the gases out the exhaust valve. A steam turbine, Fig. 20–3b, is very similar except that the reciprocating piston is replaced by a rotating turbine that resembles a paddlewheel with many sets of blades. Most of our electricity today is generated using steam turbines.[†] The material that is heated and cooled, steam in this case, is called the **working substance**. In an old-time steam engine (see page 576), the high temperature is obtained by burning wood, coal, oil, or other fuel to heat the steam.

[†]Even nuclear power plants utilize steam turbines; the nuclear fuel—uranium—serves as fuel to heat the steam.

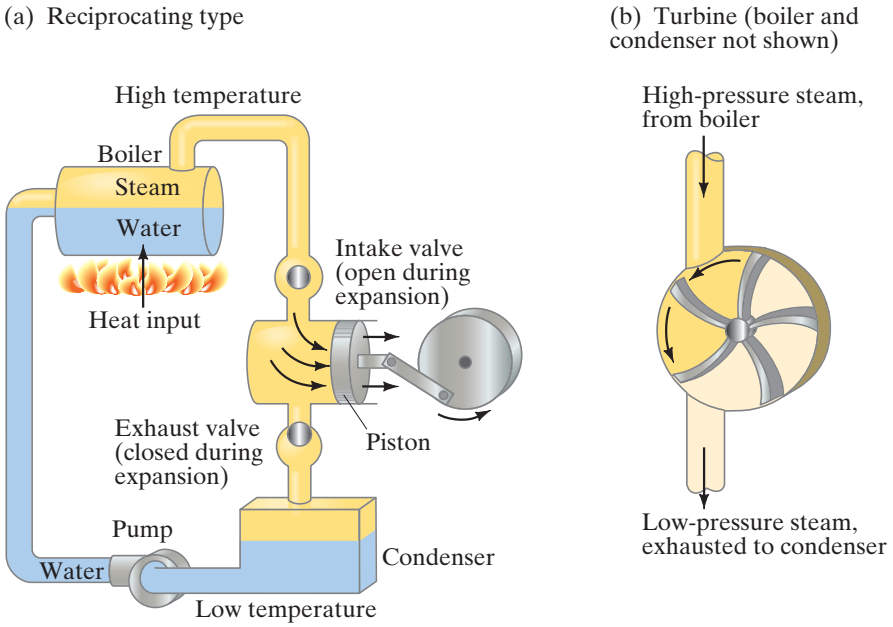


FIGURE 20–3 Steam engines.

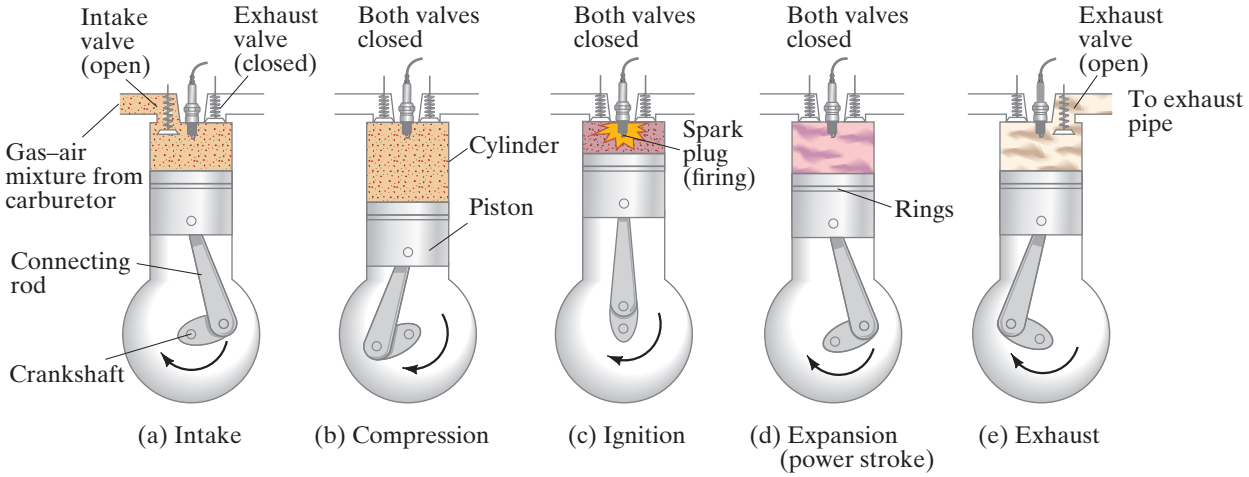


FIGURE 20–4 Four-stroke-cycle internal combustion engine: (a) the gasoline–air mixture flows into the cylinder as the piston moves down; (b) the piston moves upward and compresses the gas; (c) the brief instant when firing of the spark plug ignites the highly compressed gasoline–air mixture, raising it to a high temperature; (d) the gases, now at high temperature and pressure, expand against the piston in this, the power stroke; (e) the burned gases are pushed out to the exhaust pipe. When the piston reaches the top, the exhaust valve closes and the intake valve opens, and the whole cycle repeats. (a), (b), (d), and (e) are the four strokes of the cycle. The piston goes down and up twice per complete cycle.

In an internal combustion engine (used in most automobiles), the high temperature is achieved by burning the gasoline–air mixture in the cylinder itself (ignited by the spark plug), as described in Fig. 20–4.

Why a ΔT Is Needed to Drive a Heat Engine

To see why a *temperature difference* is required to run an engine, consider a steam engine. In the reciprocating engine, for example, suppose there were no condenser or pump (Fig. 20–3a), and that the steam was at the same temperature throughout the system. Then the pressure of the gas being exhausted would be the same as on intake. The work done by the gas *on* the piston when it expanded would equal the amount of work done *by* the piston to force the steam out the exhaust; hence, no net work would be done. Useless! In a real engine, the exhausted gas is cooled to a lower temperature and condensed so that the exhaust pressure is less than the intake pressure. Thus, the work the piston must do on the gas to expel it on the exhaust stroke is less than the work done by the gas on the piston during the intake. So a net amount of work can be obtained—but only if there is a difference of temperature (which allows a difference in pressure). Similarly, in the gas turbine (Fig. 20–3b) if the gas isn't cooled, the pressure on each side of the blades would be the same. By cooling the gas on the exhaust side, the pressure on the back side of the blade is less and hence the turbine turns.

Efficiency and the Second Law

The **efficiency**, e, of any heat engine can be defined as the ratio of the work it does, W, to the heat input at the high temperature, Q_H (Fig. 20–2), over a complete cycle:

$$e = \frac{W}{Q_H}. \qquad \textbf{(20–1a)}$$

This is a sensible definition since W is the output (what you get from the engine), whereas Q_H is what you put in and pay for in burned fuel. Since energy is conserved, the heat input Q_H must equal the work done plus the heat that flows out at the low temperature (Q_L):

$$Q_H = W + Q_L.$$

Thus $W = Q_H - Q_L$, and the efficiency of an engine can also be written as

$$e = \frac{W}{Q_H} = \frac{Q_H - Q_L}{Q_H}$$

or

$$e = 1 - \frac{Q_L}{Q_H}. \qquad \textbf{(20–1b)}$$

To give the efficiency as a percent, we multiply Eqs. 20–1 by 100. Note that e could be 1.0 (or 100%) only if Q_L were zero—that is, only if no heat were exhausted to the environment (which we will see shortly never happens).

EXAMPLE 20–1 **Car efficiency.** An automobile engine has an efficiency of 20% and produces an average of 23,000 J of mechanical work per second (= 23 kW) during operation. (*a*) How much heat input is required, and (*b*) how much heat is discharged as waste heat from this engine, per second?

APPROACH We want to find the heat input Q_H as well as the heat output Q_L, given $W = 23{,}000\ \text{J}$ each second and an efficiency $e = 0.20$. We can use the definition of efficiency, Eq. 20–1 in its various forms, to find first Q_H and then Q_L.

SOLUTION (*a*) From Eq. 20–1a, $e = W/Q_H$, we solve for Q_H:

$$Q_H = \frac{W}{e} = \frac{23{,}000\ \text{J}}{0.20} = 1.15 \times 10^5\ \text{J} = 115\ \text{kJ}.$$

The engine requires $115\ \text{kJ/s} = 115\ \text{kW}$ of heat input.

(*b*) Now we use Eq. 20–1b $(e = 1 - Q_L/Q_H)$ and solve for Q_L:

$$Q_L = (1 - e)Q_H = (0.80)115\ \text{kJ} = 92\ \text{kJ}.$$

The engine discharges heat to the environment at a rate of $92\ \text{kJ/s} = 92\ \text{kW}$.

NOTE Of the 115 kJ that enters the engine per second, only 23 kJ (20%) does useful work whereas 92 kJ (80%) is wasted as heat output.

NOTE The problem was stated in terms of energy per unit time. We could just as well have stated it in terms of power, since $1\ \text{J/s} = 1\ \text{W}$.

20–3 The Carnot Engine; Reversible and Irreversible Processes

To explore how to increase efficiency, the French scientist Sadi Carnot (1796–1832) examined the characteristics of an ideal engine, now called a **Carnot engine**. No Carnot engine actually exists, but as a theoretical idea it played an important role in the development and understanding of the second law of thermodynamics.

The idealized Carnot engine consists of four processes done in a cycle, two of which are adiabatic $(Q = 0)$ and two are isothermal $(\Delta T = 0)$. This idealized **Carnot cycle** is shown in Fig. 20–5. Each of the processes is considered to be done **reversibly**. That is, each of the processes (such as the expansion of the gases against a piston) is considered to be carried out infinitely slowly, so that the process could be considered a series of equilibrium states, and the whole process could be done in reverse with no change in the magnitude of work done or heat exchanged.

A perfectly reversible process is not possible in reality because it would require an infinite time. Reversible processes can be approached arbitrarily closely, and are important theoretically, just as the concept of an ideal gas is. A real process, on the other hand, would occur more quickly; there would be turbulence in the gas, friction would be present, and so on. Because of these factors, a real process cannot be done precisely in reverse—the turbulence would be different and the heat lost to friction would not reverse itself.[†] Thus, **real** processes are **irreversible**.

CAUTION
Real processes are always irreversible

The isothermal processes of a Carnot engine, where heats Q_H and Q_L are transferred, are assumed to be done at constant temperatures T_H and T_L. That is, the system is assumed to be in contact with idealized *heat reservoirs* (page 551) which are so large their temperatures don't change significantly when Q_H and Q_L are transferred. As we do for other heat engines, we take both Q_L and Q_H as positive.

[†]A real, irreversible process cannot be plotted on a *PV* diagram, except approximately if it approaches an ideal reversible process. But a reversible process (since it is a quasistatic series of equilibrium states) always can be plotted on a *PV* diagram. A reversible process, when done in reverse, retraces the same path on a *PV* diagram.

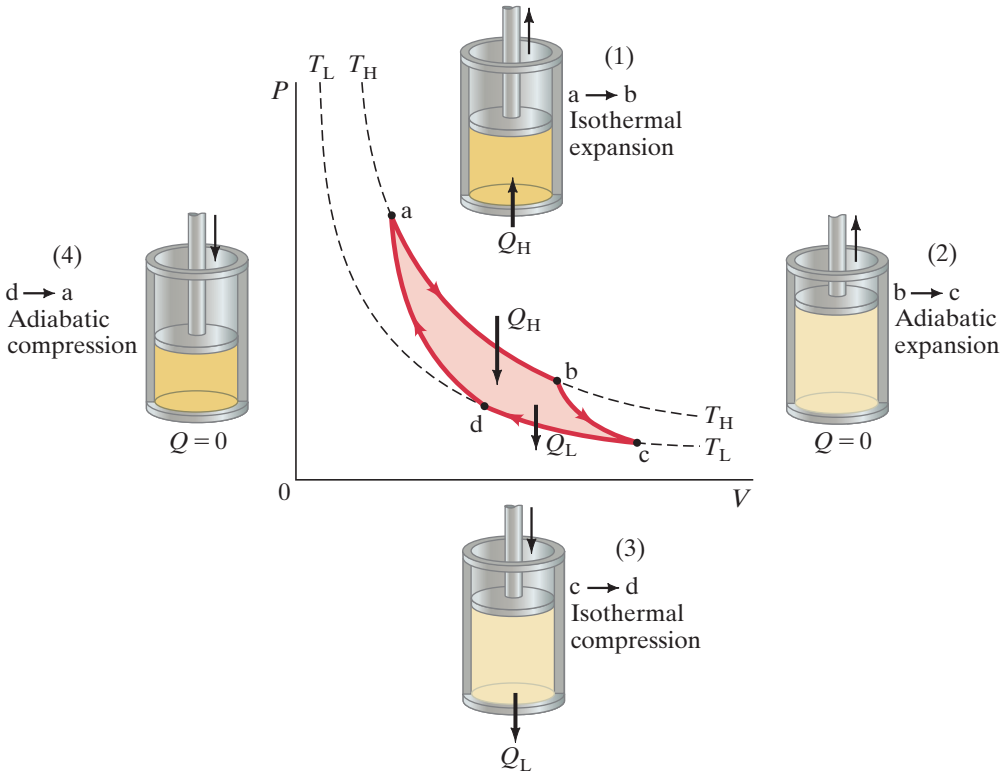


FIGURE 20–5 The Carnot cycle. Heat engines work in a cycle, and the cycle for the hypothetical Carnot engine begins at point a on this *PV* diagram for an ideal gas. (1) The gas is first expanded isothermally, with the addition of heat Q_H, along the path ab at temperature T_H. (2) Next the gas expands adiabatically from b to c—no heat is exchanged, but the temperature drops to T_L. (3) The gas is then compressed at constant temperature T_L, path cd, and heat Q_L flows out. (4) Finally, the gas is compressed adiabatically, path da, back to its original state.

The net work done in one cycle by a Carnot engine (or any other type of engine using a reversible cycle) is equal to the area enclosed by the curve representing the cycle on a *PV* diagram, the curve abcd in Fig. 20–5 (see Section 19–7, Figs. 19–11 and 19–12).

The efficiency of a Carnot engine, like any heat engine, is given by Eq. 20–1b:

$$e = 1 - \frac{Q_L}{Q_H}.$$

For a Carnot engine using an ideal gas, however, we can show that the efficiency depends only on the temperatures of the heat reservoirs, T_H and T_L. In the first isothermal process ab in Fig. 20–5, the work done by the gas is (see Eq. 19–9)

$$W_{ab} = nRT_H \ln \frac{V_b}{V_a},$$

where n is the number of moles of the ideal gas used as working substance. Because the internal energy of an ideal gas does not change when the temperature remains constant, the first law of thermodynamics ($\Delta E_{int} = Q - W$) tells us that the heat added to the gas equals the work done by the gas:

$$Q_H = nRT_H \ln \frac{V_b}{V_a}.$$

Similarly, the heat lost by the gas in the isothermal process cd is

$$Q_L = nRT_L \ln \frac{V_c}{V_d}.$$

The paths bc and da are adiabatic, so we have from Eq. 19–16:

$$P_b V_b^\gamma = P_c V_c^\gamma \quad \text{and} \quad P_d V_d^\gamma = P_a V_a^\gamma,$$

where $\gamma = C_P/C_V$ is the ratio of molar specific heats (Eq. 19–15). Also, from the ideal gas law,

$$\frac{P_b V_b}{T_H} = \frac{P_c V_c}{T_L} \quad \text{and} \quad \frac{P_d V_d}{T_L} = \frac{P_a V_a}{T_H}.$$

When we divide these last equations, term by term, into the corresponding set of equations on the line above, we obtain

$$T_H V_b^{\gamma-1} = T_L V_c^{\gamma-1} \quad \text{and} \quad T_L V_d^{\gamma-1} = T_H V_a^{\gamma-1}.$$

Next we divide the equation on the left by the one on the right and obtain

$$\left(\frac{V_b}{V_a}\right)^{\gamma-1} = \left(\frac{V_c}{V_d}\right)^{\gamma-1}.$$

So if $(V_b/V_a)^{\gamma-1} = (V_c/V_d)^{\gamma-1}$ as we just saw, then

$$\frac{V_b}{V_a} = \frac{V_c}{V_d}.$$

We insert this result into our equations for Q_H and Q_L on the previous page:

$$\frac{Q_L}{Q_H} = \frac{nRT_L \ln (V_c/V_d)}{nRT_H \ln (V_b/V_a)}$$

or

$$\frac{Q_L}{Q_H} = \frac{T_L}{T_H}. \qquad \text{[Carnot cycle]} \quad \textbf{(20–2)}$$

Hence the efficiency of a reversible Carnot engine can now be written

$$e_{\text{ideal}} = 1 - \frac{Q_L}{Q_H}$$

or

$$e_{\text{ideal}} = 1 - \frac{T_L}{T_H}. \qquad \left[\begin{array}{c}\text{Carnot (ideal) efficiency;}\\ \text{Kelvin temperatures}\end{array}\right] \quad \textbf{(20–3)}$$

The temperatures T_L and T_H are the absolute or Kelvin temperatures as measured on the ideal gas temperature scale. Thus the efficiency of a Carnot engine depends only on the temperatures T_L and T_H.

We could imagine other possible reversible cycles that could be used for an ideal reversible engine. According to a theorem stated by Carnot:

> **All reversible engines operating between the same two constant temperatures T_H and T_L have the same efficiency. Any irreversible engine operating between the same two fixed temperatures will have an efficiency less than this.**

This is known as **Carnot's theorem**. It tells us that Eq. 20–3, $e = 1 - (T_L/T_H)$, applies to any ideal reversible engine with fixed input and exhaust temperatures, T_H and T_L, and that this Eq. 20–3 expresses the fundamental upper limit to the efficiency of any heat engine. A higher efficiency would violate the second law of thermodynamics.[†] Real engines always have an efficiency lower than this because of losses due to friction and the like. Real engines that are well designed reach 60 to 80% of the Carnot efficiency.

EXAMPLE 20–2 **Steam engine efficiency.** A steam engine operates between 500°C and 270°C. What is the maximum possible efficiency of this engine?

APPROACH The maximum possible efficiency is the idealized Carnot efficiency, Eq. 20–3. We must use Kelvin temperatures.

SOLUTION We first change the temperature to kelvins by adding 273 to the given Celsius temperatures: $T_H = 773\,\text{K}$ and $T_L = 543\,\text{K}$. Then

$$e_{\text{ideal}} = 1 - \frac{543}{773} = 0.30.$$

To get the efficiency in percent, we multiply by 100. Thus, the maximum (or Carnot) efficiency is 30%. Real engines might attain 0.70 of this value, or 21%.

NOTE In this Example, the exhaust temperature is still rather high, 270°C. Steam engines can be arranged in series so that the exhaust of one engine is used as intake by a second or third engine.

[†]If an engine had a higher efficiency than Eq. 20–3, it could be used in conjunction with a Carnot engine that is made to work in reverse as a refrigerator. If W was the same for both, the net result would be a flow of heat at a low temperature to a high temperature without work being done. That would violate the Clausius statement of the second law (and also experience—i.e., the laboratory).

EXAMPLE 20–3 **A phony claim?** An engine manufacturer makes the following claims: An engine's heat input per second is 9.0 kJ at 435 K. The heat output per second is 4.0 kJ at 285 K. Do you believe these claims?

APPROACH The engine's efficiency can be calculated from the definition, Eqs. 20–1. It must be less than the maximum possible, Eq. 20–3.

SOLUTION The claimed efficiency of the engine is (Eq. 20–1b)

$$e = 1 - \frac{Q_L}{Q_H} = 1 - \frac{4.0\,\text{kJ}}{9.0\,\text{kJ}} = 0.56,$$

or 56%. However, the maximum possible efficiency is given by the Carnot efficiency, Eq. 20–3:

$$e_{\text{ideal}} = 1 - \frac{T_L}{T_H} = 1 - \frac{285\,\text{K}}{435\,\text{K}} = 0.34,$$

or 34%. The manufacturer's claims violate the second law of thermodynamics and cannot be believed.

EXERCISE A A motor is running with an intake temperature $T_H = 400$ K and an exhaust temperature $T_L = 300$ K. Which of the following is *not* a possible efficiency for the engine? (*a*) 0.10; (*b*) 0.16; (*c*) 0.24; (*d*) 0.30.

We can see from Eq. 20–3 that a 100% efficient engine is not possible. Only if the exhaust temperature, T_L, were at absolute zero would 100% efficiency be reachable. But getting to absolute zero is a practical (as well as theoretical) impossibility.[†] Because no engine can be 100% efficient, it can be stated that

SECOND LAW OF THERMODYNAMICS (Kelvin-Planck statement)

no device is possible whose sole effect is to transform a given amount of heat completely into work.

This is known as the **Kelvin-Planck statement of the second law of thermodynamics**. Figure 20–6 diagrams the ideal perfect heat engine, which can not exist. If it did, remarkable things could happen. For example, if the engine of a ship did not need a low-temperature reservoir to exhaust heat into, the ship could travel across the ocean using the vast resources of the internal energy of the ocean water. Indeed, we would have no fuel problems at all!

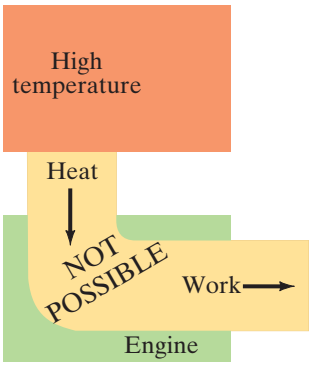


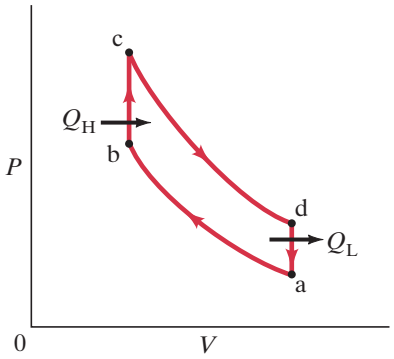
FIGURE 20–6 Diagram of an impossible perfect heat engine in which all heat input is used to do work.

EXERCISE B Return to the Chapter-Opening Question, page 576, and answer it again now. Try to explain why you may have answered differently the first time.

*The Otto Cycle

The operation of an automobile internal combustion engine (Fig. 20–4) can be approximated by a reversible cycle known as the *Otto cycle*, whose *PV* diagram is shown in Fig. 20–7. Unlike the Carnot cycle, the input and exhaust temperatures of the Otto cycle are *not* constant. Paths ab and cd are adiabatic, and paths bc and da are at constant volume. The gas (gasoline–air mixture) enters the cylinder at point a and is compressed adiabatically (compression stroke) to point b. At b ignition occurs (spark plug) and the quick burning of the gas adds heat Q_H to the system at constant volume (approximately in a real engine). The temperature and pressure rise, and then in the power stroke, cd, the gas expands adiabatically. In the exhaust stroke, da, heat Q_L is ejected to the environment (in a real engine, the gas leaves the engine and is replaced by a new mixture of air and fuel).

FIGURE 20–7 The Otto cycle.



[†]This result is known as the *third law of thermodynamics*, as discussed in Section 20–10.

EXAMPLE 20–4 **The Otto cycle.** (*a*) Show that for an ideal gas as working substance, the efficiency of an Otto cycle engine is

$$e = 1 - \left(\frac{V_a}{V_b}\right)^{1-\gamma}$$

where γ is the ratio of specific heats ($\gamma = C_P/C_V$, Eq. 19–15) and V_a/V_b is the *compression ratio*. (*b*) Calculate the efficiency for a compression ratio $V_a/V_b = 8.0$ assuming a diatomic gas like O_2 and N_2.

APPROACH We use the original definition of efficiency and the results from Chapter 19 for constant volume and adiabatic processes (Sections 19–8 and 19–9).

SOLUTION The heat exchanges take place at constant volume in the ideal Otto cycle, so from Eq. 19–11a:

$$Q_H = nC_V(T_c - T_b) \quad \text{and} \quad Q_L = nC_V(T_d - T_a).$$

Then from Eq. 20–1b,

$$e = 1 - \frac{Q_L}{Q_H} = 1 - \left[\frac{T_d - T_a}{T_c - T_b}\right].$$

To get this in terms of the compression ratio, V_a/V_b, we use the result from Section 19–9, Eq. 19–16, $PV^\gamma = \text{constant}$ during the adiabatic processes ab and cd. Thus

$$P_a V_a^\gamma = P_b V_b^\gamma \quad \text{and} \quad P_c V_c^\gamma = P_d V_d^\gamma.$$

We use the ideal gas law, $P = nRT/V$, and substitute P into these two equations

$$T_a V_a^{\gamma-1} = T_b V_b^{\gamma-1} \quad \text{and} \quad T_c V_c^{\gamma-1} = T_d V_d^{\gamma-1}.$$

Then the efficiency (see above) is

$$e = 1 - \left[\frac{T_d - T_a}{T_c - T_b}\right] = 1 - \left[\frac{T_c(V_c/V_d)^{\gamma-1} - T_b(V_b/V_a)^{\gamma-1}}{T_c - T_b}\right].$$

But processes bc and da are at constant volume, so $V_c = V_b$ and $V_d = V_a$. Hence $V_c/V_d = V_b/V_a$ and so we achieve our desired result:

$$e = 1 - \left[\frac{(V_b/V_a)^{\gamma-1}(T_c - T_b)}{T_c - T_b}\right] = 1 - \left(\frac{V_b}{V_a}\right)^{\gamma-1} = 1 - \left(\frac{V_a}{V_b}\right)^{1-\gamma}.$$

(*b*) For diatomic molecules (Section 19–8), $\gamma = C_P/C_V = 1.4$ so

$$e = 1 - (8.0)^{1-\gamma} = 1 - (8.0)^{-0.4} = 0.56.$$

Real engines do not reach this high efficiency because they do not follow perfectly the Otto cycle, plus there is friction, turbulence, heat loss, and incomplete combustion of the gases.

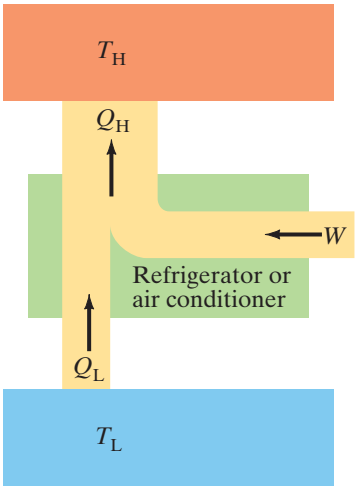


FIGURE 20–8 Schematic diagram of energy transfers for a refrigerator or air conditioner (AC).

20–4 Refrigerators, Air Conditioners, and Heat Pumps

The operating principle of refrigerators, air conditioners, and heat pumps is just the reverse of a heat engine. Each operates to transfer heat *out* of a cool environment into a warm environment. As diagrammed in Fig. 20–8, by doing work W, heat is taken from a low-temperature region, T_L (such as inside a refrigerator), and a greater amount of heat is exhausted at a high temperature, T_H (the room). Heat Q_L is removed from cooling coils *inside* the refrigerator and heat Q_H is given off by coils *outside* the rear of the refrigerator, Fig. 20–9. You can often feel this heated air coming out beneath a refrigerator. The work W is usually done by an electric motor which compresses a fluid, as illustrated in Fig. 20–9. (We assume Q_L, Q_H, and W are all positive, as in Section 20–2.)

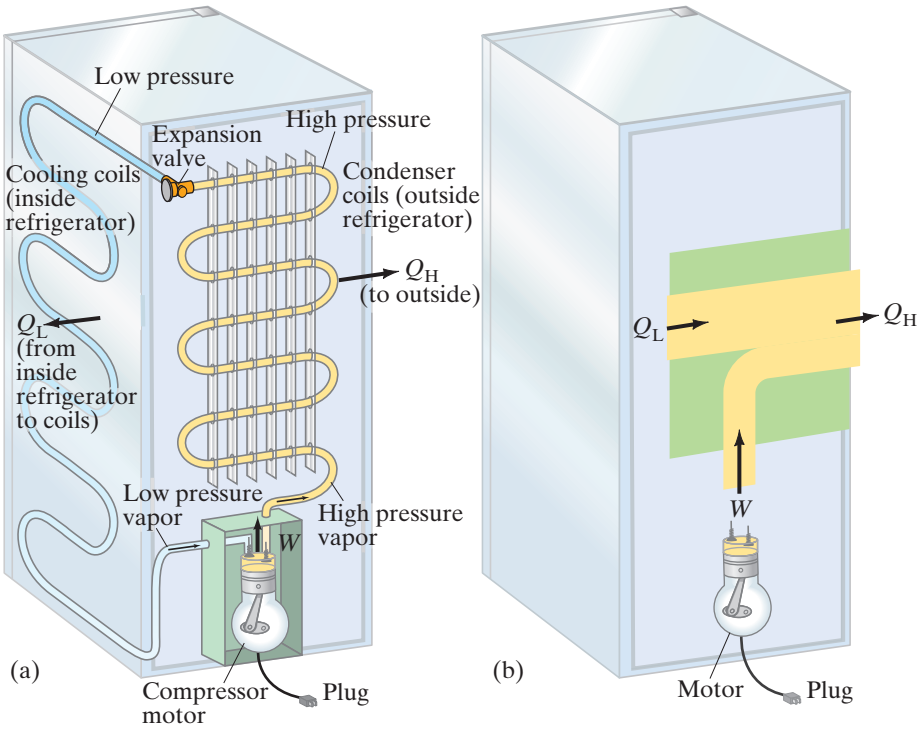


FIGURE 20–9 (a) Typical refrigerator system. The electric compressor motor forces a gas at high pressure through a heat exchanger (condenser) on the rear *outside* wall of the refrigerator, where Q_H is given off and the gas cools to become liquid. The liquid passes from a high-pressure region, via a valve, to low-pressure tubes on the *inside* walls of the refrigerator; the liquid evaporates at this lower pressure and thus absorbs heat (Q_L) from the inside of the refrigerator. The fluid returns to the compressor, where the cycle begins again. (b) Schematic diagram, like Fig. 20–8.

PHYSICS APPLIED
Refrigerator

A perfect **refrigerator**—one in which no work is required to take heat from the low-temperature region to the high-temperature region—is not possible. This is the **Clausius statement of the second law of thermodynamics**, already mentioned in Section 20–1: it can be stated formally as

SECOND LAW OF THERMODYNAMICS (Clausius statement)

> **no device is possible whose sole effect is to transfer heat from one region at a temperature T_L into a second region at a higher temperature T_H.**

To make heat flow from a low-temperature object (or region) to one at a higher temperature, work must be done. Thus, *there can be no perfect refrigerator*.

The **coefficient of performance** (COP) of a refrigerator is defined as the heat Q_L removed from the low-temperature area (inside a refrigerator) divided by the work W done to remove the heat (Fig. 20–8):

$$\text{COP} = \frac{Q_L}{W}. \qquad \left[\begin{matrix}\text{refrigerator and}\\ \text{air conditioner}\end{matrix}\right] \qquad \textbf{(20–4a)}$$

We use Q_L because it is the heat removed from inside that matters from a practical point of view. This makes sense because the more heat Q_L that can be removed from inside the refrigerator for a given amount of work, the better (more efficient) the refrigerator is. Energy is conserved, so from the first law of thermodynamics we can write $Q_L + W = Q_H$, or $W = Q_H - Q_L$ (see Fig. 20–8 or 20–9b). Then Eq. 20–4a becomes

$$\text{COP} = \frac{Q_L}{W} = \frac{Q_L}{Q_H - Q_L}. \qquad \left[\begin{matrix}\text{refrigerator and}\\ \text{air conditioner}\end{matrix}\right] \qquad \textbf{(20–4b)}$$

For an ideal refrigerator (not a perfect one, which is impossible), the best we could do would be

$$\text{COP}_{\text{ideal}} = \frac{T_L}{T_H - T_L}, \qquad \left[\begin{matrix}\text{refrigerator and}\\ \text{air conditioner}\end{matrix}\right] \qquad \textbf{(20–4c)}$$

analagous to an ideal (Carnot) engine (Eqs. 20–2 and 20–3).

PHYSICS APPLIED
Air conditioner

An **air conditioner** works very much like a refrigerator, although the actual construction details are different: an air conditioner takes heat Q_L from inside a room or building at a low temperature, and deposits heat Q_H outside to the environment at a higher temperature. Equations 20–4 also describe the coefficient of performance for an air conditioner.

EXAMPLE 20–5 **Making ice.** A freezer has a COP of 2.8 and uses 200 watts of power. How long would it take to freeze an ice-cube tray that contains 600 g of water at 0°C?

APPROACH In Eq. 20–4b, Q_L is the heat that must be transferred out of the water so it will become ice. To determine Q_L, we use the latent heat of fusion L of water and Eq. 19–4, $Q = mL$.

SOLUTION From Table 19–2, $L = 333\,\text{kJ/kg}$ for water. Hence

$$Q_L = mL = (0.600\,\text{kg})(3.33 \times 10^5\,\text{J/kg}) = 2.0 \times 10^5\,\text{J}$$

is the total energy that needs to be removed from the water. The freezer does work at the rate of 200 watts = 200 J/s = W/t, which is the work W it can do in t seconds. We solve for t: $t = W/(200\,\text{J/s})$. For W, we can also use Eq. 20–4a: $W = Q_L/\text{COP}$. Thus

$$t = \frac{W}{200\,\text{J/s}} = \frac{Q_L/\text{COP}}{200\,\text{J/s}} = \frac{2.0 \times 10^5\,\text{J}}{(2.8)(200\,\text{J/s})} = 360\,\text{s},$$

or about 6 min.

Heat naturally flows from high temperature to low temperature. Refrigerators and air conditioners do work to accomplish the opposite: to make heat flow from cold to hot. We might say they "pump" heat from cold areas to hotter areas, against the natural tendency of heat to flow from hot to cold, just as water can be pumped uphill, against the natural tendency to flow downhill. The term **heat pump** is usually reserved for a device that can heat a house in winter by using an electric motor that does work W to take heat Q_L from the outside at low temperature and delivers heat Q_H to the warmer inside of the house; see Fig. 20–10. As in a refrigerator, there is an indoor and an outdoor heat exchanger (coils of the refrigerator) and an electric compressor motor. The operating principle is like that for a refrigerator or air conditioner; but the objective of a heat pump is to heat (deliver Q_H), rather than to cool (remove Q_L). Thus, the coefficient of performance of a heat pump is defined differently than for an air conditioner because it is the heat Q_H delivered to the inside of the house that is important now:

PHYSICS APPLIED
Heat pump

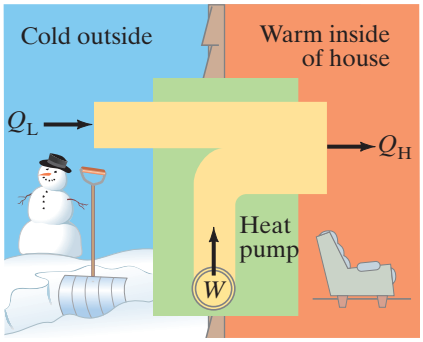


FIGURE 20–10 A heat pump uses an electric motor to "pump" heat from the cold outside to the warm inside of a house.

$$\text{COP} = \frac{Q_H}{W}. \qquad \text{[heat pump]} \quad \textbf{(20–5)}$$

The COP is necessarily greater than 1, because $W + Q_L = Q_H$. Typical heat pumps today have COP ≈ 2.5 to 3. Most heat pumps can be "turned around" and used as air conditioners in the summer.

EXAMPLE 20–6 **Heat pump.** A heat pump has a coefficient of performance of 3.0 and is rated to do work at 1500 watts. (*a*) How much heat can it add to a room per second? (*b*) If the heat pump were turned around to act as an air conditioner in the summer, what would you expect its coefficient of performance to be, assuming all else stays the same?

APPROACH We use the definitions of coefficient of performance, which are different for the two devices in (*a*) and (*b*).

SOLUTION (*a*) We use Eq. 20–5 for the heat pump, and, since our device does 1500 J of work per second, it can pour heat into the room at a rate of

$$Q_H = \text{COP} \times W = 3.0 \times 1500\,\text{J} = 4500\,\text{J}$$

per second, or at a rate of 4500 W [≈4 Btu/s]. That is, the heat pump works at 1500 W in order to drag 3000 W from the outdoors and deliver 4500 W indoors.

(*b*) If our device is turned around in summer, it can take heat Q_L from inside the house, doing 1500 J of work per second to then dump $Q_H = 4500\,\text{J}$ per second to the hot outside. Energy is conserved, so $Q_L + W = Q_H$ (see Fig. 20–10, but reverse the inside and outside of the house). Then

$$Q_L = Q_H - W = 4500\,\text{J} - 1500\,\text{J} = 3000\,\text{J}.$$

The coefficient of performance as an air conditioner would thus be (Eq. 20–4a)

$$\text{COP} = \frac{Q_L}{W} = \frac{3000\,\text{J}}{1500\,\text{J}} = 2.0.$$

NOTE The coefficients of performance are defined differently for heat pumps and air conditioners.

CAUTION
Heat pumps and air conditioners have different COP definitions

EXERCISE C The heat pump of Example 20–6 uses 1500 W of electric power to deliver 4500 W of heat. Does this sound like we're getting something for nothing? (*a*) Explain why we aren't. (*b*) Compare to the freezer of Example 20–5 whose motor uses 200 W to extract heat at what rate?

A good heat pump can sometimes be a money saver and an energy saver, depending on the cost of the unit and installation, etc. Compare, for example, our heat pump in Example 20–6 to, say, a 1500-W electric heater. We plug the heater into the wall, it draws 1500 W of electricity, and it delivers 1500 W of heat to the room. Our heat pump when plugged into the wall also draws 1500 W of electricity (which is what we pay for), but it delivers 4500 W of heat!

*SEER Rating

Cooling devices such as refrigerators and air conditioners are often given a rating known as SEER (Seasonal Energy Efficiency Ratio), which is defined as

PHYSICS APPLIED
SEER rating

$$\text{SEER} = \frac{(\text{heat removed in Btu})}{(\text{electrical input in watt-hours})},$$

as measured by averaging over varying (seasonal) conditions. The definition of the SEER is basically the same as the COP except for the (unfortunate) mixed units. Given that $1\ \text{Btu} = 1056\,\text{J}$ (see Section 19–1 and Problem 4 in Chapter 19), then a $\text{SEER} = 1$ is a COP equal to $(1\ \text{Btu}/1\ \text{W}\cdot\text{h}) = (1056\,\text{J})/(1\,\text{J/s} \times 3600\,\text{s}) = 0.29$. And a $\text{COP} = 1$ is a $\text{SEER} = 1/0.29 = 3.4$.

20–5 Entropy

Thus far we have stated the second law of thermodynamics for specific situations. What we really need is a general statement of the second law of thermodynamics that will cover all situations, including processes discussed earlier in this Chapter, processes that are not observed in nature even though they would not violate the first law of thermodynamics. It was not until the latter half of the nineteenth century that the second law of thermodynamics was finally stated in a general way—namely, in terms of a quantity called **entropy**, introduced by Clausius in the 1860s. Entropy is a state variable, as we will see, and in Section 20–7 we will see that entropy can be interpreted as a measure of the order or disorder of a system.

When we deal with entropy—as with potential energy—it is the *change* in entropy during a process that is important, not the absolute amount. According to Clausius, when an amount of heat Q is *added* to a system by a reversible process at constant temperature, the change in entropy S of the system is given by

$$\Delta S = \frac{Q}{T}, \qquad \textbf{(20–6)}$$

where T is the Kelvin temperature. If heat is lost, Q is *negative* in this equation, as we now resume our *original sign conventions* on page 549.

CAUTION
Sign conventions for Q and W (page 549) in use from now on

EXAMPLE 20–7 **Entropy change in melting.** An ice cube of mass 56 g is taken from a storage compartment at 0°C and placed in a paper cup. After a few minutes, exactly half of the mass of the ice cube has melted, becoming water at 0°C. Find the change in entropy of the 56 grams of ice/water.

APPROACH We consider the 56 g of water, initially in the form of ice, as our system. To determine the entropy change, we first must find the heat needed to melt the ice, which we do using the latent heat of fusion of water, $L = 333\ \text{kJ/kg}$ (Section 19–5). The heat Q required comes from the surroundings.

SOLUTION The heat required to melt 28 g of ice (half of the 56-g ice cube) is

$$Q = mL = (0.028\ \text{kg})(333\ \text{kJ/kg}) = 9.3\ \text{kJ}.$$

The temperature remains constant in our process, so the change in entropy using Eq. 20–6 is:

$$\Delta S = \frac{Q}{T} = \frac{9.3\ \text{kJ}}{273\ \text{K}} = 34\ \text{J/K}.$$

NOTE The change in entropy of the surroundings (cup, air) has not been computed.

If the temperature is not constant, we define entropy S by the relation

$$dS = \frac{dQ}{T}. \qquad \textbf{(20–7)}$$

Then the change in entropy of a system taken reversibly between two equilibrium states a and b is given by[†]

$$\Delta S = S_\text{b} - S_\text{a} = \int_\text{a}^\text{b} dS = \int_\text{a}^\text{b} \frac{dQ}{T}. \quad \text{[reversible process]} \qquad \textbf{(20–8)}$$

PROBLEM SOLVING
Entropy is a state variable

A careful analysis (next page) shows that the change in entropy when a system is moved by a reversible process from any state a to another state b does not depend on the process. That is, $\Delta S = S_\text{b} - S_\text{a}$ depends only on the states a and b of the system. Thus entropy (unlike heat) is a **state variable**. Any system in a given state has a mass, a temperature, a volume, a pressure, and also has a particular value of entropy. Therefore the change in entropy ΔS for a real process, which is always irreversible (page 580), can be determined by calculating ΔS for an idealized reversible process between the same two states. (We used this fact in Example 20–7.)

For now, we can see why entropy is a state variable for a Carnot cycle. In Eq. 20–2 we saw that $Q_\text{L}/Q_\text{H} = T_\text{L}/T_\text{H}$, which we rewrite as

$$\frac{Q_\text{L}}{T_\text{L}} = \frac{Q_\text{H}}{T_\text{H}}.$$

This relation shows that in the *PV* diagram for a Carnot cycle, Fig. 20–5, the entropy change $\Delta S = Q/T$ in going from state a to state c along path abc $(= Q_\text{H}/T_\text{H} + 0)$ is thus the same as going along the path adc $(0 + Q_\text{L}/T_\text{L})$. That is, the change in entropy is *path independent*—it depends only on the initial and final states of the system.

*Showing Entropy Is a State Variable

In our study of the Carnot cycle we found (Eq. 20–2) that $Q_\text{L}/Q_\text{H} = T_\text{L}/T_\text{H}$. We rewrite this as

$$\frac{Q_\text{H}}{T_\text{H}} = \frac{Q_\text{L}}{T_\text{L}}.$$

In this relation, both Q_H and Q_L are positive. But now let us recall our original convention

[†]Equation 20–8 says nothing about the value of S. It only gives the change in S. This is much like potential energy (Chapter 8). However, one form of the *third law of thermodynamics* (see also Section 20–10) states that as $T \to 0$, $S \to 0$.

as used in the first law (Section 19–6, page 549), that Q is positive when it represents a heat flow into the system (as Q_H) and negative for a heat flow out of the system (as $-Q_L$). Then this last relation becomes

$$\frac{Q_H}{T_H} + \frac{Q_L}{T_L} = 0. \qquad \text{[Carnot cycle]} \quad \textbf{(20–9)}$$

Now consider *any* reversible cycle, as represented by the smooth (oval-shaped) curve in Fig. 20–11. Any reversible cycle can be approximated as a series of Carnot cycles. Figure 20–11 shows only six—the isotherms (dashed lines) are connected by adiabatic paths for each—and the approximation becomes better and better if we increase the number of Carnot cycles. Equation 20–9 is valid for each of these cycles, so we can write

$$\Sigma \frac{Q}{T} = 0 \qquad \text{[Carnot cycles]} \quad \textbf{(20–10)}$$

for the sum of all these cycles. But note that the heat output Q_L of one cycle crosses the boundary below it and is approximately equal to the negative of the heat input, Q_H, of the cycle below it (actual equality occurs in the limit of an infinite number of infinitely thin Carnot cycles). Hence the heat flows on the inner paths of all these Carnot cycles cancel out, so the net heat transferred, and the work done, is the same for the series of Carnot cycles as for the original cycle. Hence, in the limit of infinitely many Carnot cycles, Eq. 20–10 applies to any reversible cycle. In this case Eq. 20–10 becomes

$$\oint \frac{dQ}{T} = 0, \qquad \text{[reversible cycle]} \quad \textbf{(20–11)}$$

where dQ represents an infinitesimal heat flow. The symbol $\oint$ means that the integral is taken around a closed path; the integral can be started at any point on the path such as at a or b in Fig. 20–11, and proceed in either direction. If we divide the cycle of Fig. 20–11 into two parts as indicated in Fig. 20–12, then

$$\int_{a\,\mathrm{I}}^{b} \frac{dQ}{T} + \int_{b\,\mathrm{II}}^{a} \frac{dQ}{T} = 0.$$

The first term is the integral from point a to point b along path I in Fig. 20–12, and the second term is the integral from b back to a along path II. If path II is taken in reverse, dQ at each point becomes $-dQ$, since the path is reversible. Therefore

$$\int_{a\,\mathrm{I}}^{b} \frac{dQ}{T} = \int_{a\,\mathrm{II}}^{b} \frac{dQ}{T}. \qquad \text{[reversible paths]} \quad \textbf{(20–12)}$$

The integral of dQ/T between any two equilibrium states, a and b, does not depend on the path of the process. By defining entropy as $dS = dQ/T$ (Eq. 20–7), we see from Eq. 20–12 that the change in entropy between any two states along a reversible path is *independent of the path between two points a and b*. Thus *entropy is a state variable*—its value depends only on the state of the system, and not on the process or the past history of how it got there. This is in clear distinction to Q and W which are *not* state variables; their values do depend on the processes undertaken.

Real processes are irreversible. But because entropy is a state variable, the change in entropy ΔS for an irreversible process can be determined by calculating ΔS for a reversible process between the same two states.

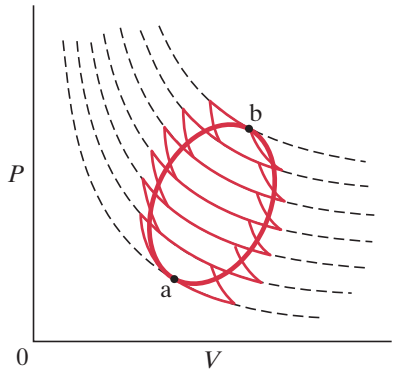
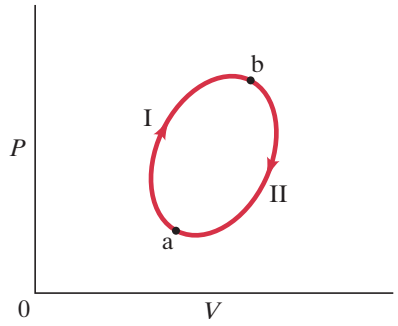


FIGURE 20–11 Any reversible cycle (the arbitrary red oval shown here) can be approximated as a series of Carnot cycles. (The dashed lines represent isotherms.)

FIGURE 20–12 The integral, $\oint dS$, of the entropy for a reversible cycle is zero. Hence the difference in entropy between states a and b, $S_b - S_a = \int_a^b dS$, is the same for path I as for path II.



20–6 Entropy and the Second Law of Thermodynamics

We have defined a new quantity, S, the entropy, which is a state variable and can be used to describe the state of the system, along with P, T, V, E_{int}, and n. But what does this rather abstract quantity have to do with the second law of thermodynamics? To answer this, let us take some examples in which we calculate the entropy changes during particular processes. But note first that Eq. 20–8 can be applied only to reversible processes. How then do we calculate $\Delta S = S_b - S_a$ for a real process that is irreversible? What we can do is this. We figure out some other *reversible* process that takes the system between the same two states, and calculate ΔS for this reversible process. This will equal ΔS for the irreversible process because ΔS depends only on the initial and final states of the system.

If the temperature varies during a process, a summation of the heat flow over the changing temperature can often be calculated using calculus or a computer. However, if the temperature change is not too great, a reasonable approximation can be made using the average value of the temperature, as indicated in the next Example.

EXAMPLE 20–8 **ESTIMATE** **Entropy change when water samples are mixed.** A sample of 50.00 kg of water at 20.00°C is mixed with 50.00 kg of water at 24.00°C. Estimate the change in entropy.

APPROACH The final temperature of the mixture will be 22.00°C, since we started with equal amounts of water. We use the specific heat of water and the methods of calorimetry (Sections 19–3 and 19–4) to determine the heat transferred. Then we use the average temperature of each sample of water to estimate the entropy change (Q/T) for the two samples, as if the processes were reversible.

SOLUTION A quantity of heat,

$$Q = mc\,\Delta T = (50.00\,\text{kg})\,(4186\,\text{J/kg}\cdot\text{C}^\circ)\,(2.00\,\text{C}^\circ) = 4.186\times10^5\,\text{J},$$

flows out of the hot water as it cools down from 24°C to 22°C, and this heat flows into the cold water as it warms from 20°C to 22°C. The total change in entropy, ΔS, will be the sum of the changes in entropy of the hot water, ΔS_H, and that of the cold water, ΔS_C:

$$\Delta S = \Delta S_H + \Delta S_C.$$

We estimate entropy changes by writing $\Delta S = Q/\overline{T}$, where $\overline{T}$ is an "average" temperature for each process, which ought to give a reasonable estimate since the temperature change is small. For the hot water we use an average temperature of 23°C (296 K), and for the cold water an average temperature of 21°C (294 K). Thus

$$\Delta S_H \approx -\frac{4.186\times10^5\,\text{J}}{296\,\text{K}} = -1414\,\text{J/K}$$

which is negative because this heat flows out (sign conventions, page 549) whereas heat is added to the cold water:

$$\Delta S_C \approx \frac{4.186\times10^5\,\text{J}}{294\,\text{K}} = 1424\,\text{J/K}.$$

The entropy of the hot water (S_H) decreases since heat flows out of the hot water. But the entropy of the cold water (S_C) increases by a greater amount. The total change in entropy is

$$\Delta S = \Delta S_H + \Delta S_C \approx -1414\,\text{J/K} + 1424\,\text{J/K} \approx 10\,\text{J/K}.$$

We see that although the entropy of one part of the system decreased, the entropy of the other part increased by a greater amount so that the net change in entropy of the whole system is positive.

We can now show in general that for an isolated system of two objects, the flow of heat from the higher-temperature (T_H) object to the lower-temperature (T_L) object always results in an increase in the total entropy. The two objects eventually come to some intermediate temperature, T_M. The heat lost by the hotter object $(Q_H = -Q,$ where Q is positive$)$ is equal to the heat gained by the colder one $(Q_L = Q)$, so the total change in entropy is

$$\Delta S = \Delta S_H + \Delta S_L = -\frac{Q}{T_{HM}} + \frac{Q}{T_{LM}},$$

where T_{HM} is some intermediate temperature between T_H and T_M for the hot object as it cools from T_H to T_M, and T_{LM} is the counterpart for the cold object. Since the temperature of the hot object is, at all times during the process, greater than that of the cold object, then $T_{HM} > T_{LM}$. Hence

$$\Delta S = Q\left(\frac{1}{T_{LM}} - \frac{1}{T_{HM}}\right) > 0.$$

One object decreases in entropy, while the other gains in entropy, but the *total* change is positive.

EXAMPLE 20–9 **Entropy changes in a free expansion.** Consider the *adiabatic free expansion* of n moles of an ideal gas from volume V_1 to volume V_2, where $V_2 > V_1$ as was discussed in Section 19–7, Fig. 19–14 (repeated below). Calculate the change in entropy (a) of the gas and (b) of the surrounding environment. (c) Evaluate ΔS for 1.00 mole, with $V_2 = 2.00\, V_1$.

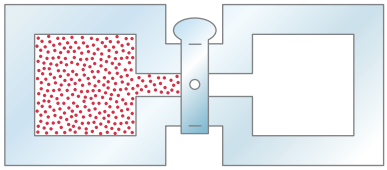


FIGURE 19–14 (Repeated.) Free expansion.

APPROACH We saw in Section 19–7 that the gas is initially in a closed container of volume V_1, and, with the opening of a valve, it expands adiabatically into a previously empty container. The total volume of the two containers is V_2. The whole apparatus is thermally insulated from the surroundings, so no heat flows into the gas, $Q = 0$. The gas does no work, $W = 0$, so there is no change in internal energy, $\Delta E_{int} = 0$, and the temperature of the initial and final states is the same, $T_2 = T_1 = T$. The process takes place very quickly, and is irreversible. Thus we cannot apply Eq. 20–8 to this process. Instead we must think of a reversible process that will take the gas from volume V_1 to V_2 at the same temperature, and use Eq. 20–8 on this reversible process to get ΔS. A reversible isothermal process will do the trick; in such a process, the internal energy does not change, so from the first law,

$$dQ = dW = P\,dV.$$

SOLUTION (a) For the gas,

$$\Delta S_{gas} = \int \frac{dQ}{T} = \frac{1}{T}\int_{V_1}^{V_2} P\,dV.$$

The ideal gas law tells us $P = nRT/V$, so

$$\Delta S_{gas} = \left(\frac{1}{T}\right)(nRT)\int_{V_1}^{V_2}\frac{dV}{V} = nR\ln\frac{V_2}{V_1}.$$

Since $V_2 > V_1$, $\Delta S_{gas} > 0$.

(b) Since no heat is transferred to the surrounding environment, there is no change of the state of the environment due to this process. Hence $\Delta S_{env} = 0$. Note that the total change in entropy, $\Delta S_{gas} + \Delta S_{env}$, is greater than zero.

(c) Since $n = 1.00$ and $V_2 = 2.00\, V_1$, then $\Delta S_{gas} = R\ln 2.00 = 5.76\ \text{J/K}$.

EXAMPLE 20–10 Heat transfer. A red-hot 2.00-kg piece of iron at temperature $T_1 = 880\,\text{K}$ is thrown into a huge lake whose temperature is $T_2 = 280\,\text{K}$. Assume the lake is so large that its temperature rise is insignificant. Determine the change in entropy (*a*) of the iron and (*b*) of the surrounding environment (the lake).

APPROACH The process is irreversible, but the same entropy change will occur for a reversible process, and we use the concept of specific heat, Eq. 19–2.

SOLUTION (*a*) We assume the specific heat of the iron is constant at $c = 450\,\text{J/kg}\cdot\text{K}$. Then $dQ = mc\,dT$ and in a quasistatic reversible process

$$\Delta S_{\text{iron}} = \int \frac{dQ}{T} = mc\int_{T_1}^{T_2}\frac{dT}{T} = mc\ln\frac{T_2}{T_1} = -mc\ln\frac{T_1}{T_2}.$$

Putting in numbers, we find

$$\Delta S_{\text{iron}} = -(2.00\,\text{kg})(450\,\text{J/kg}\cdot\text{K})\ln\frac{880\,\text{K}}{280\,\text{K}} = -1030\,\text{J/K}.$$

(*b*) The initial and final temperatures of the lake are the same, $T = 280\,\text{K}$. The lake receives from the iron an amount of heat

$$Q = mc(T_2 - T_1) = (2.00\,\text{kg})(450\,\text{J/kg}\cdot\text{K})(880\,\text{K} - 280\,\text{K}) = 540\,\text{kJ}.$$

Strictly speaking, this is an irreversible process (the lake heats up locally before equilibrium is reached), but is equivalent to a reversible isothermal transfer of heat $Q = 540\,\text{kJ}$ at $T = 280\,\text{K}$. Hence

$$\Delta S_{\text{env}} = \frac{540\,\text{kJ}}{280\,\text{K}} = 1930\,\text{J/K}.$$

Thus, although the entropy of the iron actually decreases, the *total* change in entropy of iron plus environment is positive: $1930\,\text{J/K} - 1030\,\text{J/K} = 900\,\text{J/K}$.

In each of these Examples, the entropy of our system plus that of the environment (or surroundings) either stayed constant or increased. For any *reversible* process, the total entropy change is zero. This is because any reversible process can be considered as a series of quasistatic isothermal transfers of heat ΔQ between a system and the environment, which differ in temperature only by an infinitesimal amount. Hence the change in entropy of either the system or environment is $\Delta Q/T$ and that of the other is $-\Delta Q/T$, so the total is

$$\Delta S = \Delta S_{\text{syst}} + \Delta S_{\text{env}} = 0. \qquad \text{[any reversible process]}$$

In Examples 20–8, 20–9, and 20–10, we found that the total entropy of system plus environment increases. Indeed, it has been found that for all real (irreversible) processes, the total entropy increases. No exceptions have been found. We can thus make the *general statement of the second law of thermodynamics* as follows:

The entropy of an isolated system never decreases. It either stays constant (reversible processes) or increases (irreversible processes).

Since all real processes are irreversible, we can equally well state the second law as:

SECOND LAW OF THERMODYNAMICS (general statement)

The total entropy of any system plus that of its environment increases as a result of any natural process:

$$\Delta S = \Delta S_{\text{syst}} + \Delta S_{\text{env}} > 0. \qquad \textbf{(20–13)}$$

Although the entropy of one part of the universe may decrease in any process (see the last few Examples we have done), the entropy of some other part of the universe always increases by a greater amount, so the total entropy always increases.

Now that we finally have a quantitative general statement of the second law of thermodynamics, we can see that it is an unusual law. It differs considerably

from other laws of physics, which are typically equalities (such as $F = ma$) or conservation laws (such as for energy and momentum). The second law of thermodynamics introduces a new quantity, the entropy, S, but does not tell us it is conserved. Quite the opposite. Entropy is not conserved in natural processes. Entropy always increases in time for real processes.

"Time's Arrow"

The second law of thermodynamics summarizes which processes are observed in nature, and which are not. Or, said another way, it tells us about the *direction* processes go. For the reverse of any of the processes in the last few Examples, the entropy would decrease; and we never observe those processes. For example, we never observe heat flowing spontaneously from a cold object to a hot object, the hot one getting hotter and the cold one colder. Nor do we ever observe a gas spontaneously compressing itself into a smaller volume, the reverse of Example 20–9 (gases always expand to fill their containers). Nor do we see thermal energy transform into kinetic energy of a rock so the rock rises spontaneously from the ground. Any of these processes would be consistent with the first law of thermodynamics (conservation of energy). But they are not consistent with the second law of thermodynamics, and this is why we need the second law. If you were to see a movie run backward, you would probably realize it immediately because you would see odd occurrences—such as rocks rising spontaneously from the ground, or air rushing in from the atmosphere to fill an empty balloon (the reverse of free expansion). When watching a movie or video, we are tipped off to a faked reversal of time by observing whether entropy is increasing or decreasing. Hence entropy has been called **time's arrow**, for it can tell us in which direction time is going.

20–7 Order to Disorder

The concept of entropy, as we have discussed it so far, may seem rather abstract. But we can relate it to the more ordinary concepts of *order* and *disorder*. In fact, the entropy of a system can be considered a *measure of the disorder of the system*. Then the second law of thermodynamics can be stated simply as:

SECOND LAW OF THERMODYNAMICS (general statement)

natural processes tend to move toward a state of greater disorder.

Exactly what we mean by **disorder** may not always be clear, so we now consider a few examples. Some of these will show us how this very general statement of the second law applies beyond what we usually consider as thermodynamics.

Let us look at the simple processes mentioned in Section 20–1. First, a jar containing separate layers of salt and pepper is more orderly than a jar in which the salt and pepper are all mixed up. Shaking a jar containing separate layers results in a mixture, and no amount of shaking brings the layers back again. The natural process is from a state of relative order (layers) to one of relative disorder (a mixture), not the reverse. That is, disorder increases. Second, a solid coffee cup is a more "orderly" and useful object than the pieces of a broken cup. Cups break when they fall, but they do not spontaneously mend themselves (as faked in Fig. 20–1). Again, the normal course of events is an increase of disorder.

Let us consider some processes for which we have actually calculated the entropy change, and see that an increase in entropy results in an increase in disorder (or vice versa). When ice melts to water at 0°C, the entropy of the water increases (Example 20–7). Intuitively, we can think of solid water, ice, as being more ordered than the less orderly fluid state which can flow all over the place. This change from order to disorder can be seen more clearly from the molecular point of view: the orderly arrangement of water molecules in an ice crystal has changed to the disorderly and somewhat random motion of the molecules in the fluid state.

When a hot substance is put in contact with a cold substance, heat flows from the high temperature to the low until the two substances reach the same intermediate temperature. Entropy increases, as we saw in Example 20–8. At the beginning of the process we can distinguish two classes of molecules: those with a high average kinetic energy (the hot object), and those with a low average kinetic energy (the cooler object). After the process in which heat flows, all the molecules are in one class with the same average kinetic energy; we no longer have the more orderly arrangement of molecules in two classes. Order has gone to disorder. Furthermore, the separate hot and cold objects could serve as the hot- and cold-temperature regions of a heat engine, and thus could be used to obtain useful work. But once the two objects are put in contact and reach the same temperature, no work can be obtained. Disorder has increased, because a system that has the ability to perform work must surely be considered to have a higher order than a system no longer able to do work.

When a stone falls to the ground, its macroscopic kinetic energy is transformed to thermal energy. (We noted earlier that the reverse never happens: a stone never absorbs thermal energy and rises into the air of its own accord.) This is another example of order changing to disorder. Thermal energy is associated with the disorderly random motion of molecules, but the molecules in the falling stone all have the same velocity downward in addition to their own random velocities. Thus, the more orderly kinetic energy of the stone as a whole (which could do useful work) is changed to disordered thermal energy when the stone strikes the ground. Disorder increases in this process, as it does in all processes that occur in nature.

Biological Development

PHYSICS APPLIED
Biological development

An interesting example of the increase in entropy relates to the biological development and growth of organisms. Clearly, a human being is a highly ordered organism. The development of an individual from a single cell to a grown person is a process of increasing order. Evolution too might be seen as an increase in order. Do these processes violate the second law of thermodynamics? No, they do not. In the processes of growth and evolution, and even during the mature life of an individual, waste products are eliminated. These small molecules that remain as a result of metabolism are simple molecules without much order. Thus they represent relatively higher disorder or entropy. Indeed, the total entropy of the molecules cast aside by organisms during the processes of development and growth is greater than the decrease in entropy associated with the order of the growing individual or evolving species.

20–8 Unavailability of Energy; Heat Death

In the process of heat conduction from a hot object to a cold one, we have seen that entropy increases and that order goes to disorder. The separate hot and cold objects could serve as the high- and low-temperature regions for a heat engine and thus could be used to obtain useful work. But after the two objects are put in contact with each other and reach the same uniform temperature, no work can be obtained from them. With regard to being able to do useful work, order has gone to disorder in this process.

The same can be said about a falling rock that comes to rest upon striking the ground. Before hitting the ground, all the kinetic energy of the rock could have been used to do useful work. But once the rock's mechanical kinetic energy becomes thermal energy, doing useful work is no longer possible.

Both examples illustrate an important aspect of the second law of thermodynamics:

in any natural process, some energy becomes unavailable to do useful work.

In any process, no energy is ever lost (it is always conserved). Rather, energy becomes less useful—it can do less useful work. As time goes on, **energy is degraded**, in a sense. It goes from more orderly forms (such as mechanical) eventually to the least orderly form: internal, or thermal, energy. Entropy is a factor here because the amount of energy that becomes unavailable to do work is proportional to the change in entropy during any process.[†]

[†]It can be shown that the amount of energy that becomes unavailable to do useful work is equal to $T_L \Delta S$, where T_L is the lowest available temperature and ΔS is the total increase in entropy during the process.

A natural outcome of the degradation of energy is the prediction that as time goes on, the universe is expected to approach a state of maximum disorder. Matter would become a uniform mixture, and heat would have flowed from high-temperature regions to low-temperature regions until the whole universe is at one temperature. No work could then be done. All the energy of the universe would have degraded to thermal energy. This prediction, called the **heat death** of the universe, has been much discussed, but it would lie *very* far in the future. It is a complicated subject, and some scientists question whether thermodynamics, as we now understand it, actually applies to the universe as a whole, which is at a much larger scale.

20–9 Statistical Interpretation of Entropy and the Second Law

The ideas of entropy and disorder are made clearer with the use of a statistical or probabilistic analysis of the molecular state of a system. This statistical approach, which was first applied toward the end of the nineteenth century by Ludwig Boltzmann (1844–1906), makes a clear distinction between the "macrostate" and the "microstate" of a system. The **microstate** of a system would be specified by giving the position and velocity of every particle (or molecule). The **macrostate** of a system is specified by giving the far fewer macroscopic properties of the system—the temperature, pressure, number of moles, and so on. In reality, we can know only the macrostate of a system. There are generally far too many molecules in a system to be able to know the velocity and position of every one at a given moment. Nonetheless, we can hypothesize a great many different microstates that can correspond to the *same* macrostate.

Let us take a very simple example. Suppose you repeatedly shake four coins in your hand and drop them on a table. Specifying the *number* of heads and the number of tails that appear on a given throw is the *macrostate* of this system. Specifying *each coin* as being a head or a tail is the *microstate* of the system. In Table 20–1 we see how many microstates correspond to each macrostate:

TABLE 20–1 Microstates per Macrostates (4 coins)

Macrostate	Possible Microstates (H = heads, T = tails)	Number of Microstates
4 heads	HHHH	1
3 heads, 1 tail	HHHT, HHTH, HTHH, THHH	4
2 heads, 2 tails	HHTT, HTHT, THHT, HTTH, THTH, TTHH	6
1 head, 3 tails	TTTH, TTHT, THTT, HTTT	4
4 tails	TTTT	1

A basic assumption behind the statistical approach is that *each microstate is equally probable*. Thus the number of microstates that give the same macrostate corresponds to the relative probability of that macrostate occurring. The macrostate of two heads and two tails is the most probable one in our case of tossing four coins; out of the total of 16 possible microstates, six correspond to two heads and two tails, so the probability of throwing two heads and two tails is 6 out of 16, or 38%. The probability of throwing one head and three tails is 4 out of 16, or 25%. The probability of four heads is only 1 in 16, or 6%. If you threw the coins 16 times, you might not find that two heads and two tails appear exactly 6 times, or four tails exactly once. These are only probabilities or averages. But if you made 1600 throws, very nearly 38% of them would be two heads and two tails. The greater the number of tries, the closer the percentages are to the calculated probabilities.

CAUTION
Careful use of probability

EXERCISE D In Table 20–1, what is the probability that there will be at least two heads? (*a*) $\frac{1}{2}$; (*b*) $\frac{1}{16}$; (*c*) $\frac{1}{8}$; (*d*) $\frac{3}{8}$; (*e*) $\frac{11}{16}$.

If we toss more coins—say, 100 all at the same time—the relative probability of throwing all heads (or all tails) is greatly reduced. There is only one microstate corresponding to all heads. For 99 heads and 1 tail, there are 100 microstates because each of the coins could be the one tail. The relative probabilities for other macrostates are given in Table 20–2. About 1.3×10^{30} microstates are possible.[†] Thus the relative probability of finding all heads is about 1 in 10^{30}, an incredibly unlikely event! The probability of obtaining 50 heads and 50 tails (see Table 20–2) is $(1.0 \times 10^{29})/(1.3 \times 10^{30}) = 0.08$ or 8%. The probability of obtaining anything between 45 and 55 heads is about 70%.

TABLE 20–2 Probabilities of Various Macrostates for 100 Coin Tosses

Macrostate		Number of Microstates	Probability
Heads	Tails		
100	0	1	7.9×10^{-31}
99	1	1.0×10^{2}	7.9×10^{-29}
90	10	1.7×10^{13}	1.4×10^{-17}
80	20	5.4×10^{20}	4.2×10^{-10}
60	40	1.4×10^{28}	0.011
55	45	6.1×10^{28}	0.047
50	50	1.0×10^{29}	0.077
45	55	6.1×10^{28}	0.047
40	60	1.4×10^{28}	0.011
20	80	5.4×10^{20}	4.2×10^{-10}
10	90	1.7×10^{13}	1.4×10^{-17}
1	99	1.0×10^{2}	7.9×10^{-29}
0	100	1	7.9×10^{-31}

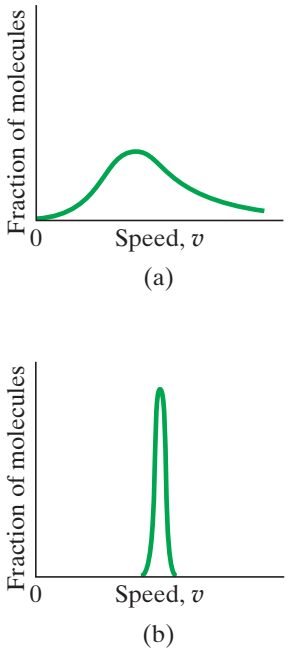


FIGURE 20–13 (a) Most probable distribution of molecular speeds in a gas (Maxwellian, or random). (b) Orderly, but highly unlikely, distribution of speeds in which all molecules have nearly the same speed.

Thus we see that as the number of coins increases, the probability of obtaining the most orderly arrangement (all heads or all tails) becomes extremely unlikely. The least orderly arrangement (half heads, half tails) is the most probable, and the probability of being within, say, 5% of the most probable arrangement greatly increases as the number of coins increases. These same ideas can be applied to the molecules of a system. For example, the most probable state of a gas (say, the air in a room) is one in which the molecules take up the whole space and move about randomly; this corresponds to the Maxwellian distribution, Fig. 20–13a (and see Section 18–2). On the other hand, the very orderly arrangement of all the molecules located in one corner of the room and all moving with the same velocity (Fig. 20–13b) is extremely unlikely.

From these examples, we see that probability is directly related to disorder and hence to entropy. That is, the most probable state is the one with greatest entropy or greatest disorder and randomness; it is also the macrostate that corresponds to the *most microstates*. Boltzmann showed, consistent with Clausius's definition $(dS = dQ/T)$, that the entropy of a system in a given (macro) state can be written

$$S = k \ln \mathcal{W}, \tag{20–14}$$

where k is Boltzmann's constant $(k = R/N_A = 1.38 \times 10^{-23}\,\text{J/K})$ and $\mathcal{W}$ is the number of microstates corresponding to the given macrostate. That is, $\mathcal{W}$ is proportional to the probability of occurrence of that state. $\mathcal{W}$ is called the **thermodynamic probability**, or, sometimes, the disorder parameter.

[†]Each coin has two possibilities, heads or tails. Then the possible number of microstates is $2 \times 2 \times 2 \times \cdots = 2^{100} \approx 1.3 \times 10^{30}$ (using a calculator or logarithms).

EXAMPLE 20–11 **Free expansion—statistical determination of entropy.** Use Eq. 20–14 to determine the change in entropy for the adiabatic *free expansion* of a gas, a calculation we did macroscopically in Example 20–9. Assume $\mathcal{W}$, the number of microstates for each macrostate, is the number of possible positions.

APPROACH We assume the number of moles is $n = 1$, and then the number of molecules is $N = nN_A = N_A$, Avogadro's number. We let the volume double, just as in Example 20–9. Because the volume doubles, the number of possible positions for each molecule doubles.

SOLUTION When the volume doubles, each molecule has two times as many positions (microstates) available. For two molecules, the number of total microstates increases by $2 \times 2 = 2^2$. For N_A molecules, the total number of microstates increases by a factor of $2 \times 2 \times 2 \times \cdots = 2^{N_A}$. That is

$$\frac{\mathcal{W}_2}{\mathcal{W}_1} = 2^{N_A}.$$

The change in entropy is, from Eq. 20–14,

$$\Delta S = S_2 - S_1 = k(\ln \mathcal{W}_2 - \ln \mathcal{W}_1) = k \ln \frac{\mathcal{W}_2}{\mathcal{W}_1} = k \ln 2^{N_A}$$
$$= kN_A \ln 2 = R \ln 2$$

which is the same result we obtained in Example 20–9.

In terms of probability, the second law of thermodynamics—which tells us that entropy increases in any process—reduces to the statement that **those processes occur which are most probable**.

There is an additional element, however. The second law in terms of probability does not *forbid* a decrease in entropy. Rather, it says the probability is extremely low. It is not impossible that salt and pepper could separate spontaneously into layers, or that a broken cup could mend itself (Fig. 20–1). It is even possible that a lake could freeze over on a hot summer day (that is, for heat to flow out of the cold lake into the warmer surroundings). But the probability for such events occurring is extremely small. In our coin examples, we saw that increasing the number of coins from 4 to 100 drastically reduced the probability of large deviations from the average, or most probable, arrangement. In ordinary systems we are not dealing with only 100 molecules, but with extremely large numbers of molecules: in 1 mol alone there are 6×10^{23} molecules. Hence the probability of deviation far from the average is incredibly tiny. For example, it has been calculated that the probability that a stone resting on the ground could transform 1 cal of thermal energy into mechanical energy and rise up into the air is much less likely than the probability that a group of monkeys typing randomly would by chance produce the complete works of Shakespeare. Put another way, the probability is less than once in the entire age of the universe; this could be what "never" means.

*20–10 Thermodynamic Temperature; Third Law of Thermodynamics

In Section 20–3 we saw for a Carnot cycle that the ratio of the heat absorbed Q_H from the high-temperature reservoir and the heat exhausted Q_L to the low-temperature reservoir is equal to the ratio of the temperatures of the two reservoirs (Eq. 20–2):

$$\frac{Q_L}{Q_H} = \frac{T_L}{T_H}.$$

This result is valid for any reversible engine and does not depend on the working substance. It can thus serve as the basis for the **Kelvin** or **thermodynamic temperature scale**.

We use this relation and the ideal gas temperature scale (Section 17–10) to complete the definition of the thermodynamic scale: we assign the value $T_{tp} = 273.16\,\text{K}$ to the triple point of water so that

$$T = (273.16\,\text{K})\left(\frac{Q}{Q_{tp}}\right),$$

where Q and Q_{tp} are the magnitudes of the heats exchanged by a Carnot engine with reservoirs at temperatures T and T_{tp}. Then the thermodynamic scale is identical to the ideal gas scale over the latter's range of validity.

Very low temperatures are difficult to obtain experimentally. The closer the temperature is to absolute zero, the more difficult it is to reduce the temperature further, and it is generally accepted that *it is not possible to reach absolute zero in any finite number of processes*. This last statement is one way to state the **third law of thermodynamics**. Since the maximum efficiency that any heat engine can have is the Carnot efficiency

$$e = 1 - \frac{T_L}{T_H},$$

and since T_L can never be zero, we see that a 100% efficient heat engine is not possible.

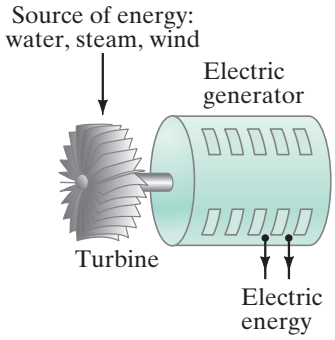
(a) 

(b) 

(c) 

FIGURE 20–14 (a) A fossil-fuel steam plant (this one uses forest waste products = biomass). (b) An array of mirrors focuses sunlight on a boiler to produce steam at a solar energy installation. (c) Large cooling towers at an electric generating plant.

FIGURE 20–15 Mechanical energy is transformed to electric energy with a turbine and generator.



PHYSICS APPLIED

Heat engines and thermal pollution

20–11 Thermal Pollution, Global Warming, and Energy Resources

Much of the energy we use in everyday life—from motor vehicles to most of the electricity produced by power plants—uses a heat engine. About $\frac{2}{3}$ of the electric energy produced in the U.S. is generated at fossil-fuel steam plants (coal, oil, or gas—see Fig. 20–14a), and they make use of a heat engine (steam engines). In electric power plants, the steam drives the turbines and generators (Fig. 20–15) whose output is electric energy. The various means to turn the turbine are shown in Table 20–3, along with some of the advantages and disadvantages of each. Even nuclear power plants use a steam engine, run on nuclear fuel. Electricity produced by falling water at dams, by windmills, or by solar cells (Fig. 20–14b) does not involve a heat engine.

The heat output Q_L from every heat engine, from power plants to cars, is referred to as **thermal pollution** because this heat (Q_L) must be absorbed by the environment—such as by water from rivers or lakes, or by the air using large cooling towers (Fig. 20–14c). When water is the coolant, this heat raises the temperature of the water, altering the natural ecology of aquatic life (warmer water holds less oxygen). In air cooling towers, the output heat Q_L raises the temperature of the atmosphere, which affects climate.

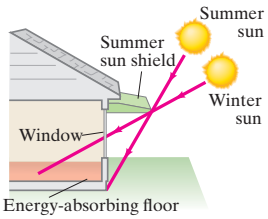
Air pollution—the chemicals released in the burning of fossil fuels in cars, power plants, heating, and industrial furnaces—gives rise to smog and other problems. Buildup of CO_2 in the Earth's atmosphere, due to the burning of fossil fuels (carbon in the fuel combines with O_2 of the air) absorbs infrared radiation the Earth naturally emits, and thus contributes to **global warming**.

Electric energy production (Table 20–3) is responsible for only about $\frac{1}{4}$ of fossil-fuel pollution. The other $\frac{3}{4}$ is due to direct burning of fossil fuels in cars, trucks, airplanes, manufacturing furnaces, and heating houses and buildings.

The **carbon footprint** of these activities refers to the negative impact on the environment due to the burning of fossil fuels releasing CO_2 (and other noxious products) into the atmosphere, contributing to **climate change**. A carbon footprint can be expressed in kg or tons of carbon (in the CO_2). Limiting the burning of fossil fuels can help reduce these problems.

And do not be fooled that electric cars do not pollute. They do. When you plug them in to charge, that electricity is produced at power plants which mostly burn very polluting fossil fuels (Table 20–3). So *your electric car is no cleaner than the power plant that supplies your electricity*. Furthermore, to make an electric car includes massive energy storage batteries which require very large amounts of energy (and pollution) to manufacture, and later to recycle (or dispose of) when worn out.

TABLE 20–3 Electric Energy Resources

Form of Electric Energy Production	% of Production (approx.) U.S.	% of Production (approx.) World	Advantages	Disadvantages
Fossil-fuel steam plants: burn coal, oil, or natural gas to boil water, producing high-pressure steam that turns a turbine of a generator (Figs. 20–3b, 20–15); uses heat engine.	64	65	We know how to build them; for now relatively inexpensive.	Air pollution; thermal pollution; limited efficiency; land devastation from extraction of raw materials (mining); global warming; accidents such as oil spills at sea or on land; limited fuel supply (estimates range from a few decades to maybe a century).
Nuclear energy:				
Fission: nuclei of uranium or plutonium atoms split ("fission") with release of energy (Chapter 42) that heats steam; uses heat engine.	19	10	Normally almost no air pollution; less contribution to global warming; relatively inexpensive. 	Thermal pollution; accidents can release damaging radioactivity; difficult disposal of radioactive by-products; possible diversion of nuclear material by terrorists; limited fuel supply.
Fusion: energy released when isotopes of hydrogen (or other small nuclei) combine or "fuse" (Chapter 42).	0	0	Relatively "clean"; vast fuel supply (hydrogen in water molecules in oceans); less contribution to global warming.	Not yet workable.
Hydroelectric: falling water turns turbines at the base of a dam. 	7	16	No heat engine needed; no air, water, or thermal pollution; relatively inexpensive; high efficiency; dams can control flooding.	Reservoirs behind dams inundate scenic or inhabited land; they fill with silt over time and become useless; dams block upstream migration of salmon and other fish for reproduction; few locations remain for new dams; drought.
Geothermal: natural steam from inside the Earth comes to the surface (hot springs, geysers, steam vents); or cold water passed down into contact with hot, dry rock is heated to steam.	<1	1	No heat engine needed; little air pollution; good efficiency; relatively inexpensive and "clean."	Few appropriate sites; small production; mineral content of spent hot water can pollute.
Wind power: 3-kW to 5-MW windmills (vanes up to 50 m long) turn a generator. 	7	5	No heat engine; no air, water, or thermal pollution; relatively inexpensive.	Large array of big windmills might affect weather and be eyesores; slows down cooling winds; hazardous to migratory birds; winds not always strong.
Solar energy:				
Active solar heating: rooftop solar panels absorb the Sun's rays, heating water in tubes for inside heating and hot water supply. (See also Fig. 20–14b.)	1	>1	No heat engine needed; no air or thermal pollution; unlimited fuel supply. 	Space limitations; may require back-up; relatively expensive; less effective when cloudy.
Passive solar heating: architectural devices—windows along southern exposure, sunshade over windows to keep Sun's rays out in summer.			No heat engine needed; no air or thermal pollution; relatively inexpensive.	Almost none, but other methods needed too. 
Solar cells (photovoltaic cells): convert sunlight directly into electricity without use of heat engine.	2 [California ≈ 20%]	3	No heat engine; thermal, air, and water pollution very low; good efficiency (>30% and improving).	Expensive; chemical pollution at manufacture; much land needed as Sun's energy not concentrated; absorption (dark color) without reemission = global warming.

PHYSICS APPLIED

Thermal pollution unavoidable
Best = use less energy!
[Avoid also plastic especially in packaging: protect our oceans.]

Thermal pollution, however, is unavoidable. Engineers can try to design and build engines that are more efficient, but they cannot surpass the Carnot efficiency and must live with T_L being at best the ambient temperature of water or air. The second law of thermodynamics tells us the limit imposed by nature. Even alternative energy sources like wind and solar cells can contribute to global warming: windmills slow down cooling winds; solar cells are "dark" and absorb more of the Sun's energy, thus warming the Earth more. The best we can do, in the light of the second law of thermodynamics, is use less energy and conserve our fuel resources.

PROBLEM SOLVING

Thermodynamics

1. Define the **system** you are dealing with; distinguish the system under study from its surroundings.
2. When applying the first law of thermodynamics, be careful of **signs** associated with **work** and **heat**. In the first law, work done *by* the system is positive; work done *on* the system is negative. Heat *added* to the system is positive, but heat *removed* from it is negative. With heat engines, we usually consider the heat intake, the heat exhausted, and the work done as positive.
3. Watch the **units** used for work and heat; work is most often expressed in joules, and heat can be in calories, kilocalories, or joules. Be consistent: choose only one unit for use throughout a given problem.
4. **Temperatures** must generally be expressed in kelvins; temperature *differences* may be expressed in C° or K.
5. **Efficiency** (or coefficient of performance) is a ratio of two energy transfers: useful output divided by required input. Efficiency (but *not* coefficient of performance) is always less than 1 in value, and hence is often stated as a percentage.
6. The **entropy** of a system increases when heat is added to the system, and decreases when heat is removed. If heat is transferred from system A to system B, the change in entropy of A is negative and the change in entropy of B is positive.

Summary

A **heat engine** is a device for changing thermal energy, by means of heat flow, partially into useful work.

The **efficiency** e of a heat engine is defined as the ratio of the work W done by the engine to the high temperature heat input Q_H. Because of conservation of energy, the work output equals $Q_H - Q_L$, where Q_L is the heat exhausted at low temperature to the environment; hence

$$e = \frac{W}{Q_H} = \frac{Q_H - Q_L}{Q_H} = 1 - \frac{Q_L}{Q_H}. \quad \textbf{(20–1)}$$

Q_H, Q_L, W, as defined for heat engines, are positive.

A **Carnot engine** consists of two isothermal and two adiabatic processes in an idealized reversible cycle.

The *upper limit* on the efficiency (the *Carnot efficiency*) can be written in terms of the higher and lower operating temperatures (in kelvins) of the engine, T_H and T_L, as

$$e_{\text{ideal}} = 1 - \frac{T_L}{T_H}. \quad \textbf{(20–3)}$$

Real (irreversible) engines always have an efficiency less than this.

The operation of **refrigerators** and **air conditioners** is the reverse of a heat engine: work is done to extract heat Q_L from a cool region and exhaust it to a region at a higher temperature. The coefficient of performance (COP) for either is

$$\text{COP} = \frac{Q_L}{W}, \quad \begin{bmatrix}\text{refrigerator or}\\ \text{air conditioner}\end{bmatrix} \quad \textbf{(20–4a)}$$

where W is the work needed to remove heat Q_L from the area with the low temperature, which is the purpose of these devices.

A **heat pump** uses work W to bring heat Q_L from the cold outside and deliver heat Q_H to warm the interior. The coefficient of performance of a heat pump is

$$\text{COP} = \frac{Q_H}{W}, \quad \text{[heat pump]} \quad \textbf{(20–5)}$$

because it is the heat Q_H delivered inside the building that counts.

The **second law of thermodynamics** can be stated in several equivalent ways:

(a) heat flows spontaneously from a hot object to a cold one, but not the reverse;
(b) there can be no 100% efficient heat engine—that is, one that can change a given amount of heat completely into work;
(c) natural processes tend to move toward a state of greater disorder or greater **entropy**.

Statement (c) is the most general statement of the second law of thermodynamics, and can be restated as: the total entropy, S, of any system plus that of its environment increases as a result of any natural process:

$$\Delta S > 0. \quad \textbf{(20–13)}$$

The change in entropy in a process that transfers heat Q at a constant temperature T is

$$\Delta S = \frac{Q}{T}. \quad \textbf{(20–6)}$$

In general for a reversible process, $\Delta S = \int dQ/T$.

Entropy is a quantitative measure of the **disorder** of a system. The second law of thermodynamics also indicates that as time goes on, energy is **degraded** to less useful forms—that is, it is less available to do useful work.

The second law of thermodynamics tells us in which direction processes tend to proceed; hence entropy is called **time's arrow**.

Entropy can be examined from a statistical point of view, considering **macrostates** (for example, P, V, T) and **microstates** (state of each molecule). The most probable processes are the ones we observe, and they are the macrostates that have the most microstates. They are the ones that increase entropy the most. Processes that violate the second law "could" occur, but only with *extremely* low probability.

[*All heat engines give rise to **thermal pollution** because they exhaust heat to the environment.]

Questions

1. Can mechanical energy ever be transformed completely into heat or internal energy? Can the reverse happen? In each case, if your answer is no, explain why not; if yes, give one or two examples.
2. Can you warm a kitchen in winter by leaving the oven door open? Can you cool the kitchen on a hot summer day by leaving the refrigerator door open? Explain.
3. Would a definition of heat engine efficiency as $e = W/Q_L$ be useful? Explain.
4. What are the high-temperature and the low-temperature areas for (*a*) an internal combustion engine, and (*b*) a steam engine? Are they, strictly speaking, heat reservoirs?
5. What exactly is a Carnot engine and why is it important? How practical is it?
6. The oceans contain a tremendous amount of thermal (internal) energy. Why, in general, is it not possible to put this energy to useful work?
7. Discuss the factors that keep real engines from reaching Carnot efficiency.
8. The expansion valve in a refrigeration system, Fig. 20–9, is crucial for cooling the fluid. Explain how the cooling occurs.
9. (*a*) Describe how heat could be added to a system reversibly. (*b*) Could you use a stove burner to add heat to a system reversibly? Explain.
10. (*a*) What happens if you remove the lid of a bottle containing chlorine gas? (*b*) Does the reverse process ever happen? Why or why not? (*c*) Can you think of two other examples of irreversibility?
11. The COPs are defined differently for heat pumps and air conditioners. Explain why.
12. A gas is allowed to expand (*a*) adiabatically and (*b*) isothermally. In each process, does the entropy increase, decrease, or stay the same? Explain.
13. Give three examples, other than those mentioned in this Chapter, of naturally occurring processes in which order goes to disorder. Discuss the observability of the reverse process.
14. Which do you think has the greater entropy, 1 kg of solid iron or 1 kg of liquid iron? Why?
15. (*a*) Make up an advertisement for a refrigerator or an air conditioner that violates the first law of thermodynamics (see Section 20–4). (*b*) Make up an ad for a car engine that violates the second law of thermodynamics (see Section 20–2).
16. Think up several processes (other than those already mentioned) that would obey the first law of thermodynamics, but, if they actually occurred, would violate the second law.
17. Suppose a lot of papers are strewn all over the floor; then you stack them neatly. Does this violate the second law of thermodynamics? Explain.
18. The first law of thermodynamics is sometimes whimsically stated as, "You can't get something for nothing," and the second law as, "You can't even break even." Explain how these statements could be equivalent to the formal statements.
19. Powdered milk is very slowly (quasistatically) added to water while being stirred. Is this a reversible process? Explain.
20. Two identical systems are taken from state a to state b by two different *irreversible* processes. Will the change in entropy for the system be the same for each process? For the environment? Answer carefully and completely.
21. It can be said that the *total change in entropy during a process is a measure of the irreversibility of the process*. Discuss why this is valid, starting with the fact that $\Delta S_{\text{total}} = \Delta S_{\text{system}} + \Delta S_{\text{environment}} = 0$ for a reversible process.
22. Entropy is often called "time's arrow" because it tells us in which direction natural processes occur. If a movie were run backward, name some processes that you might see that would tell you that time was "running backward."
23. Living organisms, as they grow, convert relatively simple food molecules into a complex structure. Is this a violation of the second law of thermodynamics? Explain your answer.

MisConceptual Questions

1. Which is possible: converting (i) 100 J of work entirely into 100 J of heat, (ii) 100 J of heat entirely into 100 J of work?
   (*a*) Only (i) is possible.
   (*b*) Only (ii) is possible.
   (*c*) Both (i) and (ii) are possible.
   (*d*) Neither (i) nor (ii) is possible.
2. A heat engine operates between a high temperature of about 600°C and a low temperature of about 300°C. What is the maximum theoretical efficiency for this engine?
   (*a*) = 100%. (*b*) ≈ 66%. (*c*) ≈ 50%. (*d*) ≈ 34%.
   (*e*) Cannot be determined from the given information.
3. On a very hot day, could you cool your kitchen by leaving the refrigerator door open?
   (*a*) Yes, but it would be very expensive.
   (*b*) Yes, but only if the humidity is below 50%.
   (*c*) No, the refrigerator would exhaust the same amount of heat into the room as it takes out of the room.
   (*d*) No, the heat exhausted by the refrigerator into the room is more than the heat the refrigerator takes out of the room.
4. Which of the following possibilities could increase the efficiency of a heat engine or an internal combustion engine?
   (*a*) Increase the temperature of the hot part of the system and reduce the temperature of the exhaust.
   (*b*) Increase the temperatures of both the hot part and the exhaust part of the system by the same amount.
   (*c*) Decrease the temperatures of both the hot part and the exhaust part of the system by the same amount.
   (*d*) Decrease the temperature of the hot part and increase the temperature of the exhaust part by the same amount.
   (*e*) None of the above; only redesigning the engine or using better gas could improve the engine's efficiency.
5. About what percentage of the heat produced by burning gasoline is turned into useful work by a typical automobile?
   (*a*) 20%.
   (*b*) 50%.
   (*c*) 80%.
   (*d*) 90%.
   (*e*) Nearly 100%.

6. An engine manufacturer makes the claim that the engine they have developed will, on each cycle, take 100 J of heat out of boiling water at 100°C, do mechanical work of 80 J, and exhaust 20 J of heat at 10°C. What, if anything, is wrong with this claim?
   (*a*) The heat exhausted must always be greater than the work done according to the second law of thermodynamics.
   (*b*) There is nothing wrong with this claim because 100 J = 20 J + 80 J.
   (*c*) The efficiency of this engine is greater than the ideal Carnot cycle efficiency.
   (*d*) This engine violates the first law of thermodynamics because 100 J + 20 J ≠ 80 J.
   (*e*) An engine would operate by taking in heat at the lower temperature and exhausting heat at the higher temperature.

7. Which statement is true regarding the entropy change of an ice cube that melts?
   (*a*) Since melting occurs at the melting point temperature, there is no temperature change so there is no entropy change.
   (*b*) Entropy increases.
   (*c*) Entropy decreases.

8. A piece of ice at 0°C melts very slowly at 0°C. Assume that the ice is in contact with a heat reservoir whose temperature is only infinitesimally greater than 0°C. What are the entropy changes of the ice cube and the heat reservoir, respectively?
   (*a*) Both entropy changes are negative.
   (*b*) Both entropy changes are 0.
   (*c*) Both entropy changes are positive.
   (*d*) The changes are the same magnitude, but the ice gains entropy and the reservoir loses entropy.
   (*e*) The changes are the same magnitude, but the ice loses entropy and the reservoir gains entropy.

9. When a gas is compressed isothermally, its entropy
   (*a*) increases.
   (*b*) decreases.
   (*c*) stays the same.

10. There are three coins on a table—one head and two tails. How many microstates would give this macrostate?
   (*a*) 1 (*b*) 2 (*c*) 3 (*d*) 6
   (*e*) None of the above.

11. Which of the processes shown in Fig. 20–16 results in the least entropy change for an ideal gas changing from state S (start) to state F (finish)?
   (*a*) Process A.
   (*b*) Process B.
   (*c*) Process C.
   (*d*) Process D.
   (*e*) All 4 processes produce the same entropy change of the gas.

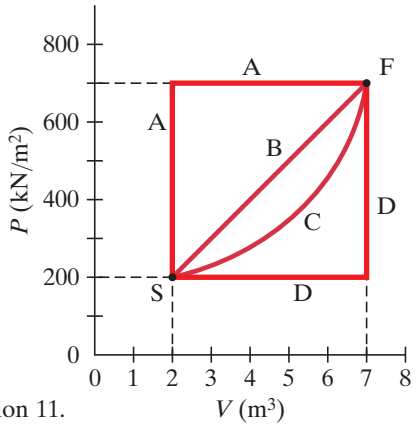


FIGURE 20–16
MisConceptual Question 11.

12. Which of the following are not subject to the second law?
   (*a*) Electric cars.
   (*b*) Electricity-producing windmills.
   (*c*) Hydroelectric power stations.
   (*d*) Nuclear power stations.
   (*e*) Coal-fired power stations.
   (*f*) None of the above. All are subject to the second law.

Problems

20–2 Heat Engines

1. (I) A heat engine exhausts 7600 J of heat while performing 2600 J of useful work. What is the efficiency of this engine?
2. (I) The exhaust temperature of a heat engine is 230°C. What is the high temperature if the Carnot efficiency is 34%?
3. (II) A certain power plant puts out 680 MW of electric power. Estimate the heat discharged per second, assuming that the plant has an efficiency of 32%.
4. (II) A typical compact car experiences a total drag force of about 350 N at 55 mi/h. If this car gets 32 miles per gallon of gasoline at this speed, and a liter of gasoline (1 gal = 3.8 L) releases about 3.2×10^7 J when burned, what is the car's efficiency?
5. (II) A four-cylinder gasoline engine has an efficiency of 0.22 and delivers 180 J of work per cycle per cylinder. If the engine runs at 25 cycles per second (1500 rpm), determine (*a*) the work done per second, and (*b*) the total heat input per second from the gasoline. (*c*) If the energy content of gasoline is 120 MJ per gallon, how far can the car go on one gallon (or on one liter), assuming 1500 rpm in high gear corresponds to a speed of 80 km/h?
6. (III) Figure 20–17 is a *PV* diagram for a reversible heat engine in which 1.0 mol of argon, a nearly ideal monatomic gas, is initially at STP (point a). Points b and c are on an isotherm at $T = 423$ K. Process ab is at constant volume, process ac at constant pressure. (*a*) Is the path of the cycle carried out clockwise or counterclockwise? (*b*) What is the efficiency of this engine?

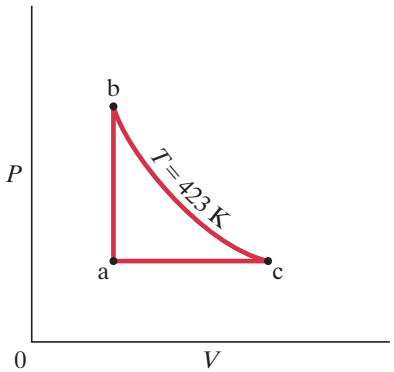


FIGURE 20–17 Problem 6.

20–3 Carnot Engine

7. (I) What is the maximum efficiency of a heat engine whose operating temperatures are 530°C and 345°C?

8. (II) Nitrogen liquefies at 77 K (Fig. 20–18) and is relatively cheap to produce. (*a*) Estimate the efficiency of an ideal engine that operated between room temperature (293 K) and liquid nitrogen as the low temperature reservoir. (*b*) Discuss the issues for why this high efficiency may not be practical.



FIGURE 20–18
Problem 8.

9. (II) A heat engine exhausts its heat at 340°C and has a Carnot efficiency of 36%. What exhaust temperature would enable it to achieve a Carnot efficiency of 45%?

10. (II) A heat engine uses a heat source at 540°C and has an ideal (Carnot) efficiency of 22%. To increase the ideal efficiency to 42%, what must be the temperature of the heat source?

11. (II) (*a*) Show that the work done by a Carnot engine is equal to the area enclosed by the Carnot cycle on a *PV* diagram, Fig. 20–5. (See Section 19–7.) (*b*) Generalize this to any reversible cycle.

12. (II) A Carnot engine's operating temperatures are 210°C and 45°C. The engine's power output is 960 W. Calculate the rate of heat output.

13. (II) A nuclear power plant operates at 65% of its maximum theoretical (Carnot) efficiency between temperatures of 660°C and 330°C. If the plant produces electric energy at the rate of 1.4 GW, how much exhaust heat is discharged per hour?

14. (II) Which will improve the efficiency of a Carnot engine more: a 10 C° increase in the high-temperature reservoir, or a 10 C° decrease in the low-temperature reservoir? Give detailed results. Can you state a generalization?

15. (II) Assume that a 65-kg hiker needs to eat 4.0×10^3 kcal of energy to supply a day's worth of metabolism ($= Q_H$). Estimate the elevation change the person can climb in one day, using only this amount of energy. As a fun and rough prediction, treat the person as an isolated heat engine, operating between the internal temperature of 37°C (98.6°F) and the ambient air temperature of 20°C.

16. (II) A Carnot engine performs work at the rate of 520 kW with an input of 950 kcal of heat per second. If the temperature of the heat source is 480°C, at what temperature is the waste heat exhausted?

17. (II) A particular car does work at the rate of about 7.0 kJ/s when traveling at a steady 21.8 m/s along a level road. This is the work done against friction. The car can travel 17 km on 1.0 L of gasoline at this speed (about 40 mi/gal). What is the minimum value for T_H if T_L is 25°C? The energy available from 1.0 L of gas is 3.2×10^7 J.

18. (II) The working substance of a certain Carnot engine is 1.0 mol of an ideal monatomic gas. During the isothermal expansion portion of this engine's cycle, the volume of the gas doubles, while during the adiabatic expansion the volume increases by a factor of 6.2. The work output of the engine is 920 J in each cycle. Compute the temperatures of the two reservoirs between which this engine operates.

19. (III) A Carnot cycle, shown in Fig. 20–5, has the following conditions: $V_a = 7.5$ L, $V_b = 15.0$ L, $T_H = 470$°C, and $T_L = 260$°C. The gas used in the cycle is 0.50 mol of a diatomic gas, $\gamma = 1.4$. Calculate (*a*) the pressures at a and b; (*b*) the volumes at c and d. (*c*) What is the work done along process ab? (*d*) What is the heat lost along process cd? (*e*) Calculate the net work done for the whole cycle. (*f*) What is the efficiency of the cycle, using the definition $e = W/Q_H$? Show that this is the same as given by Eq. 20–3.

20. (III) One mole of monatomic gas undergoes a Carnot cycle with $T_H = 350$°C and $T_L = 210$°C. The initial pressure is 8.8 atm. During the isothermal expansion, the volume doubles. (*a*) Find the values of the pressure and volume at the points a, b, c, and d of Fig. 20–5. (*b*) Determine Q, W, and ΔE_{int} for each segment of the cycle. (*c*) Calculate the efficiency of the cycle using Eqs. 20–1 and 20–3.

*21. (III) In an engine that approximates the Otto cycle (Fig. 20–7), gasoline vapor must be ignited at the end of the cylinder's adiabatic compression by the spark from a spark plug. The ignition temperature of 87-octane gasoline vapor is about 430°C. Assuming that the working gas is diatomic and enters the cylinder at 25°C, determine the maximum compression ratio of the engine.

20–4 Refrigerators, Air Conditioners, Heat Pumps

22. (I) If an ideal refrigerator keeps its contents at 2.5°C when the house temperature is 22°C, what is its COP?

23. (I) The low temperature of a freezer cooling coil is −8°C and the discharge temperature is 33°C. What is the maximum theoretical coefficient of performance?

24. (I) A Carnot heat pump supplies 3900 W to keep a room at 23°C while drawing energy from the outside air at the rate of 3400 W. What is the outside air temperature?

25. (II) What is the temperature inside an ideal refrigerator-freezer that operates with a COP = 7.0 in a 22°C room?

26. (II) A *heat pump* is used to keep a house warm at 22°C. How much work is required of the pump to deliver 2800 J of heat into the house if the outdoor temperature is (*a*) 0°C, (*b*) −15°C? Assume a COP of 3.0. (*c*) Redo for both temperatures assuming an ideal Carnot heat pump, first deriving the ideal (Carnot) coefficient of performance as

$$\text{COP} = \frac{T_H}{T_H - T_L}.$$

27. (II) (*a*) What is the coefficient of performance of an ideal heat pump that extracts heat from 6°C air outside and deposits heat inside a house at 24°C? (*b*) If this heat pump operates on 1200 W of electrical power, what is the maximum heat it can deliver into your house each hour? See Problem 26.

28. (II) (*a*) Given that the coefficient of performance of a *refrigerator* is defined (Eq. 20–4a) as

$$\text{COP} = \frac{Q_L}{W},$$

show that Eqs. 20–4c is valid for an ideal (Carnot) refrigerator,

$$\text{COP}_{\text{ideal}} = \frac{T_L}{T_H - T_L}.$$

(*b*) Write the COP in terms of the efficiency e of the reversible heat engine obtained by running the refrigerator backward. (*c*) What is the coefficient of performance for an ideal refrigerator that maintains a freezer compartment at −18°C when the condenser's temperature is 24°C?

29. (II) An ideal heat pump is used to maintain the inside temperature of a house at $T_{in} = 22°C$ when the outside temperature is T_{out}. Assume the heat pump does work at a rate of 1700 W. Also assume that the house loses heat via conduction through its walls and other surfaces at a rate given by $(650\ W/C°)(T_{in} - T_{out})$. (*a*) For what outside temperature would the heat pump have to operate at all times in order to maintain the house at an inside temperature of 22°C? (*b*) If the outside temperature is less than you just calculated, what happens? (*c*) If the outside temperature is 8°C, what percentage of the time does the heat pump have to operate in order to maintain the house at an inside temperature of 22°C?

30. (II) An ideal (Carnot) engine has an efficiency of 33%. If it were possible to run it backward as a heat pump, what would be its coefficient of performance?

31. (II) What volume of water at 0°C can a freezer make into ice cubes in 1.0 h, if the coefficient of performance of the cooling unit is 6.0 and the power input is 1.2 kilowatt?

32. (II) A central heat pump operating as an air conditioner draws 33,000 Btu per hour from a building and operates between the temperatures of 24°C and 38°C. (*a*) If its coefficient of performance is only 25% of a Carnot air conditioner, what is the effective coefficient of performance? (*b*) What is the power (kW) required of the compressor motor? (*c*) What is the power in terms of hp?

33. (II) How much less per year would it cost a family to operate a heat pump that has a coefficient of performance of 2.9 than an electric heater that costs $2100 to heat their home for a year? If the conversion to the heat pump costs $15,000, how long would it take the family to break even on heating costs? How much would the family save in 20 years?

20–5 and 20–6 Entropy

34. (I) What is the change in entropy of 280 g of steam at 100°C when it is condensed to water at 100°C?

35. (I) 1.0 kg of water is heated from 0°C to 100°C. Estimate the change in entropy of the water.

36. (I) What is the change in entropy of 1.00 m^3 of water at 0°C when it is frozen to ice at 0°C?

37. (II) A 6.8-kg box having an initial speed of 4.0 m/s slides along a rough table and comes to rest. Estimate the total change in entropy of the universe. Assume all objects are at room temperature (293 K).

38. (II) If 0.45 kg of water at 100°C is changed by a reversible process to steam at 100°C, determine the change in entropy of (*a*) the water, (*b*) the surroundings, and (*c*) the universe as a whole. (*d*) How would your answers differ if the process were irreversible?

39. (II) If 1.00 m^3 of water at 0°C is frozen and cooled to −8.0°C by being in contact with a great deal of ice at −8.0°C, estimate the total change in entropy of the process.

40. (II) A 2.8-kg piece of aluminum at 36.5°C is placed in 1.0 kg of water in a Styrofoam container at room temperature (20.0°C). Determine the net change in entropy of the system when it comes to equilibrium.

41. (II) An aluminum rod conducts 8.25 cal/s from a heat source maintained at 225°C to a large body of water at 22°C. Calculate the rate at which entropy increases in this process.

42. (II) A falling rock has kinetic energy K just before striking the ground and coming to rest. What is the total change in entropy of rock plus environment as a result of this collision?

43. (II) An ideal gas expands isothermally $(T = 440\ K)$ from a volume of 2.50 L and a pressure of 7.5 atm to a pressure of 1.0 atm. What is the entropy change for this process?

44. (II) When 2.0 kg of water at 12.0°C is mixed with 3.0 kg of water at 38.0°C in a well-insulated container, what is the change in entropy of the system? (*a*) Make an estimate; (*b*) use the integral $\Delta S = \int dQ/T$.

45. (II) A real heat engine working between heat reservoirs at 970 K and 650 K produces 550 J of work per cycle for a heat input of 2500 J. (*a*) Compare the efficiency of this real engine to that of an ideal (Carnot) engine. (*b*) Calculate the total entropy change of the universe per cycle of the real engine, and (*c*) also if the engine is ideal (Carnot).

46. (II) The temperature of 1.6 mol of an ideal diatomic gas goes from 25°C to 55°C at a constant volume. What is the change in entropy? Use $\Delta S = \int dQ/T$.

47. (II) (*a*) An ice cube of mass m at 0°C is placed in a large 20°C room. Heat flows (from the room to the ice cube) such that the ice cube melts and the liquid water warms to 20°C. The room is so large that its temperature remains nearly 20°C at all times. Calculate the change in entropy for the (water + room) system due to this process. Will this process occur naturally? (*b*) A mass m of liquid water at 20°C is placed in a large 20°C room. Heat flows (from the water to the room) such that the liquid water cools to 0°C and then freezes into a 0°C ice cube. The room is so large that its temperature remains 20°C at all times. Calculate the change in entropy for the (water + room) system due to this process. Will this process occur naturally?

48. (II) An ideal gas of n moles undergoes the reversible process ab shown in the PV diagram of Fig. 20–19. The temperature T of the gas is the same at points a and b. Determine the change in entropy of the gas due to this process. [*Hint*: Recall that entropy is a state variable.]

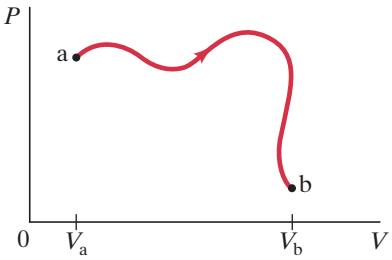


FIGURE 20–19
Problem 48.

49. (II) Calculate the change in entropy of 1.00 kg of water when it is heated from 0°C to 85°C. (*a*) Make an estimate; (*b*) use the integral $\Delta S = \int dQ/T$. (*c*) Does the entropy of the surroundings change? If so, by how much?

50. (II) A 110-g insulated aluminum cup at 35°C is filled with 150 g of water at 45°C. After a few minutes, equilibrium is reached. Determine (*a*) the final temperature, and (*b*) the total change in entropy as a result of the mixing process (use $\Delta S = \int dQ/T$).

51. (II) Two samples of an ideal gas are initially at the same temperature and pressure. They are each compressed reversibly from a volume V to volume $V/2$, one isothermally, the other adiabatically. (*a*) In which sample is the final pressure greater? (*b*) Determine the change in entropy of the gas for each process by integration. (*c*) What is the entropy change of the environment for each process? (*d*) Entropy is a state variable. Why then don't the two processes give the same ΔS?

52. (II) 1.00 mole of nitrogen (N_2) gas and 1.00 mole of argon (Ar) gas are in separate, equal-sized, insulated containers at the same temperature. The containers are then connected and the gases (assumed ideal) allowed to mix. What is the change in entropy (*a*) of the system and (*b*) of the environment? (*c*) Repeat part (*a*) but assume one container is twice as large as the other.

53. (II) (*a*) Why would you expect the total entropy change in a Carnot cycle to be zero? (*b*) Do a calculation to show that it is zero.

54. (II) Thermodynamic processes are sometimes represented on *TS* (temperature–entropy) diagrams, rather than *PV* diagrams. Determine the slope of a constant-volume process on a *TS* diagram when a system with n moles of an ideal gas with constant-volume molar specific heat C_V is at temperature T.

55. (III) The specific heat per mole of potassium at low temperatures is given by $C_V = aT + bT^3$, where $a = 2.08\,\text{mJ/mol}\cdot\text{K}^2$ and $b = 2.57\,\text{mJ/mol}\cdot\text{K}^4$. Determine (by integration) the entropy change of 0.15 mol of potassium when its temperature is lowered from 5.0 K to 1.0 K.

56. (III) Consider an ideal gas of n moles with molar specific heats C_V and C_P. (*a*) Starting with the first law, show that when the temperature and volume of this gas are changed by a reversible process, its change in entropy is given by

$$dS = nC_V\frac{dT}{T} + nR\frac{dV}{V}.$$

(*b*) Show that the expression in part (*a*) can be written as

$$dS = nC_V\frac{dP}{P} + nC_P\frac{dV}{V}.$$

(*c*) Using the expression from part (*b*), show that if $dS = 0$ for the reversible process (that is, the process is adiabatic), then $PV^\gamma = \text{constant}$, where $\gamma = C_P/C_V$.

20–8 Unavailability of Energy

57. (III) A general theorem states that the amount of energy that becomes unavailable to do useful work in any process is equal to $T_L\,\Delta S$, where T_L is the lowest temperature available and ΔS is the total change in entropy during the process. Show that this is valid in the specific cases of: (*a*) a falling rock that comes to rest when it hits the ground; (*b*) the free adiabatic expansion of an ideal gas; and (*c*) the conduction of heat, Q, from a high-temperature (T_H) reservoir to a low-temperature (T_L) reservoir. [*Hint*: In part (*c*) compare to a Carnot engine.]

58. (III) Determine the work available in a 2.2-kg block of copper at 490 K if the surroundings are at 290 K. Use results of Problem 57.

20–9 Statistical Interpretation of Entropy

59. (I) Use Eq. 20–14 to determine the entropy of each of the five macrostates listed in Table 20–1 on page 595.

60. (II) Suppose that you repeatedly shake six coins in your hand and drop them on the floor. Construct a table showing the number of microstates that correspond to each macrostate. What is the probability of obtaining (*a*) three heads and three tails and (*b*) six heads?

61. (II) (*a*) Suppose you have four coins, all with tails up. You now rearrange them so two heads and two tails are up. What was the change in entropy of the coins? (*b*) Suppose your system is the 100 coins of Table 20–2; what is the change in entropy of the coins if they are mixed randomly initially, 50 heads and 50 tails, and you arrange them so all 100 are heads? (*c*) Compare these entropy changes to ordinary thermodynamic entropy changes, such as Examples 20–7, 20–8, 20–9, and 20–10.

62. (II) Calculate the probabilities, when you throw two dice, of obtaining (*a*) a 7, and (*b*) an 11.

63. (III) Consider an isolated gas-like system consisting of a box that contains $N = 10$ distinguishable atoms, each moving at the same speed v. The number of unique ways that these atoms can be arranged so that N_L atoms are within the left-hand half of the box and N_R atoms are within the right-hand half of the box is given by

$$\frac{N!}{N_L!\,N_R!},$$

where, for example, the factorial $4! = 4\cdot3\cdot2\cdot1$ (the only exception is that $0! = 1$). Define each unique arrangement of atoms within the box to be a microstate of this system. Now imagine the following two possible macrostates: state A where all of the atoms are within the left-hand half of the box and none are within the right-hand half; and state B where the distribution is uniform (that is, there is the same number in each half). See Fig. 20–20. (*a*) Assume the system is initially in state A and, at a later time, is found to be in state B. Determine the system's change in entropy. Can this process occur naturally? (*b*) Assume the system is initially in state B and, at a later time, is found to be in state A. Determine the system's change in entropy. Can this process occur naturally?

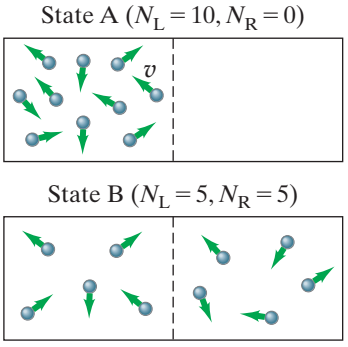


FIGURE 20–20 Problem 63.

20–11 Energy Resources

64. (I) *Photovoltaic cells* (Fig. 20–21) can produce about 170 W of electricity per square meter of surface area if directly facing the Sun. How large an area is required to supply the needs of a house that requires 28 kWh/day? Would this fit on the roof of an average house? (Assume the Sun shines about 9 h/day.)



FIGURE 20–21 Problem 64.

65. (II) Energy may be stored by pumping water to a high reservoir when demand is low and then releasing it to drive turbines (Fig. 20–15) during peak demand. Suppose water is pumped to a lake 105 m above the turbines at a rate of $1.00\times10^5\,\text{kg/s}$ for 10.0 h at night. (*a*) How much energy (kWh) is needed to do this each night? (*b*) If all this energy is released during a 14-h day, at 75% efficiency, what is the average power output?

66. (II) Water is stored in an artificial lake created by a dam (Fig. 20–22). The water depth is 48 m at the dam, and a steady flow rate of 35 m^3/s is maintained through hydroelectric turbines installed near the base of the dam. (*a*) How much electrical power can be produced? (*b*) What are the environmental implications of such a dam?



FIGURE 20–22 Problem 66. Flaming Gorge Dam on the Green River in Utah.

General Problems

67. An inventor claims to have built an engine that produces 2.00 MW of usable work while taking in 3.00 MW of thermal energy at 425 K, and rejecting 1.00 MW of thermal energy at 215 K. Is there anything fishy about his claim? Explain.

68. It has been suggested that a heat engine could be developed that made use of the temperature difference between water at the surface of the ocean and water several hundred meters deep. In the tropics, the temperatures may be 27°C and 4°C, respectively. (*a*) What is the maximum efficiency such an engine could have? (*b*) Why might such an engine be feasible in spite of the low efficiency? (*c*) Can you imagine any adverse environmental effects that might occur?

69. A heat engine takes a diatomic gas around the cycle shown in Fig. 20–23. (*a*) Using the ideal gas law, determine how many moles of gas are in this engine. (*b*) Determine the temperature at point c. (*c*) Calculate the heat input into the gas during the constant volume process from points b to c. (*d*) Calculate the work done on the gas during the isothermal compression from points a to b. (*e*) Calculate the work done by the gas during the adiabatic process from points c to a. (*f*) Determine the engine's efficiency. (*g*) What is the maximum efficiency possible for an engine working between T_a and T_c?

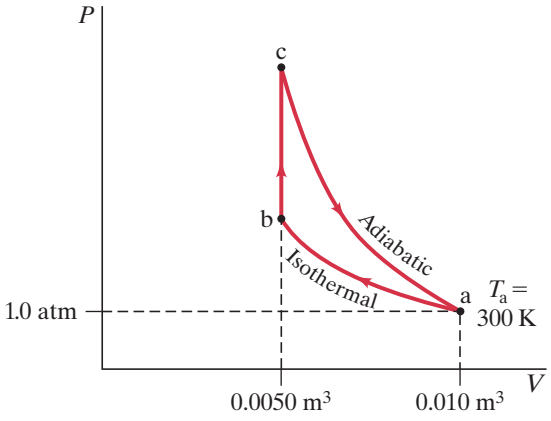


FIGURE 20–23 Problem 69.

70. Refrigeration units can be rated in "tons." A 1-ton air conditioning system can remove sufficient energy to freeze 1 ton (2000 pounds = 909 kg) of 0°C water into 0°C ice in one 24-h day. Assume the hot part of a day averages 35°C and the interior of a house is maintained at 22°C by the continuous operation of a 6-ton air conditioning system for 6 hours a day. How much does this cooling cost the homeowner per day, and per month? Assume the work done by the refrigeration unit is powered by electricity that costs \$0.13 per kWh and that the unit's coefficient of performance is only 18% of an ideal refrigerator. 1 kWh = 3.60×10^6 J.

71. (*a*) At a steam power plant, *steam engines* work in pairs, the heat output of the first one being the approximate heat input of the second. The operating temperatures of the first are 750°C and 440°C, and of the second 415°C and 240°C. If the heat of combustion of coal is 2.8×10^7 J/kg, at what rate must coal be burned if the plant is to put out 950 MW of power? Assume the efficiency of the engines is 65% of the ideal (Carnot) efficiency. (*b*) Water is used to cool the power plant. If the water temperature is allowed to increase by no more than 4.5 C°, estimate how much water must pass through the plant per hour.

72. A 38% efficient power plant puts out 850 MW of electrical power. Cooling towers take away the exhaust heat. (*a*) If the air temperature is allowed to rise 7.0 C°, estimate what volume of air (km^3) is heated per day. Will the local climate be heated significantly? (*b*) If the heated air were to form a layer 150 m thick, estimate how large an area it would cover for 24 h of operation. Assume the air has density 1.3 kg/m^3 and has specific heat of about 1.0 kJ/kg · C° at constant pressure.

73. The operation of a certain heat engine takes an ideal monatomic gas through a cycle shown as the rectangle on the *PV* diagram of Fig. 20–24. (*a*) Determine the efficiency of this engine. Let Q_H and Q_L be the total heat input and total heat exhausted during one cycle of this engine. (*b*) Compare (as a ratio) the efficiency of this engine to that of a Carnot engine operating between T_H and T_L, where T_H and T_L are the highest and lowest temperatures achieved.

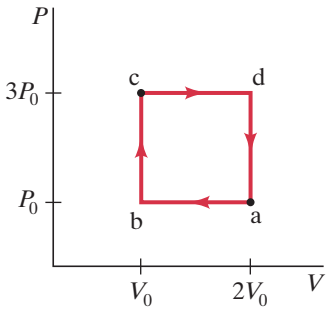


FIGURE 20–24 Problem 73.

74. A "Carnot" refrigerator (the reverse of a Carnot engine) absorbs heat from the freezer compartment at a temperature of −17°C and exhausts it into the room at 25°C. (*a*) How much work would the refrigerator do to change 0.65 kg of water at 25°C into ice at −17°C? (*b*) If the compressor output is 105 W and runs 25% of the time, how long will this take?

75. A car engine whose output power is 145 hp operates at about 15% efficiency. Assume the engine's water temperature of 85°C is its cold-temperature (exhaust) reservoir and 495°C is its thermal "intake" temperature (the temperature of the exploding gas–air mixture). (*a*) What is the ratio of its efficiency relative to its maximum possible (Carnot) efficiency? (*b*) Estimate how much power (in watts) goes into moving the car, and how much heat, in joules and in kcal, is exhausted to the air in 1.0 h.

76. 1.00 mole of an ideal monatomic gas at STP first undergoes an isothermal expansion so that the volume at b is 2.5 times the volume at a (Fig. 20–25). Next, heat is extracted at a constant volume so that the pressure drops. The gas is then compressed adiabatically back to the original state. (*a*) Calculate the pressures at b and c. (*b*) Determine the temperature at c. (*c*) Determine the work done, heat input or extracted, and the change in entropy for each process. (*d*) What is the efficiency of this cycle?

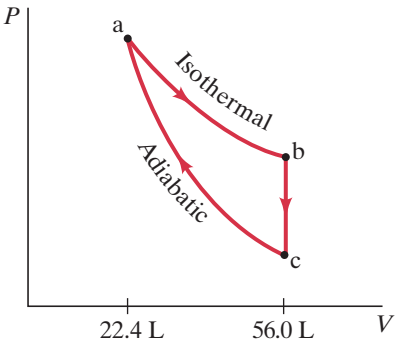


FIGURE 20–25
Problem 76.

77. Suppose a power plant delivers energy at 880 MW using steam turbines. The steam goes into the turbines superheated at 625 K and deposits its unused heat in river water at 285 K. Assume that the turbine operates as an ideal Carnot engine. (*a*) If the river flow rate is 37 m^3/s, estimate the average temperature increase of the river water immediately downstream from the power plant. (*b*) What is the entropy increase per kilogram of the downstream river water in $J/kg \cdot K$?

78. An ideal air conditioner keeps the temperature inside a room at 21°C when the outside temperature is 32°C. If 4.8 kW of power enters a room through the windows in the form of direct radiation from the Sun, how much electrical power would be saved if the windows were shaded so only 500 W enters?

79. The operation of a **diesel engine** can be idealized by the cycle shown in Fig. 20–26. Air is drawn into the cylinder during the intake stroke (not part of the idealized cycle). The air is compressed adiabatically, path ab. At point b diesel fuel is injected into the cylinder and immediately burns since the temperature is very high. Combustion is slow, and during the first part of the power stroke, the gas expands at (nearly) constant pressure, path bc. After burning, the rest of the power stroke is adiabatic, path cd. Path da corresponds to the exhaust stroke. (*a*) Show that, for a quasistatic reversible engine undergoing this cycle using an ideal gas, the ideal efficiency is

$$e = 1 - \frac{(V_a/V_c)^{-\gamma} - (V_a/V_b)^{-\gamma}}{\gamma[(V_a/V_c)^{-1} - (V_a/V_b)^{-1}]},$$

where V_a/V_b is the "compression ratio," V_a/V_c is the "expansion ratio," and γ is defined by Eq. 19–15 $\left(\gamma = \frac{C_P}{C_V}\right)$. (*b*) If $V_a/V_b = 16$ and $V_a/V_c = 4.5$, calculate the efficiency assuming the gas is diatomic (like N_2 and O_2) and ideal.

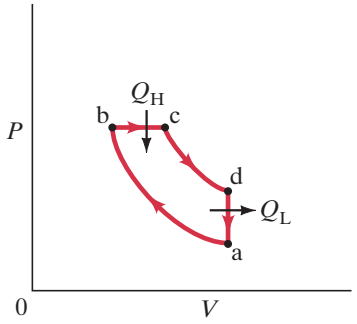


FIGURE 20–26
Problem 79.

80. Two 1100-kg cars are traveling 75 km/h in opposite directions when they collide and are brought to rest. Estimate the change in entropy of the universe as a result of this collision. Assume $T = 20°C$.

81. A cooling unit for a new freezer has an inner surface area of 8.0 m^2, and is bounded by walls 12 cm thick with a thermal conductivity of 0.050 $W/m \cdot K$. The inside must be kept at −15°C in a room that is at 22°C. The motor for the cooling unit must run no more than 18% of the time. What is the minimum power requirement of the cooling motor?

82. Suppose a heat pump has a stationary bicycle attachment that allows *you* to provide the work instead of using an electrical wall outlet. If your heat pump has a coefficient of performance of 2.0 and you can cycle at a racing pace output of about 200 W for a half hour, how much heat can you provide?

83. The **Stirling cycle**, shown in Fig. 20–27, is useful to describe some heat engines. Find the efficiency of the cycle in terms of the parameters shown, assuming an ideal monatomic gas as the working substance. The processes ab and cd are isothermal whereas bc and da are at constant volume. How does it compare to the Carnot efficiency?

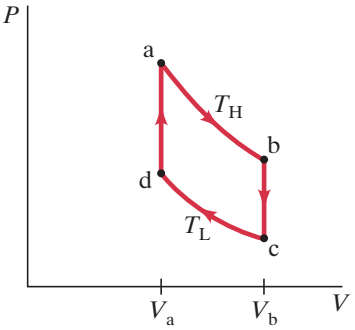


FIGURE 20–27
Problem 83.
Stirling cycle.

84. The **Brayton cycle**, depicted in the *PV* diagram of Fig. 20–28, can describe a *jet engine* gas turbine. In process ab the air–fuel mixture undergoes an adiabatic compression. This is followed, in process bc, with an isobaric (constant pressure) heating, by combustion. Process cd is an adiabatic expansion with expulsion of the products to the atmosphere. The return step, da, takes place at constant pressure. If the working gas behaves like an ideal gas, show that the efficiency of the Brayton cycle is

$$e = 1 - \left(\frac{P_b}{P_a}\right)^{\frac{1-\gamma}{\gamma}}.$$

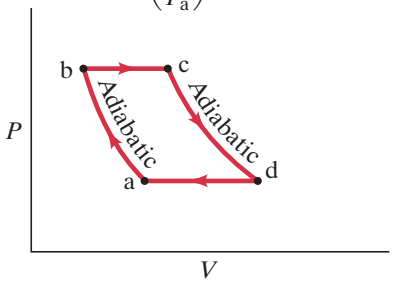


FIGURE 20–28 Problem 84. Brayton cycle.

85. A restaurant refrigerator has a coefficient of performance of 5.0. If the temperature in the kitchen outside the refrigerator is 32°C, what is the lowest temperature that could be obtained inside the refrigerator if it were ideal?

86. Thermodynamic processes can be represented not only on *PV* and *PT* diagrams, but also on a *TS* (temperature–entropy) diagram. (*a*) Draw a *TS diagram* for a Carnot cycle. (*b*) What does the area within the curve represent?

87. An aluminum can, with negligible heat capacity, is filled with 450 g of water at 0°C and then is brought into thermal contact with a similar can filled with 450 g of water at 50°C. Find the change in entropy of the system if no heat is allowed to exchange with the surroundings. Use $\Delta S = \int dQ/T$.

88. A **dehumidifier** removes water vapor from air and has been referred to as a "refrigerator with an open door." The humid air is pulled in by a fan and passes over a cold coil, whose temperature is less than the dew point, and some of the air's water condenses. After this water is extracted, the air is warmed back to its original temperature and sent into the room. In a well-designed dehumidifier, the heat that is removed by the cooling coil mostly comes from the condensation of water vapor to liquid, and this heat is used to re-warm the air. Estimate how much water is removed in 1.0 h by an ideal dehumidifier, if the temperature of the room is 25°C, the water condenses at 8°C, and the dehumidifier does work at the rate of 550 W of electrical power.

89. A bowl contains many red, orange, and green jelly beans, in equal numbers. You are to make a line of 3 jelly beans by randomly taking 3 beans from the bowl. (*a*) Construct a table showing the number of microstates that correspond to each macrostate. Then determine the probability of (*b*) all 3 beans red, and (*c*) 2 greens, 1 orange.

90. Rank the following five-card hands in order of increasing probability: (*a*) four aces and a king; (*b*) six of hearts, eight of diamonds, queen of clubs, three of hearts, jack of spades; (*c*) two jacks, two queens, and an ace; and (*d*) any hand having no two equal-value cards (no pairs, etc.). Discuss your ranking in terms of microstates and macrostates.

91. A particular 1.5-m^2 photovoltaic panel operating in direct sunlight produces electricity at 22% efficiency. The resulting electricity is used to operate an electric stove that can be used to heat water. A second system uses a 1.5-m^2 curved mirror to concentrate the Sun's energy directly onto a container of water. Estimate how long it takes each system to heat 1.0 kg of water from 25°C to 95°C. (See also Chapter 19.)

92. At low temperature the specific heat of diamond varies with absolute temperature T according to the Debye equation $C_V = 1.88 \times 10^3 \left(T/T_D\right)^3 \text{J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ where the Debye temperature for diamond is $T_D = 2230$ K. Determine the entropy change of 1.00 mol of diamond when it is heated at constant volume from 4 K to 40 K. [*Hint*: $dS = nC_V\, dT/T$, where n is the number of moles.]

93. **Trees offsetting CO_2.** Trees can help a bit to offset the buildup of CO_2 due to burning coal and other fossil fuels. CO_2 can be absorbed by tree foliage. Trees use the carbon to grow, and release O_2 into the atmosphere. Suppose a refrigerator uses 600 kWh of electricity per year (about 2×10^9 J) from a 33% efficient coal-fired power plant. Burning 1 kg of coal releases about 2×10^7 J of energy. Assume coal is all carbon, which when burned in air becomes CO_2. (*a*) How much coal is burned per year to run this refrigerator? (*b*) Assuming a forest can capture 1700 kg of carbon per hectare ($= 10{,}000\ m^2$) per year, estimate how many square meters of forest are needed to capture the carbon (in the form now of CO_2) emitted by the refrigerator in (*a*).

ANSWERS TO EXERCISES

A: (*d*).
B: (*c*).
C: (*a*) Heat Q_L comes from outside to conserve energy; (*b*) 560 W.
D: (*e*) 11/16.



This comb has acquired a static electric charge, either from passing through hair, or being rubbed by a cloth or paper towel. The electrical charge on the comb induces a polarization (separation of charge) in scraps of paper, and thus attracts them.

Our introduction to electricity in this Chapter covers conductors and insulators, and Coulomb's law which relates the force between two point charges as a function of their distance apart. We also introduce the powerful concept of electric field.

CHAPTER 21

Electric Charge and Electric Field

CHAPTER-OPENING QUESTION—Guess now!

Two identical tiny spheres have the same electric charge. If their separation is doubled, the force each exerts on the other will be

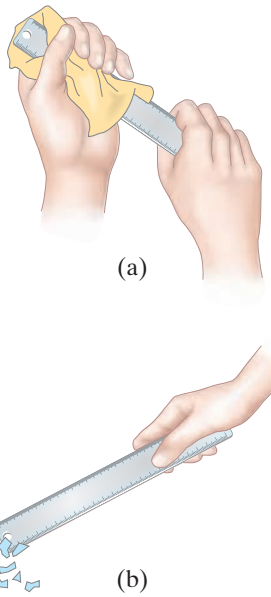
(a) half.
(b) double.
(c) four times larger.
(d) one-quarter as large.
(e) unchanged.

The word "electricity" may evoke an image of complex modern technology: lights, motors, electronics, and computers. But the electric force plays an even deeper role in our lives. According to atomic theory, electric forces between atoms and molecules hold them together to form liquids and solids, and electric forces are also involved in the metabolic processes that occur within our bodies. Many of the forces we have dealt with so far, such as elastic forces, the normal force, and friction and other contact forces (pushes and pulls), are now considered to result from electric forces acting at the atomic level. Gravity, on the other hand, is a separate force.[†]

[†]As we discussed in Section 6–7, physicists in the twentieth century came to recognize four different fundamental forces in nature: (1) gravitational force, (2) electromagnetic force (we will see later that electric and magnetic forces are intimately related), (3) strong nuclear force, and (4) weak nuclear force. The last two forces operate at the level of the nucleus of an atom. Recent theory has combined the electromagnetic and weak nuclear forces so they are now considered to have a common origin known as the electroweak force. We discuss the other forces in later Chapters.

CONTENTS

21 – 1 Static Electricity; Electric Charge and Its Conservation
21 – 2 Electric Charge in the Atom
21 – 3 Insulators and Conductors
21 – 4 Induced Charge; the Electroscope
21 – 5 Coulomb's Law
21 – 6 The Electric Field
21 – 7 Electric Field Calculations for Continuous Charge Distributions
21 – 8 Field Lines
21 – 9 Electric Fields and Conductors
21–10 Motion of a Charged Particle in an Electric Field
21–11 Electric Dipoles
*21–12 Electric Forces in Molecular Biology: DNA Structure and Replication

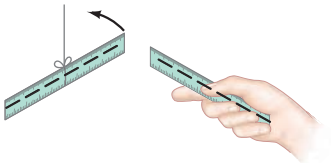


(a)

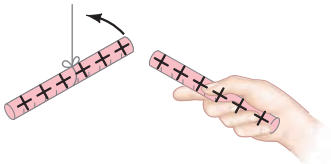
(b)

FIGURE 21–1 (a) Rub a plastic ruler with a cloth or paper towel, and (b) bring it close to some tiny pieces of paper.

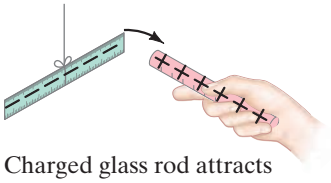
FIGURE 21–2 Like charges repel one another; unlike charges attract. (Note color coding: we color positive charged objects pink or red, and negative charges blue-green. We use these colors especially for point charges, and most of the time for real objects.)



(a) Two charged plastic rulers repel



(b) Two charged glass rods repel



(c) Charged glass rod attracts charged plastic ruler

The earliest studies on electricity date back to the ancients, but only since the late 1700s has electricity been studied in detail. We will discuss the development of ideas about electricity, including practical devices, as well as its relation to magnetism, in the next eleven Chapters.

21–1 Static Electricity; Electric Charge and Its Conservation

The word *electricity* comes from the Greek word *elektron*, which means "amber." Amber is petrified tree resin, and the ancients knew that if you rub a piece of amber with a cloth, the amber attracts small pieces of leaves or dust. A piece of hard rubber, a glass rod, or a plastic ruler rubbed with a cloth will also display this "amber effect," or **static electricity** as we call it today. You can readily pick up small pieces of paper with a plastic comb or ruler that you have just vigorously rubbed with even a paper towel. See the photo on the previous page and Fig. 21–1. You have probably experienced static electricity when combing your hair or when taking a synthetic blouse or shirt from a clothes dryer. And you may have felt a shock when you touched a metal doorknob after sliding across a car seat or walking across a synthetic carpet. In each case, an object becomes "charged" as a result of rubbing, and is said to possess a net **electric charge**.

Is all electric charge the same, or is there more than one type? In fact, there are *two* types of electric charge, as the following simple experiments show. A plastic ruler suspended by a thread is vigorously rubbed with a cloth to charge it. When a second plastic ruler, which has been charged in the same way, is brought close to the first, it is found that one ruler *repels* the other. This is shown in Fig. 21–2a. Similarly, if a rubbed glass rod is brought close to a second charged glass rod, again a repulsive force is seen to act, Fig. 21–2b. However, if the charged glass rod is brought close to the charged plastic ruler, it is found that they *attract* each other, Fig. 21–2c. The charge on the glass must therefore be different from that on the plastic. Indeed, it is found experimentally that all charged objects fall into one of two categories. Either they are attracted to the plastic and repelled by the glass; or they are repelled by the plastic and attracted to the glass. Thus there seem to be two, and only two, types of electric charge. Each type of charge repels the same type but attracts the opposite type. That is: **unlike charges attract; like charges repel**.

The two types of electric charge were referred to as *positive* and *negative* by the American statesman, philosopher, and scientist Benjamin Franklin (1706–1790). The choice of which name went with which type of charge was arbitrary. Franklin's choice set the charge on the rubbed glass rod to be positive charge, so the charge on a rubbed plastic ruler (or amber) is called negative charge. We still follow this convention today.

Franklin argued that whenever a certain amount of charge is produced on one object, an equal amount of the opposite type of charge is produced on another object. The positive and negative are to be treated *algebraically*, so during any process, the net change in the amount of charge produced is zero. For example, when a plastic ruler is rubbed with a paper towel, the plastic acquires a negative charge and the towel acquires an equal amount of positive charge. The charges are separated, but the sum of the two is zero.

This is an example of a law that is now well established: the **law of conservation of electric charge**, which states that

LAW OF CONSERVATION OF ELECTRIC CHARGE

the net amount of electric charge produced in any process is zero;

or, said another way,

no net electric charge can be created or destroyed.

If one object (or a region of space) acquires a positive charge, then an equal amount of negative charge will be found in neighboring areas or objects. No violations have ever been found, and the law of conservation of electric charge is as firmly established as those for energy and momentum.

21–2 Electric Charge in the Atom

Only within the past century has it become clear that an explanation of electricity originates inside the atom itself. In later Chapters we will discuss atomic structure and the ideas that led to our present view of the atom in more detail. But it will help our understanding of electricity if we discuss it briefly now.

A simplified model of an atom shows it as having a tiny but massive nucleus, positively charged, surrounded by one or more negatively charged electrons (Fig. 21–3). The nucleus contains protons, which are positively charged, and neutrons, which have no net electric charge. All protons and all electrons have exactly the same magnitude of electric charge; but their signs are opposite. Hence neutral atoms, having no net charge, contain equal numbers of protons and electrons. Sometimes an atom may lose one or more of its electrons, or may gain one or more electrons, in which case it will have a net positive or negative charge and is called an **ion**.

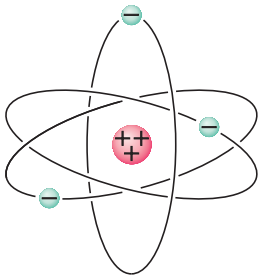
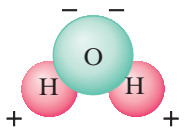


FIGURE 21–3 Simple model of the atom.

In solid materials the nuclei tend to remain close to fixed positions, whereas some of the electrons may move quite freely. When an object is *neutral*, it contains equal amounts of positive and negative charge. The charging of a solid object by rubbing can be explained by the transfer of electrons from one object to the other. When a plastic ruler becomes negatively charged by rubbing with a paper towel, electrons are transfered from the towel to the plastic, leaving the towel with a positive charge equal in magnitude to the negative charge acquired by the plastic. In liquids and gases, nuclei or ions can move as well as electrons.

Normally when objects are charged by rubbing, they hold their charge only for a limited time and eventually return to the neutral state. Where does the charge go? Usually the excess charge "leaks off" onto water molecules in the air. This is because water molecules are **polar**—that is, even though they are neutral, their charge is not distributed uniformly, Fig. 21–4. Thus the extra electrons on, say, a charged plastic ruler can "leak off" into the air because they are attracted to the positive end of water molecules. A positively charged object, on the other hand, can be neutralized by transfer of loosely held electrons from water molecules in the air. On dry days, static electricity is much more noticeable because the air contains fewer water molecules to allow leakage of charge. On humid or rainy days, it is difficult to make any object hold a net charge for long.

FIGURE 21–4 Diagram of a water molecule. Because it has opposite charges on different ends, it is called a "polar" molecule.



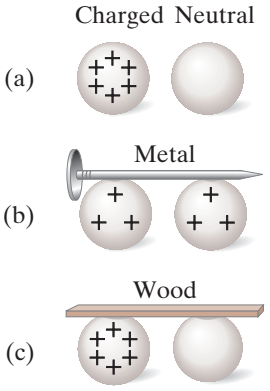
21–3 Insulators and Conductors

Suppose we have two metal spheres, one highly charged and the other electrically neutral (Fig. 21–5a). If we now place a metal object, such as a nail, so that it touches both spheres (Fig. 21–5b), the previously uncharged sphere quickly becomes charged. If, instead, we had connected the two spheres by a wooden rod or a piece of rubber (Fig. 21–5c), the uncharged ball would not become noticeably charged. Materials like the iron nail are said to be **conductors** of electricity, whereas wood and rubber are **nonconductors** or **insulators**.

Metals are generally good conductors, whereas most other materials are insulators (although even insulators conduct electricity very slightly). Nearly all natural materials fall into one or the other of these two distinct categories. However, a few materials (notably silicon and germanium) fall into an intermediate category known as **semiconductors**.

From the atomic point of view, the electrons in an insulating material are bound very tightly to the nuclei. In a good metal conductor, on the other hand, some of the electrons are bound very loosely and can move about freely within the metal (although they cannot *leave* the metal easily) and are often referred to as *free electrons* or *conduction electrons*. When a positively charged object is brought close to or touches a conductor, the free electrons in the conductor are attracted by this positively charged object and move quickly toward it. If a negatively charged object is brought close to the conductor, the free electrons in the conductor move swiftly away from it. In a semiconductor, there are many fewer free electrons, and in an insulator, almost none.

FIGURE 21–5 (a) A charged metal sphere and a neutral metal sphere. (b) The two spheres connected by a conductor (a metal nail), which conducts charge from one sphere to the other. (c) The original two spheres connected by an insulator (wood); almost no charge is conducted.



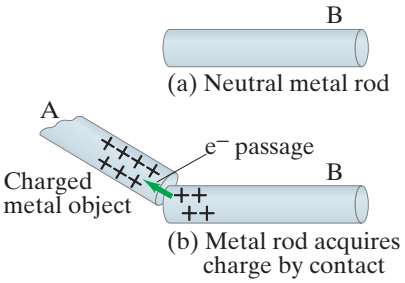


FIGURE 21–6 A neutral metal rod in (a) will acquire a positive charge if placed in contact (b) with a positively charged metal object. (Electrons move as shown by the green arrow.) This is called charging by conduction.

21–4 Induced Charge; the Electroscope

Suppose a positively charged metal object A is brought close to an uncharged metal object B. If the two touch, the free electrons in the neutral one are attracted to the positively charged object and some of those electrons will pass over to it, Fig. 21–6. Since object B, originally neutral, is now missing some of its negative electrons, it will have a net positive charge. This process is called **charging by conduction**, or "by contact," and the two objects end up with the same sign of charge.

Now suppose a positively charged object is brought close to a neutral metal rod, but does not touch it. Although the free electrons of the metal rod do not leave the rod, they still move within the metal toward the external positive charge, leaving a positive charge at the opposite end of the rod (Fig. 21–7b). A charge is said to have been *induced* at the two ends of the metal rod. No net charge has been created in the rod: charges have merely been *separated*. The net charge on the metal rod is still zero. However, if the metal is cut into two pieces, we would have two charged objects: one charged positively and one charged negatively. This is **charging by induction**.

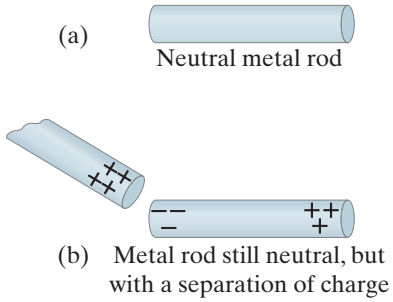


FIGURE 21–7 Charging by induction: if the rod in (b) is cut into two parts, each part will have a net charge.

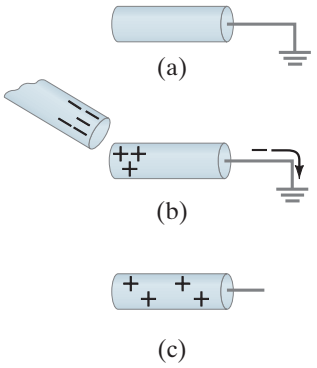
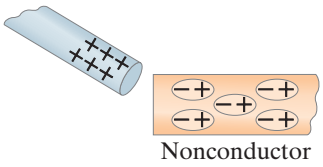


FIGURE 21–8 Inducing a charge on an object connected to ground.

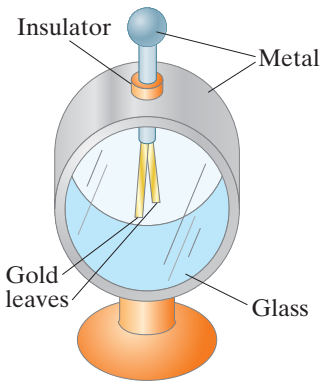
FIGURE 21–9 A charged object brought near a nonconductor causes a charge separation within the nonconductor's molecules.



Another way to induce a net charge on a metal object is to first connect it with a conducting wire to the ground (or a conducting pipe leading into the ground) as shown in Fig. 21–8a (the symbol ⏚ means connected to "ground"). The object is then said to be **grounded** or "earthed." The Earth, because it is so large and can conduct, easily accepts or gives up electrons; hence it acts like a reservoir for charge. If a charged object—say negative this time—is brought up close to the metal object, free electrons in the metal are repelled and many of them move down the wire into the Earth, Fig. 21–8b. This leaves the metal positively charged. If the wire is now cut, the metal object will have a positive induced charge on it (Fig. 21–8c). If the wire is cut *after* the negative object is moved away, the electrons would all have moved from the ground back into the metal object and it would be neutral again.

Charge separation can also be done in nonconductors. If you bring a positively charged object close to a neutral nonconductor as shown in Fig. 21–9, almost no electrons can move about freely within the nonconductor. But they can move slightly within their own atoms or molecules. Each oval in Fig. 21–9 represents a molecule (not to scale); the negatively charged electrons, attracted to the external positive charge, tend to move in its direction within their molecules. Because the negative charges in the nonconductor are nearer to the external positive charge, the nonconductor as a whole is attracted to the external positive charge (see the Chapter-Opening Photo, page 609).

FIGURE 21–10 (below) Electroscope.



An **electroscope** is a device that can be used for detecting charge. As shown in Fig. 21–10, inside a case are two movable metal leaves, often made of gold foil, attached to a metal rod with a knob on the outside. (Sometimes only one leaf is movable.)

If a positively charged object is brought close to the knob, a separation of charge is induced: electrons are attracted up into the knob, and the leaves become positively charged, Fig. 21–11a. The two leaves repel each other as shown, because they are both positively charged. If, instead, the knob is charged by conduction (touching), the whole apparatus acquires a net charge as shown in Fig. 21–11b. In either case, the greater the amount of charge, the greater the separation of the leaves.

Note that you cannot tell the sign of the charge in this way, since negative charge will cause the leaves to separate just as much as an equal amount of positive charge; in either case, the two leaves repel each other. An electroscope can, however, be used to determine the sign of the charge if it is first charged by conduction: say, negatively, as in Fig. 21–12a. Now if a negative object is brought close, as in Fig. 21–12b, more electrons are repelled and "induced" to move down into the leaves, which separate further. If a positive charge is brought close instead, the electrons are induced to flow upward, so the leaves are less negative and their separation is reduced, Fig. 21–12c.

The electroscope was used in the early studies of electricity. The same principle, aided by some electronics, is used in much more sensitive modern **electrometers**.

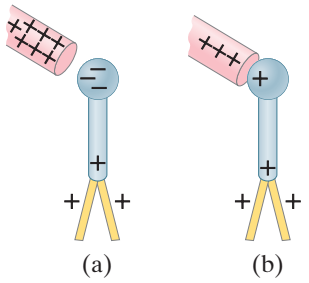
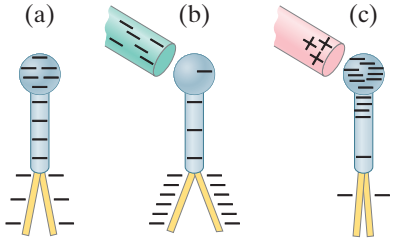


FIGURE 21–11 Electroscope charged (a) by induction, (b) by conduction.

FIGURE 21–12 A previously charged electroscope can be used to determine the sign of a charged object.



21–5 Coulomb's Law

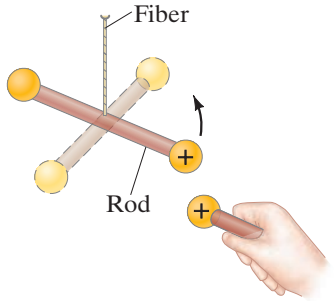
We have seen that an electric charge exerts a force of attraction or repulsion on other electric charges. What factors affect the magnitude of this force? To find an answer, the French physicist Charles Coulomb (1736–1806) investigated electric forces in the 1780s using a torsion balance (Fig. 21–13) much like that used by Cavendish for his studies of the gravitational force (Chapter 6).

Precise instruments for the measurement of electric charge were not available in Coulomb's time. Nonetheless, Coulomb was able to prepare small spheres with different magnitudes of charge in which the *ratio* of the charges was known.[†] Although he had some difficulty with induced charges, Coulomb was able to argue that the electric force one tiny charged object exerts on a second tiny charged object is directly proportional to the charge on each of them. That is, if the charge on either one of the objects is doubled, the force is doubled; and if the charge on both of the objects is doubled, the force increases to four times the original value. This was the case when the distance between the two charges remained the same. If the distance between the two charges was allowed to increase, he found that the force decreased with the *square of the distance* between them. That is, if the distance was doubled, the force fell to one-fourth of its original value. Thus, Coulomb concluded, the magnitude of the force F that one small charged object exerts on a second one is proportional to the product of the magnitude of the charge on one, Q_1, times the magnitude of the charge on the other, Q_2, and inversely proportional to the square of the distance r between them (Fig. 21–14). As an equation, we can write **Coulomb's law** as

$$F = k\frac{Q_1 Q_2}{r^2}, \qquad \text{[magnitudes]} \quad \textbf{(21–1)}$$

where k is a proportionality constant.[‡] Equation 21–1 gives the *magnitude* of the electric force that either charge exerts on the other.

FIGURE 21–13 Coulomb's apparatus: when an external charged sphere is placed close to the charged one on the suspended bar, the bar rotates slightly. The suspending fiber resists the twisting motion, and the angle of twist is proportional to the force applied. With this apparatus, Coulomb investigated how the electric force varies as a function of the magnitude of the charges and of the distance between them.



COULOMB'S LAW

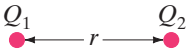


FIGURE 21–14 Coulomb's law, Eq. 21–1, gives the force between two point charges, Q_1 and Q_2, a distance r apart.

[†]Coulomb reasoned that if a charged conducting sphere is placed in contact with an identical uncharged sphere, the charge on the first would be shared equally by the two of them because of symmetry. He thus had a way to produce charges equal to $\frac{1}{2}$, $\frac{1}{4}$, and so on, of the original charge.

[‡]The validity of Coulomb's law today rests on precision measurements that are much more sophisticated than Coulomb's original experiment. The exponent 2 on r in Coulomb's law has been shown to be accurate to 1 part in 10^{16} [that is, $2 \pm (1 \times 10^{-16})$].

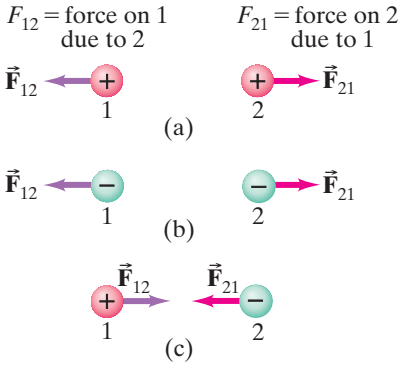


FIGURE 21–15 The direction of the static electric force one point charge exerts on another is always along the line joining the two charges, and depends on whether the charges have the same sign as in (a) and (b), or opposite signs (c).

The *direction* of the electric force *is always along the line joining the two charges*. If the two charges have the same sign, the force on either charge is directed away from the other (they repel each other). If the two charges have opposite signs, the force on one is directed toward the other (they attract). See Fig. 21–15. Notice that the force one charge exerts on the second is equal but opposite to that exerted by the second on the first, in accord with Newton's third law.

The SI unit of charge is the **coulomb** (C), and its standard definition is in terms of electric current and magnetic field, as discussed later (Section 28–3). In SI units, the constant k in Coulomb's law has the value[†]

$$k = 8.988 \times 10^9\,\mathrm{N \cdot m^2/C^2}$$

or, when we only need two significant figures,

$$k \approx 9.0 \times 10^9\,\mathrm{N \cdot m^2/C^2}.$$

We rarely encounter charges as large as a coulomb. Charges produced by rubbing ordinary objects (such as a comb or plastic ruler) are typically around a microcoulomb $(1\,\mu\mathrm{C} = 10^{-6}\,\mathrm{C})$ or less. Everyday objects that have a positive charge have a deficit of electrons, whereas negatively charged objects have an excess of electrons. The charge on one electron has a magnitude of about $1.6022 \times 10^{-19}\,\mathrm{C}$, and is negative. This is the smallest charge observed in nature, and because it is fundamental, it is given the symbol e and is often referred to as the **elementary charge**:[‡]

$$e = 1.6022 \times 10^{-19}\,\mathrm{C} \approx 1.6 \times 10^{-19}\,\mathrm{C}.$$

Note that e is defined as a positive number, so the charge on the electron is $-e$. Since an object cannot gain or lose a fraction of an electron, the net charge on any object must be an integral multiple of this charge. Electric charge is thus said to be **quantized** (existing only in discrete amounts: $1e$, $2e$, $3e$, etc.). Because e is so small, however, we normally do not notice this discreteness in macroscopic charges ($1\,\mu\mathrm{C}$ requires about 10^{13} electrons), which thus seem continuous.

The charge on a proton, a basic constituent in the nucleus of atoms (Section 21–2), is precisely $+e$, the same magnitude charge as an electron but opposite sign. (The mass of a proton is almost 2000 times greater than the electron mass.)

Coulomb's law looks a lot like the *law of universal gravitation*, $F = Gm_1m_2/r^2$, which expresses the magnitude of the gravitational force a mass m_1 exerts on a mass m_2 (Eq. 6–1). Both are **inverse square laws** $(F \propto 1/r^2)$. Both also have a proportionality to a property of each object—mass for gravity, electric charge for electricity. And both act over a distance (that is, there is no need for contact). A major difference between the two laws is that gravity is always an attractive force, whereas the electric force can be either attractive or repulsive. Electric charge comes in two types, positive and negative; gravitational mass is only positive.

The constant k in Eq. 21–1 is often written in terms of another constant, ϵ_0, called the **permittivity of free space**. It is related to k by $k = 1/4\pi\epsilon_0$. Coulomb's law can be written

COULOMB'S LAW (in terms of ϵ_0)

$$F = \frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_2}{r^2}, \qquad \textbf{(21–2)}$$

[†]In the once common cgs system of units, k is set equal to 1, and the unit of electric charge is called the *electrostatic unit* (esu) or the statcoulomb. One esu is defined as that charge, on each of two point objects 1 cm apart, that gives rise to a force of 1 dyne.

[‡]See inside front cover for precise value. (In the new SI, Section 1–4, the charge on the electron e is defined as exact, and the Coulomb defined in terms of e, details that don't concern us here.) According to the Standard Model of elementary particle physics, subnuclear particles called quarks have a smaller charge than the electron, equal to $\frac{1}{3}e$ or $\frac{2}{3}e$. Quarks have not been detected directly as isolated objects, and theory indicates that free quarks may not be detectable.

where[†]

$$\epsilon_0 = \frac{1}{4\pi k} = 8.85 \times 10^{-12}\,\mathrm{C^2/N \cdot m^2}.$$

Equation 21–2 looks more complicated than Eq. 21–1, but other fundamental equations we haven't seen yet are simpler in terms of ϵ_0 rather than k. It doesn't matter which form we use since Eqs. 21–1 and 21–2 are the same. (The precise value of ϵ_0 is given inside the front cover.)

Equations 21–1 and 21–2 apply to objects whose size is much smaller than the distance between them. Ideally, it is precise for **point charges** (spatial size negligible compared to other distances). For finite-sized objects, it is not always clear what value to use for r, particularly since the charge may not be distributed uniformly on the objects. If the two objects are spheres and the charge is known to be distributed uniformly on each, then r is the distance between the centers of the spheres (we show this in Examples 22–3 and 22–4).

Coulomb's law describes the force one electrically charged particle (Q_1) exerts on a second one (Q_2), when they are at rest. We call this the **electrostatic force** or the **Coulomb force**. Additional forces come into play when charges are in motion, and will be discussed in later Chapters. In this Chapter we discuss only charges at rest, the study of which is called **electrostatics**.

EXERCISE A Return to the Chapter-Opening Question, page 609, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 21–1 **Electric force on electron by proton.** Determine the magnitude and direction of the electric force on the electron of a hydrogen atom exerted by the single proton $(Q_2 = +e)$ that is the atom's nucleus. Assume the average distance between the revolving electron and the proton is $r = 0.53 \times 10^{-10}\,\mathrm{m}$, Fig. 21–16.

APPROACH To find the force magnitude we use Coulomb's law, $F = k\,Q_1 Q_2/r^2$ (Eq. 21–1), with $r = 0.53 \times 10^{-10}\,\mathrm{m}$. The electron and proton have the same *magnitude* of charge, e, so $Q_1 = Q_2 = 1.6 \times 10^{-19}\,\mathrm{C}$.

SOLUTION The magnitude of the force is

$$F = k\frac{Q_1 Q_2}{r^2} = \frac{(9.0 \times 10^9\,\mathrm{N \cdot m^2/C^2})(1.6 \times 10^{-19}\,\mathrm{C})(1.6 \times 10^{-19}\,\mathrm{C})}{(0.53 \times 10^{-10}\,\mathrm{m})^2}$$
$$= 8.2 \times 10^{-8}\,\mathrm{N}.$$

The direction of the force on the electron is toward the proton, because the charges have opposite signs so the force is attractive.

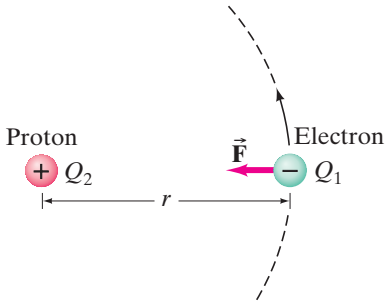


FIGURE 21–16 Example 21–1.

CONCEPTUAL EXAMPLE 21–2 **Which charge exerts the greater force?** Two positive point charges, $Q_1 = 50\,\mu\mathrm{C}$ and $Q_2 = 1\,\mu\mathrm{C}$, are separated by a distance ℓ, Fig. 21–17. Which is larger in magnitude, the force that Q_1 exerts on Q_2, or the force that Q_2 exerts on Q_1?

RESPONSE From Coulomb's law, the force on Q_1 exerted by Q_2 is

$$F_{12} = k\frac{Q_1 Q_2}{\ell^2}.$$

The force on Q_2 exerted by Q_1 is

$$F_{21} = k\frac{Q_2 Q_1}{\ell^2}$$

which is the same magnitude. The equation is symmetric with respect to the two charges, so $F_{21} = F_{12}$.

NOTE Newton's third law also tells us that these two forces must have equal magnitude.

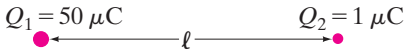


FIGURE 21–17 Example 21–2.

EXERCISE B In Example 21–2, how is the direction of F_{12} related to the direction of F_{21}?

[†]Our convention for units, such as $\mathrm{C^2/N \cdot m^2}$ for ϵ_0, means $\mathrm{m^2}$ is in the denominator. That is, $\mathrm{C^2/N \cdot m^2}$ means $\mathrm{C^2/(N \cdot m^2)}$ and does *not* mean $\mathrm{(C^2/N) \cdot m^2 = C^2 \cdot m^2/N}$.

Coulomb's law, Eq. 21–1 or 21–2, gives the force on a charge due to only *one* other charge. This Coulomb force is a vector, with magnitude and direction. If several (or many) charges are present, the *net force on any one of them will be the vector sum of the forces due to each of the others*. This **principle of superposition** is based on experiment, and tells us that electric force vectors add like any other vector. For example, if you have a system of four charges, the net force on charge 1 is the vector sum of the forces exerted on charge 1 by charges 2, 3, and 4: $\vec{\mathbf{F}}_{1\text{net}} = \vec{\mathbf{F}}_2 + \vec{\mathbf{F}}_3 + \vec{\mathbf{F}}_4$. The magnitudes of these three forces are determined from Coulomb's law, and then are added vectorially, as discussed in Chapter 3. For continuous distributions of charge, the sum becomes an integral (Section 21–7).

As in all problem solving, it is important to draw a free-body diagram (Chapter 4) for each object, showing all the forces acting *on* that object. In applying Coulomb's law, we can start with charge magnitudes only (leaving out minus signs) to get the *magnitude* of each force. Then determine separately the *direction* of the force physically (along the line joining the two particles: like charges repel, unlike charges attract), and show the force on the diagram. Finally, add all the forces on one object together as vectors to obtain the net force on that object.

PROBLEM SOLVING
Use magnitudes in Coulomb's law; find force direction from signs of charges

It can be helpful to use double subscripts on each of the forces involved. The first subscript refers to the particle *on* which the force acts; the second refers to the particle that exerts the force. For example, if we have three charges, $\vec{\mathbf{F}}_{31}$ means the force exerted *on* particle 3 *by* particle 1.

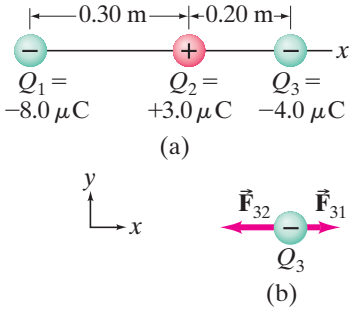


FIGURE 21–18 Example 21–3.

EXAMPLE 21–3 **Three charges in a line.** Three charged particles are arranged in a line, as shown in Fig. 21–18a. Calculate the net electrostatic force on particle 3 (the $-4.0\,\mu\text{C}$ on the right) due to the other two charges.

APPROACH The net force on particle 3 is the vector sum of the force $\vec{\mathbf{F}}_{31}$ exerted on particle 3 by particle 1 and the force $\vec{\mathbf{F}}_{32}$ exerted on particle 3 by particle 2: $\vec{\mathbf{F}} = \vec{\mathbf{F}}_{31} + \vec{\mathbf{F}}_{32}$.

SOLUTION The magnitudes of these two forces are obtained using Coulomb's law, Eq. 21–1:

$$F_{31} = k\frac{Q_3 Q_1}{r_{31}^2} = \frac{(9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(4.0 \times 10^{-6}\,\text{C})(8.0 \times 10^{-6}\,\text{C})}{(0.50\,\text{m})^2} = 1.2\,\text{N},$$

where $r_{31} = 0.50\,\text{m}$ is the distance from Q_3 to Q_1. Similarly,

$$F_{32} = k\frac{Q_3 Q_2}{r_{32}^2} = \frac{(9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(4.0 \times 10^{-6}\,\text{C})(3.0 \times 10^{-6}\,\text{C})}{(0.20\,\text{m})^2} = 2.7\,\text{N}.$$

Since we were calculating the magnitudes of the forces, we omitted the signs of the charges. But we must be aware of them to get the direction of each force. Let the line joining the particles be the x axis, and we take it positive to the right. Then, because $\vec{\mathbf{F}}_{31}$ is repulsive and $\vec{\mathbf{F}}_{32}$ is attractive, the directions of the forces are as shown in Fig. 21–18b: F_{31} points in the positive x direction (away from Q_1) and F_{32} points in the negative x direction (toward Q_2). The net force on particle 3 is then

$$F = -F_{32} + F_{31} = -2.7\,\text{N} + 1.2\,\text{N} = -1.5\,\text{N}.$$

The magnitude of the net force is 1.5 N, and it points to the left.

NOTE Charge Q_1 acts on charge Q_3 just as if Q_2 were not there (this is the principle of superposition). That is, the charge in the middle, Q_2, in no way blocks the effect of charge Q_1 acting on Q_3. Naturally, Q_2 exerts its own force on Q_3.

CAUTION
Each charge exerts its own force. No charge blocks the effect of the others

EXERCISE C Determine the magnitude and direction of the net force on charge Q_2 in Fig. 21–18a.

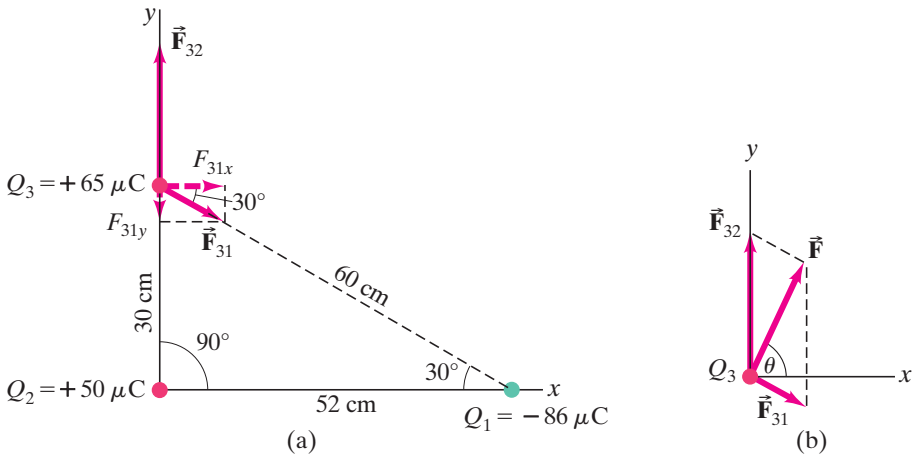


FIGURE 21–19 Determining the forces for Example 21–4. (a) The directions of the individual forces are as shown because $\vec{\mathbf{F}}_{32}$ is repulsive (the force on Q_3 is in the direction away from Q_2 because Q_3 and Q_2 are both positive) whereas $\vec{\mathbf{F}}_{31}$ is attractive (Q_3 and Q_1 have opposite signs), so $\vec{\mathbf{F}}_{31}$ points toward Q_1. (b) Adding $\vec{\mathbf{F}}_{32}$ to $\vec{\mathbf{F}}_{31}$ to obtain the net force $\vec{\mathbf{F}}$.

EXAMPLE 21–4 **Electric force using vector components.** Calculate the net electrostatic force on charge Q_3 shown in Fig. 21–19a due to the charges Q_1 and Q_2.

APPROACH We use Coulomb's law to find the magnitudes of the individual forces. The direction of each force will be along the line connecting Q_3 to Q_1 or Q_2. The forces $\vec{\mathbf{F}}_{31}$ and $\vec{\mathbf{F}}_{32}$ have the directions shown in Fig. 21–19a, since Q_1 exerts an attractive force on Q_3, and Q_2 exerts a repulsive force. Figure 21–19a also shows the x and y components of $\vec{\mathbf{F}}_{31}$. $\vec{\mathbf{F}}_{32}$ is along the $+y$ axis so has no x component. $\vec{\mathbf{F}}_{31}$ and $\vec{\mathbf{F}}_{32}$ are *not* along the same line, so to find the resultant force on Q_3 we add x and y components of $\vec{\mathbf{F}}_{31}$ and $\vec{\mathbf{F}}_{32}$ separately and perform the vector addition.

SOLUTION The magnitudes of $\vec{\mathbf{F}}_{31}$ and $\vec{\mathbf{F}}_{32}$ are (ignoring signs of the charges since we know the directions)

$$F_{31} = k\frac{Q_3 Q_1}{r_{31}^2} = \frac{(9.0\times10^9\,\mathrm{N\cdot m^2/C^2})(6.5\times10^{-5}\,\mathrm{C})(8.6\times10^{-5}\,\mathrm{C})}{(0.60\,\mathrm{m})^2} = 140\,\mathrm{N},$$

$$F_{32} = k\frac{Q_3 Q_2}{r_{32}^2} = \frac{(9.0\times10^9\,\mathrm{N\cdot m^2/C^2})(6.5\times10^{-5}\,\mathrm{C})(5.0\times10^{-5}\,\mathrm{C})}{(0.30\,\mathrm{m})^2} = 325\,\mathrm{N}.$$

(We keep 3 significant figures until the end, and then keep 2 because only 2 are given.) We determine the x and y components of $\vec{\mathbf{F}}_{31}$ (Fig. 21–19a) using trigonometric functions (as we discussed in Section 3–4):

$$F_{31x} = F_{31}\cos 30^\circ = (140\,\mathrm{N})\cos 30^\circ = 120\,\mathrm{N},$$
$$F_{31y} = -F_{31}\sin 30^\circ = -(140\,\mathrm{N})\sin 30^\circ = -70\,\mathrm{N}.$$

The force $\vec{\mathbf{F}}_{32}$ has only a y component. So the net force $\vec{\mathbf{F}}$ on Q_3 has components

$$F_x = F_{31x} = 120\,\mathrm{N},$$
$$F_y = F_{32} + F_{31y} = 325\,\mathrm{N} - 70\,\mathrm{N} = 255\,\mathrm{N}.$$

The magnitude of the net force is

$$F = \sqrt{F_x^2 + F_y^2} = \sqrt{(120\,\mathrm{N})^2 + (255\,\mathrm{N})^2} = 280\,\mathrm{N};$$

and it acts at an angle θ (see Fig. 21–19b) given by

$$\tan\theta = \frac{F_y}{F_x} = \frac{255\,\mathrm{N}}{120\,\mathrm{N}} = 2.13,$$

so $\theta = \tan^{-1}(2.13) = 65^\circ$.

NOTE Because $\vec{\mathbf{F}}_{31}$ and $\vec{\mathbf{F}}_{32}$ are not along the same line, the magnitude of $\vec{\mathbf{F}}_3$ is not equal to the sum (or difference as in Example 21–3) of the separate magnitudes.

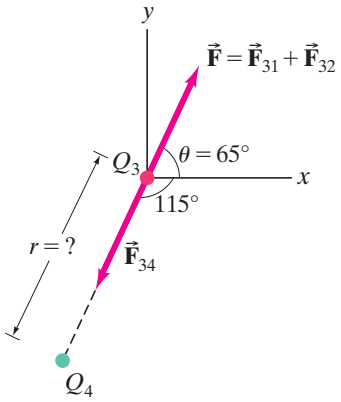


FIGURE 21–20 Example 21–5: Q_4 exerts force $(\vec{\mathbf{F}}_{34})$ that makes the net force on Q_3 zero.

CONCEPTUAL EXAMPLE 21–5 **Make the force on Q_3 zero.** In Fig. 21–19, where could you place a fourth charge, $Q_4 = -50\,\mu\text{C}$, so that the net force on Q_3 would be zero?

RESPONSE By the principle of superposition, we need a force in exactly the opposite direction to the resultant $\vec{\mathbf{F}}$ $(= \vec{\mathbf{F}}_{31} + \vec{\mathbf{F}}_{32})$ due to Q_2 and Q_1 that we calculated in Example 21–4, Fig. 21–19b. This new force must have the same magnitude, 280 N, and must point down and to the left of Q_3 in Fig. 21–19b, in the direction opposite to $\vec{\mathbf{F}}$. So Q_4 must be along this line as shown in Fig. 21–20. The distance r of Q_4 from Q_3 should be (using Coulomb's law $F_{34} = k\,Q_3\,Q_4/r^2$)

$$r = \sqrt{kQ_3Q_4/F_{34}} = \sqrt{(9.0\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(50\times10^{-6}\,\text{C})(65\times10^{-6}\,\text{C})/280\,\text{N}} = 0.32\,\text{m}.$$

EXERCISE D (*a*) Consider two point charges, $+Q$ and $-Q$, which are fixed a distance d apart. Can you find a location where a third positive charge Q could be placed so that the net electric force on this third charge is zero? (*b*) What if the first two charges were both $+Q$?

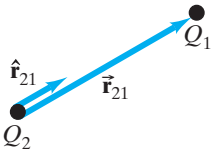
*Vector Form of Coulomb's Law

Coulomb's law can be written in vector form (as we did for Newton's law of universal gravitation in Chapter 6, Section 6–2) as

$$\vec{\mathbf{F}}_{12} = k\frac{Q_1 Q_2}{r_{21}^2}\hat{\mathbf{r}}_{21},$$

where $\vec{\mathbf{F}}_{12}$ is the vector force on charge Q_1 due to Q_2 and $\hat{\mathbf{r}}_{21}$ is the unit vector pointing from Q_2 toward Q_1. That is, $\vec{\mathbf{r}}_{21}$ points from the "source" charge (Q_2) toward the charge on which we want to know the force (Q_1). See Fig. 21–21. The charges Q_1 and Q_2 can be either positive or negative, and this will affect the direction of the electric force. If Q_1 and Q_2 have the same sign, the product $Q_1 Q_2 > 0$ and the force on Q_1 points away from Q_2: that is, it is repulsive. If Q_1 and Q_2 have opposite signs, $Q_1 Q_2 < 0$ and $\vec{\mathbf{F}}_{12}$ points toward Q_2, it is attractive.

FIGURE 21–21 Determining the force on Q_1 due to Q_2, showing the direction of the unit vector $\hat{\mathbf{r}}_{21}$.

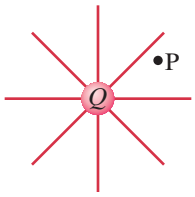


21–6 The Electric Field

Many common forces might be referred to as "contact forces," such as your hands pushing or pulling a cart, or a tennis racket hitting a tennis ball.

In contrast, both the gravitational force and the electrical force act over a distance: there is a force between two objects even when the objects are not touching. The idea of a force *acting at a distance* was a difficult one for early thinkers. Newton himself felt uneasy with this idea when he published his law of universal gravitation (Chapter 6). A helpful way to look at the situation uses the idea of the **field**, developed by the British scientist Michael Faraday (1791–1867). In the electrical case, according to Faraday, an *electric field* extends outward from every charge and permeates all of space (Fig. 21–22). If a second charge (call it q) is placed near the first charge (Q), then q feels a force exerted by the electric field that is there (say, at point P in Fig. 21–22).

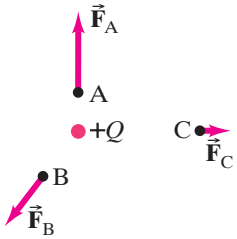
FIGURE 21–22 An electric field surrounds every charge (Q). The red lines indicate the electric field extending out from charge Q, and P is an arbitrary point.



The electric field at point P is considered to interact directly with charge q to produce the force on q.

We can in principle investigate the electric field surrounding a charge or group of charges by measuring the force on a small positive **test charge**, which is at rest. By a test charge we mean a charge so small that the force it exerts on other nearby charges does not significantly move or affect the other charges that create the field. If a tiny positive test charge q is placed at various locations in the vicinity of a single positive charge Q as shown in Fig. 21–23 (points A, B, C), the force exerted on q is as shown. The force at B is less than at A because B's distance from Q is greater (Coulomb's law); and the force at C is smaller still. In each case, the force on q is directed radially away from Q. *The electric field is defined in terms of the force on such a positive test charge.*

FIGURE 21–23 Force exerted by charge $+Q$ on a small positive test charge, q, placed at points A, B, and C.



In particular, the **electric field**, $\vec{\mathbf{E}}$, at any point in space is defined as the force $\vec{\mathbf{F}}$ exerted on a tiny positive test charge placed at that point divided by the magnitude of the test charge q:

$$\vec{\mathbf{E}} = \frac{\vec{\mathbf{F}}}{q}. \qquad [q \to 0] \qquad \textbf{(21–3)}$$

More precisely, $\vec{\mathbf{E}}$ is defined as the limit of $\vec{\mathbf{F}}/q$ as q is taken smaller and smaller, approaching zero. That is, q is so tiny that it exerts essentially no force on the other charges which created the field. From this definition (Eq. 21–3), we see that the electric field at any point in space is a vector whose direction is the direction of the force on a tiny positive test charge at that point, and whose magnitude is the *force per unit charge*. Thus $\vec{\mathbf{E}}$ has SI units of newtons per coulomb (N/C).

The reason for defining $\vec{\mathbf{E}}$ as $\vec{\mathbf{F}}/q$ (with $q \to 0$) is so that $\vec{\mathbf{E}}$ does not depend on the magnitude of the test charge q. This means that $\vec{\mathbf{E}}$ describes only the effect of the charges creating the electric field at that point.

The electric field at any point in space can be measured, based on the definition, Eq. 21–3. For simple situations involving one or several point charges, we can calculate $\vec{\mathbf{E}}$. For example, the electric field at a distance r from a single point charge Q would have magnitude

$$E = \frac{F}{q} = \frac{kqQ/r^2}{q}$$

$$E = k\frac{Q}{r^2}; \qquad [\text{single point charge}] \qquad \textbf{(21–4a)}$$

or, in terms of ϵ_0 as in Eq. 21–2 $\left(k = 1/4\pi\epsilon_0\right)$:

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}. \qquad [\text{single point charge}] \qquad \textbf{(21–4b)}$$

Notice that E is independent of the test charge q—that is, E depends only on the charge Q which produces the field, and not on the value of the test charge q. Equations 21–4 are referred to as the *electric field form of Coulomb's law*.

If we are given the electric field $\vec{\mathbf{E}}$ at a given point in space, then we can calculate the force $\vec{\mathbf{F}}$ that $\vec{\mathbf{E}}$ exerts on any charge q placed at that point (see Eq. 21–3):

$$\vec{\mathbf{F}} = q\vec{\mathbf{E}}. \qquad \textbf{(21–5)}$$

This is valid even if q is not small as long as q does not cause the charges creating $\vec{\mathbf{E}}$ to move. If q is positive, $\vec{\mathbf{F}}$ and $\vec{\mathbf{E}}$ point in the same direction. If q is negative, $\vec{\mathbf{F}}$ and $\vec{\mathbf{E}}$ point in opposite directions. See Fig. 21–24.

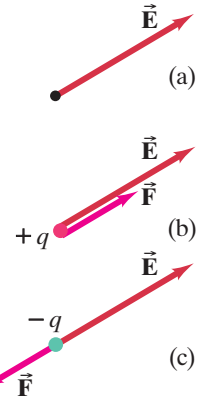


FIGURE 21–24 (a) Electric field at a given point in space. (b) Force on a positive charge at that point. (c) Force on a negative charge at that point.

EXAMPLE 21–6 **Printing and copying.** Many computer printers and photocopiers work by arranging positive charges (in the pattern to be copied) on the surface of a drum, then gently sprinkling negatively charged dry toner (ink) particles onto the drum. The toner particles temporarily stick to the pattern on the drum (Fig. 21–25). Suppose each toner particle has a mass of 9.0×10^{-16} kg and carries an average of 20 extra electrons to provide an electric charge. Assuming that the electric force on a toner particle must exceed twice its weight in order to ensure sufficient attraction, compute the required electric field strength near the surface of the drum.

APPROACH The electric force on a toner particle of charge $q = 20e$ is $F = qE$, where E is the needed electric field. This force needs to be at least as great as twice the weight (mg) of the particle.

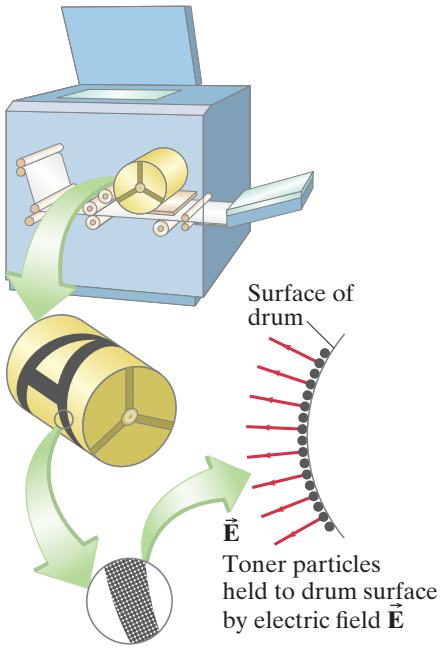
SOLUTION The minimum value of electric field satisfies the relation

$$qE = 2mg$$

where $q = 20e$. Hence

$$E = \frac{2mg}{q} = \frac{2\left(9.0 \times 10^{-16}\,\text{kg}\right)\left(9.8\,\text{m/s}^2\right)}{20\left(1.6 \times 10^{-19}\,\text{C}\right)} = 5.5 \times 10^3\,\text{N/C}.$$

FIGURE 21–25 Example 21–6



PHYSICS APPLIED
Photocopiers and printers

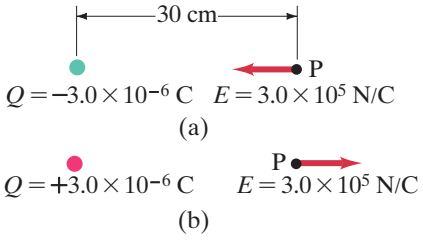


FIGURE 21–26 Example 21–7. Electric field at point P (a) due to a negative charge Q, and (b) due to a positive charge Q, each 30 cm from P.

EXAMPLE 21–7 **Electric field of a single point charge.** Calculate the magnitude and direction of the electric field at a point P which is 30 cm to the right of a point charge $Q = -3.0 \times 10^{-6}$ C.

APPROACH The magnitude of the electric field due to a single point charge is given by Eq. 21–4. The direction is found using the sign of the charge Q.

SOLUTION The magnitude of the electric field is:

$$E = k\frac{Q}{r^2} = \frac{(9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(3.0 \times 10^{-6}\,\text{C})}{(0.30\,\text{m})^2} = 3.0 \times 10^5\,\text{N/C}.$$

The direction of the electric field is *toward* the charge Q, to the left as shown in Fig. 21–26a, since we defined the direction as that of the force on a positive test charge which here would be attractive. If Q had been positive, the electric field would have pointed away, as in Fig. 21–26b.

NOTE There is no electric charge at point P. But there is an electric field there. The only charge here is Q.

This Example illustrates a general result: The electric field $\vec{\mathbf{E}}$ due to a positive charge points away from the charge, whereas $\vec{\mathbf{E}}$ due to a negative charge points toward that charge.

If the electric field at a given point in space is due to more than one charge, the individual fields (call them $\vec{\mathbf{E}}_1$, $\vec{\mathbf{E}}_2$, etc.) due to each charge are added vectorially to get the total field at that point:

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_1 + \vec{\mathbf{E}}_2 + \cdots.$$

The validity of this **superposition principle** for electric fields is fully confirmed by experiment.

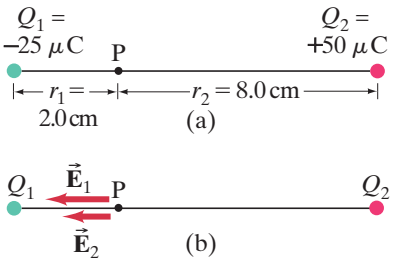


FIGURE 21–27 Example 21–8. In (b), we don't know the relative magnitudes of $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$ until we do the calculation.

EXAMPLE 21–8 **$\vec{\mathbf{E}}$ at a point between two charges.** Two point charges are separated by a distance of 10.0 cm. One has a charge of $-25\,\mu$C and the other $+50\,\mu$C. (*a*) Determine the direction and magnitude of the electric field at a point P between the two charges that is 2.0 cm from the negative charge (Fig. 21–27a). (*b*) If an electron $(\text{mass} = 9.11 \times 10^{-31}\,\text{kg})$ is placed at rest at P and then released, what will be its initial acceleration (direction and magnitude)?

APPROACH The electric field at P will be the vector sum of the fields created separately by Q_1 and Q_2. The field due to the negative charge Q_1 points toward Q_1, and the field due to the positive charge Q_2 points away from Q_2. Thus both fields point to the left as shown in Fig. 21–27b, and we can add the magnitudes of the two fields together algebraically, ignoring the signs of the charges. In (*b*) we use Newton's second law ($\Sigma\vec{\mathbf{F}} = m\vec{\mathbf{a}}$) to determine the acceleration, where $\Sigma\vec{\mathbf{F}} = q\Sigma\vec{\mathbf{E}}$.

SOLUTION (*a*) Each field is due to a point charge as given by Eq. 21–4, $E = kQ/r^2$. The total field points to the left and has magnitude

$$E = k\frac{Q_1}{r_1^2} + k\frac{Q_2}{r_2^2} = k\left(\frac{Q_1}{r_1^2} + \frac{Q_2}{r_2^2}\right)$$

$$= (9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)\left(\frac{25 \times 10^{-6}\,\text{C}}{(2.0 \times 10^{-2}\,\text{m})^2} + \frac{50 \times 10^{-6}\,\text{C}}{(8.0 \times 10^{-2}\,\text{m})^2}\right)$$

$$= 6.3 \times 10^8\,\text{N/C}.$$

(*b*) The electric field points to the left, so the electron will feel a force to the *right* since it is negatively charged. Therefore the acceleration $a = F/m$ (Newton's second law) will be to the right. The force on a charge q in an electric field E is $F = qE$ (Eq. 21–5). Hence the magnitude of the electron's initial acceleration is

$$a = \frac{F}{m} = \frac{qE}{m} = \frac{(1.60 \times 10^{-19}\,\text{C})(6.3 \times 10^8\,\text{N/C})}{9.11 \times 10^{-31}\,\text{kg}} = 1.1 \times 10^{20}\,\text{m/s}^2.$$

NOTE By considering the directions of *each* field ($\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$) before doing any calculations, we made sure our calculation could be done simply and correctly.

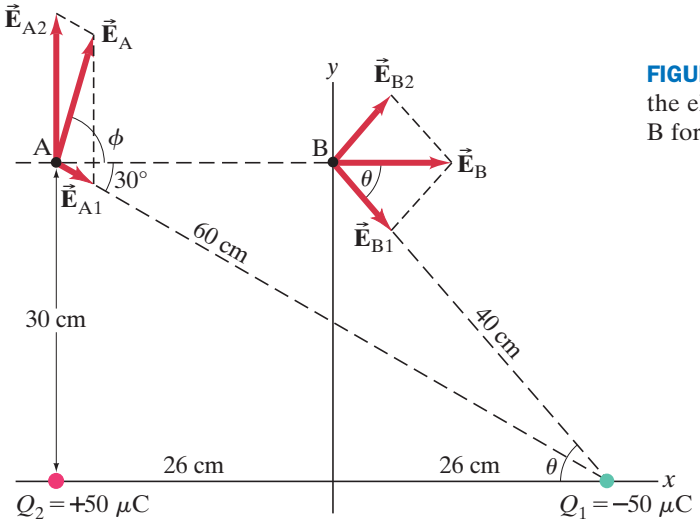


FIGURE 21–28 Calculation of the electric field at points A and B for Example 21–9.

EXAMPLE 21–9 **$\vec{E}$ above two point charges.** Calculate the total electric field (*a*) at point A and (*b*) at point B in Fig. 21–28 due to both charges, Q_1 and Q_2.

APPROACH The calculation is much like that of Example 21–4, except now we are dealing with electric fields instead of force. The electric field at point A is the vector sum of the fields $\vec{E}_{A1}$ due to Q_1, and $\vec{E}_{A2}$ due to Q_2. We find the magnitude of the field produced by each point charge, then we add their components to find the total field at point A. We do the same for point B.

SOLUTION (*a*) The magnitude of the electric field produced at point A by each of the charges Q_1 and Q_2 is given by $E = kQ/r^2$, so

$$E_{A1} = \frac{(9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(50 \times 10^{-6}\,\text{C})}{(0.60\,\text{m})^2} = 1.25 \times 10^6\,\text{N/C},$$

$$E_{A2} = \frac{(9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(50 \times 10^{-6}\,\text{C})}{(0.30\,\text{m})^2} = 5.0 \times 10^6\,\text{N/C}.$$

PROBLEM SOLVING
Ignore signs of charges and determine direction physically, showing directions on diagram

The direction of E_{A1} points from A toward Q_1 (negative charge), whereas E_{A2} points from A away from Q_2, as shown; so the total electric field at A, $\vec{E}_A$, has components

$$E_{Ax} = E_{A1}\cos 30° = 1.1 \times 10^6\,\text{N/C},$$
$$E_{Ay} = E_{A2} - E_{A1}\sin 30° = 4.4 \times 10^6\,\text{N/C}.$$

Thus the magnitude of $\vec{E}_A$ is

$$E_A = \sqrt{(1.1)^2 + (4.4)^2} \times 10^6\,\text{N/C} = 4.5 \times 10^6\,\text{N/C},$$

and its direction is ϕ given by $\tan\phi = E_{Ay}/E_{Ax} = 4.4/1.1 = 4.0$, so $\phi = 76°$.

(*b*) Because B is equidistant from the two equal charges (40 cm by the Pythagorean theorem), the magnitudes of E_{B1} and E_{B2} are the same; that is,

$$E_{B1} = E_{B2} = \frac{kQ}{r^2} = \frac{(9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(50 \times 10^{-6}\,\text{C})}{(0.40\,\text{m})^2}$$
$$= 2.8 \times 10^6\,\text{N/C}.$$

PROBLEM SOLVING
Use symmetry to save work, when possible

Also, because of the symmetry, the y components are equal and opposite, and so cancel out. Hence the total field E_B is horizontal and equals $E_{B1}\cos\theta + E_{B2}\cos\theta = 2E_{B1}\cos\theta$. From the diagram, $\cos\theta = 26\,\text{cm}/40\,\text{cm} = 0.65$. Then

$$E_B = 2E_{B1}\cos\theta = 2(2.8 \times 10^6\,\text{N/C})(0.65) = 3.6 \times 10^6\,\text{N/C},$$

and the direction of $\vec{E}_B$ is along the $+x$ direction.

NOTE We could have done part (*b*) in the same way we did part (*a*). But symmetry allowed us to solve the problem with less effort.

EXERCISE E Four charges of equal magnitude, but possibly different sign, are placed on the corners of a square. What arrangement of charges will produce an electric field with the greatest magnitude at the center of the square? (*a*) All four positive charges; (*b*) all four negative charges; (*c*) three positive and one negative; (*d*) two positive and two negative; (*e*) three negative and one positive.

PROBLEM SOLVING

Electrostatics: Electric Forces and Electric Fields

Whether you use electric field or electrostatic forces, the procedure for solving electrostatics problems is similar:

1. **Draw** a careful **diagram**—namely, a free-body diagram for each object, showing all the forces acting on that object, or showing the electric field at a point due to all significant charges present. Determine the **direction** of each force or electric field physically: like charges repel each other, unlike charges attract; fields point away from a + charge, and toward a − charge. Show and label each vector force or field on your diagram.
2. **Apply Coulomb's law** to calculate the magnitude of the force that each contributing charge exerts on a charged object, or the magnitude of the electric field each charge produces at a given point. Deal only with magnitudes of charges (leaving out minus signs), and obtain the magnitude of each force or electric field.
3. **Add vectorially** all the forces on an object, or the contributing fields at a point, to get the resultant. Use **symmetry** (say, in the geometry) whenever possible.
4. **Check** your answer. Is it **reasonable**? If a function of distance, does it give reasonable results in limiting cases?

21–7 Electric Field Calculations for Continuous Charge Distributions

In many cases we can treat charge as being distributed continuously.† We can divide up a charge distribution into infinitesimal charges dQ, each of which will act as a tiny point charge. The contribution to the electric field at a distance r from each dQ is

$$dE = \frac{1}{4\pi\epsilon_0}\frac{dQ}{r^2}. \qquad \textbf{(21–6a)}$$

Then the electric field, $\vec{\mathbf{E}}$, at any point is obtained by summing over all the infinitesimal contributions, which is the integral

$$\vec{\mathbf{E}} = \int d\vec{\mathbf{E}}. \qquad \textbf{(21–6b)}$$

Note that $d\vec{\mathbf{E}}$ is a vector (Eq. 21–6a gives its magnitude). Sometimes Eq. 21–6b is difficult to evaluate, but often we can use other techniques as we will see. (Numerical integration, discussed in Appendix C, can also be used in many cases.)

EXAMPLE 21–10 **A ring of charge.** A thin, ring-shaped object of radius a holds a total charge $+Q$ distributed uniformly around it. Determine the electric field at a point P on its axis, a distance x from the center. See Fig. 21–29. Let λ be the charge per unit length (C/m).

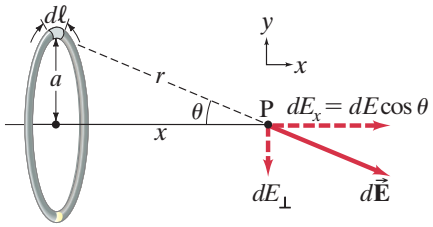


FIGURE 21–29 Example 21–10.

†Because we believe there is a minimum charge (e), the treatment here is only for convenience; it is nonetheless useful and accurate since e is usually very much smaller than macroscopic charges.

APPROACH AND SOLUTION We explicitly follow the steps of the Problem Solving Strategy on page 622.

1. **Draw** a careful **diagram**, which is Fig. 21–29. The **direction** of the electric field due to one infinitesimal length $d\ell$ of the charged ring is shown in Fig. 21–29.
2. **Apply Coulomb's law**. The electric field, $d\vec{\mathbf{E}}$, due to this particular segment of the ring of length $d\ell$ has magnitude

$$dE = \frac{1}{4\pi\epsilon_0}\frac{dQ}{r^2}.$$

The whole ring has circumference of $2\pi a$, so the charge on a length $d\ell$ is

$$dQ = Q\left(\frac{d\ell}{2\pi a}\right) = \lambda\, d\ell$$

where $\lambda = Q/2\pi a$ is the charge per unit length (or the linear charge density). Now we write dE as

$$dE = \frac{1}{4\pi\epsilon_0}\frac{\lambda\, d\ell}{r^2}.$$

3. Add **vectorially** and use **symmetry**: The vector $d\vec{\mathbf{E}}$ has components dE_x along the x axis and $dE_\perp$ perpendicular to the x axis (Fig. 21–29). We are going to sum (integrate) around the entire ring. We note that an equal-length segment $d\ell$, diametrically opposite the $d\ell$ shown, will produce a $d\vec{\mathbf{E}}$ whose component perpendicular to the x axis will just cancel the $dE_\perp$ shown. This is true for all segments of the ring, so by symmetry $\vec{\mathbf{E}}$ will have zero y components, and so we need only sum the x components, dE_x. The total field is then

PROBLEM SOLVING
Use symmetry when possible

$$E = E_x = \int dE_x = \int dE\cos\theta = \frac{1}{4\pi\epsilon_0}\lambda\int\frac{d\ell}{r^2}\cos\theta.$$

Because $r^2 = x^2 + a^2$, where we treat x as fixed (not a variable), and $\cos\theta = x/r$, we have

$$E = \frac{\lambda}{(4\pi\epsilon_0)}\frac{x}{(x^2+a^2)^{\frac{3}{2}}}\int_0^{2\pi a} d\ell = \frac{1}{4\pi\epsilon_0}\frac{\lambda x(2\pi a)}{(x^2+a^2)^{\frac{3}{2}}} = \frac{1}{4\pi\epsilon_0}\frac{Qx}{(x^2+a^2)^{\frac{3}{2}}}.$$

4. To **check reasonableness**, note that at great distances, $x \gg a$, this result reduces to $E = Q/(4\pi\epsilon_0 x^2)$. We would expect this result because at great distances the ring would appear to be a point charge ($1/r^2$ dependence). Also note that our result gives $E = 0$ at $x = 0$, as we might expect because all components will cancel at the center of the circle.

PROBLEM SOLVING
Check result by noting that at a great distance the ring looks like a point charge

Note in this Example three important problem-solving techniques that can be used elsewhere: (1) using symmetry to reduce the complexity of the problem; (2) expressing the charge dQ in terms of a charge density (here linear, $\lambda = Q/2\pi a$); and (3) checking the answer at the limit of large r, which serves as an indication (but not proof) of the correctness of the answer—if the result does not check at large r, your result has to be wrong.

PROBLEM SOLVING
Use symmetry, charge density, and values at $r = 0$ and ∞

CONCEPTUAL EXAMPLE 21–11 **Charge at the center of a ring.** Imagine a small positive charge placed at the center of a nonconducting ring carrying a uniformly distributed negative charge. Is the positive charge in equilibrium if it is displaced slightly from the center along the axis of the ring, and if so is it stable? What if the small charge is negative? Neglect gravity, as it is much smaller than the electrostatic forces.

RESPONSE A positive charge at the center of the ring is in equilibrium because there is no net force on it, by *symmetry*. If the positive charge is moved away from the center of the negative ring along the axis in either direction, the attractive force acts back towards the center of the ring and so the charge is in *stable* equilibrium. A negative charge at the center of the ring feels no net force, but is in *unstable* equilibrium because if it is moved along the ring's axis, the repulsive force acts away from the ring, pushing the charge farther away.

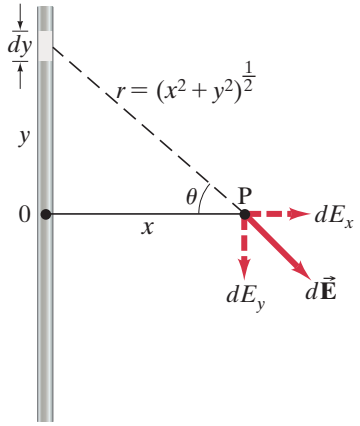


FIGURE 21–30 Example 21–12.

EXAMPLE 21–12 Long line of charge. Determine the magnitude of the electric field at any point P a distance x from the midpoint 0 of a very long straight wire of uniformly distributed positive charge, Fig. 21–30. Assume x is much smaller than the length of the wire, and let λ be the charge per unit length (C/m).

APPROACH We set up a coordinate system so the wire is on the y axis with origin 0 as shown. A segment of wire dy has charge $dQ = \lambda\, dy$. The field $d\vec{\mathbf{E}}$ at point P due to this length dy of wire (at y) has magnitude

$$dE = \frac{1}{4\pi\epsilon_0}\frac{dQ}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{\lambda\, dy}{(x^2 + y^2)},$$

where $r = (x^2 + y^2)^{\frac{1}{2}}$ as shown in Fig. 21–30. The vector $d\vec{\mathbf{E}}$ has components dE_x and dE_y as shown, where $dE_x = dE\cos\theta$ and $dE_y = dE\sin\theta$.

SOLUTION Because 0 is at the midpoint of the wire, the y component of $\vec{\mathbf{E}}$ will be zero since there will be equal contributions to $E_y = \int dE_y$ from above and below point 0:

$$E_y = \int dE\sin\theta = 0.$$

Thus we have

$$E = E_x = \int dE\cos\theta = \frac{\lambda}{4\pi\epsilon_0}\int \frac{\cos\theta\, dy}{x^2 + y^2}.$$

The integration here is over y, along the wire, with x treated as constant. We must now write θ as a function of y, or y as a function of θ. We do the latter: since $y = x\tan\theta$, then $dy = x\, d\theta/\cos^2\theta$. Furthermore, because $\cos\theta = x/\sqrt{x^2 + y^2}$, then $1/(x^2 + y^2) = \cos^2\theta/x^2$ and our integrand above is $(\cos\theta)(x\, d\theta/\cos^2\theta)(\cos^2\theta/x^2) = \cos\theta\, d\theta/x$. Hence

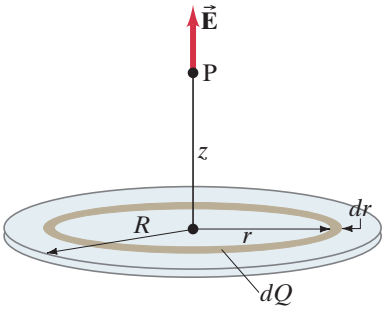
$$E = \frac{\lambda}{4\pi\epsilon_0}\frac{1}{x}\int_{-\pi/2}^{\pi/2}\cos\theta\, d\theta = \frac{\lambda}{4\pi\epsilon_0 x}(\sin\theta)\Big|_{-\pi/2}^{\pi/2} = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{x},$$

where we have assumed the wire is extremely long in both directions ($y \to \pm\infty$) which corresponds to the limits $\theta = \pm\pi/2$. Thus the field near a long straight wire of uniform charge decreases inversely as the first power of the distance from the wire. Compare to a point charge where E decreases as the *square* of the distance.

NOTE This result, obtained for an infinite wire, is a good approximation for a wire of finite length as long as x is small compared to the distance of P from the ends of the wire.

EXAMPLE 21–13 Uniformly charged disk. Charge is distributed uniformly over a thin circular disk of radius R. The charge per unit area (C/m^2) is σ. Calculate the electric field at a point P on the axis of the disk, a distance z above its center, Fig. 21–31.

FIGURE 21–31 Example 21–13; a uniformly charged thin flat disk of radius R.



APPROACH We can think of the disk as a set of concentric rings. We can then apply the result of Example 21–10 to each of these rings, and then sum over all the rings.

SOLUTION For the ring of radius r shown in Fig. 21–31, the electric field has magnitude (see result of Example 21–10)

$$dE = \frac{1}{4\pi\epsilon_0}\frac{z\, dQ}{(z^2 + r^2)^{\frac{3}{2}}}$$

where we have written dE (instead of E) for this thin ring of total charge dQ. The ring has area $(dr)(2\pi r)$ and charge per unit area $\sigma = dQ/(2\pi r\, dr)$, which we call the surface charge density. We solve this for dQ $(= \sigma\, 2\pi r\, dr)$ and insert it in the equation above for dE:

$$dE = \frac{1}{4\pi\epsilon_0}\frac{z\sigma 2\pi r\, dr}{(z^2 + r^2)^{\frac{3}{2}}} = \frac{z\sigma r\, dr}{2\epsilon_0(z^2 + r^2)^{\frac{3}{2}}}.$$

Now we integrate[†] over all the rings, starting at $r = 0$ out to the largest with $r = R$:

$$E = \frac{z\sigma}{2\epsilon_0}\int_0^R \frac{r\,dr}{(z^2 + r^2)^{\frac{3}{2}}} = \frac{z\sigma}{2\epsilon_0}\left[-\frac{1}{(z^2 + r^2)^{\frac{1}{2}}}\right]_0^R = \frac{\sigma}{2\epsilon_0}\left[1 - \frac{z}{(z^2 + R^2)^{\frac{1}{2}}}\right].$$

This gives the magnitude of $\vec{\mathbf{E}}$ at any point z along the axis of the disk. The direction of each $d\vec{\mathbf{E}}$ due to each ring is along the z axis (as in Example 21–10), and therefore the direction of $\vec{\mathbf{E}}$ is along z. If Q (and σ) are positive, $\vec{\mathbf{E}}$ points away from the disk; if Q (and σ) are negative, $\vec{\mathbf{E}}$ points toward the disk.

If the radius of the disk in Example 21–13 is much greater than the distance of our point P from the disk (i.e., $z \ll R$) then we can obtain a very useful result: the second term in the solution above becomes very small, so

$$E = \frac{\sigma}{2\epsilon_0}. \qquad \text{[infinite plane]} \quad \textbf{(21–7)}$$

This result is valid for any point above (or below) an infinite plane of any shape holding a uniform surface charge density σ (charge per unit area). It is also valid for points close to a finite plane, as long as the point is close to the plane compared to the distance to the edges of the plane. Thus the field near a large uniformly charged plane is uniform, and directed outward if the plane is positively charged.

It is interesting to compare here the distance dependence of the electric field (1) due to a point charge $(E \sim 1/r^2)$, (2) due to a very long uniform line of charge $(E \sim 1/r)$, and (3) due to a very large uniform plane of charge (E does not depend on r).

CAUTION
Comparing dependence of E on distance

EXAMPLE 21–14 Two parallel plates. Two large parallel plates are very thin and are separated by a distance d which is small compared to their height and width. One plate carries a positive charge, with uniform charge density σ, and the other carries a uniform negative charge density, $-\sigma$. Determine the electric field between the plates near their center where the field is assumed to be uniform.

APPROACH From Eq. 21–7, each plate sets up an electric field of magnitude $\sigma/2\epsilon_0$. The field due to the positive plate points away from that plate whereas the field due to the negative plate points toward that plate. See Fig. 21–32, which is a zoom-in view of the central portion of the two plates.

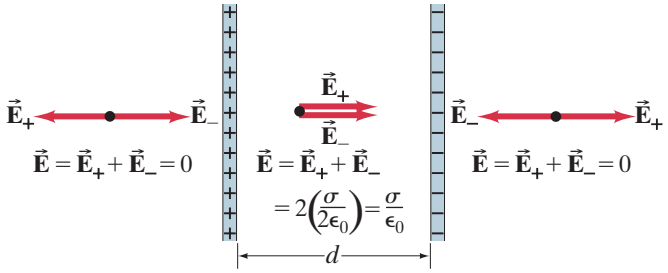


FIGURE 21–32 Example 21–14. Two large flat parallel plates very close together, oppositely charged. Shown here is only the central portion for calculating the uniform electric field between the plates.

SOLUTION In the region between the plates, the fields add together as shown:

$$E = E_+ + E_- = \frac{\sigma}{2\epsilon_0} + \frac{\sigma}{2\epsilon_0} = \frac{\sigma}{\epsilon_0}. \quad \left[\begin{array}{c}\text{2 large flat plates, oppositely}\\ \text{charged, close together}\end{array}\right] \quad \textbf{(21–8)}$$

The field is uniform, since the plates are very large compared to their separation, so this result is valid for any point, whether near one or the other of the plates, or midway between them as long as the point is far from the ends. Outside the plates, the fields cancel,

$$E = E_+ + E_- = \frac{\sigma}{2\epsilon_0} - \frac{\sigma}{2\epsilon_0} = 0,$$

as shown in Fig. 21–32. These results are valid ideally for infinitely large plates; they are a good approximation for finite plates if the separation is much less than the dimensions of the plate and for points not too close to the edge.

NOTE These useful and extraordinary results illustrate the *principle of superposition* and its great power.

[†]You should confirm this integral by taking the derivation of the result and seeing that it equals the integrand. You might also take a look at Appendix B.

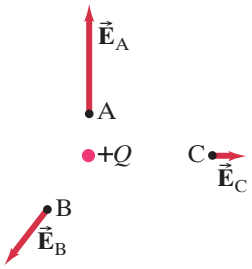


FIGURE 21–33 Electric field vector at three points, due to one point charge Q.

FIGURE 21–34 Electric field lines near (a) a single positive point charge, (b) a single negative point charge.

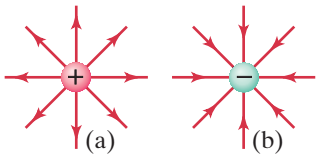
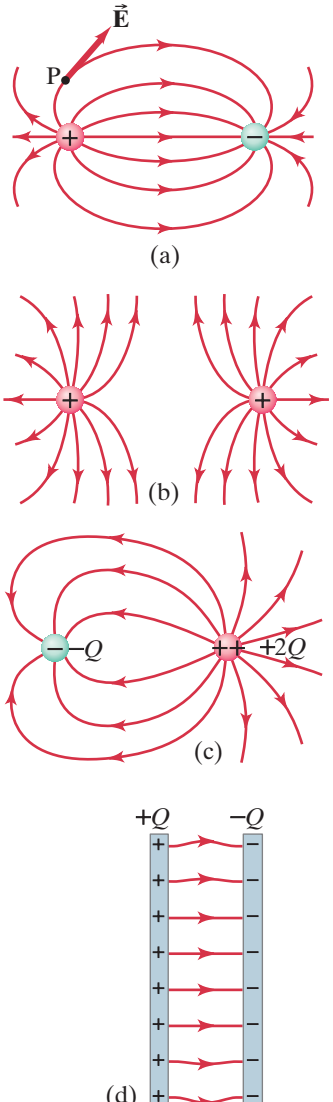


FIGURE 21–35 Electric field lines for four arrangements of charges.



21–8 Field Lines

Since the electric field is a vector, it is sometimes referred to as a **vector field**. We could indicate the electric field with arrows at various points in a given situation, such as at A, B, and C in Fig. 21–33. The directions of $\vec{\mathbf{E}}_A$, $\vec{\mathbf{E}}_B$, and $\vec{\mathbf{E}}_C$ are the same as for the forces shown earlier in Fig. 21–23, but the magnitudes (arrow lengths) are different since we divide $\vec{\mathbf{F}}$ by q to get $\vec{\mathbf{E}}$. However, the relative magnitudes of $\vec{\mathbf{E}}_A$, $\vec{\mathbf{E}}_B$, and $\vec{\mathbf{E}}_C$ are the same as for the forces since we divide by the same q each time. To indicate the electric field in such a way at *many* points, however, would result in many arrows, which would become cluttered and confusing. To avoid this, we use another technique, that of field lines.

To visualize the electric field, we draw a series of lines to indicate the direction of the electric field at various points in space. These **electric field lines** (or **lines of force**) are drawn to indicate the direction of the force due to the given field on a positive test charge. The lines of force due to a single isolated positive charge are shown in Fig. 21–34a, and for a negative charge in Fig. 21–34b. In part (a) the lines point radially outward from the charge, and in part (b) they point radially inward toward the charge because that is the direction the force would be on a positive test charge in each case (as in Fig. 21–26). Only a few representative lines are shown. We could draw lines in between those shown since the electric field exists there as well. We can draw the lines so that the *number of lines starting on a positive charge, or ending on a negative charge, is proportional to the magnitude of the charge*. Notice that nearer the charge, where the electric field is greater $(F \propto 1/r^2)$, the lines are closer together. This is a general property of electric field lines: *the closer together the lines are, the stronger the electric field in that region*. In fact, field lines can be drawn so that the *number of lines crossing unit area perpendicular to* $\vec{\mathbf{E}}$ *is proportional to the magnitude of the electric field*.

Figure 21–35a shows the electric field lines due to two equal charges of opposite sign, a combination known as an **electric dipole**. The electric field lines are curved in this case and are directed from the positive charge to the negative charge. The direction of the electric field at any point is tangent to the field line at that point as shown by the vector arrow $\vec{\mathbf{E}}$ at point P. To satisfy yourself that this is the correct pattern for the electric field lines, you can make a few calculations such as those done in Example 21–9 for just this case (see Fig. 21–28). Figure 21–35b shows the electric field lines for two equal positive charges, and Fig. 21–35c for unequal charges, $-Q$ and $+2Q$. Note that twice as many lines leave $+2Q$, as enter $-Q$ (number of lines is proportional to magnitude of Q). Finally, in Fig. 21–35d, we see the field lines between two parallel plates carrying equal but opposite charges. Notice that the electric field lines between the two plates start out perpendicular to the surface of the metal plates (Example 21–14 and the next Section) and go directly from one plate to the other, as we expect because a positive test charge placed between the plates would feel a strong repulsion from the positive plate and a strong attraction to the negative plate. The field lines between two close plates are parallel and equally spaced in the central region, but fringe outward near the edges. Thus, in the central region, the electric field has the same magnitude at all points, and we can write (see Example 21–14)

$$E = \text{constant} = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 A}. \quad \left[\begin{array}{l}\text{between two closely spaced,}\\ \text{oppositely charged, parallel plates}\end{array}\right] \qquad \textbf{(21–8)}$$

where $Q = \sigma A$ is the charge on one plate of area A. The fringing of the field near the edges can often be ignored, particularly if the separation of the plates is small compared to their height and width. We summarize the properties of field lines as:

1. Electric field lines indicate the direction of the electric field; the field points in the direction tangent to the field line at any point.
2. The lines are drawn so that the magnitude of the electric field, E, is proportional to the number of lines crossing unit area perpendicular to the lines. The closer together the lines, the stronger the field.
3. Electric field lines start on positive charges and end on negative charges; and the number starting or ending is proportional to the magnitude of the charge.

Also note that field lines never cross. Why not? Because it would not make sense for the electric field to have two directions at the same point.

Our diagrams for electric field lines, as Figs. 21–34 and 21–35, are 2-dimensional. It is hard to show them in 3 dimensions, but the Chapter-opening drawing for Chapter 22, page 643, is an attempt to suggest field lines in 3-D.

Gravitational Field

The field concept can also be applied to the gravitational force (Chapter 6). Thus we can say that a **gravitational field** exists for every object that has mass. One object attracts another by means of the gravitational field. The Earth, for example, can be said to possess a gravitational field (Fig. 21–36) which is responsible for the gravitational force on objects. The force that keeps Earth satellites and the Moon revolving around Earth is exerted by the gravitational field produced at their position by the Earth. The *gravitational field* is defined as the *force per unit mass*. The magnitude of the Earth's gravitational field at any point above the Earth's surface is GM_E/r^2, where M_E is the mass of the Earth, r is the distance of the point from the Earth's center, and G is the gravitational constant (Chapter 6). At the Earth's surface, r is the radius of the Earth and the gravitational field is equal to g, the acceleration due to gravity. Beyond the Earth, the gravitational field can be calculated at any point as a sum of terms due to Earth, Sun, Moon, and other bodies that contribute significantly.

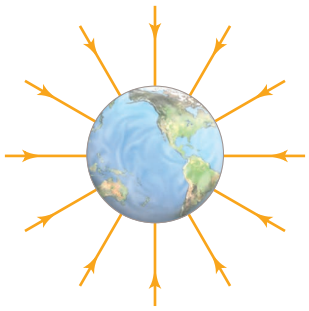


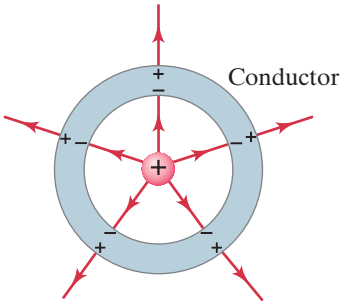
FIGURE 21–36 The Earth's gravitational field, which at any point is directed toward the Earth's center (the force on any mass points toward the Earth's center).

21–9 Electric Fields and Conductors

We now discuss some properties of conductors. First, *the electric field inside a conductor is zero in the static situation*—that is, when the charges are at rest. If there were an electric field within a conductor, there would be a force on the free electrons. The electrons would move until they reached positions where the electric field, and therefore the electric force on them, did become zero.

This reasoning has some interesting consequences. For one, *any net charge on a conductor distributes itself on the surface*. For a negatively charged conductor, you can imagine that the negative charges repel one another and race to the surface to get as far from one another as possible. Another consequence is the following. Suppose that a positive charge Q is surrounded by an isolated uncharged metal conductor whose shape is a spherical shell, Fig. 21–37. Because there can be no field within the metal, the lines leaving the central positive charge must end on negative charges on the inner surface of the metal shell. That is, the encircled charge $+Q$ induces an equal amount of negative charge, $-Q$, on the inner surface of the spherical shell. Since the shell is neutral, a positive charge of the same magnitude, $+Q$, must exist on the outer surface of the shell. Thus, although no field exists in the metal shell itself, an electric field exists outside of it, as shown in Fig. 21–37, as if the metal were not even there.

FIGURE 21–37 A charge inside a neutral spherical metal shell induces charge on the shell's surfaces. The electric field exists even beyond the shell, but not within the conducting shell itself.



A related property of static electric fields and conductors is that *the electric field is always perpendicular to the surface outside of a conductor*. If there were a component of $\vec{\mathbf{E}}$ parallel to the surface (Fig. 21–38), it would exert a force on free electrons at the surface, causing the electrons to move along the surface until they reached positions where no net force was exerted on them parallel to the surface—that is, until the electric field was perpendicular to the surface. [These results will be confirmed in the next Chapter, Section 22–3, Example 22–8.]

These properties apply only to conductors. Inside a nonconductor, which does not have free electrons, a static electric field can exist as we will see in the next Chapter (Examples 22–4 and 22–5). Also, the electric field outside a *non*conductor does not necessarily make an angle of 90° to the surface.

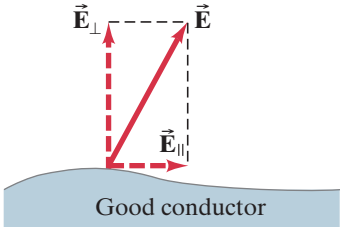


FIGURE 21–38 If the electric field $\vec{\mathbf{E}}$ at the surface of a conductor had a component parallel to the surface, $\vec{\mathbf{E}}_\parallel$, the latter would accelerate electrons into motion (no longer static). In the static case, $\vec{\mathbf{E}}_\parallel$ must be zero, and the electric field must be perpendicular to the conductor's surface: $\vec{\mathbf{E}} = \vec{\mathbf{E}}_\perp$.

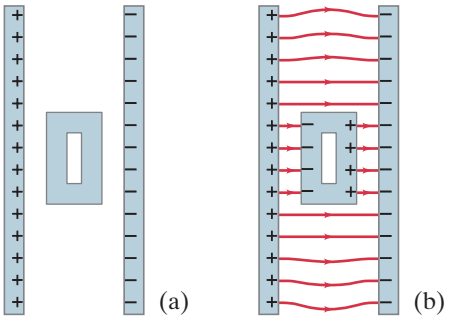


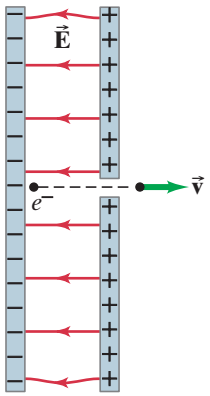
FIGURE 21–39 Example 21–15.

PHYSICS APPLIED
Electrical shielding

FIGURE 21–40 High-voltage "Van de Graaff" generators create strong electric fields in the vicinity of the "Faraday cage" below. The strong field accelerates stray electrons in the atmosphere to the high kinetic energy needed to knock electrons out of air atoms, causing an avalanche of charge which flows to (or from) the metal cage. The metal cage protects the person inside it.



FIGURE 21–41 Example 21–16.



CONCEPTUAL EXAMPLE 21–15 **Shielding, and safety in a storm.** A neutral hollow metal box is placed between two parallel charged plates as shown in Fig. 21–39a. What is the field like inside the box?

RESPONSE If our metal box had been solid, and not hollow, free electrons in the box would have redistributed themselves along the surface until all their individual fields would have canceled each other inside the box. The net field inside the box would have been zero. For a hollow box, the external field is not changed since the electrons in the metal can move just as freely as before to the surface. Hence the field inside the hollow metal box is also zero, and the field lines are shown in Fig. 21–39b. A conducting box is an effective device for shielding delicate instruments and electronic circuits from unwanted external electric fields. We also can see that a relatively safe place to be during a lightning storm is inside a parked car, surrounded by metal. See also Fig. 21–40, where a person inside a porous "cage" is protected from a strong electric discharge. (It is not safe in a lightning storm to be near a tree which can conduct, or out in the open where you are taller than the surroundings.)

21–10 Motion of a Charged Particle in an Electric Field

If an object having an electric charge q is at a point in space where the electric field is $\vec{\mathbf{E}}$, the force on the object exerted by that electric field is given by

$$\vec{\mathbf{F}} = q\vec{\mathbf{E}}$$

(see Eq. 21–5). In the past few Sections we have seen how to determine $\vec{\mathbf{E}}$ for some particular situations. Now let us suppose we know $\vec{\mathbf{E}}$ and we want to find the force on a charged object and the object's subsequent motion. (We assume no other forces act.)

EXAMPLE 21–16 **Electron accelerated by electric field.** An electron (mass $m = 9.1 \times 10^{-31}\,\text{kg}$) is accelerated in the uniform field $\vec{\mathbf{E}}\,(E = 2.0 \times 10^4\,\text{N/C})$ between two parallel charged plates. The separation of the plates is 1.5 cm. The electron is accelerated from rest near the negative plate and passes through a tiny hole in the positive plate, Fig. 21–41. (*a*) With what speed does it leave the hole? (*b*) Show that the gravitational force can be ignored. Assume the hole is so small that it does not affect the uniform field between the plates.

APPROACH We can obtain the electron's velocity using the kinematic equations of Chapter 2, after first finding its acceleration from Newton's second law, $F = ma$. The magnitude of the force on the electron is $F = qE$ and is directed to the right.

SOLUTION (*a*) The magnitude of the electron's acceleration is

$$a = \frac{F}{m} = \frac{qE}{m}.$$

Between the plates $\vec{\mathbf{E}}$ is uniform so the electron undergoes uniformly accelerated motion with acceleration

$$a = \frac{(1.6 \times 10^{-19}\,\text{C})(2.0 \times 10^4\,\text{N/C})}{(9.1 \times 10^{-31}\,\text{kg})} = 3.5 \times 10^{15}\,\text{m/s}^2.$$

It travels a distance $x = 1.5 \times 10^{-2}\,\text{m}$ before reaching the hole, and since its initial speed was zero, we can use the kinematic equation, $v^2 = v_0^2 + 2ax$ (Eq. 2–12c), with $v_0 = 0$:

$$v = \sqrt{2ax} = \sqrt{2(3.5 \times 10^{15}\,\text{m/s}^2)(1.5 \times 10^{-2}\,\text{m})} = 1.0 \times 10^7\,\text{m/s}.$$

There is no electric field outside the plates, so after passing through the hole, the electron moves with this speed, which is now constant.

(*b*) The magnitude of the electric force on the electron is

$$qE = (1.6 \times 10^{-19}\,\text{C})(2.0 \times 10^{4}\,\text{N/C}) = 3.2 \times 10^{-15}\,\text{N}.$$

The gravitational force is

$$mg = (9.1 \times 10^{-31}\,\text{kg})(9.8\,\text{m/s}^2) = 8.9 \times 10^{-30}\,\text{N},$$

which is 10^{14} times smaller! Note that the electric field due to the electron does not come into it (a particle cannot exert a force on itself).

EXAMPLE 21–17 **Electron moving perpendicular to $\vec{\mathbf{E}}$.** Suppose an electron traveling with speed v_0 enters a uniform electric field $\vec{\mathbf{E}}$, which is at right angles to $\vec{\mathbf{v}}_0$ as shown in Fig. 21–42. Describe its motion by giving the equation of its path while in the electric field. Ignore gravity.

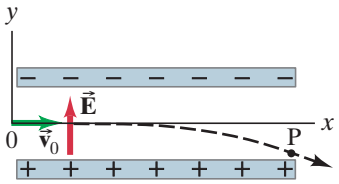


FIGURE 21–42 Example 21–17.

APPROACH Again we use Newton's second law, with $F = qE$, and the kinematic equations from Chapter 2.

SOLUTION When the electron enters the electric field (at $x = y = 0$) it has velocity $\vec{\mathbf{v}}_0 = v_0\hat{\mathbf{i}}$ in the x direction. The electric field $\vec{\mathbf{E}}$, pointing vertically upward, imparts a uniform vertical acceleration to the electron of

$$a_y = \frac{F}{m} = \frac{qE}{m} = -\frac{eE}{m},$$

where we set $q = -e$ for the electron. The electron's vertical position is given by Eq. 2–12b,

$$y = \frac{1}{2}a_y t^2 = -\frac{eE}{2m}t^2$$

since the motion is at constant acceleration. The horizontal position is given by

$$x = v_0 t$$

since $a_x = 0$. We eliminate t between these two equations and obtain

$$y = -\frac{eE}{2mv_0^2}x^2,$$

which is the equation of a parabola (just as in projectile motion, Section 3–7).

21–11 Electric Dipoles

The combination of two equal charges of opposite sign, $+Q$ and $-Q$, separated by a distance ℓ, is referred to as an **electric dipole**. The quantity $Q\ell$ is called the **dipole moment** and is represented[†] by the symbol p. The dipole moment can be considered to be a vector $\vec{\mathbf{p}}$, of magnitude $Q\ell$, that points from the negative to the positive charge as shown in Fig. 21–43. Many molecules, such as the diatomic molecule CO, have a dipole moment (C has a small positive charge and O a small negative charge of equal magnitude), and are referred to as **polar molecules**.

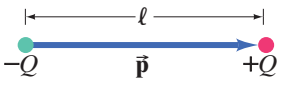


FIGURE 21–43 A dipole consists of equal but opposite charges, $+Q$ and $-Q$, separated by a distance ℓ. The dipole moment is $\vec{\mathbf{p}} = Q\vec{\boldsymbol{\ell}}$ and points from the negative to the positive charge (violet arrow).

[†]Be careful not to confuse this p for dipole moment with p for momentum.

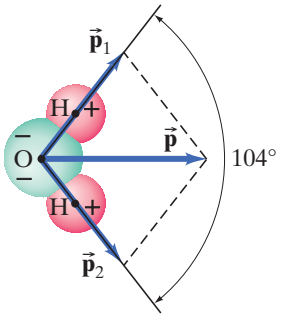
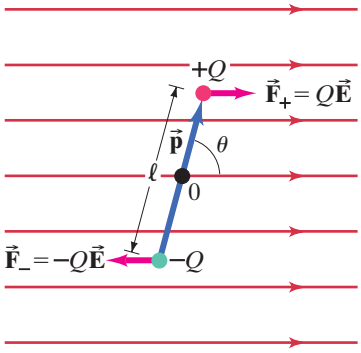


FIGURE 21–44 In the water molecule (H_2O), the electrons spend more time around the oxygen atom than around the two hydrogen atoms. The net dipole moment $\vec{\mathbf{p}}$ can be considered as the vector sum of two dipole moments $\vec{\mathbf{p}}_1$ and $\vec{\mathbf{p}}_2$ that point from the O toward each H as shown: $\vec{\mathbf{p}} = \vec{\mathbf{p}}_1 + \vec{\mathbf{p}}_2$.

FIGURE 21–45 An electric dipole in a uniform electric field.



Even though a molecule as a whole may be neutral, there can be a separation of charge that results from an uneven sharing of electrons by the two atoms.[†] (Symmetric diatomic molecules, like O_2, have no dipole moment.) The water molecule, with its uneven sharing of electrons (O is negative, the two H are positive), also has a dipole moment—see Fig. 21–44.

Dipole in an External Field

First let us consider a dipole, of dipole moment $p = Q\ell$, that is placed in a uniform electric field $\vec{\mathbf{E}}$, as shown in Fig. 21–45. If the field is uniform, the force $Q\vec{\mathbf{E}}$ on the positive charge and the force $-Q\vec{\mathbf{E}}$ on the negative charge result in no net force on the dipole. There will, however, be a *torque* on the dipole (Fig. 21–45) which has magnitude (calculated about the center, 0, of the dipole)

$$\tau = QE\frac{\ell}{2}\sin\theta + QE\frac{\ell}{2}\sin\theta = pE\sin\theta. \qquad \textbf{(21–9a)}$$

This can be written in vector notation as

$$\vec{\boldsymbol{\tau}} = \vec{\mathbf{p}} \times \vec{\mathbf{E}}. \qquad \textbf{(21–9b)}$$

The effect of the torque is to try to turn the dipole so $\vec{\mathbf{p}}$ is parallel to $\vec{\mathbf{E}}$. The work done on the dipole by the electric field to change the angle θ from θ_1 to θ_2 is (see Eq. 10–22)

$$W = \int_{\theta_1}^{\theta_2} \tau\, d\theta.$$

We need to write the torque as $\tau = -pE\sin\theta$ because its direction is opposite to the direction of increasing θ (right-hand rule). Then

$$W = \int_{\theta_1}^{\theta_2} \tau\, d\theta = -pE\int_{\theta_1}^{\theta_2} \sin\theta\, d\theta = pE\cos\theta\Big|_{\theta_1}^{\theta_2} = pE(\cos\theta_2 - \cos\theta_1).$$

Positive work done by the field decreases the potential energy, U, of the dipole in this field. (Recall the relation between work and potential energy, Eq. 8–4, $\Delta U = -W$.) If we choose $U = 0$ when $\vec{\mathbf{p}}$ is perpendicular to $\vec{\mathbf{E}}$ (that is, choosing $\theta_1 = 90°$ so $\cos\theta_1 = 0$), and setting $\theta_2 = \theta$, then

$$U = -W = -pE\cos\theta = -\vec{\mathbf{p}}\cdot\vec{\mathbf{E}}. \qquad \textbf{(21–10)}$$

If the electric field is *not* uniform, the force on the $+Q$ of the dipole may not have the same magnitude as on the $-Q$, so there may be a net force as well as a torque.

EXAMPLE 21–18 Dipole in a field. The dipole moment of a water molecule is $6.1 \times 10^{-30}\ \text{C}\cdot\text{m}$. A water molecule is placed in a uniform electric field with magnitude $2.0 \times 10^5\ \text{N/C}$. (*a*) What is the magnitude of the maximum torque that the field can exert on the molecule? (*b*) What is the potential energy when the torque is at its maximum? (*c*) In what position will the potential energy take on its greatest value? Why is this different than the position where the torque is maximum?

APPROACH The torque is given by Eq. 21–9 and the potential energy by Eq. 21–10.

SOLUTION (*a*) From Eq. 21–9 we see that τ is maximized when θ is 90°. Then $\tau = pE = (6.1 \times 10^{-30}\ \text{C}\cdot\text{m})(2.0 \times 10^5\ \text{N/C}) = 1.2 \times 10^{-24}\ \text{m}\cdot\text{N}$.

(*b*) The potential energy for $\theta = 90°$ is zero (Eq. 21–10). Note that the potential energy is negative for smaller values of θ, so U is not a minimum for $\theta = 90°$.

(*c*) The potential energy U will be a maximum when $\cos\theta = -1$ in Eq. 21–10, so $\theta = 180°$, meaning $\vec{\mathbf{E}}$ and $\vec{\mathbf{p}}$ are antiparallel. The potential energy is maximized when the dipole is oriented so that it has to rotate through the largest angle, 180°, to reach the equilibrium position at $\theta = 0°$. The torque, on the other hand, is maximized when the electric forces are perpendicular to $\vec{\mathbf{p}}$.

[†]The value of the separated charges may be a fraction of e (say $\pm 0.2e$ or $\pm 0.4e$) but note that such charges do not violate what we said about e being the smallest charge. These charges less than e cannot be isolated and merely represent how much time electrons spend around one atom or the other. The charge on any normal (macroscopic) object is always an integer times $+e$ or $-e$.

Electric Field Produced by a Dipole

We have just seen how an external electric field affects an electric dipole. Now let us suppose that there is no external field, and we want to determine the electric field produced *by* the dipole. For brevity, we restrict ourselves to points that are on the perpendicular bisector of the dipole, such as point P in Fig. 21–46 which is a distance r above the midpoint of the dipole. Note that r in Fig. 21–46 is not the distance from either charge to point P; the latter distance is $(r^2 + \ell^2/4)^{\frac{1}{2}}$ and this is what must be used in Eq. 21–4. The total field at P is

$$\vec{\mathbf{E}} = \vec{\mathbf{E}}_+ + \vec{\mathbf{E}}_-,$$

where $\vec{\mathbf{E}}_+$ and $\vec{\mathbf{E}}_-$ are the fields due to the + and − charges respectively. The magnitudes E_+ and E_- are equal:

$$E_+ = E_- = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2 + \ell^2/4}.$$

Their y components cancel at point P (*symmetry* again), so the magnitude of the total field $\vec{\mathbf{E}}$ is

$$E = 2E_+\cos\phi = \frac{1}{2\pi\epsilon_0}\left(\frac{Q}{r^2 + \ell^2/4}\right)\frac{\ell}{2(r^2 + \ell^2/4)^{\frac{1}{2}}}$$

or, setting $Q\ell = p$,

$$E = \frac{1}{4\pi\epsilon_0}\frac{p}{(r^2 + \ell^2/4)^{\frac{3}{2}}}. \quad \left[\begin{array}{c}\text{on perpendicular bisector}\\ \text{of dipole}\end{array}\right] \quad \textbf{(21–11)}$$

Far from the dipole, $r \gg \ell$, this reduces to

$$E = \frac{1}{4\pi\epsilon_0}\frac{p}{r^3}. \quad \left[\begin{array}{c}\text{on perpendicular bisector}\\ \text{of dipole; } r \gg \ell\end{array}\right] \quad \textbf{(21–12)}$$

So the field decreases more rapidly for a dipole than for a single point charge $(1/r^3$ versus $1/r^2)$, which we expect since at large distances the two opposite charges appear so close together as to neutralize each other. This $1/r^3$ dependence also applies for points not on the perpendicular bisector (see Problem 64).

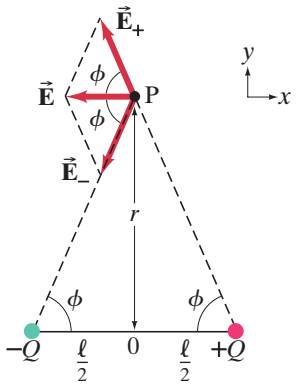


FIGURE 21–46 Electric field due to an electric dipole.

*21–12 Electric Forces in Molecular Biology: DNA Structure and Replication

The study of the structure and functioning of a living cell at the molecular level is known as *molecular biology*. It is an important area for application of physics. The interior of every biological cell is mainly water. We can imagine a cell as a thick soup of molecules continually in motion (kinetic theory, Chapter 18), colliding with one another with various amounts of kinetic energy. These molecules interact with one another because of the *electrostatic force* between molecules.

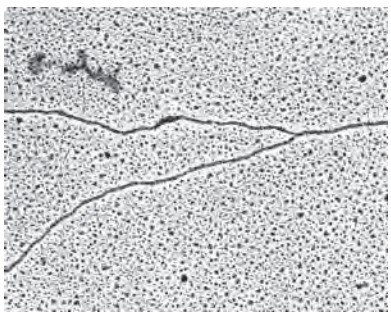
Indeed, cellular processes are now considered to be the result of *random ("thermal") molecular motion plus the ordering effect of the electrostatic force*. As an example, we look at DNA structure and replication. The picture we present is a model of what happens based on physical theories and experiment.

The genetic information that is passed on from generation to generation in all living cells is contained in the chromosomes, which are made up of genes. Each gene contains the information needed to produce a particular type of protein molecule, and that information is built into the principal molecule of a chromosome, DNA (deoxyribonucleic acid), Fig. 21–47. DNA molecules are made up of many small molecules known as nucleotide bases which are each *polar* (Section 21–2) due to unequal sharing of electrons. There are four types of nucleotide bases in DNA: adenine (A), cytosine (C), guanine (G), and thymine (T).

PHYSICS APPLIED

Inside a cell: kinetic theory plus electrostatic force

FIGURE 21–47 Image of DNA replicating, made by a transmission electron microscope.



PHYSICS APPLIED
DNA structure

The DNA of a chromosome generally consists of two long DNA strands wrapped about one another in the shape of a "double helix." The genetic information is contained in the specific order of the four bases (A, C, G, T) along each strand. As shown in Fig. 21–48, the two strands are attracted by electrostatic forces—that is, by the attraction of positive charges to negative charges that exist on parts of the molecules. We see in Fig. 21–48a that an A (adenine) on one strand is always opposite a T (thymine) on the other strand; similarly, a G (guanine) is always opposite a C (cytosine). This important ordering effect occurs because the shapes of A, T, C, and G are such that a T fits closely only into an A, and a G into a C. Only in the case of this close proximity of the charged portions can the electrostatic force (recall Coulomb's law, $F \propto 1/r^2$) be great enough to hold them together even for a short time (Fig. 21–48b), forming what are referred to as "weak bonds."

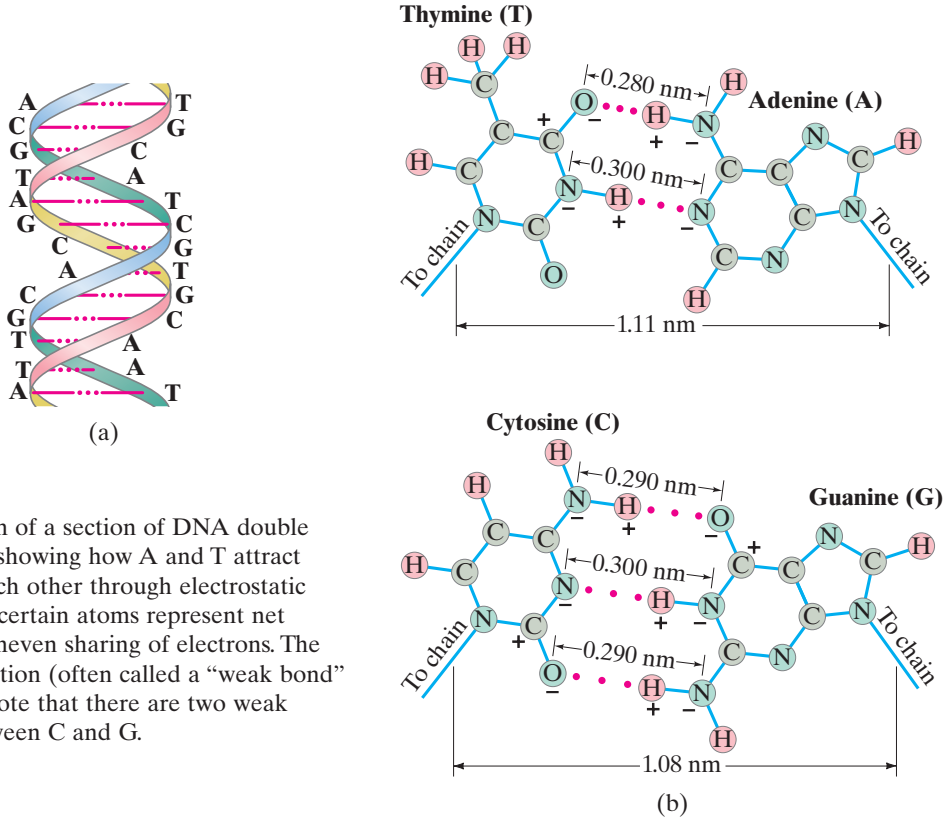


FIGURE 21–48 (a) Schematic diagram of a section of DNA double helix. (b) "Close-up" view of the helix, showing how A and T attract each other and how G and C attract each other through electrostatic forces. The + and − signs indicated on certain atoms represent net charges, usually a fraction of e due to uneven sharing of electrons. The red dots indicate the electrostatic attraction (often called a "weak bond" or "hydrogen bond"—Section 40–3). Note that there are two weak bonds between A and T, and three between C and G.

The electrostatic force between A and T, and between C and G, exists because these molecules have charged parts. These charges are due to some electrons in each of these molecules spending more time orbiting one atom than another. For example, the electron normally on the H atom of adenine (upper part of Fig. 21–48b) spends some of its time orbiting the adjacent N atom (more on this in Chapter 40), so the N has a net negative charge and the H is left with a net positive charge. This H^+ atom of adenine is then attracted to the O^- atom of thymine. These net + and − charges usually have magnitudes of a fraction of e (charge on the electron) such as $0.2e$ or $0.4e$. (This is what we mean by "polar" molecules.)

[When H^+ is involved, the weak bond it can make with a nearby negative charge, such as O^-, is relatively strong (partly because H^+ is so small) and is referred to as a **hydrogen bond** (Section 40–3).]

PHYSICS APPLIED
DNA replication

When the DNA replicates (duplicates) itself just before cell division, the arrangement of A opposite T and G opposite C is crucial for ensuring that the genetic information is passed on accurately to the next generation, Fig. 21–49. The two strands of DNA separate (with the help of enzymes, which also operate via the electrostatic force), leaving the charged parts of the bases exposed.

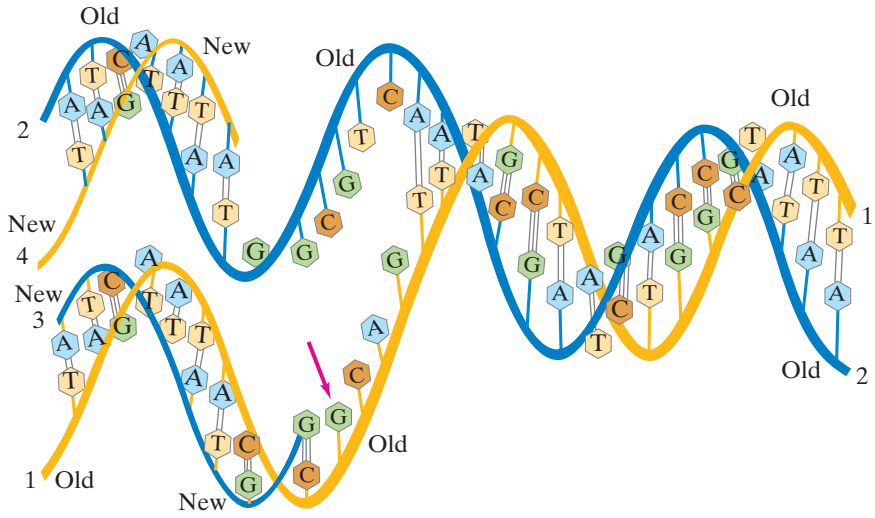


FIGURE 21–49 Replication of DNA.

Once replication starts, let us see how the correct order of bases occurs by looking at the G molecule indicated by the red arrow in Fig. 21–49. Many unattached nucleotide bases of all four kinds are bouncing around in the cellular fluid, and the only type that will experience enough attraction to our G, if it comes close to it, will be a C. The charges on the other three bases can not get close enough to those on the G to provide a significant attractive force—remember that the electrostatic (Coulomb) force decreases rapidly with distance $(\propto 1/r^2)$. Because the G does not attract an A, T, or G appreciably, an A, T, or G will be knocked away by collisions with other molecules in the cellular fluid before enzymes can attach it to the growing chain (number 3 in Fig. 21–49). But the electrostatic force will often hold a C opposite our G long enough so that an enzyme can attach the C to the growing end of the new chain. Thus electrostatic forces are responsible for selecting the bases in the proper order during replication. Note in Fig. 21–49 that the new number 4 strand has the same order of bases as the old number 1 strand; and the new number 3 strand is the same as the old number 2. So the two new double helixes, 1–3 and 2–4, are identical to the original 1–2 helix. This is how the genetic information is passed on accurately to the next generation. Rarely, errors are made, possibly causing a *mutation* (a noticeable change in the expression of that gene).

This process of DNA replication is often presented as if it occurred in clockwork fashion—as if each molecule knew its role and went to its assigned place. But this is not the case. The forces of attraction are rather weak and become significant only when charged parts of two molecules have "complementary shapes," meaning they can get close enough so that the electrostatic force $(\propto 1/r^2)$ is strong enough to form weak bonds. If the molecular shapes are not just right, there is almost no electrostatic attraction, which is why there are so few mistakes. Thus, out of the random motion of the molecules, the electrostatic force acts to bring order out of chaos.

The random (thermal) velocities of molecules in a cell affect *cloning*. When a bacterial cell divides, the two new bacteria have nearly identical DNA. Even if the DNA were perfectly identical, the two new bacteria would not end up behaving in exactly the same way. Long protein, DNA, and RNA molecules get bumped into different shapes, and even the expression[†] of genes can thus be different. Loosely held parts of large molecules such as a methyl group (CH_3) can also be knocked off by a strong collision with another molecule. Hence, cloned organisms are not identical, even when their DNA is identical. Indeed, there can not really be genetic determinism.

[†]The separate genes of a DNA double helix can be covered by protein molecules, keeping those genes from being "expressed"—that is, translated into the proteins they code for (see Section 40–3).

Summary

There are two kinds of **electric charge**, positive and negative. These designations are to be taken algebraically—that is, any charge is plus or minus so many coulombs (C), in SI units.

Electric charge is **conserved**: if a certain amount of one type of charge is produced in a process, an equal amount of the opposite type is also produced; thus the *net* charge produced is zero.

According to atomic theory, electricity originates in the atom, which consists of a positively charged nucleus surrounded by negatively charged electrons. Each electron has a charge $-e = -1.6 \times 10^{-19}$ C.

Electric **conductors** are those materials in which many electrons are relatively free to move, whereas electric **insulators** or **nonconductors** are those in which very few electrons are free to move.

An object is negatively charged when it has an excess of electrons, and positively charged when it has less than its normal number of electrons. The charge on any object is a whole number times $+e$ or $-e$. That is, charge is **quantized**.

An object can become charged by rubbing (in which electrons are transferred from one material to another), **by conduction** (which is transfer of charge from one charged object to another by touching), or **by induction** (the separation of charge within an object because of the close approach of another charged object but without touching).

Electric charges exert a force on each other. If two charges are of opposite types, one positive and one negative, they each exert an attractive force on the other. If the two charges are the same type, each repels the other.

The magnitude of the force one point charge exerts on another is proportional to the product of their charges, and inversely proportional to the square of the distance between them:

$$F = k\frac{Q_1 Q_2}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_2}{r^2}; \qquad \textbf{(21–1, 21–2)}$$

this is **Coulomb's law**. In SI units, k is often written as $1/4\pi\epsilon_0$.

We think of an **electric field** as existing in space around any charge or group of charges. The force on another charged object is then said to be due to the electric field present at that charge's location.

The *electric field*, $\vec{\mathbf{E}}$, at any point in space due to one or more charges, is defined as the force per unit charge that would act on a tiny positive test charge q placed at that point:

$$\vec{\mathbf{E}} = \frac{\vec{\mathbf{F}}}{q}. \qquad \textbf{(21–3)}$$

The magnitude of the electric field a distance r from a point charge Q is

$$E = k\frac{Q}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_2}{r^2} \qquad \textbf{(21–4)}$$

The total electric field at a point in space is equal to the vector sum of the individual fields due to each contributing charge. This is the **principle of superposition**.

Electric fields are represented by **electric field lines** that start on positive charges and end on negative charges. Their direction indicates the direction the force would be on a tiny positive test charge placed at each point. The lines can be drawn so that the number per unit area is proportional to the magnitude of E.

The static electric field inside a conductor is zero, and the electric field lines just outside a charged conductor are perpendicular to its surface.

An **electric dipole** is a combination of two equal but opposite charges, $+Q$ and $-Q$, separated by a distance ℓ. The **dipole moment** is $p = Q\ell$. A dipole placed in a uniform electric field feels no net force but does feel a net torque (unless $\vec{\mathbf{p}}$ is parallel to $\vec{\mathbf{E}}$). The electric field produced by a dipole decreases as the third power of the distance r from the dipole $(E \propto 1/r^3)$ for r large compared to ℓ.

[*In the replication of DNA, the electrostatic force plays a crucial role in selecting the proper molecules so that the genetic information is passed on accurately from generation to generation.]

Questions

1. If you charge a pocket comb by rubbing it with a silk scarf, how can you determine if the comb is positively or negatively charged?
2. Why does a shirt or blouse taken from a clothes dryer sometimes cling to your body?
3. Explain why fog or rain droplets tend to form around ions or electrons in the air.
4. A positively charged rod is brought close to a neutral tiny piece of paper, which it attracts. Draw a diagram showing the separation of charge in the paper, and explain why attraction occurs.
5. Why does a plastic ruler that has been rubbed with a cloth have the ability to pick up small pieces of paper? Why is this difficult to do on a humid day?
6. Contrast the *net charge* on a conductor to the "free charges" in the conductor.
7. Figures 21–7 and 21–8 show how a charged rod placed near an uncharged metal object can attract (or repel) electrons. There are a great many electrons in the metal, yet only some of them move as shown. Why not all of them?
8. When an electroscope is charged, the two leaves repel each other and remain at an angle. What balances the electric force of repulsion so that the leaves don't separate further?

9. The balloon in Fig. 21–50 was rubbed on a student's hair. Explain why the water drip curves instead of falling vertically.



FIGURE 21–50 Question 9.

10. What happens to the separation of the leaves of an electroscope (Fig. 21–10) when the charging object is removed from an electroscope (*a*) charged by induction and (*b*) charged by conduction? (*c*) Is it possible to tell whether the electroscope in Fig. 21–12a has been charged by induction or by conduction? If so, which way was it charged?

11. The form of Coulomb's law is very similar to that for Newton's law of universal gravitation. What are the differences between these two laws? Compare also gravitational mass and electric charge.

12. We are not normally aware of the gravitational or electric force between two ordinary objects. What is the reason in each case? Give an example where we are aware of each one and why.

13. What experimental observations, mentioned in the text, rule out the possibility that the numerator in Coulomb's law contains the sum $(Q_1 + Q_2)$ rather than the product $Q_1 Q_2$?

14. When a charged ruler attracts small pieces of paper, sometimes a piece jumps quickly away after touching the ruler. Explain.

15. Explain why the test charges we use when measuring electric fields must be small.

16. When determining an electric field, must we use a *positive* test charge, or would a negative one do as well? Explain.

17. Draw the electric field lines surrounding two negative electric charges a distance ℓ apart.

18. Assume that the two opposite charges in Fig. 21–35a are 12.0 cm apart. Consider the magnitude of the electric field 2.5 cm from the positive charge. On which side of this charge—top, bottom, left, or right—is the electric field the strongest? The weakest?

19. Consider the electric field at the three points indicated by the letters A, B, and C in Fig. 21–51. First draw an arrow at each point indicating the direction of the net force that a positive test charge would experience if placed at that point, then list the letters in order of *decreasing* field strength (strongest first). Explain.

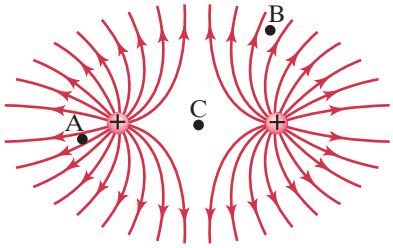


FIGURE 21–51 Question 19.

20. Why can electric field lines never cross?

21. Show, using the three properties of field lines given in Section 21–8, page 626, that the electric field lines starting or ending on a single point charge must be symmetrically spaced around the charge.

22. Given two point charges, of magnitude Q and $2Q$, a distance ℓ apart, is there a point along the straight line that passes through them where $E = 0$? Give answers for when their signs are (*a*) opposite, (*b*) the same. If yes, state roughly where this point will be.

23. Suppose the ring of Fig. 21–29 has a uniformly distributed negative charge Q. What are the magnitude and direction of $\vec{\mathbf{E}}$ at point P?

24. Consider a small positive test charge located on an electric field line at some point, such as point P in Fig. 21–35a. Is the direction of the velocity and/or acceleration of the test charge along this line? Discuss.

25. We wish to determine the electric field at a point P near a positively charged metal sphere (a good conductor). We do so by bringing a small test charge, q_0, to this point P and measure the force F_0 on it. Will F_0/q_0 at that point be greater than, less than, or equal to the electric field $\vec{\mathbf{E}}$ compared to before the test charge was present? Explain.

26. In what ways does the electron motion in Example 21–17 resemble projectile motion (Section 3–7)? In which ways does it not?

27. Explain why there can be a net force on an electric dipole placed in a nonuniform electric field.

28. Describe the motion of the dipole shown in Fig. 21–45 if it is released from rest at the position shown. Assume (*a*) uniform $\vec{\mathbf{E}}$, (*b*) $\vec{\mathbf{E}}$ increasing in magnitude going downward.

MisConceptual Questions

1. $Q_1 = -0.10\,\mu\text{C}$ is located at the origin. $Q_2 = +0.10\,\mu\text{C}$ is located on the positive x axis at $x = 1.0\,\text{m}$. Which of the following is true of the force on Q_1 due to Q_2?
(*a*) It is attractive and directed in the $+x$ direction.
(*b*) It is attractive and directed in the $-x$ direction.
(*c*) It is repulsive and directed in the $+x$ direction.
(*d*) It is repulsive and directed in the $-x$ direction.

2. Swap the positions of Q_1 and Q_2 of MisConceptual Question 1. Which of the following is true of the force on Q_1 due to Q_2?
(*a*) It does not change.
(*b*) It changes from attractive to repulsive.
(*c*) It changes from repulsive to attractive.
(*d*) It changes from the $+x$ direction to the $-x$ direction.
(*e*) It changes from the $-x$ direction to the $+x$ direction.

3. Fred the lightning bug has a mass m and a charge $+q$. Jane, his lightning-bug wife, has a mass of $\frac{3}{4}m$ and a charge $-2q$. Because they have charges of opposite sign, they are attracted to each other. Which is attracted more to the other, and by how much?
(*a*) Fred, twice as much.
(*b*) Jane, twice as much.
(*c*) Fred, four times as much.
(*d*) Jane, four times as much.
(*e*) They are attracted to each other by the same amount.

4. Figure 21–52 shows electric field lines due to a point charge. What can you say about the field at point 1 compared with the field at point 2?
(*a*) The field at point 2 is larger, because point 2 is on a field line.
(*b*) The field at point 1 is larger, because point 1 is not on a field line.
(*c*) The field at point 1 is zero, because point 1 is not on a field line.
(*d*) The field at point 1 is larger, because the field lines are closer together in that region.

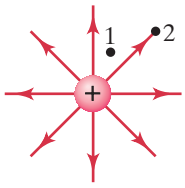


FIGURE 21–52
MisConceptual Question 4.

5. A negative point charge is in an electric field created by a positive point charge. Which of the following is true?
(*a*) The field points away from the positive charge, and the force on the negative charge is in the same direction as the field.
(*b*) The field points away from the positive charge, and the force on the negative charge is in the opposite direction to the field.
(*c*) The field points toward the positive charge, and the force on the negative charge is in the same direction as the field.
(*d*) The field points toward the positive charge, and the force on the negative charge is in the opposite direction to the field.

6. A small object with charge $+Q$ is placed a distance d from another small object with charge $+4Q$. Is it possible to place a third charge on the line between them such that the net force on the third charge will be zero?
(*a*) Yes, but only if the third charge is positive.
(*b*) Yes, but only if the third charge is negative.
(*c*) Yes, independent of the sign of the third charge.
(*d*) No, it is not possible.

7. Refer to Fig. 21–35d. If the two charged plates were moved until they are half the distance shown without changing the charge on the plates, the electric field near the center of the plates would
(*a*) remain almost exactly the same.
(*b*) increase by a factor of 2.
(*c*) increase, but not by a factor of 2.
(*d*) decrease by a factor of 2.
(*e*) decrease, but not by a factor of 2.

8. A proton and an electron are fixed at rest 1.0 m apart, and then let go. Where do they meet?
(*a*) Halfway between their original positions.
(*b*) Closer to the electron's original position.
(*c*) Closer to the proton's original position.

9. To be safe during a lightning storm, it is best to be
(*a*) in the middle of a grassy meadow.
(*b*) inside a metal car.
(*c*) next to a tall tree in a forest.
(*d*) inside a wooden building.
(*e*) on a metal observation tower.

10. Which are the worst places in MisConceptual Question 9?

11. Which vector best represents the direction of the electric field at the fourth corner of the square due to the three charges shown in Fig. 21–53?

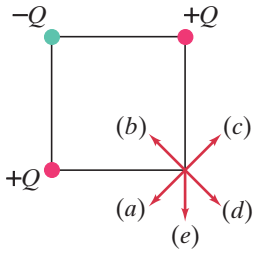


FIGURE 21–53
MisConceptual Question 11.

12. A small metal ball hangs from the ceiling by an insulating thread. The ball is attracted to a positively charged rod held near the ball. The charge of the ball must be
(*a*) positive.
(*b*) negative.
(*c*) neutral.
(*d*) positive or neutral.
(*e*) negative or neutral.

Problems

21–5 Coulomb's Law

$[1\,\text{mC} = 10^{-3}\,\text{C},\ 1\,\mu\text{C} = 10^{-6}\,\text{C},\ 1\text{nC} = 10^{-9}\,\text{C},\ 1\text{pC} = 10^{-12}\,\text{C}.]$

1. (I) What is the magnitude of the electric force of attraction between an iron nucleus $(q = +26e)$ and its innermost electron if the distance between them is $1.5 \times 10^{-12}\,\text{m}$?

2. (I) How many electrons make up a charge of $-58.0\,\mu\text{C}$?

3. (I) What is the magnitude of the force a $+25\,\mu\text{C}$ charge exerts on a $+2.5\,\text{mC}$ charge 22 cm away?

4. (I) What is the repulsive electrical force between two protons $3.5 \times 10^{-15}\,\text{m}$ apart from each other in an atomic nucleus?

5. (II) When an object such as a plastic comb is charged by rubbing it with a cloth, the net charge is typically a few microcoulombs. If that charge is $2.4\,\mu\text{C}$, by what percentage does the mass of a 9.0-g comb change during charging?

6. (II) Two charged dust particles exert a force of $4.6 \times 10^{-2}\,\text{N}$ on each other. What will be the force if they are moved so they are only one-eighth as far apart?

7. (II) Two small charged spheres are 6.22 cm apart. They are moved, and the force each exerts on the other is found to have tripled. How far apart are they now?

8. (II) A person scuffing her feet on a wool rug on a dry day accumulates a net charge of $-34\,\mu C$. How many excess electrons does she get, and by how much does her mass increase?

9. (II) What is the total charge of all the electrons in a 12-kg bar of gold? What is the net charge of the bar? (Gold has 79 electrons per atom and an atomic mass of 197 u.)

10. (II) Compare the electric force holding the electron in orbit $(r = 0.53 \times 10^{-10}\,\text{m})$ around the proton nucleus of the hydrogen atom, with the gravitational force between the same electron and proton. Give the ratio of these two forces.

11. (II) Two positive point charges are a fixed distance apart. The sum of their charges is Q_T. What charge must each have in order to (*a*) maximize the electric force between them, and (*b*) minimize it?

12. (II) Particles of charge $+82$, $+48$, and $-95\,\mu C$ are placed in a line (Fig. 21–54). The center one is 0.35 m from each of the others. Calculate the net force on each charge due to the other two.

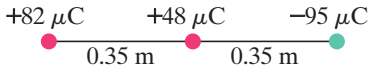


FIGURE 21–54 Problem 12.

13. (II) Three positive particles of equal charge, $+17.0\,\mu C$, are located at the corners of an equilateral triangle of side 15.0 cm (Fig. 21–55). Calculate the magnitude and direction of the net force on each particle due to the other two.

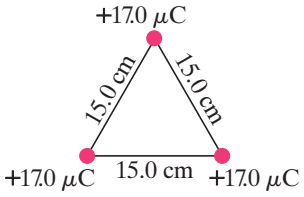


FIGURE 21–55
Problem 13.

14. (II) Two small nonconducting spheres have a total charge of $82.0\,\mu C$. (*a*) When placed 28.0 cm apart, the force each exerts on the other is 12.0 N and is repulsive. What is the charge on each? (*b*) What if the force is attractive?

15. (II) A charge of 6.45 mC is placed at each corner of a square 0.100 m on a side. Determine the magnitude and direction of the force on each charge.

16. (II) Two negative and two positive point charges (magnitude $Q = 4.15$ mC) are placed on opposite corners of a square as shown in Fig. 21–56. Determine the magnitude and direction of the force on each charge.

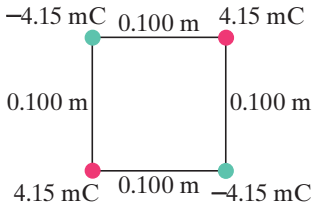


FIGURE 21–56
Problem 16.

17. (II) At each corner of a square of side ℓ there are point charges of magnitude Q, $2Q$, $3Q$, and $4Q$ (Fig. 21–57). Determine the magnitude and direction of the force on the charge $2Q$.

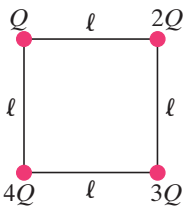


FIGURE 21–57
Problem 17.

18. (II) A large electroscope is made with "leaves" that are 66-cm-long wires with tiny 21-g spheres at the ends. When charged, nearly all the charge resides on the spheres. If the wires each make a 26° angle with the vertical (Fig. 21–58), what total charge Q must have been applied to the electroscope? Ignore the mass of the wires.

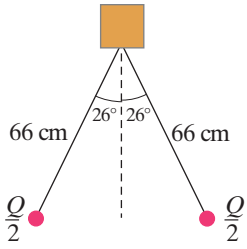


FIGURE 21–58
Problem 18.

19. (II) A charge Q is transferred from an initially uncharged plastic ball to an identical ball 21 cm away. The force of attraction is then 17 mN. How many electrons were transferred from one ball to the other?

20. (III) Two charges, $-2Q$ and $-3Q$, are a distance ℓ apart. These two charges are free to move but do not because there is a third (fixed) charge nearby. What must be the magnitude of the third charge and its placement in order for the first two to be in equilibrium?

21. (III) Two positive charges $+Q$ are affixed rigidly to the x axis, one at $x = +d$ and the other at $x = -d$. A third charge $+q$ of mass m, which is constrained to move only along the x axis, is displaced from the origin by a small distance $s \ll d$ and then released from rest. (*a*) Show that (to a good approximation) $+q$ will execute simple harmonic motion and determine an expression for its oscillation period T. (*b*) If these three charges are each singly ionized sodium atoms $(q = Q = +e)$ at the equilibrium spacing $d = 3 \times 10^{-10}\,\text{m}$ typical of the atomic spacing in a solid, find T in picoseconds.

21–6 to 21–8 Electric Field, Field Lines

22. (I) Determine the magnitude and direction of the electric force on an electron in a uniform electric field of strength 2260 N/C that points due east.

23. (I) A proton is released in a uniform electric field, and it experiences an electric force of $1.68 \times 10^{-14}\,\text{N}$ toward the south. Find the magnitude and direction of the electric field.

24. (I) Determine the magnitude and direction of the electric field 27.4 cm directly above an isolated $33.0 \times 10^{-6}\,\text{C}$ charge.

25. (I) The electric force on a $+3.70$-μC charge is $\vec{\mathbf{F}} = (7.22 \times 10^{-4}\,\text{N})\,\hat{\mathbf{j}}$. What is the electric field at the position of the charge?

26. (II) What is the electric field at a point when the force on a -1.25-μC charge placed at that point is $\vec{\mathbf{F}} = (4.0\,\hat{\mathbf{i}} - 5.2\,\hat{\mathbf{j}}) \times 10^{-3}\,\text{N}$?

27. (II) Determine the magnitude of the acceleration experienced by an electron in an electric field of 756 N/C. How does the direction of the acceleration depend on the direction of the field at that point?

28. (II) Determine the magnitude and direction of the electric field at a point midway between a -9.2 pC and a $+5.8$ pC charge 6.0 cm apart. Assume no other charges are nearby.

29. (II) What is the electric field strength at a point in space where a proton experiences an acceleration of 2.4 million "g's"?

30. (II) A very long uniformly charged wire (linear charge density $\lambda = 2.5\,\text{C/m}$) lies along the x axis in Fig. 21–59. A small charged sphere ($Q = -2.0\,\text{C}$) is at the point $x = 0\,\text{cm}$, $y = -5.0\,\text{cm}$. What is the electric field at the point $x = 7.0\,\text{cm}$, $y = 7.0\,\text{cm}$? $\vec{\mathbf{E}}_{\text{wire}}$ and $\vec{\mathbf{E}}_{\text{Q}}$ represent fields due to the long wire and the charge Q, respectively.

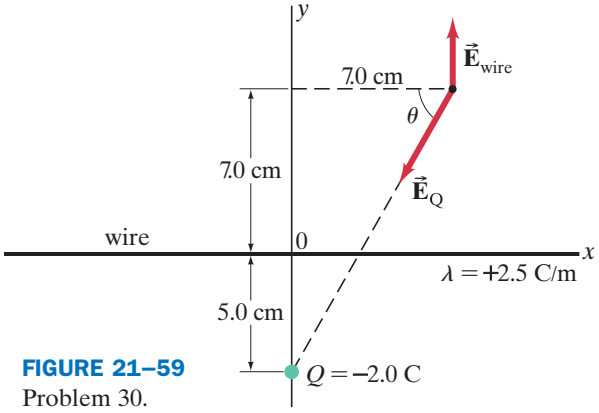


FIGURE 21–59 Problem 30.

31. (II) The electric field midway between two equal but opposite point charges is 326 N/C, and the distance between the charges is 16.0 cm. What is the magnitude of the charge on each?

32. (II) Calculate the electric field at one corner of a square 1.22 m on a side if the other three corners are occupied by $3.25 \times 10^{-6}\,\text{C}$ charges.

33. (II) Calculate the electric field at the center of a square 42.5 cm on a side if one corner is occupied by a $-33.8\,\mu\text{C}$ charge and the other three are occupied by $-22.0\,\mu\text{C}$ charges.

34. (II) Determine the direction and magnitude of the electric field at the point P in Fig. 21–60. The charges are separated by a distance $2a$, and point P is a distance x from the midpoint between the two charges. Express your answer in terms of Q, x, a, and k.

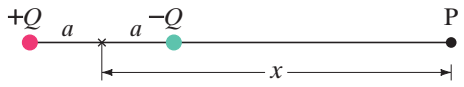


FIGURE 21–60 Problem 34.

35. (II) Two point charges, $Q_1 = -32\,\mu\text{C}$ and $Q_2 = +45\,\mu\text{C}$, are separated by a distance of 12 cm. The electric field at the point P (see Fig. 21–61) is zero. How far from Q_1 is P?

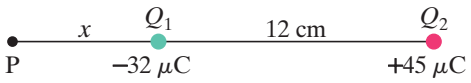


FIGURE 21–61 Problem 35.

36. (II) A very thin line of charge lies along the x axis from $x = -\infty$ to $x = +\infty$. Another similar line of charge lies along the y axis from $y = -\infty$ to $y = +\infty$. Both lines have a uniform charge per length λ. Determine the resulting electric field magnitude and direction (relative to the x axis) at a point (x, y) in the first quadrant of the xy plane.

37. (II) (a) Determine the electric field $\vec{\mathbf{E}}$ at the origin 0 in Fig. 21–62 due to the two charges at A and B. (b) Repeat, but let the charge at B be reversed in sign.

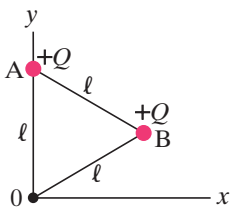


FIGURE 21–62 Problem 37.

38. (II) Draw, approximately, the electric field lines emanating from a uniformly charged straight wire whose length ℓ is not great. The spacing between lines near the wire should be much less than ℓ. [*Hint*: Also consider points very far from the wire up to 4ℓ.]

39. (II) Two parallel circular rings of radius R have their centers on the x axis separated by a distance ℓ as shown in Fig. 21–63. If each ring carries a uniformly distributed charge Q, find the electric field, $\vec{\mathbf{E}}(x)$, at points along the x axis.

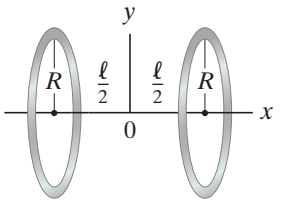


FIGURE 21–63 Problem 39.

40. (II) You are given two unknown point charges, Q_1 and Q_2. At a point on the line joining them, one-third of the way from Q_1 to Q_2, the electric field is zero (Fig. 21–64). What is the ratio Q_1/Q_2?

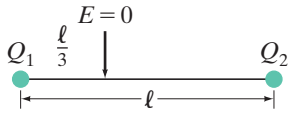


FIGURE 21–64 Problem 40.

41. (II) Use Coulomb's law to determine the magnitude and direction of the electric field at points A and B in Fig. 21–65 due to the two positive charges ($Q = 4.7\,\mu\text{C}$) shown. Are your results consistent with Fig. 21–35b?

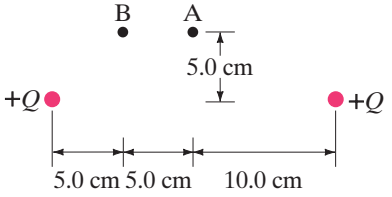


FIGURE 21–65 Problem 41.

42. (II) (a) Two equal charges Q are positioned at points $(x = \ell,\ y = 0)$ and $(x = -\ell,\ y = 0)$. Determine the electric field as a function of y for points along the y axis. (b) Show that the field is a maximum at $y = \pm\ell/\sqrt{2}$.

43. (II) At what position, $x = x_{\text{M}}$, is the magnitude of the electric field along the axis of the ring of Example 21–10 a maximum?

44. (II) Estimate the electric field at a point 2.40 cm perpendicular to the midpoint of a uniformly charged 2.00-m-long thin wire carrying a total charge of 7.45 μC.

45. (II) The uniformly charged straight wire in Fig. 21–30 has the length ℓ, where point 0 is at the midpoint. Show that the field at point P, a perpendicular distance x from 0, is given by

$$E = \frac{\lambda}{2\pi\epsilon_0}\frac{\ell}{x(\ell^2 + 4x^2)^{1/2}},$$

where λ is the charge per unit length.

46. (II) Determine the direction and magnitude of the electric field at the point P shown in Fig. 21–66. The two point charges are separated by a distance of $2a$. Point P is on the perpendicular bisector of the line joining the charges, a distance x from the midpoint between them. Express your answers in terms of Q, x, a, and k.

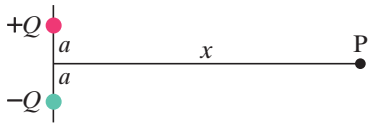


FIGURE 21–66 Problem 46.

47. (III) Use the result given in Problem 45 to find the electric field (magnitude and direction) a distance z above the center of a square loop of wire, each of whose sides has length ℓ and uniform charge per length λ (Fig. 21–67).

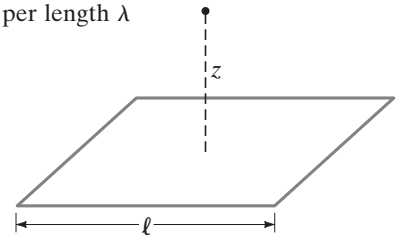


FIGURE 21–67
Problem 47.

48. (III) A thin rod bent into the shape of an arc of a circle of radius R carries a uniform charge per unit length λ. The arc subtends a total angle $2\theta_0$, symmetric about the x axis, as shown in Fig. 21–68. Determine the electric field $\vec{\mathbf{E}}$ at the origin 0.

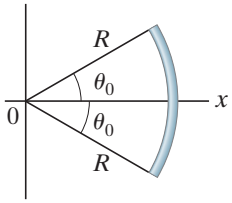


FIGURE 21–68
Problem 48.

49. (III) Suppose a uniformly charged wire has one end at point 0 and rises vertically along the positive y axis to a length ℓ. (*a*) Determine the components of the electric field E_x and E_y at point $(x, 0)$. That is, calculate $\vec{\mathbf{E}}$ near one end of a long wire, in the plane perpendicular to the wire. (*b*) If the wire extends from $y = 0$ to $y = \infty$, so that $\ell = \infty$, show that $\vec{\mathbf{E}}$ makes a 45° angle to the horizontal for any x in the plane where $y = 0$. [*Hint*: See Example 21–12 and Fig. 21–30.]

50. (III) Suppose in Example 21–12 that $x = 0.250\,\text{m}$, $Q = 2.85\,\mu\text{C}$, and that the uniformly charged wire is only 6.50 m long and extends along the y axis from $y = -4.00\,\text{m}$ to $y = +2.50\,\text{m}$. (*a*) Calculate E_x and E_y at point P. (*b*) Determine what the error would be if you simply used the result of Example 21–12, $E = \lambda/2\pi\epsilon_0 x$. Express this error as $(E_x - E)/E$ and E_y/E, and as percent.

51. (III) A thin rod of length ℓ carries a total charge Q distributed uniformly along its length. See Fig. 21–69. Determine the electric field along the axis of the rod starting at one end—that is, find $E(x)$ for $x \geq 0$ in Fig. 21–69.

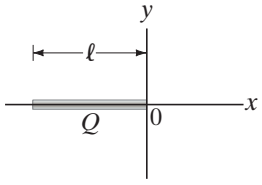


FIGURE 21–69 Problem 51.

52. (III) *Uniform plane of charge*. Charge is distributed uniformly over a large square plane of side ℓ, as shown in Fig. 21–70. The charge per unit area (C/m^2) is σ. Determine the electric field at a point P a distance z above the center of the plane, in the limit $\ell \to \infty$. [*Hint*: Divide the plane into long narrow strips of width dy, and use the result of Example 21–12; then sum the fields due to each strip to get the total field at P.]

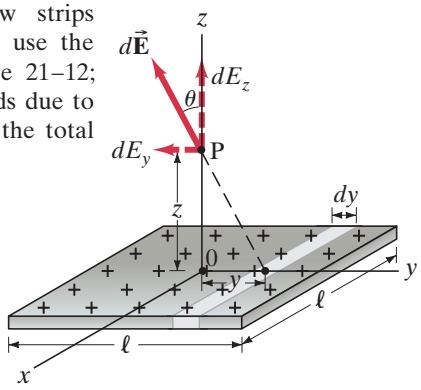


FIGURE 21–70
Problem 52.

21–10 Motion of Charges in an Electric Field

53. (II) An electron with speed $v_0 = 4.16 \times 10^6\,\text{m/s}$ is traveling parallel to a uniform electric field of magnitude $E = 8.65 \times 10^3\,\text{N/C}$. (*a*) How far will the electron travel before it stops? (*b*) How much time will elapse before it returns to its starting point?

54. (II) An electron has an initial velocity $\vec{\mathbf{v}}_0 = (8.0 \times 10^4\,\text{m/s})\,\hat{\mathbf{j}}$. It enters a region where $\vec{\mathbf{E}} = (2.0\hat{\mathbf{i}} + 7.0\hat{\mathbf{j}}) \times 10^4\,\text{N/C}$. (*a*) Determine the vector acceleration of the electron. (*b*) At what angle θ is it moving (relative to its initial direction) at $t = 1.0\,\text{ns}$?

55. (II) An electron moving to the right at $7.5 \times 10^5\,\text{m/s}$ enters a uniform electric field parallel to its direction of motion. If the electron is to be brought to rest in the space of 5.0 cm, (*a*) what direction is required for the electric field, and (*b*) what is the strength of the field?

56. (II) At what angle will the electrons in Example 21–17 leave the uniform electric field at the end of the parallel plates (point P in Fig. 21–42)? Assume the plates are 4.6 cm long, $E = 5.0 \times 10^3\,\text{N/C}$, and $v_0 = 1.00 \times 10^7\,\text{m/s}$. Ignore fringing of the field.

57. (II) An electron is traveling through a uniform electric field. The field is constant and given by $\vec{\mathbf{E}} = (2.00 \times 10^{-11}\,\text{N/C})\,\hat{\mathbf{i}} - (1.20 \times 10^{-11}\,\text{N/C})\,\hat{\mathbf{j}}$. At $t = 0$, the electron is at the origin and traveling in the x direction with a speed of 1.90 m/s. What is its position 3.00 s later?

58. (II) A positive charge q is placed at the center of a circular ring of radius R. The ring carries a uniformly distributed negative charge of total magnitude $-Q$. (*a*) If the charge q is displaced from the center a small distance x as shown in Fig. 21–71, show that it will undergo simple harmonic motion when released. (*b*) If its mass is m, what is its period?

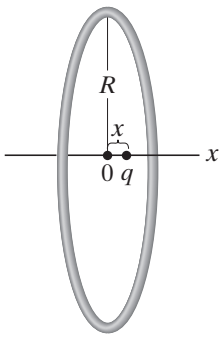


FIGURE 21–71
Problem 58.

21–11 Electric Dipoles

59. (II) A dipole consists of charges $+e$ and $-e$ separated by 0.68 nm. It is in an electric field $E = 2.2 \times 10^4$ N/C. (*a*) What is the value of the dipole moment? (*b*) What is the torque on the dipole when it is perpendicular to the field? (*c*) What is the torque on the dipole when it is at an angle of 45° to the field? (*d*) What is the work required to rotate the dipole from being oriented parallel to the field to being antiparallel to the field?

60. (II) The HCl molecule has a dipole moment of about 3.4×10^{-30} C · m. The two atoms are separated by about 1.0×10^{-10} m. (*a*) What is the net charge on each atom? (*b*) Is this equal to an integral multiple of e? If not, explain. (*c*) What maximum torque would this dipole experience in a 2.5×10^4 N/C electric field? (*d*) How much energy would be needed to rotate one molecule 45° from its equilibrium position of lowest potential energy?

61. (II) An electric dipole, of dipole moment p and moment of inertia I, is placed in a uniform electric field $\vec{\mathbf{E}}$. (*a*) If displaced by an angle θ as shown in Fig. 21–45 and released, under what conditions will it oscillate in simple harmonic motion? (*b*) What will be its frequency?

62. (II) Suppose both charges in Fig. 21–46 (for a dipole) are positive. (*a*) Show that the field on the perpendicular bisector, for $r \gg \ell$, is given by $(1/4\pi\epsilon_0)(2Q/r^2)$. (*b*) Explain why the field decreases as $1/r^2$ here for 2 positive charges, whereas for a dipole it decreases as $1/r^3$.

63. (III) Suppose a dipole $\vec{\mathbf{p}}$ is placed in a nonuniform electric field $\vec{\mathbf{E}} = E\hat{\mathbf{i}}$ that points along the x axis. If $\vec{\mathbf{E}}$ depends only on x, show that the net force on the dipole is

$$\vec{\mathbf{F}} = \left(\vec{\mathbf{p}} \cdot \frac{d\vec{\mathbf{E}}}{dx}\right)\hat{\mathbf{i}},$$

where $d\vec{\mathbf{E}}/dx$ is called the gradient of the field in the x direction.

64. (III) For the dipole shown in Fig. 21–46, let the x axis be horizontal along the line joining the $-Q$ and $+Q$ charges, with origin midway between the two charges. (*a*) Show that the electric field at any point x ($x \gg \ell$) has magnitude

$$E = \frac{1}{4\pi\epsilon_0}\frac{2p}{x^3}.$$

(*b*) In what direction does $\vec{\mathbf{E}}$ point?

*21–12 DNA

***65.** (III) The two strands of the helix-shaped DNA molecule are held together by electrostatic forces as shown in Fig. 21–48. Assume that the net average charge (due to electron sharing) indicated on H and N atoms has magnitude $0.2e$ and on the indicated C and O atoms is $0.4e$. Assume also that atoms on each molecule are separated by 1.0×10^{-10} m. Estimate the net force between (*a*) a thymine and an adenine, and (*b*) a cytosine and a guanine. For each bond (red dots) consider only the three atoms in a line (two atoms on one molecule, one atom on the other). (*c*) Estimate the total force for a DNA molecule containing 10^5 pairs of such molecules. Assume half are A–T pairs and half are C–G pairs.

General Problems

66. How close must two electrons be if the magnitude of the electric force between them is equal to the weight of either at the Earth's surface?

67. Given that the human body is mostly made of water, estimate the total amount of positive charge in a 72-kg person.

68. A water droplet of radius 0.018 mm remains stationary in the air. If the downward-directed electric field of the Earth is 150 N/C, how many excess electron charges must the water droplet have?

69. Estimate the net force between the CO group and the HN group shown in Fig. 21–72. The C and O have charges $\pm 0.40e$, and the H and N have charges $\pm 0.20e$, where $e = 1.6 \times 10^{-19}$ C. [*Hint*: Do not include the "internal" forces between C and O, or between H and N.]

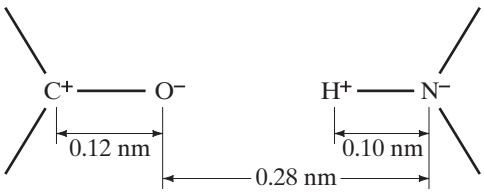


FIGURE 21–72 Problem 69.

70. Suppose that electrical attraction, rather than gravity, was responsible for holding the Moon in orbit around the Earth. If equal and opposite charges Q were placed on the Earth and the Moon, what should be the value of Q to maintain the present orbit? Use data given on the inside front cover of this book. Treat the Earth and Moon as point particles.

71. In a simple model of the hydrogen atom, the electron revolves in a circular orbit around the proton with a speed of 2.2×10^6 m/s. Determine the radius of the electron's orbit. [*Hint*: See Chapter 5 on circular motion.]

72. A positive point charge $Q_1 = 2.5 \times 10^{-5}$ C is fixed at the origin of coordinates, and a negative point charge $Q_2 = -5.0 \times 10^{-6}$ C is fixed to the x axis at $x = +3.2$ m. Find the location of the place(s) along the x axis where the electric field due to these two charges is zero.

73. When clothes are removed from a dryer, a 40-g sock is stuck to a sweater, by "static electricity". Estimate the minimum attractive force between the sock and the sweater. Then estimate the minimum charge on the sock and the sweater. Assume the charging came entirely from the sock rubbing against the sweater so that they have equal and opposite charges, and approximate the sweater as a flat sheet of uniform charge.

74. A small lead sphere carrying an electric charge $Q = -3.00 \times 10^{-6}$ C is suspended vertically from an ideal spring (spring constant $k = 115$ N/m) as in Fig. 21–73. The total mass of the sphere is 0.650 kg, and its center lies 15.0 cm above a tabletop when in equilibrium. The sphere is pulled down 5.00 cm below equilibrium, and then it is released. Using what you know about harmonic oscillation, write an expression for the electric field strength as a function of time that would be measured at the point on the tabletop (P) directly below the sphere.

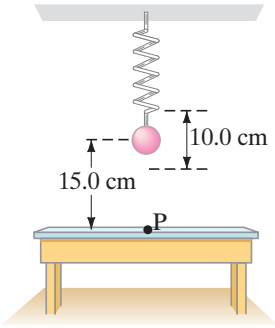


FIGURE 21–73
Problem 74.

75. Dry air will break down and generate a spark if the electric field exceeds about 3×10^6 N/C. How much charge could be packed onto a small green pea (diameter 0.75 cm) before the pea spontaneously discharges?

76. Two point charges, $Q_1 = -6.7\,\mu$C and $Q_2 = 2.6\,\mu$C, are located between two oppositely charged parallel plates, as shown in Fig. 21–74. The two charges are separated by a distance of $x = 0.47$ m. Assume that the electric field produced by the charged plates is uniform and equal to $E = 53{,}000$ N/C. Calculate the net electrostatic force on Q_1 and give its direction.

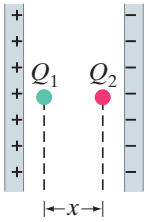


FIGURE 21–74
Problem 76.

77. Packing material made of pieces of foamed polystyrene can easily become charged and stick to each other. Given that the density of this material is about 35 kg/m^3, estimate how much charge might be on a 2.0-cm-diameter foamed polystyrene sphere, assuming the electric force between two spheres stuck together is equal to the weight of one sphere.

78. One type of *electric quadrupole* consists of two dipoles placed end to end with their negative charges (say) overlapping; that is, in the center is $-2Q$ flanked (on a line) by a $+Q$ to either side (Fig. 21–75). Determine the electric field $\vec{\mathbf{E}}$ at points along the perpendicular bisector and show that E decreases as $1/r^4$ for $r \gg \ell$. Measure r from the $-2Q$ charge.

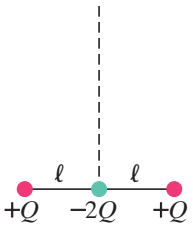


FIGURE 21–75
Problem 78.

79. Suppose electrons enter a uniform electric field midway between two plates (1.0 cm apart) at an angle θ_0 to the horizontal, as shown in Fig. 21–76. The path is symmetrical, so they leave at the same angle θ_0 and just barely miss the top plate. What is θ_0? Ignore fringing of the field.

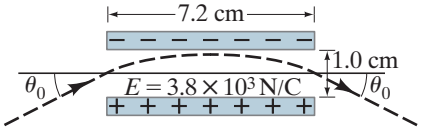


FIGURE 21–76 Problem 79.

80. An electron moves in a circle of radius r around a very long uniformly charged wire in a vacuum chamber, as shown in Fig. 21–77. The linear charge density on the wire is $\lambda = 0.18\,\mu$C/m. (*a*) What is the electric field at the electron (magnitude and direction) in terms of r and λ? (*b*) What is the speed of the electron?

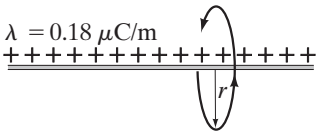


FIGURE 21–77
Problem 80.

81. Three very large square charged planes are arranged as shown (on edge) in Fig. 21–78. From left to right, the planes have charge densities per unit area of $-0.50\,\mu$C/m^2, $+0.25\,\mu$C/m^2, and $-0.35\,\mu$C/m^2. Find the total electric field (direction and magnitude) at the points A, B, C, and D. Assume the plates are much larger than the distance AD.

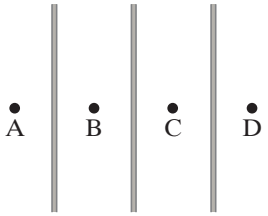


FIGURE 21–78
Problem 81.

82. A point charge ($m = 1.5$ gram) at the end of an insulating cord of length 55 cm is observed to be in equilibrium in a uniform horizontal electric field of 9500 N/C, when the pendulum's position is as shown in Fig. 21–79, with the charge 12 cm above the lowest (vertical) position. If the field points to the right in Fig. 21–79, determine the magnitude and sign of the point charge.

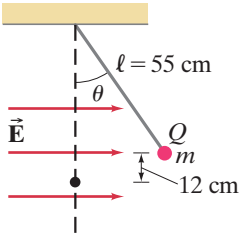


FIGURE 21–79
Problem 82.

83. Four equal positive point charges, each of charge 5.8 μC, are at the corners of a square of side 9.2 cm. What charge should be placed at the center of the square so that all charges are at equilibrium? Is this a stable or an unstable equilibrium (Section 12–4) in the plane?

84. Two small, identical conducting spheres A and B are a distance R apart; each carries the same charge Q. (*a*) What is the force sphere B exerts on sphere A? (*b*) An identical sphere with zero charge, sphere C, makes contact with sphere B and is then moved very far away. What is the net force now acting on sphere A? (*c*) Sphere C is brought back and now makes contact with sphere A and is then moved far away. What is the force now on sphere A?

85. A point charge of mass 0.145 kg, and net charge $+3.40\,\mu$C, hangs at rest at the end of an insulating cord above a large sheet of charge. The horizontal sheet of fixed uniform charge creates a uniform vertical electric field in the vicinity of the point charge. The tension in the cord is measured to be 5.18 N. (*a*) Calculate the magnitude and direction of the electric field due to the sheet of charge (Fig. 21–80). (*b*) What is the surface charge density σ (C/m^2) on the sheet?

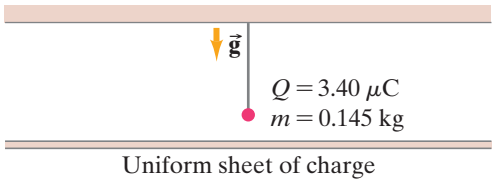


FIGURE 21–80 Problem 85.

86. A one-dimensional row of positive ions, each with charge $+Q$ and separated from its neighbors by a distance d, occupies the right-hand half of the x axis. That is, there is a $+Q$ charge at $x = 0$, $x = +d$, $x = +2d$, $x = +3d$, and so on out to ∞. (*a*) If an electron is placed at the position $x = -d$, determine F, the magnitude of the electric force that this row of charges exerts on the electron. (*b*) If the electron is instead placed at $x = -3d$, what is the value of F? [*Hint*: The infinite sum $\sum_{n=1}^{n=\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$, where n is a positive integer.]

87. (*a*) How many electrons are there in 1.00 mole of carbon atoms? (*b*) Two electrically neutral carbon spheres, each containing 1.00 mole of carbon, are separated by 18.0 cm (center to center). What fraction of electrons would have to be transferred from one sphere to the other for the electric force and the gravitational force between the spheres to be equal? Assume uniform charge distribution.

88. Near the surface of the Earth, there is a downward electric field of 150 N/C and a downward gravitational field of 9.8 N/kg. A charged 1.0-kg mass is observed to fall with acceleration $8.0\ \text{m/s}^2$. Determine the magnitude and sign of its charge.

89. Draw, approximately, the electric field lines about two point charges, $+Q$ and $-3Q$, which are a distance ℓ apart.

90. (*a*) The electric field near the Earth's surface has magnitude of about 150 N/C. What is the acceleration experienced by an electron due to this electric field near the surface of the Earth? (*b*) What about a proton? (*c*) Calculate the ratio of each acceleration to $g = 9.8\ \text{m/s}^2$.

91. Suppose the charge Q on the ring of Fig. 21–29 was all distributed uniformly on only the upper half of the ring, and no charge was on the lower half. Determine the electric field $\vec{\mathbf{E}}$ at P in terms of x, a, Q, and λ. (Take y vertically upward.)

92. A thin glass rod is a semicircle of radius R, Fig. 21–81. A charge is nonuniformly distributed along the semicircle with a linear charge density given by $\lambda = \lambda_0 \sin\theta$, where λ_0 is a positive constant. Point P is at the center of the semicircle. (*a*) Find the electric field $\vec{\mathbf{E}}$ (magnitude and direction) at point P. [*Hint:* Remember $\sin(-\theta) = -\sin\theta$, so the two halves of the rod are oppositely charged.] (*b*) Determine the acceleration (magnitude and direction) of an electron placed at point P, assuming $R = 1.0$ cm and $\lambda_0 = 1.0\ \mu\text{C/m}$.

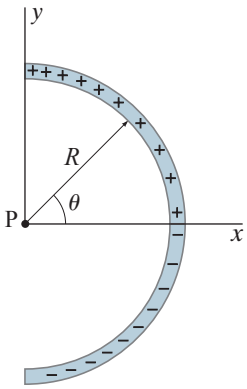
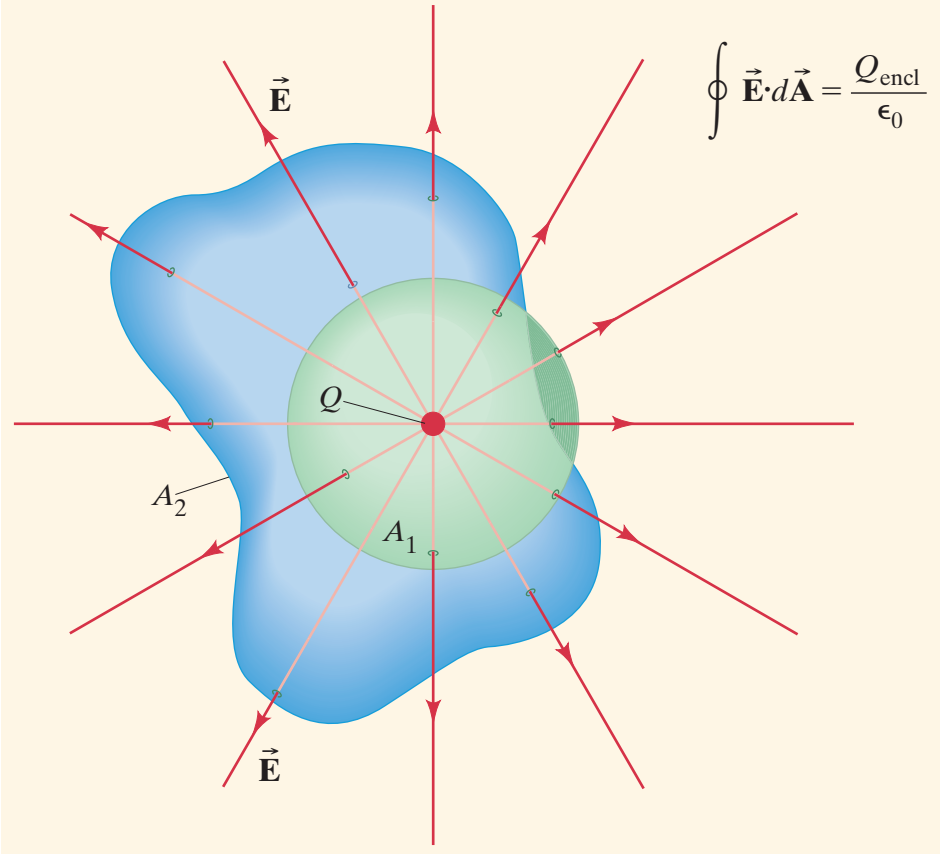


FIGURE 21–81 Problem 92.

ANSWERS TO EXERCISES

A: (*d*).
B: Opposite.
C: 0.3 N, to the right.
D: (*a*) No; (*b*) yes, midway between them.
E: (*d*), if the two + charges are not at opposite corners.



Gauss's law is an elegant relation between electric charge and electric field. It is more general than Coulomb's law. Gauss's law involves an integral of the electric field $\vec{\mathbf{E}}$ at each point on a closed surface. The surface is only imaginary, and we choose the shape and placement of the surface so that evaluating the integral is doable. In this drawing, two different 3-D surfaces are shown (one green, one blue), both enclosing a point charge Q. Gauss's law states that the product $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$, where $d\vec{\mathbf{A}}$ is an infinitesimal area of the surface, integrated over the entire surface, equals the charge enclosed by the surface Q_{encl} divided by ϵ_0. Both surfaces here enclose the same charge Q. Hence $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$ will give the same result for both surfaces.

CHAPTER 22

Gauss's Law

CHAPTER-OPENING QUESTION—Guess now!

A nonconducting sphere has a uniform charge density throughout. How does the magnitude of the electric field vary inside with distance from the center?

(a) The electric field is zero throughout.
(b) The electric field is constant but nonzero throughout.
(c) The electric field increases linearly from the center to the outer edge.
(d) The electric field decreases linearly from the center to the outer edge.
(e) The electric field increases exponentially from the center to the outer edge.
(d) The electric field decreases quadratically from the center to the outer edge.

The great mathematician Karl Friedrich Gauss (1777–1855) developed an important relation, now known as Gauss's law, which we develop and discuss in this Chapter. It is a statement of the relation between electric charge and electric field and is a more general and elegant form of Coulomb's law.

We can, in principle, determine the electric field due to any given distribution of electric charge using Coulomb's law. The total electric field at any point will be the vector sum (or integral) of contributions from all charges present (see Eq. 21–6). Except for some simple cases, the sum or integral can be quite complicated to evaluate. For situations in which an analytic solution (such as we carried out in the Examples of Sections 21–6 and 21–7) is not possible, a computer can be used.

In some cases, however, the electric field due to a given charge distribution can be calculated more easily or more elegantly using Gauss's law, as we shall see in this Chapter. But the major importance of Gauss's law is that it gives us additional insight into the nature of electrostatic fields, and a more general relationship between charge and field.

Before discussing Gauss's law itself, we first discuss the concept of *flux*.

CONTENTS

22–1 Electric Flux
22–2 Gauss's Law
22–3 Applications of Gauss's Law
*22–4 Experimental Basis of Gauss's and Coulomb's Laws

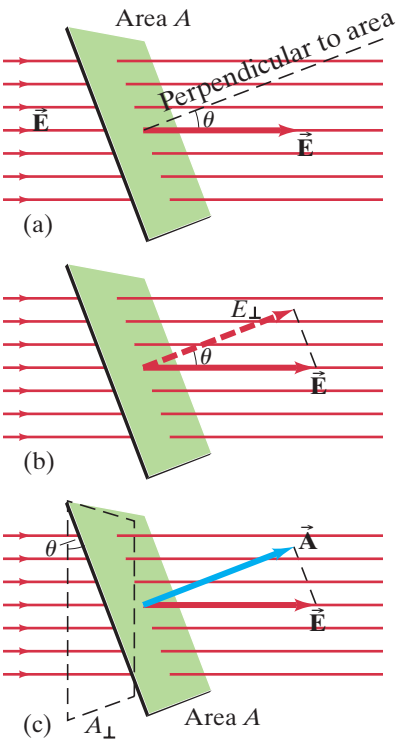


FIGURE 22–1 (a) A uniform electric field $\vec{\mathbf{E}}$ passing through a flat area A. (b) $E_\perp = E\cos\theta$ is the component of $\vec{\mathbf{E}}$ perpendicular to the plane of area A. (c) $A_\perp = A\cos\theta$ is the projection (dashed) of the area A perpendicular to the field $\vec{\mathbf{E}}$.

FIGURE 22–2 Electric flux through a curved surface. One small area of the surface, $\Delta\vec{\mathbf{A}}_i$, is indicated.

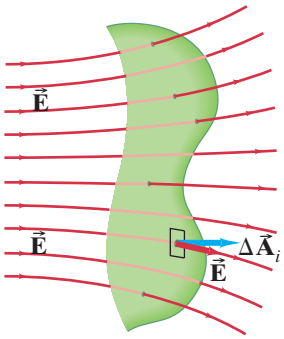
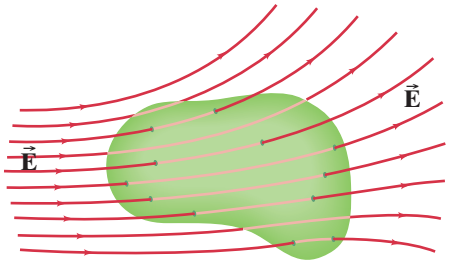


FIGURE 22–3 (below) Electric flux through a closed surface.



22–1 Electric Flux

Gauss's law involves the concept of **electric flux**, which refers to the electric field passing through a given area. For a uniform electric field $\vec{\mathbf{E}}$ passing through an area A, as shown in Fig. 22–1a, the electric flux Φ_E is defined as

$$\Phi_E = EA\cos\theta,$$

where θ is the angle between the electric field direction and a line drawn perpendicular to the area. The flux can be written equivalently as

$$\Phi_E = E_\perp A = EA_\perp = EA\cos\theta, \qquad [\vec{\mathbf{E}} \text{ uniform}] \quad \textbf{(22–1a)}$$

where $E_\perp = E\cos\theta$ is the component of $\vec{\mathbf{E}}$ along the perpendicular to the area (Fig. 22–1b) and, similarly, $A_\perp = A\cos\theta$ is the projection of the area A perpendicular to the field $\vec{\mathbf{E}}$ (Fig. 22–1c).

The area A of a surface can be represented by a vector $\vec{\mathbf{A}}$ whose magnitude is A and whose direction is perpendicular to the surface, as shown in Fig. 22–1c. The angle θ is the angle between $\vec{\mathbf{E}}$ and $\vec{\mathbf{A}}$, so the electric flux can also be written

$$\Phi_E = \vec{\mathbf{E}}\cdot\vec{\mathbf{A}}. \qquad [\vec{\mathbf{E}} \text{ uniform}] \quad \textbf{(22–1b)}$$

Electric flux can be interpreted in terms of field lines. We mentioned in Section 21–8 that field lines can always be drawn so that the number (N) passing through unit area perpendicular to the field $(A_\perp)$ is proportional to the magnitude of the field (E): that is, $E \propto N/A_\perp$. Hence,

$$N \propto EA_\perp = \Phi_E,$$

so the flux through an area is proportional to the number of lines passing through that area.

EXAMPLE 22–1 **Electric flux.** Calculate the electric flux through the rectangle shown in Fig. 22–1a. The rectangle is 10 cm by 20 cm, the electric field is uniform at 200 N/C, and the angle θ is 30°.

APPROACH We use the definition of flux, $\Phi_E = \vec{\mathbf{E}}\cdot\vec{\mathbf{A}} = EA\cos\theta$.

SOLUTION The electric flux is

$$\Phi_E = (200\,\text{N/C})(0.10\,\text{m} \times 0.20\,\text{m})\cos 30° = 3.5\,\text{N}\cdot\text{m}^2/\text{C}.$$

EXERCISE A Which of the following would cause a change in the electric flux through a circle lying in the xz plane where the electric field is $(10\,\text{N})\hat{\mathbf{j}}$? (*a*) Changing the magnitude of the electric field. (*b*) Changing the size of the circle. (*c*) Tipping the circle so that it is lying in the xy plane. (*d*) All of the above. (*e*) None of the above.

In the more general case, when the electric field $\vec{\mathbf{E}}$ is not uniform and the surface is not flat, Fig. 22–2, we divide up the chosen surface into n tiny areas, $\Delta A_1, \Delta A_2, \cdots, \Delta A_n$. We choose the division so that each ΔA_i is small enough that (1) it can be considered flat, and (2) the electric field varies so little over this small area that it can be considered uniform. Then the *total* electric flux through the entire surface is approximately

$$\Phi_E \approx \sum_{i=1}^{n} \vec{\mathbf{E}}_i\cdot\Delta\vec{\mathbf{A}}_i,$$

where $\vec{\mathbf{E}}_i$ is the field passing through $\Delta\vec{\mathbf{A}}_i$. In the limit as we let $\Delta\vec{\mathbf{A}}_i \to 0$, the sum becomes an integral over the entire surface and the relation becomes mathematically exact:

$$\Phi_E = \int \vec{\mathbf{E}}\cdot d\vec{\mathbf{A}}. \qquad \textbf{(22–2)}$$

Gauss's law involves the *total* flux through a *closed* surface—a surface of any shape that completely encloses a volume of space, such as that shown in Fig. 22–3. In this case, the net flux through the enclosing surface is given by

$$\Phi_E = \oint \vec{\mathbf{E}}\cdot d\vec{\mathbf{A}}, \qquad \textbf{(22–3)}$$

where the integral sign is written $\oint$ to indicate that the integral is over the value of $\vec{\mathbf{E}}$ at every point on a *closed* surface: one that fully encloses a volume.

Up to now we have not been concerned with an ambiguity in the direction of the vector $\vec{\mathbf{A}}$ or $d\vec{\mathbf{A}}$ that represents a surface. For example, in Fig. 22–1c, the vector $\vec{\mathbf{A}}$ could point upward and to the right (as shown) or downward to the left and still be perpendicular to the surface. For a closed surface, we define (arbitrarily) the direction of $\vec{\mathbf{A}}$, or of $d\vec{\mathbf{A}}$, to point *outward* from the enclosed volume, Fig. 22–4. For an electric field line leaving the enclosed volume (on the right in Fig. 22–4), the angle θ between $\vec{\mathbf{E}}$ and $d\vec{\mathbf{A}}$ must be less than $\pi/2$ $(= 90°)$; hence $\cos\theta > 0$. For a line entering the volume (on the left in Fig. 22–4) $\theta > \pi/2$; hence $\cos\theta < 0$. Thus, flux *entering* the enclosed volume is *negative* $\left(\int E\cos\theta\, dA < 0\right)$, whereas flux *leaving* the volume is *positive*. Consequently, Eq. 22–3 gives the net flux *out* of the volume. If Φ_E is negative, there is a net flux *into* the volume.

In Figs. 22–3 and 22–4, each field line that enters the volume also leaves the volume. Hence $\Phi_E = \oint \vec{\mathbf{E}}\cdot d\vec{\mathbf{A}} = 0$. There is no net flux into or out of this enclosed surface. The flux, $\oint \vec{\mathbf{E}}\cdot d\vec{\mathbf{A}}$, will be nonzero only if one or more lines start or end within the surface. Since electric field lines start and stop only on electric charges, the flux will be nonzero only if the surface encloses a net charge. For example, the surface labeled A_1 in Fig. 22–5 encloses a positive charge and there is a net outward flux through this surface $(\Phi_E > 0)$. In fact *all* the electric field lines are *leaving* surface A_1. The surface A_2 encloses an equal magnitude negative charge and there is a net inward flux $(\Phi_E < 0)$. For the configuration shown in Fig. 22–6, some lines enter the surface, others leave. But the net flux through the surface shown is negative (count the lines). The value of Φ_E depends on the charge enclosed by the surface, and this is what Gauss's law is all about.

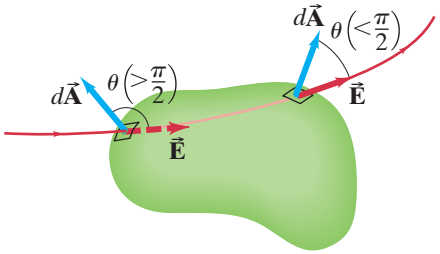


FIGURE 22–4 The direction of an element of area $d\vec{\mathbf{A}}$ is taken to point outward from an enclosed surface.

FIGURE 22–5 An electric dipole. Flux through surface A_1 is positive. Flux through A_2 is negative.

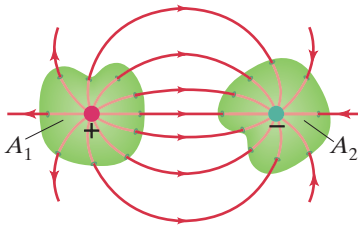
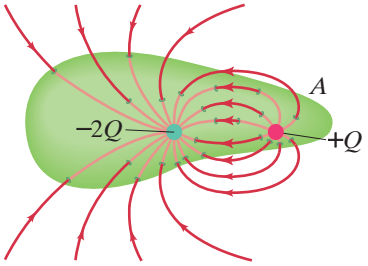


FIGURE 22–6 (below) Net flux through surface A is negative.



22–2 Gauss's Law

The precise relation between the electric flux through a closed surface and the net charge Q_{encl} enclosed within that surface is given by **Gauss's law**:

$$\oint \vec{\mathbf{E}}\cdot d\vec{\mathbf{A}} = \frac{Q_{\text{encl}}}{\epsilon_0}, \qquad \textbf{(22–4)}$$

where ϵ_0 is the same constant (permittivity of free space) that appears in Coulomb's law, $\epsilon_0 = 8.85 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2$. The integral on the left is over the value of $\vec{\mathbf{E}}$ on any closed surface, and we choose that imaginary surface for our convenience in any given situation. The charge Q_{encl} is the net charge *enclosed* by that surface. It doesn't matter where or how the charge is distributed inside the surface. Any charge outside this surface must not be included. A charge outside the chosen surface may affect the position of the electric field lines, but will not affect the net number of lines entering or leaving the surface. For example, Q_{encl} for the imaginary gaussian surface A_1 in Fig. 22–5 would be the positive charge enclosed by A_1; the negative charge on the right does contribute to the electric field at A_1 but it is *not* enclosed by surface A_1 and so is *not* included in Q_{encl}.

Now let us see how Gauss's law is related to Coulomb's law. First, we show that Coulomb's law follows from Gauss's law. In Fig. 22–7 we have a single isolated charge Q. For our "gaussian surface," we choose an imaginary sphere of radius r centered on the charge. Because Gauss's law is supposed to be valid for any surface, we have chosen one that will make our calculation easy. Because of the *symmetry* of this (imaginary) sphere about the charge at its center, we know (1) that $\vec{\mathbf{E}}$ must have the same magnitude at any point on this spherical surface, and (2) that $\vec{\mathbf{E}}$ points radially outward (inward for a negative charge) perpendicular to the surface and parallel to $d\vec{\mathbf{A}}$, an element of the surface area. Hence, we write the integral in Gauss's law as

$$\oint \vec{\mathbf{E}}\cdot d\vec{\mathbf{A}} = \oint E\, dA = E\oint dA = E(4\pi r^2)$$

since the magnitude of $\vec{\mathbf{E}}$ is the same at all points on this spherical surface, and the surface area of a sphere of radius r is $4\pi r^2$. Then Gauss's law becomes, with $Q_{\text{encl}} = Q$,

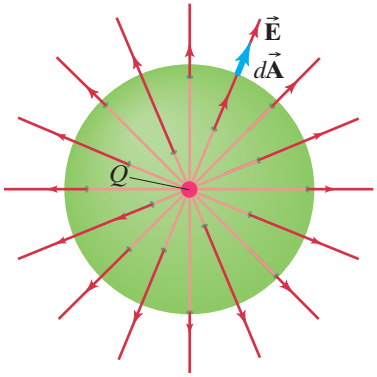
$$\frac{Q}{\epsilon_0} = \oint \vec{\mathbf{E}}\cdot d\vec{\mathbf{A}} = E(4\pi r^2).$$

So

$$E = \frac{Q}{4\pi\epsilon_0 r^2},$$

which is the electric field form of Coulomb's law, Eq. 21–4b.

FIGURE 22–7 A single point charge Q at the center of an imaginary sphere of radius r (our "gaussian surface"—that is, the closed surface we choose to use for applying Gauss's law in this case).



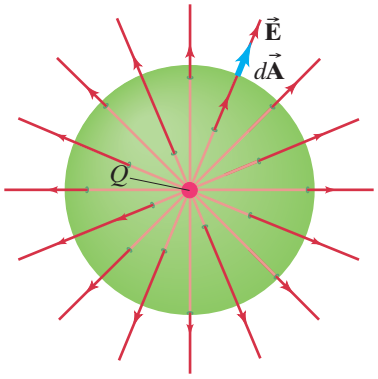


FIGURE 22–7 (Repeated.) A single point charge Q at the center of an imaginary sphere of radius r (our "gaussian surface"—that is, the closed surface we choose to use for applying Gauss's law in this case).

FIGURE 22–8 A single point charge surrounded by a spherical surface, A_1 (green color), and an irregular surface, A_2 (blue).

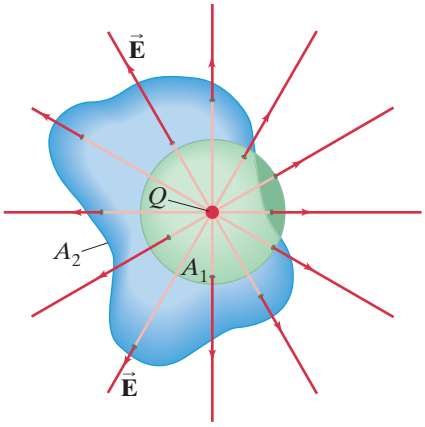
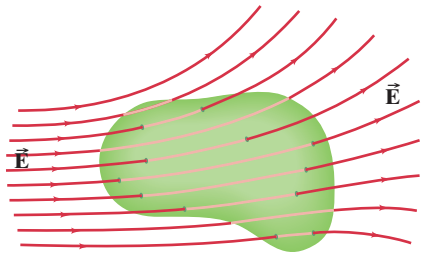


FIGURE 22–9 Electric flux through a closed surface. (Same as Fig. 22–3.) No electric charge is enclosed by this surface $(Q_{\text{encl}} = 0)$.



Now let us do the reverse, and derive Gauss's law from Coulomb's law for static electric charges.† First we consider a single point charge Q surrounded by an imaginary spherical surface as in Fig. 22–7 (and shown again, green, in Fig. 22–8). Coulomb's law tells us that the electric field at the spherical surface is $E = (1/4\pi\epsilon_0)(Q/r^2)$. Reversing the argument we just used, we have

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \oint \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}\,dA = \frac{Q}{4\pi\epsilon_0 r^2}(4\pi r^2) = \frac{Q}{\epsilon_0}.$$

This is Gauss's law, with $Q_{\text{encl}} = Q$, and we derived it for the special case of a spherical surface enclosing a point charge at its center. But what about some other surface, such as the irregular surface labeled A_2 in Fig. 22–8? The same number of field lines (due to our charge Q) pass through surface A_2, as pass through the spherical surface, A_1. Therefore, because the flux through a surface is proportional to the number of lines through it as we saw in Section 22–1, the flux through A_2 is the same as through A_1:

$$\oint_{A_2} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \oint_{A_1} \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q}{\epsilon_0}.$$

Hence, we can expect that

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q}{\epsilon_0}$$

would be valid for *any* surface surrounding a single point charge Q.

Finally, let us look at the case of more than one charge. For each charge, Q_i, enclosed by the chosen surface,

$$\oint \vec{\mathbf{E}}_i \cdot d\vec{\mathbf{A}} = \frac{Q_i}{\epsilon_0},$$

where $\vec{\mathbf{E}}_i$ refers to the electric field produced by Q_i alone. By the superposition principle for electric fields (Section 21–6), the total field $\vec{\mathbf{E}}$ is equal to the sum of the fields due to each separate charge, $\vec{\mathbf{E}} = \Sigma\vec{\mathbf{E}}_i$. Hence

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \oint (\Sigma\vec{\mathbf{E}}_i) \cdot d\vec{\mathbf{A}} = \Sigma\frac{Q_i}{\epsilon_0} = \frac{Q_{\text{encl}}}{\epsilon_0},$$

where $Q_{\text{encl}} = \Sigma Q_i$ is the total net charge enclosed within the surface. Thus we see, based on this simple argument, that Gauss's law follows from Coulomb's law for any distribution of static electric charge enclosed within a closed surface of any shape.

The derivation of Gauss's law from Coulomb's law is valid for electric fields produced by static electric charges. We will see later that electric fields can also be produced by changing magnetic fields. Coulomb's law cannot be used to describe such electric fields. But Gauss's law *is* found to hold also for electric fields produced in any of these ways. Hence *Gauss's law is a more general law than Coulomb's law*. It holds for any electric field whatsoever.

Even for the case of static electric fields that we are considering in this Chapter, it is important to recognize that $\vec{\mathbf{E}}$ on the left side of Gauss's law is not necessarily due only to the charge Q_{encl} that appears on the right. For example, in Fig. 22–9 there is an electric field $\vec{\mathbf{E}}$ at all points on the imaginary gaussian surface, but it is not due to the charge enclosed by the surface (which is $Q_{\text{encl}} = 0$ in this case). The electric field $\vec{\mathbf{E}}$ which appears on the left side of Gauss's law is the *total* electric field at each point, on the gaussian surface chosen, not just that due to the charge Q_{encl}, which appears on the right side. Gauss's law has been found to be valid for the total field at any surface. It tells us that any *difference* between the input and output flux of the electric field over any surface is due to charge within that surface.

†Note that Gauss's law would look more complicated in terms of the constant $k = 1/4\pi\epsilon_0$ that we originally used in Coulomb's law (Eqs. 21–1 or 21–4a), $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = 4\pi kQ$. We prefer ϵ_0 to k so as to make Gauss's law simpler, which is considered more general than Coulomb's law.

CONCEPTUAL EXAMPLE 22–2 **Flux from Gauss's law.** Consider the two imaginary gaussian surfaces, A_1 and A_2, shown in Fig. 22–10. The only charge present is the charge Q at the center of surface A_1. What is the net flux through each surface, A_1 and A_2?

RESPONSE The surface A_1 encloses the charge $+Q$. By Gauss's law, the net flux through A_1 is then Q/ϵ_0. For surface A_2, the charge $+Q$ is outside the surface. Surface A_2 encloses zero net charge, so the net electric flux through A_2 is zero, by Gauss's law. Note that all field lines that enter the volume enclosed by surface A_2 also leave it.

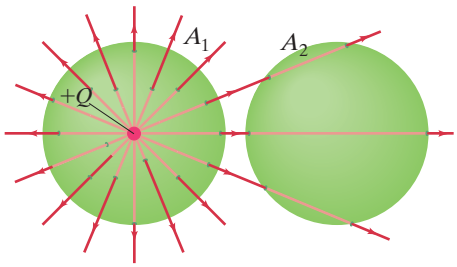


FIGURE 22–10 Example 22–2. Two gaussian surfaces.

EXERCISE B A point charge Q is at the center of a spherical gaussian surface A. When a second charge Q is placed just outside A, the total flux through this spherical surface A is (*a*) unchanged, (*b*) doubled, (*c*) halved, (*d*) none of these.

EXERCISE C Three 2.95 μC charges are in a small box. What is the net flux leaving the box? (*a*) $3.3 \times 10^{12}\,\mathrm{N \cdot m^2/C}$, (*b*) $3.3 \times 10^{5}\,\mathrm{N \cdot m^2/C}$, (*c*) $1.0 \times 10^{12}\,\mathrm{N \cdot m^2/C}$, (*d*) $1.0 \times 10^{6}\,\mathrm{N \cdot m^2/C}$, (*e*) $6.7 \times 10^{6}\,\mathrm{N \cdot m^2/C}$.

22–3 Applications of Gauss's Law

Gauss's law is a very compact and elegant way to write the relation between electric charge and electric field. It also offers a simple way to determine the electric field when the charge distribution is simple and/or possesses a high degree of *symmetry*. In order to apply Gauss's law, however, we must choose the "gaussian" surface very carefully (for the integral on the left side of Gauss's law) so we can determine $\vec{\mathbf{E}}$. We normally try to think of a surface that has just the symmetry needed so that E will be constant on all or on parts of its surface. Sometimes we choose a surface so the flux through part of the surface is zero.

EXAMPLE 22–3 **Charged spherical conducting shell.** A thin spherical shell of radius r_0 possesses a total net charge Q that is uniformly distributed on it, Fig. 22–11. Determine the electric field at points (*a*) outside the shell, and (*b*) inside the shell. (*c*) What if the conductor were a solid sphere?

APPROACH Because the charge is distributed symmetrically, the electric field must also be *symmetric*. Thus the field outside the shell must be directed radially outward (inward if $Q < 0$) and must depend only on r, not on angle (spherical coordinates).

SOLUTION (*a*) We choose our imaginary gaussian surface as a sphere of radius r $(r > r_0)$ outside and concentric with our charged shell, shown in Fig. 22–11 as the dashed green circle A_1. Then, by symmetry, the electric field will have the same magnitude at all points on this gaussian surface. Because $\vec{\mathbf{E}}$ is perpendicular to this surface, the cosine of the angle between $\vec{\mathbf{E}}$ and $d\vec{\mathbf{A}}$ is always 1. Gauss's law then gives (with $Q_{\text{encl}} = Q$ in Eq. 22–4)

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = E(4\pi r^2) = \frac{Q}{\epsilon_0},$$

where $4\pi r^2$ is the surface area of our imaginary sphere (gaussian surface) of radius r. Thus

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}. \qquad [r > r_0]$$

We see that the field outside a uniformly charged spherical shell is the same as if *all the charge were concentrated at the center as a point charge*.

(*b*) Inside the shell, the electric field must also be symmetric. So E must again have the same value at all points on an imaginary spherical gaussian surface (A_2 in Fig. 22–11) concentric with our charged shell. Thus E can be factored out of the integral and, with $Q_{\text{encl}} = 0$ because the charge inside surface A_2 is zero, we have

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = E\oint dA = E(4\pi r^2) = 0.$$

Hence

$$E = 0 \qquad [r < r_0]$$

inside a uniform spherical shell of charge (as claimed in Section 21–9).

(*c*) These same results also apply to a uniformly charged solid spherical conductor, because all the charge would lie in a thin layer at the surface.

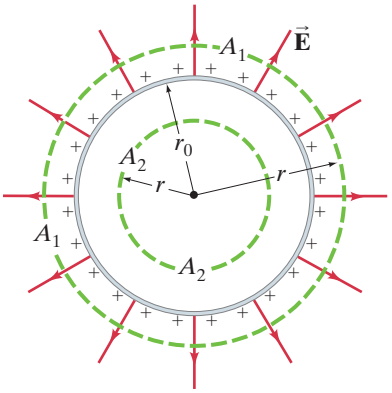


FIGURE 22–11 Cross-sectional drawing of a thin spherical shell (grey) of radius r_0, carrying a net charge Q uniformly distributed. A_1 and A_2 represent two gaussian surfaces we use to determine $\vec{\mathbf{E}}$. Example 22–3.

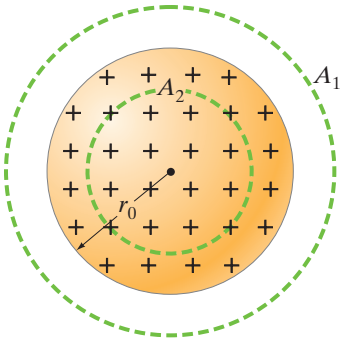


FIGURE 22–12 A solid sphere of uniform charge density. Example 22–4.

EXAMPLE 22–4 Solid sphere of charge. An electric charge Q is distributed uniformly throughout a nonconducting sphere of radius r_0, Fig. 22–12. Determine the electric field (*a*) outside the sphere $(r > r_0)$ and (*b*) inside the sphere $(r < r_0)$.

APPROACH Since the charge is distributed symmetrically in the sphere, the electric field at all points must again be *symmetric*. $\vec{\mathbf{E}}$ depends only on r and is directed radially outward (or inward if $Q < 0$).

SOLUTION (*a*) For our gaussian surface we choose a sphere of radius r $(r > r_0)$, labeled A_1 in Fig. 22–12. Since E depends only on r, Gauss's law gives, with $Q_{\text{encl}} = Q$,

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = E(4\pi r^2) = \frac{Q}{\epsilon_0}$$

or

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}.$$

Again, the field outside a spherically symmetric distribution of charge is the same as that for a point charge of the same magnitude located at the center of the sphere.

(*b*) Inside the sphere, we choose for our gaussian surface a concentric sphere of radius r $(r < r_0)$, labeled A_2 in Fig. 22–12. From symmetry, the magnitude of $\vec{\mathbf{E}}$ is the same at all points on A_2, and $\vec{\mathbf{E}}$ is perpendicular to the gaussian surface, so

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = E(4\pi r^2).$$

We must equate this to $Q_{\text{encl}}/\epsilon_0$ where Q_{encl} is the charge enclosed by A_2. Q_{encl} is not the total charge Q but only a portion of it. We define the **charge density**, ρ_{E}, as the charge per unit volume $(\rho_{\text{E}} = dQ/dV)$, and here we are given that $\rho_{\text{E}} =$ constant. So the charge enclosed by the gaussian surface A_2, a sphere of radius r, is

$$Q_{\text{encl}} = \left(\frac{\frac{4}{3}\pi r^3 \rho_{\text{E}}}{\frac{4}{3}\pi r_0^3 \rho_{\text{E}}}\right) Q = \frac{r^3}{r_0^3} Q.$$

Hence, from Gauss's law,

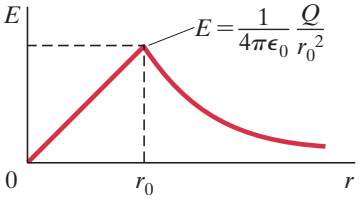
$$E(4\pi r^2) = \frac{Q_{\text{encl}}}{\epsilon_0} = \frac{r^3}{r_0^3}\frac{Q}{\epsilon_0}$$

or

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r_0^3} r. \qquad [r < r_0]$$

Thus the field increases linearly with r, until $r = r_0$. It then decreases as $1/r^2$, as plotted in Fig. 22–13.

FIGURE 22–13 Magnitude of the electric field as a function of the distance r from the center of a uniformly charged solid sphere, Example 22–4.



EXERCISE D Return to the Chapter-Opening Question, page 643, and answer it again now. Try to explain why you may have answered differently the first time.

Examples 22–3 and 22–4a tell us that

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2} \qquad \left[\begin{array}{c}\text{outside uniformly charged}\\ \text{sphere or spherical shell}\end{array}\right]$$

outside a sphere or spherical shell carrying uniform charge Q. The electric field outside a spherically symmetric distribution of charge is thus the same as for a point charge of the same magnitude at the center of the sphere, confirming what we said in Section 21–5, and applies for both a uniformly charged conductor or nonconductor.

The *gravitational force* is also a $1/r^2$ force, so this same result applies: The gravitational force exerted by a uniform sphere is the same as if all the mass were at the center, as stated at the end of Section 6–1, and derived by integration in Appendix E. Furthermore, the last equation in Example 22–4 tells us that the *gravitational field* at a point *inside* the Earth is $\propto r$, if we assume uniform mass density (again see Appendix E).

The results in Example 22–4 would have been difficult to obtain from Coulomb's law by integrating over the sphere (as we do for gravity in Appendix E). Using Gauss's law and the *symmetry* of the situation, these results are obtained rather easily, and show the great power of Gauss's law. Gauss's law holds for any surface, but is practically useful mainly when the charge distribution has a high degree of symmetry. Carrying out the integral in Gauss's law can be difficult. We try to *choose* a simple surface on which E = constant, so the integration is simple.

EXAMPLE 22–5 **Nonuniformly charged solid sphere.** Suppose the charge density of the solid sphere in Fig. 22–12, Example 22–4, is given by $\rho_E = \alpha r^2$, where α is a constant. (*a*) Find α in terms of Q and r_0, the total charge on the sphere and its radius. (*b*) Find the electric field as a function of r inside the sphere.

APPROACH We divide the sphere up into concentric thin shells of thickness dr as shown in Fig. 22–14, and (*a*) integrate $Q = \int \rho_E \, dV$ and (*b*) use Gauss's law.

SOLUTION (*a*) A thin shell of radius r and thickness dr (Fig. 22–14) has volume $dV = 4\pi r^2 \, dr$. The total charge is given by

$$Q = \int \rho_E \, dV = \int_0^{r_0} (\alpha r^2)(4\pi r^2 \, dr) = 4\pi\alpha \int_0^{r_0} r^4 \, dr = \frac{4\pi\alpha}{5} r_0^5.$$

Thus $\alpha = 5Q/4\pi r_0^5$.

(*b*) To find E inside the sphere at distance r from its center, we apply Gauss's law to an imaginary sphere of radius r which will enclose a charge

$$Q_{\text{encl}} = \int_0^r \rho_E \, dV = \int_0^r (\alpha r^2)\, 4\pi r^2 \, dr = \int_0^r \left(\frac{5Q}{4\pi r_0^5} r^2\right) 4\pi r^2 \, dr = Q \frac{r^5}{r_0^5}.$$

By *symmetry*, E will be the same at all points on the surface of a sphere of radius r, so Gauss's law gives

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q_{\text{encl}}}{\epsilon_0}, \qquad \text{so} \qquad (E)(4\pi r^2) = Q \frac{r^5}{\epsilon_0 r_0^5},$$

giving

$$E = \frac{Qr^3}{4\pi\epsilon_0 r_0^5}.$$

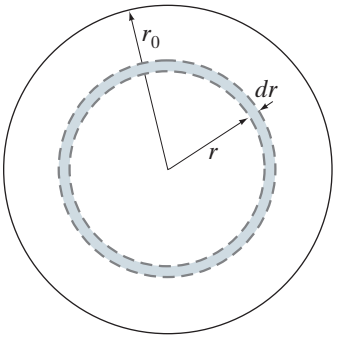


FIGURE 22–14 Example 22–5.

EXAMPLE 22–6 **Long uniform line of charge.** A very long straight wire possesses a uniform positive charge per unit length, λ. Calculate the electric field at points near (but outside) the wire, far from the ends.

APPROACH Because of the *symmetry*, we expect the field to be directed radially outward and to depend only on the perpendicular distance, R, from the wire. Because of the cylindrical symmetry, the field will be the same at all points on an imaginary gaussian surface that is a dashed green cylinder with the wire along its axis, Fig. 22–15. $\vec{\mathbf{E}}$ is perpendicular to the curved surface at all points. For Gauss's law, we need a closed surface, so we include the flat ends of the cylinder. Since $\vec{\mathbf{E}}$ is parallel to the ends, there is no flux through the ends (the cosine of the angle between $\vec{\mathbf{E}}$ and $d\vec{\mathbf{A}}$ on the ends is $\cos 90° = 0$).

SOLUTION For our chosen gaussian surface Gauss's law gives

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = E(2\pi R\ell) = \frac{Q_{\text{encl}}}{\epsilon_0} = \frac{\lambda\ell}{\epsilon_0},$$

where ℓ is the length of our chosen gaussian surface ($\ell \ll$ length of wire), and $2\pi R$ is its circumference. Hence

$$E = \frac{1}{2\pi\epsilon_0} \frac{\lambda}{R}.$$

NOTE This is the same result we found in Example 21–12 using Coulomb's law (we used x there instead of R), but here it took much less effort. Again we see the great power of Gauss's law.†

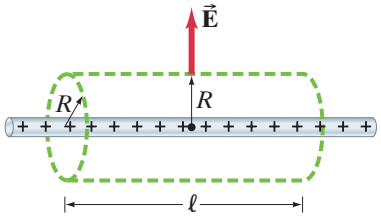


FIGURE 22–15 Calculation of $\vec{\mathbf{E}}$ due to a very long line of charge. Example 22–6. (Note: We use R for the distance from an axis—cylindrical symmetry—and r for the distance from a point. See Fig. 10–2.)

†But note that the method of Example 21–12 allows calculation of E also for a short line of charge by using the appropriate limits for the integral, whereas Gauss's law is not readily adapted.

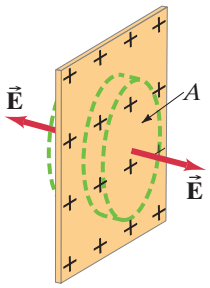


FIGURE 22–16 Calculation of the electric field outside a large uniformly charged nonconducting plane surface. Example 22–7.

EXAMPLE 22–7 **Infinite plane of charge.** Charge is distributed uniformly, with a surface charge density σ (σ = charge per unit area = dQ/dA), over a very large but very thin nonconducting flat plane surface. Determine the electric field at points near the plane.

APPROACH We choose as our imaginary gaussian surface a small closed cylinder whose axis is perpendicular to the plane and which extends through the plane as shown dashed in green in Fig. 22–16. Because of the symmetry, we expect $\vec{\mathbf{E}}$ to be directed perpendicular to the plane on both sides as shown, and to be uniform over the end caps of the cylinder, each of whose area is A.

SOLUTION Since no flux passes through the curved sides of our chosen cylindrical surface, all the flux is through the two end caps. So Gauss's law gives

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = 2EA = \frac{Q_{\text{encl}}}{\epsilon_0} = \frac{\sigma A}{\epsilon_0},$$

where $Q_{\text{encl}} = \sigma A$ is the charge enclosed by our gaussian cylinder. The electric field is then

$$E = \frac{\sigma}{2\epsilon_0}.$$

NOTE This is the same result we obtained much more laboriously in Chapter 21, Eq. 21–7. The field is uniform for points far from the ends of the plane, and close to its surface.

EXAMPLE 22–8 **Electric field near any conducting surface.** Show that the magnitude of the electric field just outside the surface of a good conductor of any shape is given by

$$E = \frac{\sigma}{\epsilon_0},$$

where σ is the surface charge density (Q/A) on the conductor's surface at that point.

APPROACH We choose as our gaussian surface a small cylindrical box (like a tiny pillbox), as we did in the previous Example. We choose the cylinder to be very small in height, so that one of its circular ends is just above the conductor (Fig. 22–17). The other end is just below the conductor's surface, and the very short sides are perpendicular to it.

SOLUTION The electric field is zero inside a conductor and is perpendicular to the surface just outside it (Section 21–9), so electric flux passes only through the outside end of our cylindrical box; no flux passes through the very short sides or through the inside end of our gaussian box. We choose the area A (of the flat cylinder end above the conductor surface) small enough so that E is essentially uniform over it. Then Gauss's law gives

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = EA = \frac{Q_{\text{encl}}}{\epsilon_0} = \frac{\sigma A}{\epsilon_0},$$

so that

$$E = \frac{\sigma}{\epsilon_0}. \qquad \text{[at surface of conductor]} \quad \textbf{(22–5)}$$

NOTE This useful result applies for a conductor of any shape.

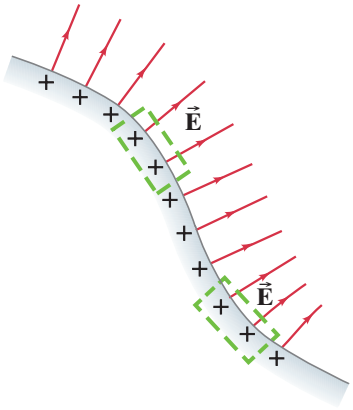


FIGURE 22–17 Electric field near the surface of a conductor. Two small cylindrical boxes are shown dashed. Either one can serve as our gaussian surface. Example 22–8.

CAUTION
When is $E = \sigma/\epsilon_0$ and when is $E = \sigma/2\epsilon_0$

Why is it that the field outside a large plane nonconductor is $E = \sigma/2\epsilon_0$ (Example 22–7) whereas outside a conductor (Example 22–8) it is $E = \sigma/\epsilon_0$? The reason for the factor of 2 comes from how we *define* charge per unit area σ. For a thin flat nonconductor, Fig. 22–16, the charge is distributed throughout the volume (not only on the surface, as for a conductor). Perhaps more simply, we see in Fig. 22–16 that the electric field appears on *both* sides of the thin nonconductor. But for a conductor, the field appears on only one side, Fig. 22–17 (being zero inside the conductor).

Example 22–8 gives us, in a simple way, the field between the two parallel plates we discussed in Fig. 21–32, Example 21–14. If the plates are large compared to their separation, the field lines are perpendicular to the plates and, except near the edges, they are parallel to each other. Therefore the electric field (see Fig. 22–18, which shows a similar very thin gaussian surface as Fig. 22–17) is also

$$E = \frac{\sigma}{\epsilon_0} \qquad \left[\begin{array}{l}\text{between two closely spaced,}\\ \text{oppositely charged, parallel plates}\end{array}\right]$$

which is Eq. 21–8, pages 625 and 626.

We saw in Section 21–9 that the static electric field inside any conductor must be zero, even if it has a net charge. (Otherwise, the free charges in the conductor would move—until the net force on each charge, and hence $\vec{\mathbf{E}}$, were zero.) We also mentioned that any net electric charge on a conductor must all reside on its outer surface. This is readily shown using Gauss's law. Consider a conductor of any shape, carrying a net charge Q, such as that shown in Fig. 22–19. We choose the gaussian surface (dashed) so that it all lies just below the surface of the conductor and encloses the whole volume of the conductor, but still *inside* the conductor. The electric field is zero at all points on this gaussian surface because it is inside the conductor. From Gauss's law, Eq. 22–4, the net charge within the surface must be zero. Thus *there can be no net charge within the conductor, and any net charge must lie on the surface.*

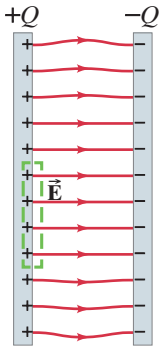
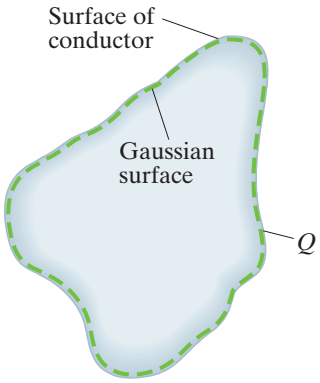


FIGURE 22–18 The electric field between two closely spaced parallel plates is uniform and equal to $E = \sigma/\epsilon_0$.

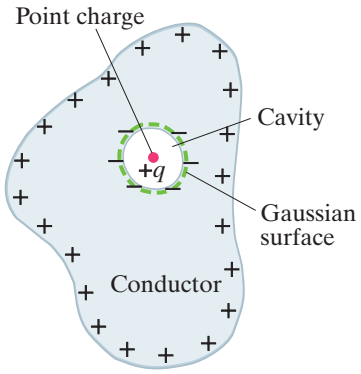
FIGURE 22–19 An insulated charged conductor of arbitrary shape, showing a gaussian surface (dashed) just below the surface of the conductor.



CONCEPTUAL EXAMPLE 22–9 **Conductor with charge inside a cavity.** Suppose a conductor carries a net charge $+Q$ and contains a cavity. (*a*) Inside the cavity resides a point charge $+q$. What can you say about the charges on the inner and outer surfaces of the conductor? (*b*) What if $q = 0$ (cavity empty)?

RESPONSE (*a*) As shown in Fig. 22–20, a gaussian surface (dashed, green) just inside the conductor surrounding the cavity must contain zero net charge ($E = 0$ in a conductor). Thus a net charge of $-q$ must exist on the cavity surface. The conductor itself carries a net charge $+Q$, so the outer surface of the conductor must carry a total charge equal to $Q + q$. These results apply to a cavity of any shape. (*b*) If the cavity is empty, $q = 0$, there is no net charge at the surface of the cavity, and the outer surface of the conductor carries charge Q.

FIGURE 22–20 Example 22–9.



EXERCISE E Which of the following statements about Gauss's law is correct? (*a*) If we know the charge enclosed by a surface, we always know the electric field everywhere at the surface. (*b*) When finding the electric field with Gauss's law, we always use a sphere for the gaussian surface. (*c*) If we know the total flux through a surface, we also know the total charge inside the surface. (*d*) We can only use Gauss's law if the electric field is constant in space.

PROBLEM SOLVING

Gauss's Law for Symmetric Charge Distributions

1. First identify the **symmetry** of the charge distribution: spherical, cylindrical, planar. This identification should suggest a gaussian surface for which $\vec{\mathbf{E}}$ will be constant and/or zero on all or on parts of the surface: a sphere for spherical symmetry, a cylinder for cylindrical symmetry and a small cylinder or "pillbox" for planar symmetry.
2. Draw the appropriate gaussian surface making sure it passes through the point where you want to know the electric field.
3. Use the symmetry of the charge distribution to determine the direction of $\vec{\mathbf{E}}$ at points on the gaussian surface.
4. Evaluate the flux, $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$. With an appropriate gaussian surface, the dot product $\vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$ should be zero or equal to $\pm E\, dA$, with the magnitude of E being constant over all or parts of the surface.
5. Determine the charge *enclosed* by the gaussian surface. Remember, it is the enclosed charge that matters. Ignore all the charge outside the gaussian surface.
6. Equate the flux to the enclosed charge and solve for E.

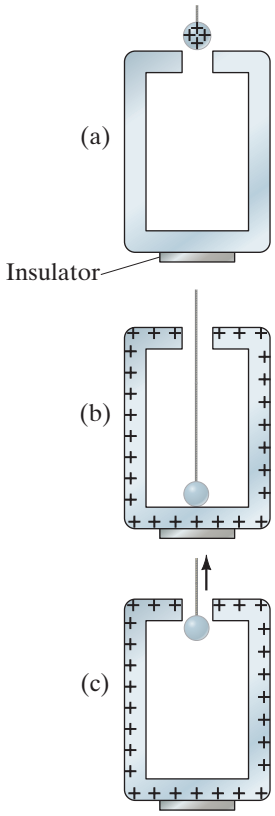


FIGURE 22–21 (a) A charged conductor (metal ball) is lowered into an insulated metal can (a good conductor) carrying zero net charge. (b) The charged ball is touched to the can and all of its charge quickly flows to the outer surface of the can. (c) When the ball is then removed, it is found to carry zero net charge.

*22–4 Experimental Basis of Gauss's and Coulomb's Laws

Gauss's law predicts that any net charge on a conductor must lie only on its surface. But is this true in real life? Let us see how it can be verified experimentally. And in confirming this prediction of Gauss's law, Coulomb's law is also confirmed since the latter follows from Gauss's law, as we saw in Section 22–2. Indeed, the earliest observation that charge resides only on the outside of a conductor was recorded by Benjamin Franklin some 30 years before Coulomb stated his law.

A simple experiment is illustrated in Fig. 22–21. A metal can with a small opening at the top rests on an insulator. The can, a conductor, is initially uncharged (Fig. 22–21a). A charged metal ball (also a conductor) is lowered by an insulating thread into the metal can, and is allowed to touch the can (Fig. 22–21b). The ball and metal can now form a single conductor. Gauss's law, as discussed above, predicts that all the charge will flow to the outer surface of the can. (The flow of charge in such situations does not occur instantaneously, but the time involved is usually negligible.) These predictions are confirmed in experiments by (1) connecting an electroscope to the can, which will show that the can is charged, and (2) connecting an electroscope to the ball after it has been withdrawn from the can (Fig. 22–21c), which will show that the ball carries zero charge.

The precision with which Coulomb's and Gauss's laws hold can be stated quantitatively by writing Coulomb's law as

$$F = k\frac{Q_1 Q_2}{r^{2+\delta}}.$$

For a perfect inverse-square law, $\delta = 0$. The most recent and precise experiments give $\delta = (1.0 \pm 1.2) \times 10^{-16}$. Thus Coulomb's and Gauss's laws are found to be valid to an extremely high precision!

Summary

The **electric flux** passing through a flat area A for a uniform electric field $\vec{\mathbf{E}}$ is

$$\Phi_E = \vec{\mathbf{E}} \cdot \vec{\mathbf{A}}. \qquad \textbf{(22–1b)}$$

If the field is not uniform, the flux is determined by integrating over the area:

$$\Phi_E = \int \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}. \qquad \textbf{(22–2)}$$

The direction of the vector $\vec{\mathbf{A}}$ or $d\vec{\mathbf{A}}$ is chosen to be perpendicular to the surface whose area is A or dA, and points outward from an enclosed surface. The flux through a surface is proportional to the number of field lines passing through it.

Gauss's law states that the net flux passing through any closed surface is equal to the net charge Q_{encl} enclosed by the surface divided by ϵ_0:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q_{\text{encl}}}{\epsilon_0}. \qquad \textbf{(22–4)}$$

Gauss's law can in principle be used to determine the electric field due to a given charge distribution, but its usefulness is mainly limited to a small number of cases, usually where the charge distribution displays much symmetry. The real importance of Gauss's law is that it is a more general and elegant statement (than Coulomb's law) for the relation between electric charge and electric field. It is one of the basic equations of electromagnetism.

Questions

1. If the electric flux through a closed surface is zero, is the electric field necessarily zero at all points on the surface? Explain. What about the converse: If $\vec{\mathbf{E}} = 0$ at all points on the surface is the flux through the surface zero?
2. Is the electric field $\vec{\mathbf{E}}$ in Gauss's law, $\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = Q_{\text{encl}}/\epsilon_0$, created only by the charge Q_{encl}?
3. What can you say about the flux through a closed surface that encloses an electric dipole?
4. The electric field $\vec{\mathbf{E}}$ is zero at all points on a closed surface; is there necessarily no net charge within the surface? If a surface encloses zero net charge, is the electric field necessarily zero at all points on the surface?
5. Define gravitational flux in analogy to electric flux. Are there "sources" and "sinks" for the gravitational field as there are for the electric field? Discuss.
6. Would Gauss's law be helpful in determining the electric field due to an electric dipole?
7. A spherical basketball (a nonconductor) is given a charge Q distributed uniformly over its surface. What can you say about the electric field inside the ball? A person now steps on the ball, collapsing it, and forcing most of the air out without altering the charge. What can you say about the field inside now?
8. In Example 22–6, it may seem that the electric field calculated is due only to the charge on the wire that is enclosed by the cylinder chosen as our gaussian surface. In fact, the entire charge along the whole length of the wire contributes to the field. Explain how the charge outside the cylindrical gaussian surface of Fig. 22–15 contributes to E at the gaussian surface. [*Hint*: Compare to what the field would be due to a short wire.]
9. Suppose the line of charge in Example 22–6 extended only a short way beyond the ends of the cylinder shown in Fig. 22–15. How would the result of Example 22–6 be altered?
10. A point charge Q is surrounded by a spherical surface of radius r_0, whose center is at C. Later, the charge is moved to the right a distance $\frac{1}{2}r_0$, but the sphere remains where it was, Fig. 22–22. Is the electric flux Φ_E through the sphere changed? Is the electric field at the surface of the sphere changed? For each "yes" answer, describe the change.

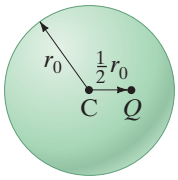


FIGURE 22–22 Question 10.

11. A solid conductor carries a net positive charge Q. There is a hollow cavity within the conductor, at whose center is a negative point charge $-q$ (Fig. 22–23). What is the charge on (*a*) the outer surface of the conductor and (*b*) the inner surface of the conductor's cavity?

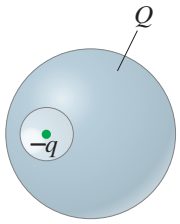


FIGURE 22–23 Question 11.

12. A point charge q is placed at the center of the cavity of a thin metal shell which is neutral. Will a charge Q placed outside the shell feel an electric force? Explain.
13. A small charged ball is inserted into a balloon. The balloon is then blown up slowly. Describe how the flux through the balloon's surface changes as the balloon is blown up. Consider both the total flux and the flux per unit surface area of the balloon.

MisConceptual Questions

1. A point charge is surrounded by a spherical gaussian surface of radius r. If the sphere is replaced by a cube of side r, Φ_E will be
   (*a*) larger.
   (*b*) smaller.
   (*c*) the same.
2. A charge Q is placed on a hollow metal ball. We saw in Chapter 21 that the charge is all on the surface of the ball because metal is a conductor. How does the charge distribute itself on the ball?
   (*a*) Half on the inside surface and half on the outside surface.
   (*b*) Part on each surface in inverse proportion to the two radii.
   (*c*) Part on each surface but with a more complicated dependence on the radii than in answer (*b*).
   (*d*) All on the inside surface.
   (*e*) All on the outside surface.
3. Which statement below is correct for using Gauss's law to find the electric field at any point on the surface of the non-conducting object shown in Fig. 22–24 below?
   (*a*) Gauss's law doesn't hold for this object as there is insufficient symmetry.
   (*b*) Gauss's law is true but not useful for this situation.
   (*c*) Gauss's law is useful in this situation; just draw an imaginary gaussian sphere just inside the shape.
   (*d*) Gauss's law is useful in this situation; just draw an imaginary gaussian sphere just outside the shape.

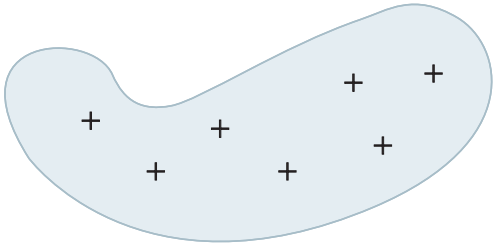


FIGURE 22–24 MisConceptual Question 3.

4. If the Earth has a large charge $+Q$ uniformly distributed over its surface, it generates an electric field of strength E at its surface. If this same charge were uniformly distributed over the Moon, what electric field strength would exist at the Moon's surface? The Moon's radius is about one-quarter that of the Earth.
(*a*) $16E$. (*b*) $4E$. (*c*) E.
(*d*) $\frac{1}{4}E$. (*e*) $\frac{1}{16}E$.

5. A $+Q$ point charge is located at the center of a hollow metal sphere. The sphere is also charged, carrying an equal but opposite total charge of $-Q$. Consider the region between the point charge and the inside of the spherical shell. Which statement is true?
(*a*) The electric field is only due to the point charge ($E = kQ/r^2$ pointing out).
(*b*) The electric field is zero.
(*c*) The electric field is twice that due to the point charge.
(*d*) The electric field is larger than for the point charge, but not twice as large.

6. For the situation described in MisConceptual Question 5, consider the region outside the spherical metal shell. Which statement is true?
(*a*) The electric field is just that due to the point charge ($E = kQ/r^2$ pointing out).
(*b*) The electric field is kQ/r^2 pointing in.
(*c*) The electric field points in, but is not kQ/r^2.
(*d*) The electric field is zero.
(*e*) The electric field points out, but is not kQ/r^2.

7. Two charged flies are buzzing around inside a sealed uncharged paper bag. The total electric flux exiting the bag is zero. Which statement is true?
(*a*) The total flux cannot be zero since the flies are charged.
(*b*) The electric field at the surface of the bag is zero.
(*c*) The flies have equal and opposite charges.
(*d*) The induced charge on the bag itself cancels out the charge on the flies to give zero flux.

8. A point charge, $+Q$, is at the center of a metal cubical box of side S. The metal box carries a net charge $-3Q$. Why can't you use Gauss's law to show that the electric field at all points just outside the box is given by $E = -2Q/(S^2\epsilon_0)$?
(*a*) Gauss's law only applies to spherical and cylindrical surfaces.
(*b*) The answer is really $E = -2Q/(6S^2\epsilon_0)$.
(*c*) The enclosed charge is $4Q$.
(*d*) The electric field is not uniform across a side of the box.
(*e*) $E = -2Q/(4\pi S^2\epsilon_0)$.

9. What charge would you put in Gauss's law in order to calculate the electric field at point A in Fig. 22–25? The outer conducting shell has a total charge of +2 C. The middle layer of the sphere is an insulator, and the innermost sphere is hollow with a charge of −3 C at the center.
(*a*) 0 C. (*b*) 2 C.
(*c*) −3 C. (*d*) −1 C.
(*e*) Gauss's law won't work for this situation.

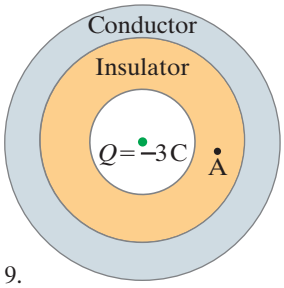


FIGURE 22–25
MisConceptual Question 9.

10. A nonconducting sphere of radius r_0 contains a net charge $+Q$, distributed uniformly. Which statements are not true regarding the electric field due to the charge on this sphere?
(*a*) The electric field above the surface of the sphere is uniform, nonzero, and points outward.
(*b*) The electric field outside the sphere is constant since the charge enclosed by a gaussian sphere concentric with the charged sphere is constant for any $r > r_0$.
(*c*) The electric field is 0 at the center of the sphere.
(*d*) The electric field is constant inside the sphere because for concentric gaussian spheres inside the charged sphere, the surface area and volume enclosed both increase with r.
(*e*) At $r = r_0/2$, the electric field has half the value of the electric field on the surface of the sphere.

Problems

22–1 Electric Flux

1. (I) A uniform electric field of magnitude 6.4×10^2 N/C passes through a circle of radius 13 cm. What is the electric flux through the circle when its face is (*a*) perpendicular to the field lines, (*b*) at 45° to the field lines, and (*c*) parallel to the field lines?

2. (I) The Earth possesses an electric field of (average) magnitude 150 N/C near its surface. The field points radially inward. Calculate the net electric flux outward through a spherical surface surrounding, and just beyond, the Earth's surface.

3. (II) A cube of side ℓ is placed in a uniform field E_0 with edges parallel to the field lines. (*a*) What is the net flux through the cube? (*b*) What is the flux through each of its six faces?

4. (II) A uniform field $\vec{\mathbf{E}}$ is parallel to the axis of a hollow hemisphere of radius r, Fig. 22–26. (*a*) What is the electric flux through the hemispherical surface? (*b*) What is the result if $\vec{\mathbf{E}}$ is instead vertical, perpendicular to the axis and still uniform?

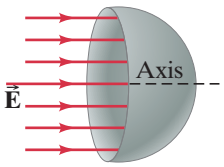


FIGURE 22–26 Problem 4.

22–2 Gauss's Law

5. (I) The total electric flux out from a cubical box of side 21.0 cm is 1.85×10^3 N·m²/C. What charge is enclosed by the box?

6. (I) Figure 22–27 shows five closed surfaces that surround various charges in a plane, as indicated. Determine the electric flux through each surface, S_1, S_2, S_3, S_4, and S_5. The surfaces are flat "pillbox" surfaces that extend only slightly above and below the plane in which the charges lie.

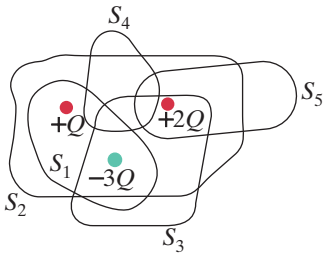


FIGURE 22–27 Problem 6.

7. (II) In Fig. 22–28, two objects, O_1 and O_2, have charges $+1.0\,\mu C$ and $-2.0\,\mu C$, respectively, and a third object, O_3, is electrically neutral. (*a*) What is the electric flux through the surface A_1 that encloses all the three objects? (*b*) What is the electric flux through the surface A_2 that encloses the third object only?

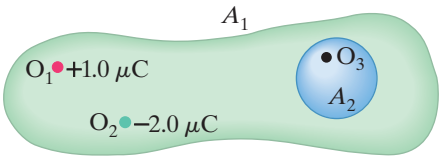


FIGURE 22–28 Problem 7.

8. (II) A thin circular ring of charge with uniform linear charge density (as in Fig. 21–29) is completely enclosed by an imaginary hollow donut shape. An exact copy of the ring is completely enclosed by an imaginary hollow sphere. What is the ratio of the flux out of the donut shape to that out of the sphere?

9. (II) In a certain region of space, the electric field is constant in direction (say horizontal, in the x direction), but its magnitude decreases from $E = 650\,N/C$ at $x = 0$ to $E = 370\,N/C$ at $x = 25\,m$. Determine the charge within a cubical box of side $\ell = 25\,m$, where the box is oriented so that four of its sides are parallel to the field lines (Fig. 22–29).

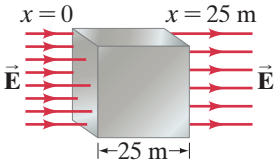


FIGURE 22–29 Problem 9.

10. (II) A point charge Q is placed at the center of a cube of side ℓ. What is the flux through one face of the cube?

11. (II) A 25.0-cm-long uniformly charged plastic rod is sealed inside a plastic bag. The total electric flux leaving the bag is $7.3 \times 10^5\,N \cdot m^2/C$. What is the linear charge density on the rod?

22–3 Applications of Gauss's Law

12. (I) Draw the electric field lines around a negatively charged metal egg.

13. (I) The field just outside a 3.50-cm-radius metal sphere is $5.85 \times 10^2\,N/C$ and points toward the sphere. What charge resides on the sphere?

14. (I) Starting from the result of Example 22–3, show that the electric field just outside a uniformly charged spherical conductor is $E = \sigma/\epsilon_0$, consistent with Example 22–8.

15. (I) A long thin wire, hundreds of meters long, carries a uniformly distributed charge of $-8.4\,\mu C$ per meter of length. Estimate the magnitude and direction of the electric field at points (*a*) 5.0 m, and (*b*) 1.5 m, perpendicular from the center of the wire.

16. (I) A metal globe of the Earth has 1.50 mC of charge put on it at the north pole. Then -3.00 mC of charge is applied to the south pole. Draw the field lines for this system after it has come to equilibrium.

17. (II) A point charge Q rests at the center of an uncharged thin spherical conducting shell. (See Fig. 22–30.) What is the electric field E as a function of r (*a*) for r less than the inner radius of the shell, (*b*) inside the shell, and (*c*) beyond the shell? (*d*) How does the shell affect the field due to Q alone? (*e*) How does the charge Q affect the shell?

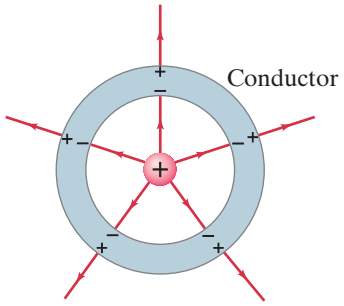


FIGURE 22–30 Problem 17.

18. (II) A solid metal sphere of radius 3.00 m carries a total charge of $-4.80\,\mu C$. What is the magnitude of the electric field at a distance from the sphere's center of (*a*) 0.250 m, (*b*) 2.90 m, (*c*) 3.10 m, and (*d*) 8.00 m? How would the answers differ if the sphere were (*e*) a thin shell, or (*f*) a solid nonconductor uniformly charged throughout?

19. (II) A 15.0-cm-diameter nonconducting sphere carries a total charge of $3.25\,\mu C$ distributed uniformly throughout its volume. Graph the electric field E as a function of the distance r from the center of the sphere from $r = 0$ to $r = 30.0$ cm.

20. (II) A flat square sheet of thin aluminum foil, 25 cm on a side, carries a uniformly distributed 245 nC charge. What, approximately, is the electric field (*a*) 1.0 cm above the center of the sheet and (*b*) 15 m above the center of the sheet?

21. (II) A spherical cavity of radius 4.50 cm is at the center of a metal sphere of radius 18.0 cm. A point charge $Q = 6.50\,\mu C$ rests at the very center of the cavity, whereas the metal conductor carries no net charge. Determine the electric field at a point (*a*) 3.00 cm from the center of the cavity, (*b*) 6.00 cm from the center of the cavity, (*c*) 30.0 cm from the center.

22. (II) A solid nonconducting sphere of radius 12.0 cm is made of two layers. The innermost section has a radius of 6.0 cm and a uniform charge density of $-4.0\,C/m^3$. The outer layer has a uniform charge density of $+9.0\,C/m^3$. Determine the electric field for (*a*) $0 < r < 6.0$ cm, (*b*) $6.0\text{ cm} < r < 12.0$ cm, and (*c*) $12.0\text{ cm} < r < 50.0$ cm. (*d*) Plot the magnitude of the electric field for $0 < r < 50.0$ cm. Is the field continuous at the edges of the layers?

23. (II) A solid metal cube has a spherical cavity right at its center. At the center of the cavity there is a point charge $Q = +8.00\,\mu C$. The metal cube carries a net charge $q = -6.40\,\mu C$ (not including Q). Determine (*a*) the total charge on the surface of the spherical cavity and (*b*) the total charge on the outer surface of the cube.

24. (II) Two large, flat metal plates are separated by a distance that is very small compared to their height and width. The conductors are given equal but opposite uniform surface charge densities $\pm\sigma$. Ignore edge effects and use Gauss's law to show (*a*) that for points far from the edges, the electric field between the plates is $E = \sigma/\epsilon_0$ and (*b*) that outside the plates on either side the field is zero. (*c*) How would your results be altered if the two plates were nonconductors? (See Fig. 22–31).

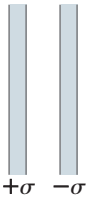


FIGURE 22–31
Problems 24, 25, and 26.

25. (II) Suppose the two conducting plates in Problem 24 have the *same* sign and magnitude of charge. What then will be the electric field (*a*) between them and (*b*) outside them on either side? (*c*) What if the plates are nonconducting?

26. (II) The electric field between two parallel square metal plates is 130 N/C. The plates are 0.65 m on a side and are separated by 3.0 cm. What is the charge on each plate (assume equal and opposite)? Neglect edge effects.

27. (II) Two thin concentric spherical shells of radii r_1 and r_2 $(r_1 < r_2)$ contain uniform surface charge densities σ_1 and σ_2, respectively (see Fig. 22–32). Determine the electric field for (*a*) $0 < r < r_1$, (*b*) $r_1 < r < r_2$, and (*c*) $r > r_2$. (*d*) Under what conditions will $E = 0$ for $r > r_2$? (*e*) Under what conditions will $E = 0$ for $r_1 < r < r_2$? Neglect the thickness of the shells.

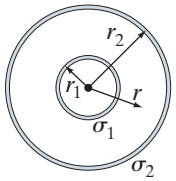


FIGURE 22–32 Two spherical shells (Problem 27).

28. (II) A spherical rubber balloon carries a total charge Q uniformly distributed on its surface. At $t = 0$ the nonconducting balloon has radius r_0 and the balloon is then slowly blown up so that r increases linearly to $2r_0$ in a time T. Determine the electric field as a function of time (*a*) just outside the balloon surface and (*b*) at $r = 3.2r_0$.

29. (II) Suppose the nonconducting sphere of Example 22–4 has a spherical cavity of radius r_1 centered at the sphere's center (Fig. 22–33). Assuming the charge Q is distributed uniformly in the "shell" (between $r = r_1$ and $r = r_0$), determine the electric field as a function of r for (*a*) $0 < r < r_1$, (*b*) $r_1 < r < r_0$, and (*c*) $r > r_0$.

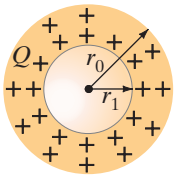


FIGURE 22–33
Problems 29, 30, 31, and 44.

30. (II) Suppose in Fig. 22–33, Problem 29, there is also a charge $-q$ at the center of the cavity. Determine the electric field for (*a*) $0 < r < r_1$, (*b*) $r_1 < r < r_0$, and (*c*) $r > r_0$.

31. (II) Suppose the thick spherical shell of Problem 29 is a conductor. It carries a total net charge Q and at its center there is a point charge $+q$. What total charge is found on (*a*) the inner surface of the shell and (*b*) the outer surface of the shell? Determine the electric field for (*c*) $0 < r < r_1$, (*d*) $r_1 < r < r_0$, and (*e*) $r > r_0$.

32. (II) Suppose that at the center of the cavity inside the shell (charge Q) of Fig. 22–11 (and Example 22–3), there is a point charge q $(\neq \pm Q)$. Determine the electric field for (*a*) $0 < r < r_0$ and for (*b*) $r > r_0$. What are your answers if (*c*) $q = Q$ and (*d*) $q = -Q$?

33. (II) A long cylindrical shell of radius R_0 and length ℓ $(R_0 \ll \ell)$ possesses a uniform surface charge density (charge per unit area) σ (Fig. 22–34). Determine the electric field at points R: (*a*) outside the cylinder $(R > R_0)$ and (*b*) inside the cylinder $(0 < R < R_0)$; assume the points are far from the ends and not too far from the shell $(R \ll \ell)$. (*c*) Compare to the result for a long line of charge, Example 22–6. Neglect the thickness of shell.

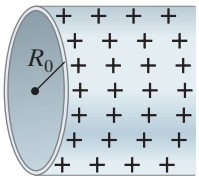


FIGURE 22–34
Problem 33.

34. (II) A very long solid nonconducting cylinder of radius R_0 and length ℓ $(R_0 \ll \ell)$ possesses a uniform volume charge density $\rho_E\,(\mathrm{C/m^3})$, Fig. 22–35. Determine the electric field at points (*a*) outside the cylinder $(R > R_0)$ and (*b*) inside the cylinder $(R < R_0)$. Do only for points far from the ends and for which $R \ll \ell$.

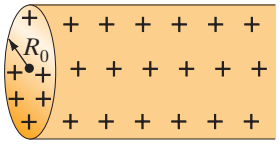


FIGURE 22–35
Problem 34.

35. (II) A thin cylindrical shell of radius R_1 is surrounded by a second concentric cylindrical shell of radius R_2 (Fig. 22–36). The inner shell has a total charge $+Q$ and the outer shell $-Q$. Assuming the length ℓ of the shells is much greater than R_1 or R_2, determine the electric field as a function of R (the perpendicular distance from the common axis of the cylinders) for (*a*) $0 < R < R_1$, (*b*) $R_1 < R < R_2$, and (*c*) $R > R_2$. (*d*) What is the kinetic energy of an electron if it moves between (and concentric with) the shells in a circular orbit of radius $(R_1 + R_2)/2$? Neglect thickness of shells.

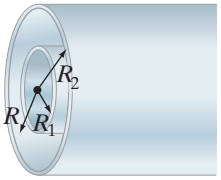


FIGURE 22–36
Problems 35, 36, and 37.

36. (II) A thin cylindrical shell of radius $R_1 = 4.5$ cm is surrounded by a second cylindrical shell of radius $R_2 = 9.0$ cm, as in Fig. 22–36. Both cylinders are 5.0 m long and the inner one carries a total charge $Q_1 = -0.88$ nC and the outer one $Q_2 = +1.56$ nC. For points far from the ends of the cylinders, determine the electric field at a radial distance R from the central axis of (*a*) 3.0 cm, (*b*) 7.0 cm, and (*c*) 12.0 cm.

37. (II) (*a*) If an electron $(m = 9.1 \times 10^{-31}\,\mathrm{kg})$ escaped from the surface of the inner cylinder in Problem 36 (Fig. 22–36) with negligible speed, what would be its speed when it reached the outer cylinder? (*b*) If a proton $(m = 1.67 \times 10^{-27}\,\mathrm{kg})$ revolves in a circular orbit of radius $R = 7.0$ cm about the axis (i.e., between the cylinders), what must be its speed?

38. (II) A nonconducting sphere of radius r_0 is uniformly charged with volume charge density ρ_E. It is surrounded by a concentric metal (conducting) spherical shell of inner radius r_1 and outer radius r_2, which carries a net charge $+Q$. Determine the resulting electric field in the regions (*a*) $0 < r < r_0$, (*b*) $r_0 < r < r_1$, (*c*) $r_1 < r < r_2$, and (*d*) $r > r_2$ where the radial distance r is measured from the center of the nonconducting sphere.

39. (II) A very long solid nonconducting cylinder of radius R_1 is uniformly charged with a volume charge density ρ_E. It is surrounded by a concentric cylindrical tube of inner radius R_2 and outer radius R_3 as shown in Fig. 22–37, and it too carries a uniform charge density ρ_E. Determine the electric field as a function of the distance R from the center of the cylinders for (*a*) $0 < R < R_1$, (*b*) $R_1 < R < R_2$, (*c*) $R_2 < R < R_3$, and (*d*) $R > R_3$. (*e*) If $\rho_E = 15\,\mu C/m^3$ and $R_1 = \frac{1}{2}R_2 = \frac{1}{3}R_3 = 5.0$ cm, plot E as a function of R from $R = 0$ to $R = 20.0$ cm. Assume the cylinders are very long compared to R_3.

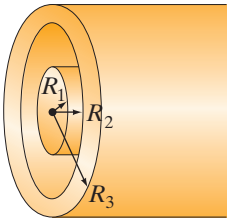


FIGURE 22–37
Problem 39.

40. (II) A very long solid nonconducting cylinder of radius R_1 is uniformly charged with charge density ρ_E. It is surrounded by a cylindrical metal (conducting) tube of inner radius R_2 and outer radius R_3, which has no net charge (cross-sectional view shown in Fig. 22–38). If the axes of the two cylinders are parallel, but displaced from each other by a distance d, determine the resulting electric field in the region $R > R_3$, where the radial distance R is measured from the metal cylinder's axis. Assume $d < (R_2 - R_1)$.

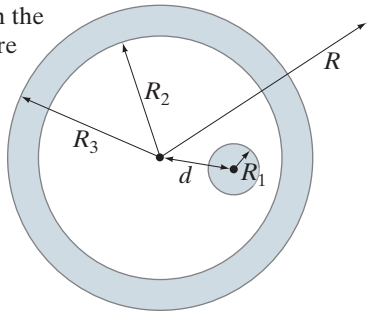


FIGURE 22–38
Problem 40.

41. (II) A flat disk-like ring (inner radius R_0, outer radius $4R_0$) is uniformly charged. In terms of the total charge Q, determine the electric field on the axis at points (*a*) $0.25R_0$ and (*b*) $75R_0$ from the center of the ring. [*Hint*: The ring can be replaced with two oppositely charged superposed disks.]

42. (II) An uncharged solid conducting sphere of radius r_0 contains two spherical cavities of radii r_1 and r_2, respectively. Point charge Q_1 is then placed within the cavity of radius r_1 and point charge Q_2 is placed within the cavity of radius r_2 (Fig. 22–39). Determine the resulting electric field (magnitude and direction) at locations outside the solid sphere ($r > r_0$), where r is the distance from its center.

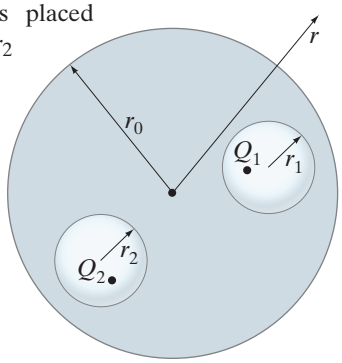


FIGURE 22–39
Problem 42.

43. (III) A very large (i.e., assume infinite) flat slab of nonconducting material has thickness d and a uniform volume charge density $+\rho_E$. (*a*) Show that a uniform electric field exists outside of this slab. Determine its magnitude E and its direction (relative to the slab's surface). (*b*) As shown in Fig. 22–40, the slab is now aligned so that one of its surfaces lies on the line $y = x$. At time $t = 0$, a pointlike particle (mass m, charge $+q$) is located at position $\vec{\mathbf{r}} = +y_0\,\hat{\mathbf{j}}$ and has velocity $\vec{\mathbf{v}} = v_0\,\hat{\mathbf{i}}$. Show that the particle will collide with the slab if $v_0 \geq \sqrt{\sqrt{2}qy_0\rho_E d/\epsilon_0 m}$. Ignore gravity.

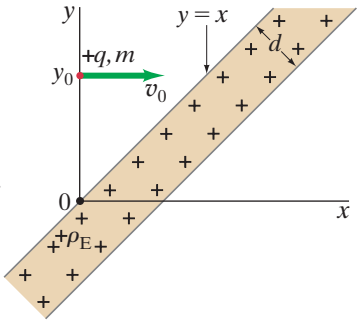


FIGURE 22–40
Problem 43.
Cross-section of slab.

44. (III) Suppose the density of charge between r_1 and r_0 of the hollow sphere of Problem 29 (Fig. 22–33) varies as a function of r as $\rho_E = \rho_0 r_1/r$. Determine the electric field as a function of r for (*a*) $0 < r < r_1$, (*b*) $r_1 < r < r_0$, and (*c*) $r > r_0$. (*d*) Plot E versus r from $r = 0$ to $r = 2r_0$, assuming $r_1 = 0.50\,r_0$.

45. (III) Suppose two thin flat nonconducting plates measure $1.0\text{ m} \times 1.0\text{ m}$ and are separated by 5.0 mm. They are oppositely charged with $\pm 25\,\mu C$. (*a*) Estimate the total force exerted by one plate on the other (ignore edge effects). (*b*) How much work would be required to move the plates from 5.0 mm apart to 1.00 cm apart?

46. (III) A flat slab of nonconducting material (Fig. 22–41) carries a uniform charge per unit volume, ρ_E. The slab has thickness d which is small compared to the height and breadth of the slab. Determine the electric field as a function of x (*a*) inside the slab and (*b*) outside the slab (at distances much less than the slab's height or breadth). Take the origin at the center of the slab.

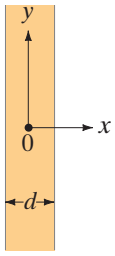


FIGURE 22–41
Problem 46.

47. (III) A flat slab of nonconducting material has thickness $2d$, which is small compared to its height and breadth. Define the x axis to be along the direction of the slab's thickness with the origin at the center of the slab (Fig. 22–42). If the slab carries a volume charge density $\rho_E(x) = -\rho_0$ in the region $-d \leq x < 0$, and $\rho_E(x) = +\rho_0$ in the region $0 < x \leq +d$, determine the electric field $\vec{\mathbf{E}}$ as a function of x in the regions (*a*) outside the slab, (*b*) $0 < x \leq +d$, and (*c*) $-d \leq x < 0$. Let ρ_0 be a positive constant.

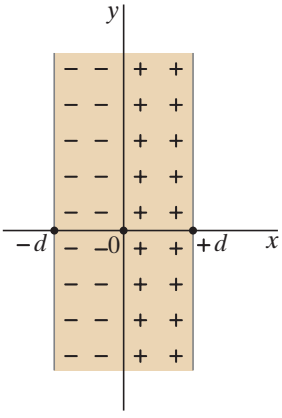


FIGURE 22–42
Problem 47.

48. (III) An extremely long, solid nonconducting cylinder has a radius R_0. The charge density within the cylinder is a function of the distance R from the axis, given by $\rho_E(R) = \rho_0(R/R_0)^2$. What is the electric field everywhere inside and outside the cylinder (far away from the ends) in terms of ρ_0 and R_0?

49. (III) Charge is distributed within a solid sphere of radius r_0 in such a way that the charge density is a function of the radial position within the sphere of the form: $\rho_E(r) = \rho_0(r/r_0)$. If the total charge within the sphere is Q (and positive), what is the electric field everywhere within the sphere in terms of Q, r_0, and the radial position r?

General Problems

50. A point charge produces an electric flux of $+235\,\text{N}\cdot\text{m}^2/\text{C}$ through a gaussian sphere of radius 45.0 cm centered on the charge. (*a*) What is the flux through a gaussian sphere with a radius 32.5 cm? (*b*) What is the magnitude and sign of the charge?

51. A point charge of 8.20 nC is located at the origin and a second charge of -5.00 nC is located on the x axis at $x = 2.75$ cm. Calculate the electric flux through a sphere centered at the origin with radius 1.00 m. Repeat the calculation for a sphere of radius 2.00 m.

52. The Earth is surrounded by an electric field, pointing inward at every point, of magnitude $E \approx 150\,\text{N/C}$ near the Earth's surface. (*a*) What is the net charge on the Earth? (*b*) How many excess electrons per square meter on the Earth's surface does this correspond to?

53. Write Gauss's law for the gravitational field $\vec{\mathbf{g}}$ (see Section 6–8).

54. A solid nonconducting sphere of radius r_0 has a total charge Q which is distributed according to $\rho_E = br$, where ρ_E is the charge per unit volume, or charge density (C/m^3), and b is a constant. Determine (*a*) b in terms of Q, (*b*) the electric field at points inside the sphere, and (*c*) the electric field at points outside the sphere.

55. A cube of side ℓ has one corner at the origin of coordinates, and extends along the positive x, y, and z axes. Suppose the electric field in this region is given by $\vec{\mathbf{E}} = (ay + b)\hat{\mathbf{j}}$. Determine the charge inside the cube.

56. Three large but thin charged sheets are parallel to each other as shown in Fig. 22–43. Sheet I has a total surface charge density of $7.5\,\text{nC/m}^2$, sheet II has a charge density of $-3.0\,\text{nC/m}^2$, and sheet III has a charge density of $6.0\,\text{nC/m}^2$. Estimate the force per unit area on each sheet, in N/m^2.

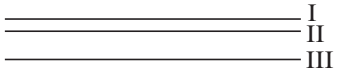


FIGURE 22–43 Problem 56.

57. A point charge Q is placed a distance $r_0/2$ above the surface of an imaginary spherical surface of radius r_0 (Fig. 22–44). (*a*) What is the total electric flux through the sphere? (*b*) What range of values does E have at the surface of the sphere? (*c*) Is $\vec{\mathbf{E}}$ perpendicular to the sphere at all points? (*d*) Is Gauss's law useful for obtaining E at the surface of the sphere?

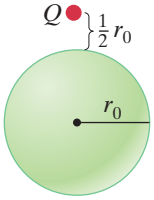


FIGURE 22–44 Problem 57.

58. A point charge Q is on the axis of a short cylinder at its center. The diameter of the cylinder is equal to its length ℓ (Fig. 22–45). What is the total flux through the curved sides of the cylinder? [*Hint*: First calculate the flux through the ends.]

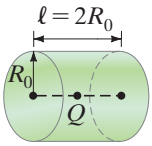


FIGURE 22–45 Problem 58.

59. The neutral hydrogen atom can be modeled as a positive point charge $+1.6 \times 10^{-19}$ C surrounded by a distribution of negative charge with volume density given by $\rho_E(r) = -Ae^{-2r/a_0}$ where $a_0 = 0.53 \times 10^{-10}$ m is called the *Bohr radius*, and $e = 2.718\cdots$ is the base of the natural log. (*a*) Determine the constant A so that the total amount of negative charge is -1.6×10^{-19} C. (*b*) What is the net charge inside a sphere of radius a_0? (*c*) What is the strength of the electric field at a distance a_0 from the nucleus? [*Hint*: Do not confuse the exponential number e with the elementary charge e which uses the same symbol but has a completely different meaning and value $(e = 1.6 \times 10^{-19}\,\text{C})$.]

60. A very large thin plane has uniform surface charge density σ. Touching it on the right (Fig. 22–46) is a long wide slab of thickness d with uniform volume charge density ρ_E. Determine the electric field (*a*) to the left of the plane, (*b*) to the right of the slab, and (*c*) everywhere inside the slab.

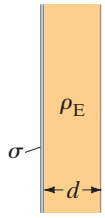


FIGURE 22–46 Problem 60.

61. A sphere of radius r_0 carries a uniform volume charge density ρ_E (Fig. 22–47). A spherical cavity of radius $r_0/2$ is then scooped out and left empty, as shown. (*a*) What is the magnitude and direction of the electric field at point A? (*b*) What is the direction and magnitude of the electric field at point B? Points A and C are at the centers of the respective spheres. [*Hint*: use the principle of superposition; subtract one sphere from the other.]

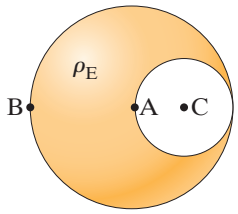


FIGURE 22–47 Problem 61.

62. Dry air will break down and generate a spark if the electric field exceeds about $3 \times 10^6\,\text{N/C}$. How much charge could be packed onto the surface of a green pea (diameter 0.75 cm) before the pea spontaneously discharges?

63. Three very large sheets are separated by equal distances of 15.0 cm (Fig. 22–48). The first and third sheets are very thin and nonconducting and have charge per unit area σ of $+5.00\,\mu\text{C/m}^2$ and $-5.00\,\mu\text{C/m}^2$ respectively. The middle sheet is conducting but has no net charge. (*a*) What is the electric field inside the middle sheet? What is the electric field (*b*) between the left and middle sheets, and (*c*) between the middle and right sheets? (*d*) What is the charge density on the surface of the middle sheet facing the left sheet, and (*e*) on the surface facing the right sheet?

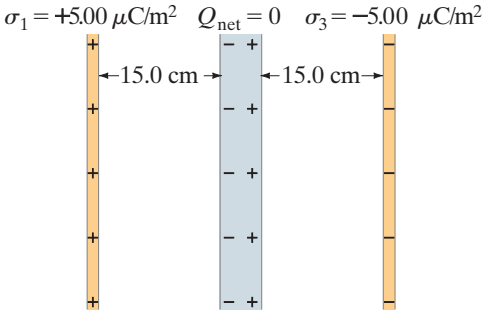


FIGURE 22–48 Problem 63.

64. In a cubical volume, 0.50 m on a side, the electric field is

$$\vec{\mathbf{E}} = E_0\left(1 + \frac{z}{a}\right)\hat{\mathbf{i}} + E_0\left(\frac{z}{a}\right)\hat{\mathbf{j}}$$

where $E_0 = 0.125\,\text{N/C}$ and $a = 0.50\,\text{m}$. The cube has its sides parallel to the coordinate axes, Fig. 22–49. Determine the net charge within the cube.

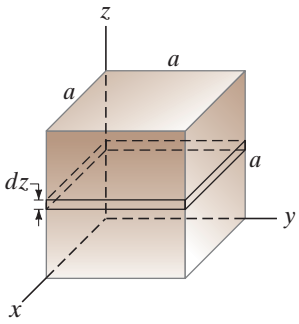


FIGURE 22–49
Problem 64.

65. A conducting spherical shell (Fig. 22–50) has inner radius = 10.0 cm, outer radius = 15.0 cm, and has a $+3.0\,\mu\text{C}$ point charge at the center. A charge of $-3.0\,\mu\text{C}$ is put on the conductor. (*a*) Where on the conductor does the $-3.0\,\mu\text{C}$ end up? (*b*) What is the electric field both inside and outside the shell?

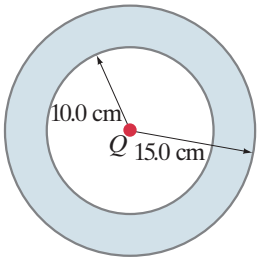


FIGURE 22–50 Problem 65.

66. A hemisphere of radius R is placed in a charge-free region of space where a uniform electric field exists of magnitude E directed perpendicular to the hemisphere's circular base (Fig. 22–51). (*a*) Using the definition of Φ_E through an "open" surface, calculate (via explicit integration) the electric flux through the hemisphere. [*Hint*: In Fig. 22–51 you can see that, on the surface of a sphere, the infinitesimal area located between the angles θ and $\theta + d\theta$ is $dA = (2\pi R\sin\theta)(R\,d\theta) = 2\pi R^2\sin\theta\,d\theta$.] (*b*) Choose an appropriate gaussian surface and use Gauss's law to much more easily obtain the same result for the electric flux through the hemisphere.

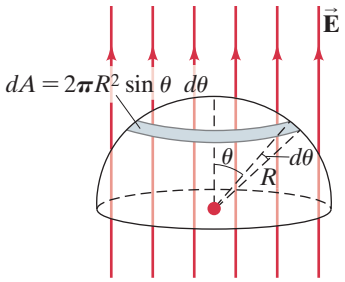


FIGURE 22–51 Problem 66.

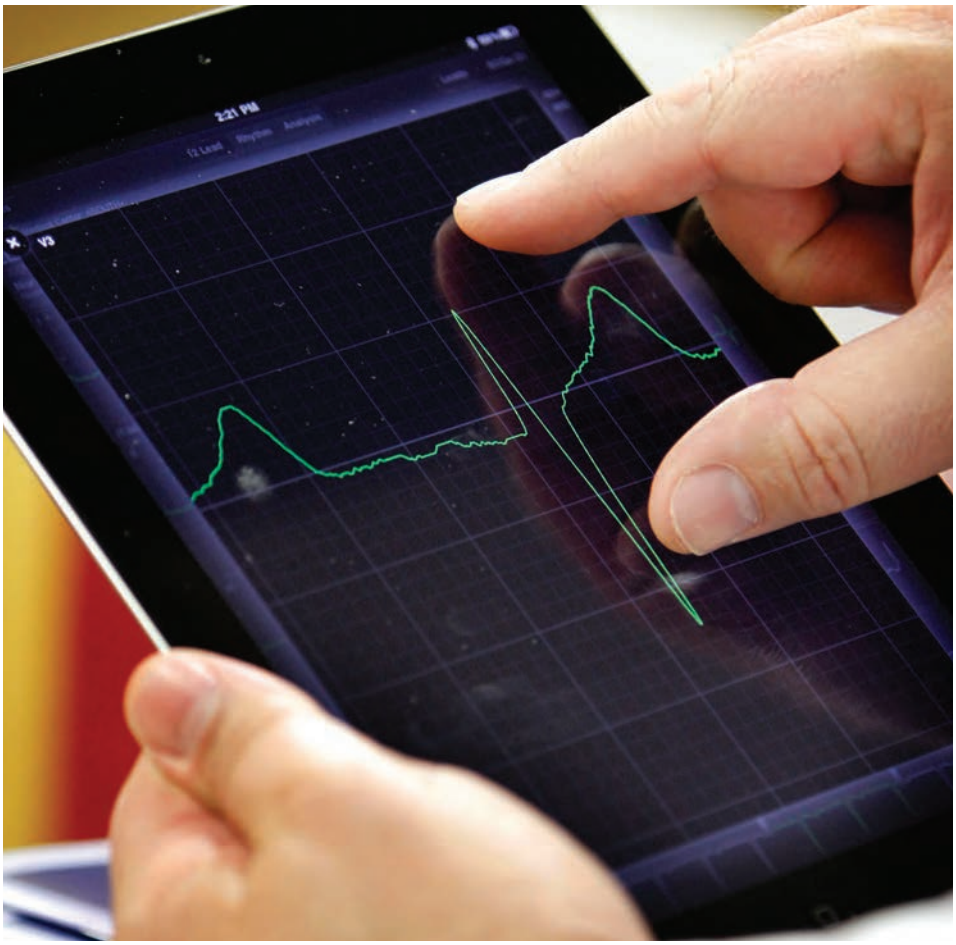
ANSWERS TO EXERCISES

A: (*d*).
B: (*a*).
C: (*d*).
D: (*c*).
E: (*c*).

We are used to voltage in our lives—a 12-volt car battery, 110 V or 220 V at home, 1.5-volt flashlight batteries, and so on. Here we see displayed the voltage produced across a human heart, known as an electrocardiogram (ECG).

Voltage is the same as electric potential difference between two points. Electric potential is defined as the potential energy per unit charge.

We discuss voltage and its relation to electric field, as well as the electric potential due to point charges and charge distributions. Also discussed are digital voltages, binary numbers and digital electronics, TV and computer monitors, and digital TV.



CHAPTER 23

Electric Potential

CONTENTS

23 – 1 Electric Potential Energy and Potential Difference
23 – 2 Relation between Electric Potential and Electric Field
23 – 3 Electric Potential Due to Point Charges
23 – 4 Potential Due to Any Charge Distribution
23 – 5 Equipotential Lines and Surfaces
23 – 6 Potential Due to Electric Dipole; Dipole Moment
23 – 7 $\vec{\mathbf{E}}$ Determined from V
23 – 8 Electrostatic Potential Energy; the Electron Volt
23 – 9 Digital; Binary Numbers; Signal Voltage
*23–10 TV and Computer Monitors
*23–11 Electrocardiogram (ECG or EKG)

CHAPTER-OPENING QUESTION—Guess now!

When two positively charged small spheres are pushed toward each other, what happens to their electric potential energy?

(a) It remains unchanged.
(b) It decreases.
(c) It increases.
(d) There is no potential energy in this situation.

We saw in Chapters 7 and 8 that the concept of energy was extremely valuable in dealing with the subject of mechanics. For one thing, energy is found to be a conserved quantity and is thus an important tool for understanding nature. We also saw that many Problems could be solved using energy even when a knowledge of the forces involved was not possible, or when using Newton's laws would have been too difficult.

The energy perspective is useful also in the study of electric forces and electricity. It not only extends the law of conservation of energy, but it gives us another way to view electrical phenomena. The energy concept is also a tool in solving Problems more easily in many cases than by using forces and electric fields.

23–1 Electric Potential Energy and Potential Difference

Electric Potential Energy

To apply conservation of energy to electric phenomena, we need to define electric potential energy as we did for other types of potential energy. We saw in Chapter 8 that potential energy can be defined only for a conservative force. The work done by a conservative force in moving an object between any two positions is independent of the path taken. The electrostatic force between any two charges (Eq. 21–1, $F = kQ_1 Q_2/r^2$) is conservative because the dependence on position is just like the gravitational force, $1/r^2$, which we saw in Section 8–7 is conservative. Hence we can define potential energy U for the electrostatic force.

We saw in Chapter 8 that the change in potential energy between any two points, a and b, equals the negative of the work done by the conservative force on an object as it moves from point a to point b: $\Delta U = -W$.

Thus we define the change in electric potential energy, $U_b - U_a$, when a point charge q moves from some point a to another point b, as the negative of the work done by the electric force on the charge as it moves from point a to point b. For example, consider the electric field between two equally but oppositely charged parallel plates; we assume their separation is small compared to their width and height, so the field $\vec{\mathbf{E}}$ will be uniform over most of the region, Fig. 23–1. Now consider a tiny positive point charge q placed at the point "a" very near the positive plate as shown. The magnitude of this charge q is so small that it has no effect on $\vec{\mathbf{E}}$. If this charge q at point a is released, the electric force will do work on the charge and accelerate it toward the negative plate. The work W done by the electric field E to move the charge a distance d is (using Eq. 21–5, $F = qE$)

$$W = Fd = qEd \qquad [\text{uniform } \vec{\mathbf{E}}]$$

The change in electric potential energy equals the negative of the work done by the electric force:

$$U_b - U_a = -W = -qEd \qquad [\text{uniform } \vec{\mathbf{E}}] \quad \textbf{(23–1)}$$

for this case of uniform electric field $\vec{\mathbf{E}}$. In the case illustrated, the potential energy decreases (ΔU is negative); and as the charged particle accelerates from point a to point b in Fig. 23–1, the particle's kinetic energy K increases—by an equal amount. In accord with the conservation of energy, electric potential energy is transformed into kinetic energy, and the total energy is conserved. Note that the positive charge q has its greatest potential energy at point a, near the positive plate.[†] The reverse is true for a negative charge: its potential energy is greatest near the negative plate.

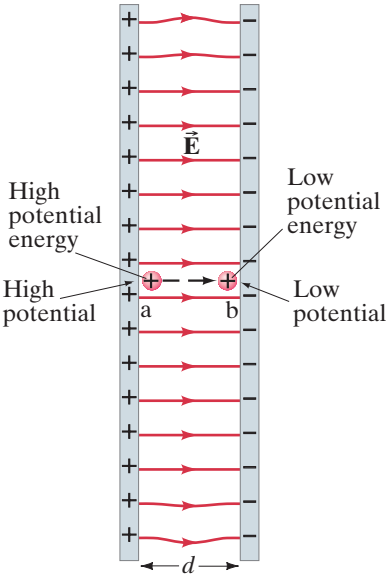


FIGURE 23–1 Work is done by the electric field $\vec{\mathbf{E}}$ in moving the positive charge from position a to position b.

Electric Potential and Potential Difference

In Chapter 21, we defined the electric field as the force per unit charge. Similarly, it is useful to define the **electric potential** (or simply the **potential** when "electric" is understood) as the *electric potential energy per unit charge*. Electric potential is given the symbol V. If we put a tiny positive test charge q in an electric field, it will have electric potential energy U_a at some arbitrary point a (relative to some zero potential energy); then the electric potential V_a at this point is defined as

$$V_a = \frac{U_a}{q}. \qquad \textbf{(23–2a)}$$

As we discussed in Chapter 8, only differences in potential energy are physically meaningful. Hence only the **difference in potential**, or the **potential difference**, between two points a and b (such as in Fig. 23–1) is measurable. When the electric force does positive work on a charge, the kinetic energy increases and the potential energy decreases.

[†]At point a, the positive charge q has its greatest ability to do work (on some other object or system).

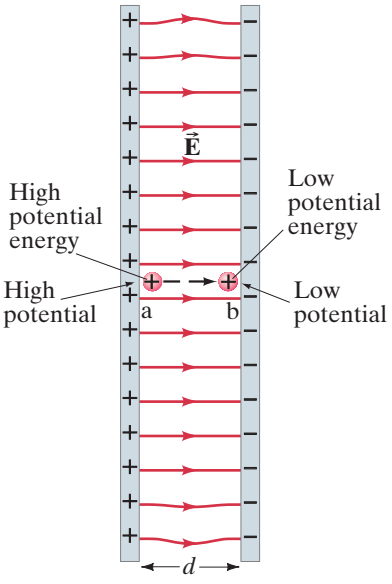
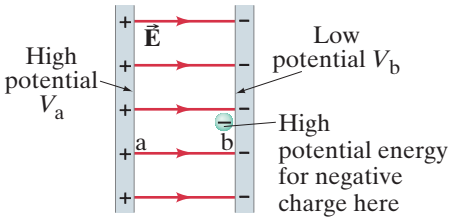


FIGURE 23–1 (Repeated.) Work is done by the electric field $\vec{\mathbf{E}}$ in moving the positive charge from position a to position b.

FIGURE 23–2 Central part of Fig. 23–1, showing a negative point charge near the negative plate, where its potential energy is high (but the potential $V = V_b$ is low). Example 23–1.



⚠ CAUTION

A negative charge has high potential energy when potential V is low

The difference in potential energy, $U_b - U_a$, is equal to the negative of the work, W_{ba}, done by the electric field to move the charge from a to b; so the potential difference V_{ba} is

$$V_{ba} = \Delta V = V_b - V_a = \frac{U_b - U_a}{q} = -\frac{W_{ba}}{q}. \qquad \textbf{(23–2b)}$$

Note that electric potential, like electric field, does not depend on our test charge q. V depends on the other charges that create the field, not on the test charge q; q acquires potential energy by being in the potential V due to the other charges.

We can see from our definition that the positive plate in Fig. 23–1 is at a higher potential than the negative plate. Thus a positively charged object moves naturally from a high potential to a low potential.

The unit of electric potential, and of potential difference, is joules/coulomb and is given a special name, the **volt**, in honor of Alessandro Volta (1745–1827) who is best known for inventing the electric battery. The volt is abbreviated V, so 1 V = 1 J/C. Potential difference, since it is measured in volts, is often referred to as **voltage**. (Be careful not to confuse V for volts with italic V for voltage.)

If we wish to speak of the potential V_a at some point a, we must be aware that V_a depends on where the potential is chosen to be zero, just as in mechanics (Section 8–2). The zero for electric potential in a given situation can be chosen arbitrarily, as for any potential energy, because only differences in potential energy can be measured. Often the ground, or a conductor connected directly to the ground (the Earth), is taken as zero potential, and other potentials are given with respect to ground. (Thus, at a point where the voltage is 50 V, the difference of potential between it and ground is 50 V.) In other cases, we sometimes choose the potential to be zero at an infinite distance $(r = \infty)$.

CONCEPTUAL EXAMPLE 23–1 **A negative charge.** Suppose a negative charge, such as an electron, is placed near the negative plate in Fig. 23–1, at point b, shown here in Fig. 23–2. If the electron is free to move, will its electric potential energy increase or decrease?

RESPONSE An electron released at point b will be attracted to the positive plate. As the electron accelerates toward the positive plate, its kinetic energy increases, so its potential energy *decreases*: $U_a < U_b$ and $\Delta U = U_a - U_b < 0$. But note that the electron moves from point b at low potential to point a at higher potential: $V_{ab} = V_a - V_b > 0$. (Potentials V_a and V_b are due to the charges on the plates, not due to the electron.) The signs of ΔU and ΔV are opposite because of the negative charge of the electron.

NOTE A positive charge placed next to the negative plate at b would stay there, with no acceleration. A positive charge tends to move from high potential to low, whereas a negative charge moves from low to high potential.

Because the electric potential difference is defined as the potential energy difference per unit charge, then the change in *potential energy* of a charge q when it moves from point a to point b is

$$\Delta U = U_b - U_a = q(V_b - V_a) = qV_{ba}. \qquad \textbf{(23–3)}$$

That is, if an object with charge q moves through a potential difference V_{ba}, its potential energy changes by an amount qV_{ba}. For example, if the potential difference between the two plates in Fig. 23–1 is 6 V, then a +1 C charge moved from point b to point a will gain (1 C)(6 V) = 6 J of electric potential energy. (And it will lose 6 J of electric potential energy if it moves from a to b.) Similarly, a +2 C charge will gain $\Delta U = (2\text{ C})(6\text{ V}) = 12\text{ J}$, and so on. Thus, electric potential difference is a measure of how much energy an electric charge can acquire in a given situation. And, since energy is the ability to do work, the electric potential difference is also a measure of how much *work* a given charge can do. The exact amount of energy or work depends *both* on the potential difference and on the charge.

To better understand electric potential, let's make a comparison to the gravitational case when a rock falls from the top of a cliff. The greater the height, h, of a cliff, the more potential energy $(=mgh)$ the rock has at the top of the cliff, relative to the bottom, and the more kinetic energy it will have when it reaches the bottom. The actual amount of kinetic energy it will acquire, and the amount of work it can do, depends both on the height of the cliff and the mass m of the rock. A large rock and a small rock can be at the same height h (Fig. 23–3a) and thus have the same "gravitational potential," but the larger rock has the greater potential energy (it has more mass). The electrical case is similar (Fig. 23–3b): the potential energy change, or the work that can be done, depends both on the potential difference (corresponding to the height of the cliff) and on the charge (corresponding to mass), Eq. 23–3. But note a significant difference: electric charge comes in two types, + and −, whereas gravitational mass is always +.

Sources of electrical energy such as batteries and electric generators are meant to maintain a potential difference. The actual amount of energy transformed by such a device depends on how much charge flows, as well as the potential difference (Eq. 23–3). For example, consider an automobile headlight connected to a 12.0-V battery. The amount of energy transformed (into light and thermal energy) is proportional to how much charge flows, which in turn depends on how long the light is on. If over a given period of time 5.0 C of charge flows through the light, the total energy transformed is $(5.0\,\text{C})(12.0\,\text{V}) = 60\,\text{J}$. If the headlight is left on twice as long, 10.0 C of charge will flow and the energy transformed is $(10.0\,\text{C})(12.0\,\text{V}) = 120\,\text{J}$. Table 23–1 presents some typical voltages.

EXAMPLE 23–2 **Electron in a vacuum tube.** Suppose an electron is accelerated from rest through a potential difference $V_b - V_a = V_{ba} = +5000\,\text{V}$ (Fig. 23–4). (*a*) What is the change in electric potential energy of the electron? What is (*b*) the kinetic energy, and (*c*) the speed of the electron $(m = 9.1 \times 10^{-31}\,\text{kg})$ as a result of this acceleration?

APPROACH The electron, accelerated toward the positive plate, will change in potential energy by an amount $\Delta U = qV_{ba}$ (Eq. 23–3). The loss in potential energy will equal its gain in kinetic energy (energy conservation).

SOLUTION (*a*) The charge on an electron is $q = -e = -1.6 \times 10^{-19}\,\text{C}$. Therefore its change in potential energy is

$$\Delta U = qV_{ba} = (-1.6 \times 10^{-19}\,\text{C})(+5000\,\text{V}) = -8.0 \times 10^{-16}\,\text{J}.$$

The minus sign indicates that the potential energy decreases. The potential difference, V_{ba}, has a positive sign because the final potential V_b is higher than the initial potential V_a. Negative electrons are attracted toward a positive electrode (or plate) and repelled away from a negative electrode.

(*b*) The potential energy lost by the electron becomes kinetic energy K. From conservation of energy (Eq. 8–9a), $\Delta K + \Delta U = 0$, so

$$\begin{aligned} \Delta K &= -\Delta U \\ \tfrac{1}{2}mv^2 - 0 &= -q(V_b - V_a) = -qV_{ba}, \end{aligned}$$

where the initial kinetic energy is zero since we are given that the electron started from rest. So the final kinetic energy $K = -qV_{ba} = 8.0 \times 10^{-16}\,\text{J}$.

(*c*) In the equation just above we solve for v:

$$v = \sqrt{-\frac{2qV_{ba}}{m}} = \sqrt{-\frac{2(-1.6 \times 10^{-19}\,\text{C})(5000\,\text{V})}{9.1 \times 10^{-31}\,\text{kg}}} = 4.2 \times 10^{7}\,\text{m/s}.$$

NOTE The electric potential energy does not depend on the mass, only on the charge and voltage. The speed *does* depend on m.

EXERCISE A Instead of the electron in Example 23–2, suppose a proton $(m = 1.67 \times 10^{-27}\,\text{kg})$ was accelerated from rest by a potential difference $V_{ba} = -5000\,\text{V}$. What would be the proton's (*a*) change in potential energy, and (*b*) final speed?

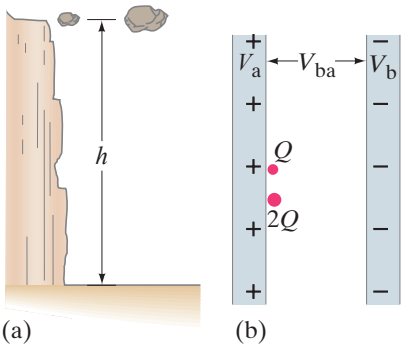


FIGURE 23–3 (a) Two rocks are at the same height. The larger rock has more potential energy. (b) Two positive charges have the same electric potential. The $2Q$ charge has more potential energy.

TABLE 23–1 Some Typical Potential Differences (Voltages)

Source	Voltage (approx.)
Thundercloud to ground	10^8 V
High-voltage power line	10^5–10^6 V
Automobile ignition	10^4 V
Household outlet	10^2 V
Automobile battery	12 V
Flashlight battery (AA, AAA, C, D)	1.5 V
Resting potential across nerve membrane	10^{-1} V
Potential changes on skin (ECG and EEG)	10^{-4} V

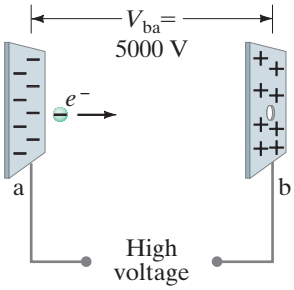


FIGURE 23–4 Electron accelerated, Example 23–2.

23–2 Relation between Electric Potential and Electric Field

The effects of any charge distribution can be described either in terms of electric field or in terms of electric potential. Electric potential is often easier to use because it is a scalar, whereas electric field is a vector. There is a direct connection between the electric potential produced by a given arrangement of charges and the electric field due to those charges, which we now examine.

We start by recalling the relation between a conservative force $\vec{\mathbf{F}}$ and the potential energy U associated with that force. As discussed in Section 8–2, the difference in potential energy between any two points in space, a and b, is given by Eq. 8–4:

$$U_b - U_a = -\int_a^b \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}},$$

where $d\vec{\boldsymbol{\ell}}$ is an infinitesimal increment of displacement, and the integral is taken along any path in space from point a to point b. (We call this a "line integral", as mentioned already in Section 7–3.) For the electrical case, we are more interested in the potential difference, given by Eq. 23–2b, $V_{ba} = V_b - V_a = (U_b - U_a)/q$, rather than in the potential energy itself. Also, the electric field $\vec{\mathbf{E}}$ at any point in space is defined as the force per unit charge (Eq. 21–3): $\vec{\mathbf{E}} = \vec{\mathbf{F}}/q$. Putting these two relations in the above equation gives us

$$V_{ba} = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}. \qquad \textbf{(23–4a)}$$

This is the general relation between electric field and potential difference. See Fig. 23–5. If we are given the electric field due to some arrangement of electric charge, we can use Eq. 23–4a to determine V_{ba}.

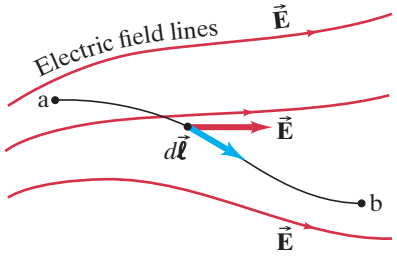


FIGURE 23–5 To find V_{ba} in a nonuniform electric field $\vec{\mathbf{E}}$, we integrate $\vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$ from point a to point b.

A simple special case is a uniform field. In Fig. 23–1, for example, a path parallel to the electric field lines from point a at the positive plate to point b at the negative plate gives (since $\vec{\mathbf{E}}$ and $d\vec{\boldsymbol{\ell}}$ are in the same direction at each point),

$$V_{ba} = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -E\int_a^b d\ell = -Ed$$

or

$$V_{ba} = -Ed \qquad \text{[only if } E \text{ is uniform]} \qquad \textbf{(23–4b)}$$

where d is the distance, parallel to the field lines, between points a and b. Be careful not to use Eq. 23–4b unless you are sure the electric field is uniform.

From either of Eqs. 23–4 we can see that the units for electric field can be written as volts per meter (V/m) as well as newtons per coulomb (N/C). These are equivalent because $1\,\text{N/C} = 1\,\text{N}\cdot\text{m/C}\cdot\text{m} = 1\,\text{J/C}\cdot\text{m} = 1\,\text{V/m}$. The minus sign in Eq. 23–4b tells us that $\vec{\mathbf{E}}$ points in the direction of decreasing potential V.

EXAMPLE 23–3 **Electric field obtained from voltage.** Two parallel plates are charged to produce a potential difference of 50 V. If the separation between the plates is 0.050 m, determine the magnitude of the electric field in the space between the plates (Fig. 23–6).

APPROACH We can apply Eq. 23–4b to obtain the magnitude of E, if we assume E is uniform.

SOLUTION The electric field magnitude is $E = V_{ba}/d = (50\,\text{V}/0.050\,\text{m}) = 1000\,\text{V/m}$.

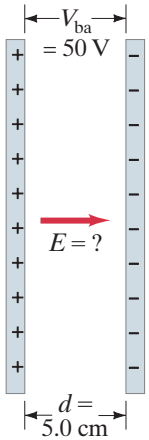


FIGURE 23–6 Example 23–3.

EXAMPLE 23–4 **Charged conducting sphere.** Determine the potential at a distance r from the center of a charged conducting sphere of radius r_0 for (*a*) $r > r_0$, (*b*) $r = r_0$, (*c*) $r < r_0$. The total charge on the sphere is Q.

APPROACH The charge Q is distributed over the surface of the sphere since it is a conductor. We saw in Example 22–3 that the electric field outside a conducting sphere is

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2} \qquad [r > r_0]$$

and points radially outward (inward if $Q < 0$). Since we know $\vec{\mathbf{E}}$, we can start by using Eq. 23–4a.

SOLUTION (*a*) We use Eq. 23–4a and integrate along a radial line with $d\vec{\ell}$ parallel to $\vec{\mathbf{E}}$ (Fig. 23–7) between two points which are distances r_a and r_b from the sphere's center, and we set $d\ell = dr$:

$$V_b - V_a = -\int_{r_a}^{r_b}\vec{\mathbf{E}}\cdot d\vec{\ell} = -\frac{Q}{4\pi\epsilon_0}\int_{r_a}^{r_b}\frac{dr}{r^2}$$

$$= \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_b} - \frac{1}{r_a}\right).$$

If we let $V = 0$ for $r = \infty$ (let's choose $V_b = 0$ at $r_b = \infty$), then at any other point r (for $r > r_0$) we have

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}. \qquad [r > r_0]$$

We will see in the next Section that this same equation applies for the potential a distance r from a single point charge. (In fact we already have that result here if we let $r_0 \to 0$ in Fig. 23–7.) Thus the electric potential outside a spherical conductor with a uniformly distributed charge is the same as if all the charge were at its center.

(*b*) As r approaches r_0, we see that

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r_0} \qquad [r = r_0]$$

at the surface of the conductor.

(*c*) For points within the conductor, $E = 0$. Thus the integral, $\int\vec{\mathbf{E}}\cdot d\vec{\ell}$, between $r = r_0$ and any point within the conductor gives zero change in V. Hence V is constant within the conductor:

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r_0}. \qquad [r \le r_0]$$

The whole conductor, not just its surface, is at this same potential. Plots of both E and V as a function of r are shown in Fig. 23–8 for a positively charged conducting sphere.

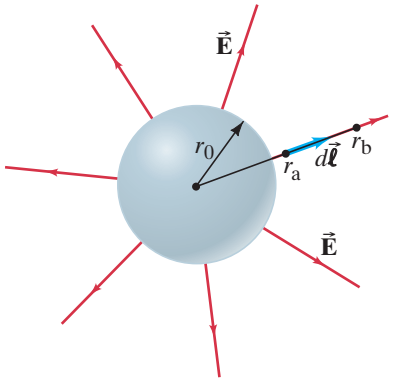
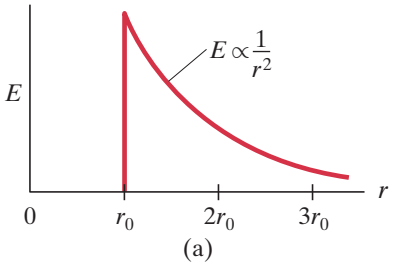
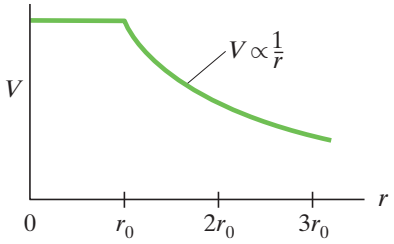


FIGURE 23–7 Example 23–4. Integrating $\vec{\mathbf{E}}\cdot d\vec{\ell}$ for the field outside a spherical conductor.

FIGURE 23–8 (a) E versus r, and (b) V versus r, for a positively charged solid conducting sphere of radius r_0 (the charge distributes itself evenly on the surface); r is the distance from the center of the sphere. Example 23–4.



(a)



(b)

EXAMPLE 23–5 **Breakdown voltage.** When very high voltages are present, air can become ionized due to the high electric fields. Any free electrons in the air (produced by cosmic rays, for example) can be accelerated to sufficient kinetic energy to knock electrons out of O_2 and N_2 molecules of the air. The air then becomes conducting and the high voltage cannot be maintained as charge flows. This **breakdown** of air occurs when the electric field exceeds about $E_{BD} = 3 \times 10^6\,\text{V/m}$ (*a*) Show that the breakdown voltage V_{BD} for a spherical conductor in air is proportional to the radius of the sphere, and (*b*) estimate the breakdown voltage in air for a sphere of diameter 1.0 cm.

APPROACH The electric potential at the surface of a spherical conductor of radius r_0 (Example 23–4), and the electric field just outside its surface, are

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r_0} \quad \text{and} \quad E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r_0^2}.$$

SOLUTION (*a*) We combine these two equations and obtain

$$V_{BD} = r_0 E_{BD}. \quad \text{[at surface of spherical conductor]} \quad \textbf{(23–4c)}$$

(*b*) For $r_0 = 5 \times 10^{-3}\,\text{m}$, the breakdown voltage in air is

$$V_{BD} = (5 \times 10^{-3}\,\text{m})(3 \times 10^6\,\text{V/m}) \approx 15{,}000\,\text{V}.$$

CAUTION
Breakdown voltage is smaller at sharp points because $V \propto r_0$.

When high voltages are present, a glow known as a **corona discharge** may be seen around sharp points due to the high electric fields at these points with small r_0 (Eq. 23–4c), which can ionize air molecules. When electrons recombine with their molecules, light is emitted. Such breakdown of air is the source of lightning, the spark of a car's spark plug, and even short sparks between your fingers and a doorknob after you walk across a synthetic rug, or slide across a car seat (which can result in a significant transfer of charge to you). **Lightning rods**, with their sharp tips, are intended to ionize the surrounding air when a storm cloud is near, and to provide a conduction path to the ground in order to discharge a dangerous high-voltage cloud slowly, over a period of time. Thus lightning rods, connected to the ground, are intended to draw electric charge off threatening clouds before a large buildup of charge results in a swift destructive lightning bolt.

EXERCISE B On a dry day, a person can become electrically charged by rubbing against rugs and other ordinary objects. Suppose you feel a small shock as you reach for a metal doorknob, noting that the shock occurs along with a tiny spark when your hand is about 3.0 mm from the doorknob. As a rough estimate, use Eq. 23–4b to estimate the potential difference between your hand and the doorknob. (*a*) 9 V; (*b*) 90 V; (*c*) 900 V; (*d*) 9000 V.

23–3 Electric Potential Due to Point Charges

The electric potential at a distance r from a single point charge Q can be derived directly from Eq. 23–4a, $V_b - V_a = -\int \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$. The electric field due to a single point charge has magnitude (Eq. 21–4)

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2} \quad \text{or} \quad E = k\frac{Q}{r^2}$$

(where $k = 1/4\pi\epsilon_0 = 8.99 \times 10^9\,\text{N} \cdot \text{m}^2/\text{C}^2$), and is directed radially outward from a positive charge (inward if $Q < 0$). We take the integral in Eq. 23–4a along a (straight) field line (Fig. 23–9) from point a, a distance r_a from Q, to point b, a distance r_b from Q. Then $d\vec{\boldsymbol{\ell}}$ will be parallel to $\vec{\mathbf{E}}$ and $d\ell = dr$. Thus

$$V_b - V_a = -\int_{r_a}^{r_b} \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{Q}{4\pi\epsilon_0}\int_{r_a}^{r_b}\frac{1}{r^2}\,dr = \frac{1}{4\pi\epsilon_0}\left(\frac{Q}{r_b} - \frac{Q}{r_a}\right).$$

As mentioned earlier, only differences in potential have physical meaning.

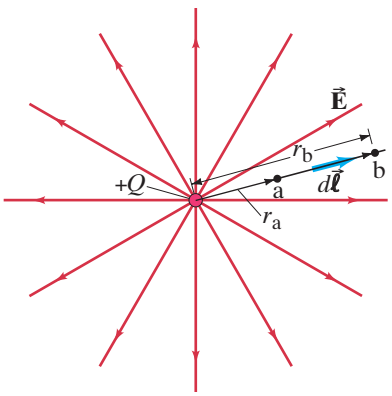


FIGURE 23–9 We integrate Eq. 23–4a along the straight line (shown in black) from point a to point b. The line ab is parallel to a field line.

We are free, therefore, to choose the value of the potential at some one point to be whatever we please. It is common to choose the potential to be zero at infinity (let $V_b = 0$ at $r_b = \infty$). Then the electric potential V at a distance r from a single point charge is

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} = k\frac{Q}{r}. \qquad \left[\begin{array}{c}\text{single point charge;}\\ V = 0 \text{ at } r = \infty\end{array}\right] \quad \textbf{(23–5)}$$

We can think of V here as representing the absolute potential at a distance r from the charge Q, where $V = 0$ at $r = \infty$; or we can think of V as the potential difference between r and infinity. Notice that the potential V decreases with the first power of the distance, whereas the electric field (Eq. 21–4) decreases as the *square* of the distance. The potential near a positive charge is large, and it decreases toward zero at very large distances (Fig. 23–10). For a negative charge, the potential is negative and increases toward zero at large distances (Fig. 23–11). Equation 23–5 is sometimes called the **Coulomb potential** (it has its origin in Coulomb's law).

In Example 23–4 we found that the potential due to a uniformly charged sphere is given by the same relation, Eq. 23–5, for points outside the sphere. Thus we see that the potential outside a uniformly charged sphere is the same as if all the charge were concentrated at its center.

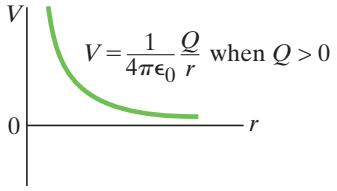
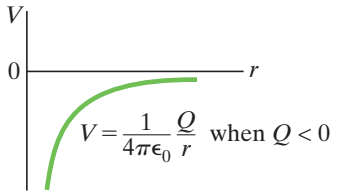


FIGURE 23–10 Potential V as a function of distance r from a single point charge Q when the charge is positive.

CAUTION

$V \propto \frac{1}{r}$, $E \propto \frac{1}{r^2}$ for a point charge

FIGURE 23–11 Potential V as a function of distance r from a single point charge Q when the charge is negative.



EXERCISE C What is the potential at a distance of 3.0 cm from a point charge $Q = -2.0 \times 10^{-9}$ C? (*a*) 600 V; (*b*) 60 V; (*c*) 6 V; (*d*) −600 V; (*e*) −60 V; (*f*) −6 V.

EXAMPLE 23–6 **Work required to bring two positive charges close together.** What minimum work must be done by an external force to bring a charge $q = 3.00\,\mu$C from a great distance away (take $r = \infty$) to a point 0.500 m from a charge $Q = 20.0\,\mu$C?

APPROACH To find the work we cannot simply multiply the force times distance because the force is not constant (it is proportional to $1/r^2$). Instead we can set the change in potential energy equal to the (positive of the) work required of a different *external* force (Section 8–2), and Eq. 23–3: $W = \Delta U = q(V_b - V_a)$. We get the potentials V_b and V_a using Eq. 23–5.

SOLUTION The external work required is equal to the change in potential energy:

$$W = q(V_b - V_a)$$
$$= q\left(\frac{kQ}{r_b} - \frac{kQ}{r_a}\right),$$

where $r_b = 0.500$ m and $r_a = \infty$. The right-hand term within the parentheses is zero $(1/\infty = 0)$ so

$$W = (3.00 \times 10^{-6}\,\text{C})\frac{(8.99 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(2.00 \times 10^{-5}\,\text{C})}{(0.500\,\text{m})} = 1.08\,\text{J}.$$

NOTE We could not use Eq. 23–4b here because it applies *only* to uniform fields. But we did use Eq. 23–3 because it is always valid.

CAUTION

We cannot use $W = Fd$ when F is not constant

EXERCISE D What work is required to bring a charge $q = 3.00\,\mu$C originally a distance of 1.50 m from a charge $Q = 20.0\,\mu$C until it is 0.50 m away?

To determine the electric field at points near a collection of two or more point charges requires adding up the electric fields due to each charge. Since the electric field is a vector, this can be time consuming or complicated. To find the electric potential at a point due to a collection of point charges is far easier, because the electric potential is a scalar, and you only need to add numbers (with appropriate signs) without concern for direction. This is a major advantage in using electric potential for solving Problems.

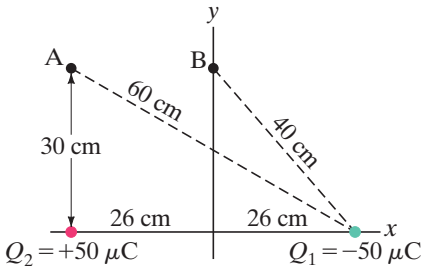


FIGURE 23–12 Example 23–7. (See also Example 21–9, Fig. 21–28.)

CAUTION
Potential is a scalar and has no components

EXAMPLE 23–7 **Potential above two charges.** Calculate the electric potential (*a*) at point A in Fig. 23–12 due to the two charges shown, and (*b*) at point B. [This is the same situation as Example 21–9, Fig. 21–28 in Chapter 21, where we calculated the electric field at these points.]

APPROACH The total potential at point A (or at point B) is the algebraic sum of the potentials at that point due to each of the two charges Q_1 and Q_2. The potential due to each single charge is given by Eq. 23–5. We do not have to worry about directions because electric potential is a scalar quantity. But we do have to keep track of the signs of charges.

SOLUTION (*a*) We add the potentials at point A due to each charge Q_1 and Q_2, and we use Eq. 23–5 for each:

$$\begin{aligned} V_A &= V_{A2} + V_{A1} \\ &= k\frac{Q_2}{r_{2A}} + k\frac{Q_1}{r_{1A}} \end{aligned}$$

where $r_{1A} = 60\,\text{cm}$ and $r_{2A} = 30\,\text{cm}$. Then

$$\begin{aligned} V_A &= \frac{(9.0\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(5.0\times10^{-5}\,\text{C})}{0.30\,\text{m}} \\ &\quad + \frac{(9.0\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(-5.0\times10^{-5}\,\text{C})}{0.60\,\text{m}} \\ &= 1.50\times10^6\,\text{V} - 0.75\times10^6\,\text{V} \\ &= 7.5\times10^5\,\text{V}. \end{aligned}$$

(*b*) At point B, $r_{1B} = r_{2B} = 0.40\,\text{m}$, so

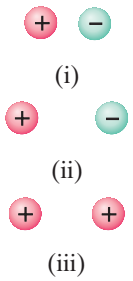
$$\begin{aligned} V_B &= V_{B2} + V_{B1} \\ &= \frac{(9.0\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(5.0\times10^{-5}\,\text{C})}{0.40\,\text{m}} \\ &\quad + \frac{(9.0\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(-5.0\times10^{-5}\,\text{C})}{0.40\,\text{m}} \\ &= 0\,\text{V}. \end{aligned}$$

NOTE The two terms in the sum in (*b*) cancel for any point equidistant from Q_1 and Q_2 $(r_{1B} = r_{2B})$. Thus the potential will be zero everywhere on the plane equidistant between the two opposite charges. This plane where V is constant is called an "equipotential surface."

Simple summations like these can be performed for any number of point charges.

CONCEPTUAL EXAMPLE 23–8 **Potential energies.** Consider two isolated charges, Q_1 and Q_2. This pair of charges is shown in three different configurations in Fig. 23–13. (*a*) Which set has a positive potential energy? (*b*) Which set has the most negative potential energy? (*c*) Which set requires the most work to separate the charges to infinity? Assume the charges all have the same magnitude.

FIGURE 23–13 Example 23–8.



RESPONSE The potential energy equals the work required to bring the two charges near each other, starting at a great distance (∞). Assume the left $(+)$ charge Q_1 is already there. To bring a second charge Q_2 close to the first charge from a great distance away (∞) requires external work

$$W_{\text{ext}} = Q_2V = k\frac{Q_1Q_2}{r}$$

where r is the final distance between them.

Thus the potential energy U of the two charges is

$$U = k\frac{Q_1 Q_2}{r}.$$

(*a*) Set (iii) has a positive potential energy because the charges have the same sign. (*b*) Both (i) and (ii) have opposite signs of charge and negative potential energy. Because r is smaller in (i), U is most negative for (i). (*c*) Set (i) will require the most work for separation to infinity. The more negative the potential energy, the more work required to separate the charges and bring the potential energy up to zero $(r = \infty)$, as in Fig. 23–11.

EXERCISE E Return to the Chapter-Opening Question, page 660, and answer it again now. Try to explain why you may have answered differently the first time.

23–4 Potential Due to Any Charge Distribution

If we know the electric field in a region of space due to any distribution of electric charge, we can determine the difference in potential between two points in the region using Eq. 23–4a, $V_{ba} = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$. In many cases we don't know $\vec{\mathbf{E}}$ as a function of position, and it may be difficult to calculate. We can calculate the potential V due to a given charge distribution in another way, using the potential due to a single point charge, Eq. 23–5:

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r},$$

where we choose $V = 0$ at $r = \infty$. Then we can sum over all the charges. If we have n individual point charges, the potential at some point a (relative to $V = 0$ at $r = \infty$) is

$$V_a = \sum_{i=1}^{n} V_i = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{n}\frac{Q_i}{r_{ia}}, \qquad \textbf{(23–6a)}$$

where r_{ia} is the distance from the i^{th} charge (Q_i) to the point a. (We already used this approach in Example 23–7 for two charges.) If the charge distribution can be considered continuous, then

$$V = \frac{1}{4\pi\epsilon_0}\int\frac{dq}{r}, \qquad \textbf{(23–6b)}$$

where r is the distance from a tiny element of charge, dq, to the point where V is being determined.

EXAMPLE 23–9 **Potential due to a ring of charge.** A thin circular ring of radius R has a uniformly distributed charge Q. Determine the electric potential at a point P on the axis of the ring a distance x from its center, Fig. 23–14.

APPROACH We integrate over the ring using Eq. 23–6b.

SOLUTION Each point on the ring is equidistant from point P, and this distance is $(x^2 + R^2)^{\frac{1}{2}}$. So the potential at P is:

$$V = \frac{1}{4\pi\epsilon_0}\int\frac{dq}{r} = \frac{1}{4\pi\epsilon_0}\frac{1}{(x^2 + R^2)^{\frac{1}{2}}}\int dq = \frac{1}{4\pi\epsilon_0}\frac{Q}{(x^2 + R^2)^{\frac{1}{2}}}.$$

NOTE For points very far away from the ring, $x \gg R$, this result reduces to $(1/4\pi\epsilon_0)(Q/x)$, the potential of a point charge, as we should expect. This sort of check does not prove our result is correct, but it does give us very high confidence that we did not make a blunder.

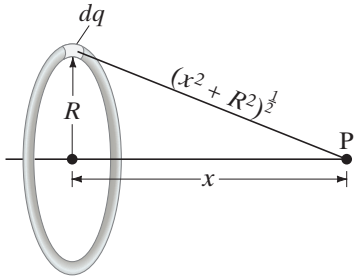


FIGURE 23–14 Example 23–9. Calculating the potential at point P, a distance x from the center of a uniform ring of charge.

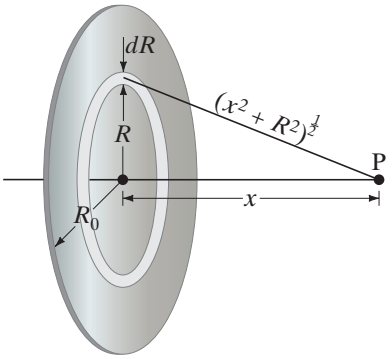


FIGURE 23–15 Example 23–10. Calculating the electric potential at point P on the axis of a uniformly charged thin disk.

EXAMPLE 23–10 **Potential due to a charged disk.** A thin flat disk, of radius R_0, has a uniformly distributed charge Q, Fig. 23–15. Determine the potential at a point P on the axis of the disk, a distance x from its center.

APPROACH We divide the disk into thin rings of radius R and thickness dR and use the result of Example 23–9 to sum over the disk.

SOLUTION The charge Q is distributed uniformly, so the charge contained in each concentric ring is proportional to its area. The disk has area $A = \pi R_0^2$ and each thin ring has area $dA = (2\pi R)(dR)$. Hence

$$\frac{dq}{Q} = \frac{dA}{A} = \frac{2\pi R\, dR}{\pi R_0^2}$$

so

$$dq = Q\frac{(2\pi R)(dR)}{\pi R_0^2} = \frac{2QR\, dR}{R_0^2}.$$

Then the potential at P, using Eq. 23–6b in which r is replaced by $(x^2 + R^2)^{\frac{1}{2}}$, is

$$V = \frac{1}{4\pi\epsilon_0}\int \frac{dq}{(x^2+R^2)^{\frac{1}{2}}} = \frac{2Q}{4\pi\epsilon_0 R_0^2}\int_0^{R_0} \frac{R\, dR}{(x^2+R^2)^{\frac{1}{2}}} = \frac{Q}{2\pi\epsilon_0 R_0^2}\left(x^2+R^2\right)^{\frac{1}{2}}\Big|_{R=0}^{R=R_0}$$

$$= \frac{Q}{2\pi\epsilon_0 R_0^2}\left[(x^2+R_0^2)^{\frac{1}{2}} - x\right] = \frac{Qx}{2\pi\epsilon_0 R_0^2}\left[\left(1+\frac{R_0^2}{x^2}\right)^{\frac{1}{2}} - 1\right]$$

NOTE To check our result, consider $x \gg R_0$. Our formula reduces to (we use the binomial expansion—Appendix A):

$$V \approx \frac{Qx}{2\pi\epsilon_0 R_0^2}\left[\left(1 + \frac{1}{2}\frac{R_0^2}{x^2}\right) - 1\right] = \frac{Q}{4\pi\epsilon_0 x}.$$

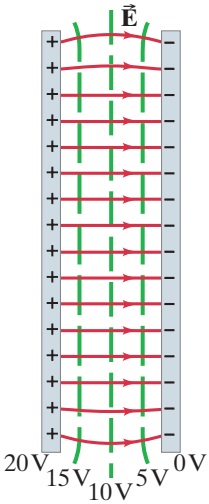
This is the formula for a point charge, as we expect.

23–5 Equipotential Lines and Surfaces

The electric potential can be represented by drawing **equipotential lines** or, in three dimensions, **equipotential surfaces**. An equipotential surface is one on which all points are at the same potential. That is, the potential difference between any two points on the surface is zero, so no work is required to move a charge from one point on the surface to another. An *equipotential surface must be perpendicular to the electric field* at any point. If this were not so—that is, if there were a component of $\vec{\mathbf{E}}$ parallel to the surface—it would require work to move the charge along the surface against this component of $\vec{\mathbf{E}}$; and this would contradict the idea that it is an *equi*potential surface. This can also be seen from Eq. 23–4a, $\Delta V = -\int \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$. On a surface where V is constant, $\Delta V = 0$, so we must have either $\vec{\mathbf{E}} = 0$, $d\vec{\boldsymbol{\ell}} = 0$, or $\cos\theta = 0$ where θ is the angle between $\vec{\mathbf{E}}$ and $d\vec{\boldsymbol{\ell}}$. Thus in a region where $\vec{\mathbf{E}}$ is not zero, the path $d\vec{\boldsymbol{\ell}}$ along an equipotential must have $\cos\theta = 0$, meaning $\theta = 90°$ and $\vec{\mathbf{E}}$ is perpendicular to the equipotential.

The fact that the electric field lines and equipotential surfaces are mutually perpendicular helps us locate the equipotentials when the electric field lines are known. In a normal two-dimensional drawing, we show equipotential *lines*, which are the intersections of equipotential surfaces with the plane of the drawing. In Fig. 23–16, a few of the equipotential lines are drawn (dashed green lines) for the electric field (red lines) between two parallel plates at a potential difference of 20 V. The negative plate is arbitrarily chosen to be zero volts and the potential of each equipotential line is indicated. Note that $\vec{\mathbf{E}}$ points toward lower values of V.

FIGURE 23–16 Equipotential lines (the green dashed lines) between oppositely charged parallel plates are always perpendicular to the electric field lines (solid red lines).



EXAMPLE 23–11 **Point charge equipotential surfaces.** For a single point charge with $Q = 4.0 \times 10^{-9}\,\text{C}$, sketch the equipotential surfaces (or lines in a plane containing the charge) corresponding to $V_1 = 10\,\text{V}$, $V_2 = 20\,\text{V}$, and $V_3 = 30\,\text{V}$.

APPROACH The electric potential V depends on the distance r from the charge (Eq. 23–5).

SOLUTION The electric field for a positive point charge is directed radially outward. Since the equipotential surfaces must be perpendicular to the lines of electric field, they will be spherical in shape, centered on the point charge, Fig. 23–17. From Eq. 23–5 we have $r = (1/4\pi\epsilon_0)(Q/V)$, so that for $V_1 = 10\,\text{V}$, $r_1 = (9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(4.0 \times 10^{-9}\,\text{C})/(10\,\text{V}) = 3.6\,\text{m}$, for $V_2 = 20\,\text{V}$, $r_2 = 1.8\,\text{m}$, and for $V_3 = 30\,\text{V}$, $r_3 = 1.2\,\text{m}$, as shown.

NOTE The equipotential surface with the largest potential is closest to the positive charge. How would this change if Q were negative?

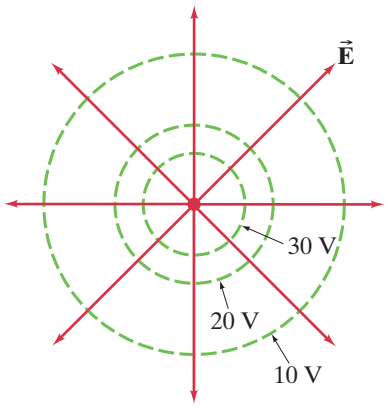
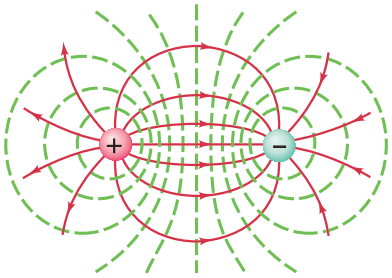


FIGURE 23–17 Example 23–11. Electric field lines and equipotential surfaces for a point charge.

The equipotential lines for the case of two equal but oppositely charged particles are shown in Fig. 23–18 as green dashed lines. (This combination of equal + and − charges is called an "electric dipole," as we saw in Section 21–11).

Unlike electric field lines, which start and end on electric charges, equipotential lines and surfaces are always continuous and never end, and so continue beyond the borders of Figs. 23–16 and 23–18. A useful analogy for equipotential lines is a topographic map: the contour lines are essentially gravitational equipotential lines (Fig. 23–19).

FIGURE 23–18 Equipotential lines (green, dashed) are always perpendicular to the electric field lines (solid red) shown here for two equal but oppositely charged particles (an electric dipole).



We saw in Section 21–9 that there can be no electric field within a conductor in the static case, for otherwise the free electrons would feel a force and would move, and thus not be static. Indeed, the *entire volume of a conductor must be at the same potential in the static case*. The surface of a conductor is then an equipotential surface. (If it weren't, the free electrons at the surface would move, since whenever there is a potential difference between two points, free charges will move.) This is fully consistent with our result in Section 21–9 that the electric field at the surface of a conductor must be perpendicular to the surface.



FIGURE 23–19 A topographic map (here, a portion of the Sierra Nevada in California) shows continuous contour lines, each of which is at a fixed height above sea level. Here they are at 80 ft (25 m) intervals. The contour lines of constant elevation are gravitational equipotential lines. If you walk along one contour line, you neither climb nor descend. If you cross lines, and especially if you climb perpendicular to the lines, you will be changing your gravitational potential (rapidly, if the lines are close together, meaning steep uphill).

23–6 Potential Due to Electric Dipole; Dipole Moment

Two equal point charges Q, of opposite sign, separated by a distance ℓ, are called an **electric dipole**. The electric field lines and equipotential surfaces for a dipole were shown in Fig. 23–18. Because electric dipoles occur often in physics, as well as in other disciplines such as molecular biology, it is useful to examine them more closely.

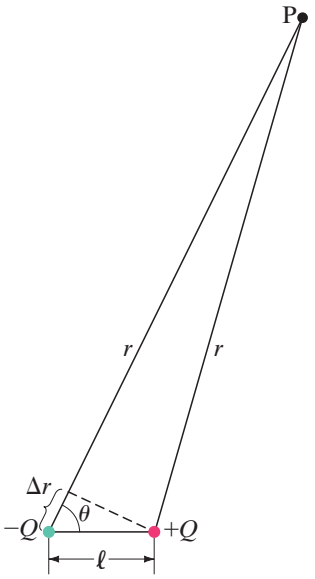


FIGURE 23–20 Electric dipole. Calculation of potential V at point P.

PHYSICS APPLIED
Dipoles in molecular biology

TABLE 23–2 Dipole Moments of Selected Molecules

Molecule	Dipole Moment (C · m)
$H_2^{(+)}O^{(-)}$	6.1×10^{-30}
$H^{(+)}Cl^{(-)}$	3.4×10^{-30}
$N^{(-)}H_3^{(+)}$	5.0×10^{-30}
$>N^{(-)}—H^{(+)}$	$\approx 3.0^{\dagger} \times 10^{-30}$
$>C^{(+)}=O^{(-)}$	$\approx 8.0^{\dagger} \times 10^{-30}$

†These last two groups often appear on larger molecules; hence the value for the dipole moment will vary somewhat, depending on the rest of the molecule.

The electric potential at an arbitrary point P due to a dipole, Fig. 23–20, is the sum of the potentials due to each of the two charges (we take $V = 0$ at $r = \infty$):

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} + \frac{1}{4\pi\epsilon_0}\frac{(-Q)}{(r+\Delta r)} = \frac{1}{4\pi\epsilon_0}Q\left(\frac{1}{r} - \frac{1}{r+\Delta r}\right) = \frac{Q}{4\pi\epsilon_0}\frac{\Delta r}{r(r+\Delta r)},$$

where r is the distance from P to the positive charge and $r + \Delta r$ is the distance to the negative charge. This equation becomes simpler if we consider points P whose distance from the dipole is much larger than the separation of the two charges—that is, for $r \gg \ell$. From Fig. 23–20 we see that $\Delta r \approx \ell \cos\theta$. Thus for $r \gg \Delta r = \ell \cos\theta$, in the equation above we can neglect Δr in the denominator as compared to r. Therefore, we obtain

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q\ell\cos\theta}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{p\cos\theta}{r^2}, \quad [\text{dipole; } r \gg \ell] \quad \textbf{(23–7)}$$

where quantity $p = Q\ell$ is called the **dipole moment**. We see that the potential decreases as the *square* of the distance from the dipole, whereas for a single point charge the potential decreases with the first power of the distance (Eq. 23–5). It is not surprising that the potential should fall off faster for a dipole; when you are far from a dipole, the two equal but opposite charges appear so close together as to tend to neutralize each other.

In many molecules, even though they are electrically neutral, the electrons spend more time in the vicinity of one atom than another, which results in a separation of charge. Such molecules have a dipole moment and are called **polar molecules**. We already saw that water (Fig. 21–4) is a polar molecule, and we have encountered others in our discussion of molecular biology (Section 21–12).

Table 23–2 gives the dipole moments for several molecules. The + and − signs indicate on which atoms these charges lie. The last two entries are a part of many organic molecules and play an important role in molecular biology. A dipole moment has units of coulomb-meters (C · m), although for molecules a smaller unit called a *debye* is sometimes used: 1 debye $= 3.33 \times 10^{-30}$ C · m.

23–7 $\vec{\mathbf{E}}$ Determined from V

We can use Eq. 23–4a, $V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$, to determine the difference in potential between two points if the electric field is known in the region between those two points. By inverting Eq. 23–4a, we can write the electric field in terms of the potential. Then the electric field can be determined from a knowledge of V. Let us see how to do this.

We write Eq. 23–4a in differential form as

$$dV = -\vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -E_\ell\, d\ell,$$

where dV is the infinitesimal difference in potential between two points a distance $d\ell$ apart, and E_ℓ is the component of the electric field in the direction of the infinitesimal displacement $d\vec{\boldsymbol{\ell}}$. We can then write

$$E_\ell = -\frac{dV}{d\ell}. \quad \textbf{(23–8)}$$

Thus *the component of the electric field in any direction is equal to the negative of the rate of change of the electric potential with distance in that direction*. The quantity $dV/d\ell$ is called the **gradient** of V in a particular direction. If the direction is not specified, the term *gradient* refers to that direction in which V changes most rapidly; this would be the direction of $\vec{\mathbf{E}}$ at that point,† so we can write

$$E = -\frac{dV}{d\ell}. \quad [\text{if } d\vec{\boldsymbol{\ell}} \text{ is parallel to } \vec{\mathbf{E}}]$$

†A component of a vector has magnitude less than, or at most equal to, the magnitude of the vector: $E_\ell \le E$. Thus $|dV/d\ell|$ has its maximum value when E_ℓ in Eq. 23-8 equals E.

If $\vec{\mathbf{E}}$ is written as a function of x, y, and z, and we let ℓ refer to the x, y, and z axes, then Eq. 23–8 becomes

$$E_x = -\frac{\partial V}{\partial x}, \qquad E_y = -\frac{\partial V}{\partial y}, \qquad E_z = -\frac{\partial V}{\partial z}. \tag{23–9}$$

Here, $\partial V/\partial x$ is the "partial derivative" of V with respect to x, with y and z held constant.[†] For example, if $V(x, y, z) = (2\,\text{V/m}^2)x^2 + (8\,\text{V/m}^3)y^2z + (2\,\text{V/m}^2)z^2$, then

$$E_x = -\partial V/\partial x = -(4\,\text{V/m}^2)\,x,$$
$$E_y = -\partial V/\partial y = -(16\,\text{V/m}^3)yz,$$

and

$$E_z = -\partial V/\partial z = -(8\,\text{V/m}^3)y^2 - (4\,\text{V/m}^2)z.$$

EXAMPLE 23–12 **$\vec{\mathbf{E}}$ for ring and disk.** Use electric potential to determine the electric field at point P on the axis of (*a*) a circular ring of charge (Fig. 23–14) and (*b*) a uniformly charged disk (Fig. 23–15).

APPROACH We obtained V as a function of x in Examples 23–9 and 23–10, so we find E by taking derivatives (Eqs. 23–9).

SOLUTION (*a*) From Example 23–9, the potential V along the axis of the circle is

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{(x^2 + R^2)^{\frac{1}{2}}}.$$

Then

$$E_x = -\frac{\partial V}{\partial x} = \frac{1}{4\pi\epsilon_0}\frac{Qx}{(x^2 + R^2)^{\frac{3}{2}}}.$$

This is the same result we obtained in Chapter 21, Example 21–10. ($E_y = E_z = 0$ by symmetry, but also because V does not depend on y or z so $\partial V/\partial y = \partial V/\partial z = 0$.)

(*b*) From Example 23–10,

$$V = \frac{Q}{2\pi\epsilon_0 R_0^2}\left[(x^2 + R_0^2)^{\frac{1}{2}} - x\right],$$

so

$$E_x = -\frac{\partial V}{\partial x} = \frac{Q}{2\pi\epsilon_0 R_0^2}\left[1 - \frac{x}{(x^2 + R_0^2)^{\frac{1}{2}}}\right].$$

(Again $E_y = E_z = 0$ by symmetry.) For points very close to the disk, $x \ll R_0$, the electric field can be approximated by

$$E \approx \frac{Q}{2\pi\epsilon_0 R_0^2} = \frac{\sigma}{2\epsilon_0}$$

where $\sigma = Q/\pi R_0^2$ is the surface charge density. The direction of $\vec{\mathbf{E}}$ is perpendicular to the disk. We also obtained these results in Chapter 21, Example 21–13 and Eq. 21–7.

If we compare this last Example with Examples 21–10 and 21–13, we see that here, as for many charge distributions, it is easier to calculate V first, and then $\vec{\mathbf{E}}$ from Eq. 23–9, rather than to calculate $\vec{\mathbf{E}}$ due to each charge from Coulomb's law. This is because V due to many charges is a scalar sum, whereas $\vec{\mathbf{E}}$ is a vector sum.

[†]Equation 23–9 can be written as a vector equation,

$$\vec{\mathbf{E}} = -\text{grad}\,V = -\vec{\boldsymbol{\nabla}}V = -\left(\hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}\right)V$$

where the symbol $\vec{\boldsymbol{\nabla}}$ is called the *del operator* or *gradient operator*: $\vec{\boldsymbol{\nabla}} = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}$.

23–8 Electrostatic Potential Energy; the Electron Volt

Suppose a point charge q is moved between two points in space, a and b, where the electric potential due to other charges is V_a and V_b, respectively. The change in electrostatic potential energy of q in the field of these other charges is, according to Eq. 23–2b,

$$\Delta U = U_b - U_a = q(V_b - V_a).$$

Now suppose we have a system of several point charges. What is the electrostatic potential energy of the system? It is most convenient to choose the electric potential energy to be zero when the charges are very far (ideally infinitely far) apart. A single point charge, Q_1, in isolation, has no potential energy, because if there are no other charges around, no electric force can be exerted on it. If a second point charge Q_2 is brought close to Q_1, the potential due to Q_1 at the position of this second charge is

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q_1}{r_{12}},$$

where r_{12} is the distance between Q_1 and Q_2. The potential energy of the two charges, relative to $V = 0$ at $r = \infty$, is

$$U = Q_2 V = \frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_2}{r_{12}}. \qquad \textbf{(23–10)}$$

This represents the work that needs to be done by an external force to bring Q_2 from infinity $(V = 0)$ to a distance r_{12} from Q_1. It is also the negative of the work needed to separate them to infinity.

If the system consists of three charges, the total potential energy will be the work needed to bring all three together. Equation 23–10 represents the work needed to bring Q_2 close to Q_1; to bring a third charge Q_3 so that it is a distance r_{13} from Q_1 and r_{23} from Q_2 requires work equal to

$$\frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_3}{r_{13}} + \frac{1}{4\pi\epsilon_0}\frac{Q_2 Q_3}{r_{23}}.$$

So the potential energy of a system of three point charges is

$$U = \frac{1}{4\pi\epsilon_0}\left(\frac{Q_1 Q_2}{r_{12}} + \frac{Q_1 Q_3}{r_{13}} + \frac{Q_2 Q_3}{r_{23}}\right). \qquad [V = 0 \text{ at } r = \infty]$$

For a system of four charges, the potential energy would contain six such terms, and so on. (Caution must be used when making such sums to avoid double counting of the different pairs.)

The Electron Volt, a Unit of Energy

The joule is a very large unit for dealing with energies of electrons, atoms, or molecules. For this reason, the unit **electron volt** (eV) is often used. One electron volt is defined as the energy acquired by a particle carrying a charge e, the magnitude of the charge on an electron $(q = e)$, as a result of moving through a potential difference of 1 V. The charge on an electron has magnitude 1.6022×10^{-19} C, and the change in potential energy equals qV. So 1 eV is equal to $(1.6022 \times 10^{-19}\ \text{C})\ (1.00\ \text{V}) = 1.6022 \times 10^{-19}$ J:

$$1\ \text{eV} = 1.6022 \times 10^{-19}\ \text{J} \approx 1.60 \times 10^{-19}\ \text{J}.$$

An electron that accelerates through a potential difference of 1000 V will lose 1000 eV of potential energy and thus gain 1000 eV or 1 keV (kiloelectron volt)

of kinetic energy. On the other hand, if a particle with a charge equal to twice the magnitude of the charge on the electron $(=2e=3.20\times10^{-19}\,\text{C})$ moves through a potential difference of 1000 V, its kinetic energy will change by $2000\,\text{eV}=2\,\text{keV}$.

Although the electron volt is handy for *stating* the energies of molecules and elementary particles, it is not a proper SI unit. For calculations (unless other units are not involved), electron volts must be converted to joules using the conversion factor just given.

EXERCISE F What is the kinetic energy of a He^{2+} ion released from rest and accelerated through a potential difference of 1.0 kV? (*a*) 1000 eV, (*b*) 500 eV, (*c*) 2000 eV, (*d*) 4000 eV, (*e*) 250 eV.

EXAMPLE 23–13 **Disassembling a hydrogen atom.** Calculate the work needed to "disassemble" a hydrogen atom. Assume that the proton $(Q_1=+e)$ and electron $(Q_2=-e)$ are initially separated by a distance equal to the "average" radius of the hydrogen atom in its ground state, $0.529\times10^{-10}\,\text{m}$, and that they end up an infinite distance apart from each other.

APPROACH The work necessary will be equal to the total energy, kinetic plus potential, of the electron and proton as an atom, compared to their total energy when infinitely far apart.

SOLUTION From Eq. 23–10 we have initially

$$U = \frac{1}{4\pi\epsilon_0}\frac{Q_1Q_2}{r} = \frac{1}{4\pi\epsilon_0}\frac{(e)(-e)}{r}$$

$$= \frac{-(8.99\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(1.60\times10^{-19}\,\text{C})^2}{(0.529\times10^{-10}\,\text{m})}$$

$$= -27.2(1.60\times10^{-19})\,\text{J} = -27.2\,\text{eV}.$$

This represents the potential energy. The total energy must include also the kinetic energy of the electron moving in an orbit of radius $r=0.529\times10^{-10}\,\text{m}$. From $F=ma$, using Coulomb's law for F and centripetal acceleration for a, we have

$$\frac{1}{4\pi\epsilon_0}\left(\frac{e^2}{r^2}\right) = \frac{mv^2}{r}.$$

Then

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}\left(\frac{1}{4\pi\epsilon_0}\right)\frac{e^2}{r}$$

which equals $-\frac{1}{2}U$ (as calculated above), so $K=+13.6\,\text{eV}$. The total energy initially is $E=K+U=13.6\,\text{eV}-27.2\,\text{eV}=-13.6\,\text{eV}$. To separate a stable hydrogen atom into a proton and an electron at rest very far apart ($U=0$ at $r=\infty$, $K=0$ because $v=0$) requires $+13.6\,\text{eV}$. This is, in fact, the measured ionization energy for hydrogen.

NOTE We quoted the kinetic energy as 13.6 eV. But if we wanted to determine the speed of the electron in SI units, we would have to first convert this 13.6 eV to joules (J).

NOTE To treat atoms properly, we need to use quantum theory (Chapters 37 to 39). But our "classical" calculation does give the correct answer here.

EXERCISE G The kinetic energy of a 1000-kg automobile traveling 20 m/s (70 km/h) would be about (*a*) 100 GeV, (*b*) 1000 TeV, (*c*) 10^6 TeV, (*d*) 10^{12} TeV, (*e*) 10^{18} TeV.

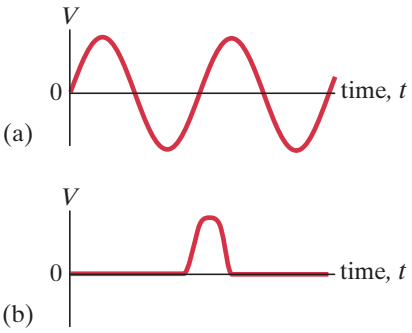


FIGURE 23–21 Two kinds of signal voltage: (a) sinusoidal, (b) a pulse, both analog. Many other shapes are possible.

23–9 Digital; Binary Numbers; Signal Voltage

Batteries and a wall socket are meant to provide a steady **supply voltage** as power to operate a flashlight, lamps, electric appliances, and other electric and electronic devices.

A **signal voltage**, on the other hand, carries **information** of some sort, and is intended to affect something else. A signal voltage varies in time and can also be very brief. A sound such as a pure tone, which may be sinusoidal as we discussed in Chapters 15 and 16 (see Figs. 15–3, 16–3, and 16–13), will produce an output voltage from a high quality microphone that is also sinusoidal (Fig. 23–21a). That signal voltage is amplified and reaches a loudspeaker, making it produce the sound we hear. A signal voltage is sometimes a simple pulse (Fig. 23–21b; see also Figs. 15–2 and 15–18), and often acts to change some aspect of an electronic device.

Signal voltages are sent to cell phones ("I've got signal"), to computers from the Internet, to radios or sound systems, and to TV sets with the information about the sound and picture. Not long ago, signal voltages were mostly **analog**—the voltage varied continuously, as in Fig. 23–21. The electronics that process or amplify an analog signal (as from a microphone) are meant to give as accurate a reproduction as possible for its output (to a loudspeaker).

Today, television and computer signals are **digital**, which are more complicated. They must be encoded so that their meaning is clear, and use a binary number system to represent a numerical value, such as a voltage that gives the brightness of a TV screen pixel. In a normal number, such as 609, there are *ten* choices for each digit—from 0 to 9—and normal numbers are called **decimal** (Latin for ten). Then 609 means: $6 \times 100 + 0 \times 10 + 9 \times 1$. In a **binary** number, each digit or **bit**† has only *two* possibilities, 0 or 1 (sometimes referred to as "off" or "on").‡ In binary, 0001 means "one," 0010 means 2, 0011 means 3, and 1101 means $(1\times2^3 + 1\times2^2 + 0\times2^1 + 1\times2^0) = 8 + 4 + 0 + 1 = 13$ in decimal. See Table 23–3. Any binary value can be represented by a voltage pattern something like that shown in Fig. 23–22.

TABLE 23–3
Binary to Decimal

Binary† number	Decimal number
00000000	0
00000001	1
00000010	2
00000011	3
00000100	4
00000111	7
00001000	8
00100101	37
11111111	255

†Note that we start counting from right to left: the 1's digit is on the far right, then the 2's, the 4's, the 8's, the 16's, the 32's, the 64's, and the 128's.

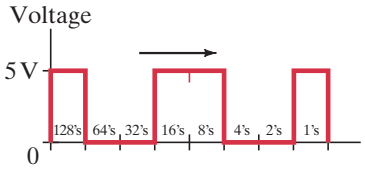


FIGURE 23–22 A digital signal: This sequence would represent 10011001 or 153 (= 128 + 0 + 0 + 16 + 8 + 0 + 0 + 1).

A "1" is a positive voltage such as +5 V, whereas a "0" could be represented as 0 V. For example, the brightness signal that goes to each of the millions of tiny picture elements or "subpixels" of a TV or computer screen (Fig. 23–31, Section 23–10), is contained in a **byte**. One byte is 8 bits, which means

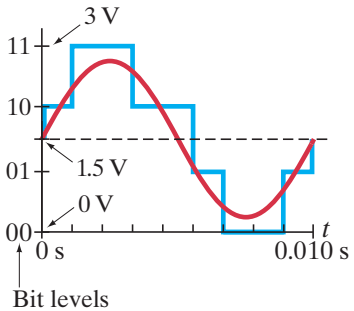
each byte of 8 bits allows $2^8 = 256$ possibilities

(that is: 0, 1, 2, up to 255) or 256 shades for each of 3 colors: red, green, blue. The full color of each pixel (the three subpixel colors) has $(256)^3 = 17 \times 10^6$ possibilities.

Digital television signals, which we discuss in the next Section, are usually transmitted over the air at a **transmission rate** or **bit-rate** at about 19 Mb/s = 19 megabits per second for normal HD (high definition). So 19×10^6 bits pass a given point per second, or one bit every 53 nanoseconds. We could write this in terms of bytes as 2.4 MB/s, where for bytes we use capital B.

When an analog signal, such as the pure sine wave of Fig. 23–21a, is converted to digital (**analog-to-digital converter, ADC**), the digital signal may look like the blue squared-off curve of Fig. 23–23. The digital signal has a limited number of discrete values. The difference between the original continuous analog signal

FIGURE 23–23 The red analog sine wave, which is at a 100-Hz frequency (1 wavelength is done in 0.010 s), has been converted to a 2-bit (4 level) digital signal (blue). (Bit depth = 2, sample rate = 900 per second.)



†"Bit" is a contraction of "**b**inary digi**t**", leaving out the 8 intervening letters.

‡*Morse code*, used in the telegraph for over a century (and still around) is binary: dots and dashes. The letter A is · —, an S is · · ·, and an O is — — —. The universal emergency distress signal SOS is · · · — — — · · ·, three dots and three dashes. They may be short and long *sounds*, by voice or from a loudspeaker and heard by an ear. Other possibilities are light flashes or flags.

and its digital approximation is called the **quantization error** or **quantization loss**. To minimize that loss, there are two important factors: (i) the **resolution** or **bit depth**, which is the number of bits or values for the voltage of each sample (= measurement); (ii) the **sampling rate**, which is the number of times per second the original analog voltage is measured or "sampled." (Audio CDs are sampled at 44.1 kHz with a bit depth of 16 bits per sample.)

Consider a digital approximation for the 100-Hz sine wave shown in Figure 23–23. The digital approximation uses (i) a 0 to 3-V, 2-bit depth, measuring only 4 possible voltages (00, 01, 10, 11, meaning 0, 1, 2, 3 in decimal); and (ii) a sampling rate of 9 samples in one cycle or one wavelength, which in one second becomes $(9)(100\,\text{cycles/s}) = 900\,\text{Hz}$, or 900 samples/second. This bit depth (2) and sampling rate (900 Hz) are both of very poor quality. For high quality reproduction, a greater bit depth and higher sampling rate are needed, which requires more data to be transmitted and stored.

For audio CDs, the sampling rate is 44.1 kHz (44,100 samplings every second) and 16-bit resolution, meaning each sampled voltage can have $2^{16} \approx 65{,}000$ different voltage levels between, let's say, 0 and 5 volts. See Fig. 23–24 for details. Audio *recording* often uses 24-bit $(2^{24} \approx 17 \times 10^6$ voltage levels) and 96 kHz (or 192 kHz) to get a better approximation of the original analog signal (on super-CDs, audio DVDs, Blu-ray, or solid-state memory), but must be transferred down to 44.1 kHz and 16-bit to produce ordinary CDs. But iPods and MP3 players have much lower sampling rates and much less detail, which some listeners can notice.

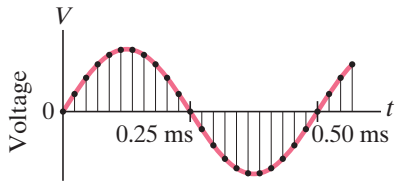
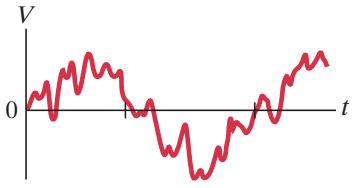


FIGURE 23–24 The sine wave shown could represent the analog electric signal from a microphone due to a pure 2000-Hz tone. (See Chapters 15 and 16.) The analog-to-digital electronics **samples** the signal—that is, measures and records the signal's voltage at intervals, many times per second. Each black dot on the curve represents the voltage measured (sampled) at that moment in time. The sampling rate in this diagram is 44,100 each second, or 44.1 kHz, like a CD. That is, a sample is taken every $(1\,\text{s})/44{,}100 = 0.000023\,\text{s} = 0.023\,\text{ms}$. In 0.50 ms, as shown here for one wavelength, 22 samples (black dots) are taken. This is an alternate way to represent sampling compared to Fig. 23–23, and shows that we cannot see any changes that might happen between the samplings of the voltage (dots).

Figure 23–24 gives some details about a pure 2000-Hz sound sampled at 44.1 kHz. Normal musical sounds are a complex summation of many such sine waves of different frequencies and amplitudes. A simple summation was shown in Figs. 15–16 and 16–13. Another example is shown in Fig. 23–25, where we can see that the fine details may be missed by a digital conversion. Look at Fig. 23–24: if that signal was instead 20,000 Hz (highest frequency of human hearing), it would be sampled only about two times per wavelength. Both those samples might be zero volts—obviously missing the entire waveform. Over many wavelengths, it might eventually reproduce the waveform somewhat well. But many sounds may change in less than a millisecond, like the initial attack of a piano note or plucked guitar string. Many audiophiles hear the difference between an analog vinyl record and its release as a CD at 44.1 kHz, and even more difference if MP3.

FIGURE 23–25 This type of complex signal is much more typical than the pure sine wave of Fig. 23–24. Sampling may not catch all the details, especially because the waveform is changing very fast in time.



An analog signal, such as from a microphone or analog radio, can go directly to a loudspeaker, but a digital signal cannot. Digital audio signals must be converted back to analog (**digital-to-analog converters, DAC**) before being sent to a loudspeaker or headset. Even in a TV set, the digital video signals are converted to analog voltages before addressing the color pixels (Section 23–10), although the picture itself might be said to be digital since it is made up of separate pixels.

Digital photographs are made up of millions of "pixels" to produce a sharp image that is not "pixelated" or blurry. Also important (and complicated) are the number of bits provided for colors, plus the ability of the sensors (Chapter 33) to sustain a wide range of brightnesses from dim to bright light conditions.

Digital data has some real advantages: for one, it can be **compressed**, in the sense that repeated information can be reduced so that less memory space is required—fewer bits and bytes. For example, adjacent "pixels" of a photograph that includes a blue sky may be essentially identical. If 500 almost identical pixels can be coded as identical, that takes up less memory (or "size") than to specify all the 500 pixels individually.

Compression schemes, like **jpeg** for photos, lose some information and the resulting picture is less sharp. In audio, MP3 players use about one-tenth the information that a CD does, but many listeners don't notice. Compression is one reason that more data or "information" can be transmitted digitally for a given "bandwidth". **Bandwidth** is the fixed range of frequencies allotted to each radio or TV station; or for digital transmission, it sets the upper limits for the **bit-rate**, the number of bits transmitted per second. Another advantage of digital is that data transmission rate, such as for TV, can be manipulated. For example, a TV station can use all its bandwidth for one very HD program, or instead can transmit two or three programs in lower definition. Both compression and low definition reduce quality.

Table 23–4 summarizes digital values for various audio devices. Also included are the audio of TV and DVD, but not the video which involves more complicated parameters.[†]

TABLE 23–4 Digital Values, Audio (values typical)

	Bit depth	Sampling rate	Transmission rate
Cell phone	8–18	8–22 kHz	<1–100 Mb/s
MP3	8	44.1 kHz	128–320 kb/s
CD	16	44.1 kHz	1.41 Mb/s
TV	24	48 kHz	19 Mb/s
DVD	24	48 kHz	3–10 Mb/s
Blu-ray disc	24	96–192 kHz	36–54 Mb/s
WiFi	—	—	50–1000 Mb/s

*Noise, and Digital Transmission

Digital information transmission has another advantage. Any distortion or unwanted (external) electrical signal that intrudes from outside, broadly called **noise**, can badly corrupt an analog signal: Fig. 23–26a shows a time-varying analog signal, and Fig. 23–26b shows nasty outside noise interfering with it. But a digital signal would still be readable (Figs. 23–26c and d), unless the noise is very large, on the order of half the bit signal itself.

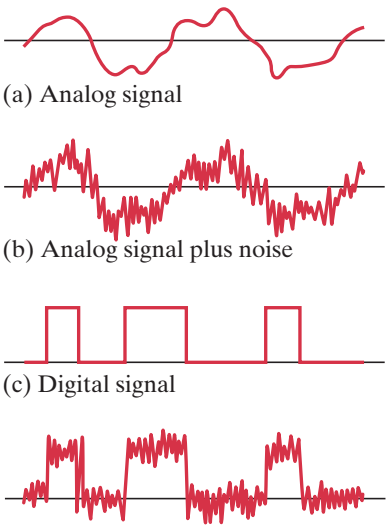


FIGURE 23–26 (a) Original analog signal and (b) the same signal dirtied up by outside signals (= noise). A digital signal (c) is still readable (d) without error if the noise is not too great.

Electronic noise is like weeds in a carefully tended garden: an unwanted external voltage and electric field, or any unintended alteration in the original signal. Noise has a number of sources. Any electronic signal involves electric charges whose electric field can affect charges in another nearby signal (**crosstalk** or **interference**). External fields, as from high voltage wires, motors, or fluorescent lamps, can produce noise. **Thermal noise** refers to random motion of electrons, much like the "thermal motion" of the molecules in a gas (Sections 17–1 and 18–1, especially Eq. 18–4). Also important, moving electrons can encounter defects in the wire or other material in which they are moving, which can alter the signal. Cell phone reception when far from a cell tower can become seriously degraded by such noise or interference.

Noise can affect a digital transmission signal (or "data link") that is traveling in a copper wire, optical fiber, or through space as an electromagnetic wave such as WiFi, Bluetooth, and broadcast radio and TV (discussed in Chapter 31). The signal may be a continuous **data stream**, or may be a series of "packets" or "blocks" of data. Noise can cause **bit-flips** (bit changed from a 1 to a 0, or a 0 to a 1). Another type of digital error is a **bit deletion**, or **insertion** of an (unwanted) bit.[‡] A single such error can make all subsequent bits be in the wrong position, and thus fully corrupt the signal.

Digital **error detection** schemes exist. An early one was the use of a **parity** bit: for example, an extra bit was added to each byte, and this ninth bit is 1 if

[†]The specifications (Table 23–4) for cell phones, broadcast TV, and radio, are set by the government, because "airwaves are for everyone". Cable TV, satellite TV, and streaming specifications are set by each company providing service; strange, because the last two also use the airwaves (the government only specifies the carrier frequency range allowed in the electromagnetic spectrum, Chapter 31, Fig. 31–12). These forms of TV vary a lot: often they are poorer than broadcast TV (partly depending on your home network), but some companies offer higher quality picture and sound, perhaps at a high price.

[‡]One way bit deletions and insertions can occur is because the sending and receiving of a signal requires a precision clock at each end, since each bit must have a precise duration. But no two clocks can ever be perfectly in synch over long stretches of a bit stream. The slight difference in the two clocks can result in a **bit slip**, either a bit insertion or deletion.

there are an odd number of bits in that byte (such as 10110011 + 1) and a zero if the byte has an even number of bits (such as 10110010 + 0). For this code, the sum of 1's is always even. If the sum comes out odd, there is an error in that byte.

Error correction schemes exist, some involving added bits, but none is absolutely perfect. If an error is detected in a data packet, the electronics may send back a message requesting the corrupt packet to be resent or, when a byte is detected to be erroneous, that byte is replaced by a "good guess," such as the previous value or an "average" of the value just before and just after, called Forward Error Correction (FEC).

The quality of an analog signal may be considered reasonably good if its **signal-to-noise ratio** (S/N) is above, perhaps, about 60 dB.[†] (The decibel unit was discussed in Section 16–3 and is a type of energy comparison, Eq. 16–6.) Digital systems have S/N ratios greater than 100 dB. Digital signals are more often specified by the **bit-error-rate** (**BER**). A BER of one bit error in a trillion, or 1 in 10^9 bits (also written as $\text{BER} = 1 \times 10^{-9}$), is considered sufficient for TV and 1×10^{-13} ethernet transmission. Decent cell phone signals may have a BER of 10^{-3} (one bit error per thousand bits).

*23–10 TV and Computer Monitors

The first television receivers used a **cathode ray tube** (**CRT**), but by 2010 they had been replaced mainly by **liquid crystal displays** (**LCD**). Nonetheless, an understanding of how a CRT works is useful in physics (see Section 27–7 on the discovery of the electron).

*CRT

The operation of a CRT depends on **thermionic emission**, discovered by Thomas Edison (1847–1931). Consider a voltage applied to two small electrodes inside an evacuated glass "tube" as shown in Fig. 23–27: the **cathode** is negative, and the **anode** is positive. If the cathode is heated (usually by an electric current) so that it becomes hot and glowing, it is found that negative charges leave the cathode and flow to the positive anode. These negative charges are now called electrons, but originally they were called **cathode rays** because they seemed to come from the cathode (more detail in Section 27–7 on the discovery of the electron).

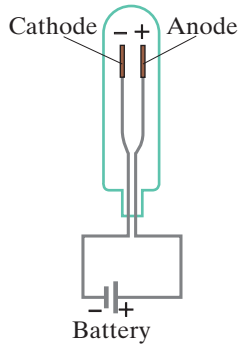


FIGURE 23–27 If the cathode inside the evacuated glass tube is heated to glowing (by an electric current, not shown), negatively charged "cathode rays" (= electrons) are "boiled off" and flow across to the anode (+), to which they are attracted.

PHYSICS APPLIED
CRT

Figure 23–28 is a simplified sketch of a CRT which is contained in an evacuated glass tube. A beam of electrons, emitted by the heated cathode, is accelerated by the high-voltage anode and passes through a small hole in that anode.

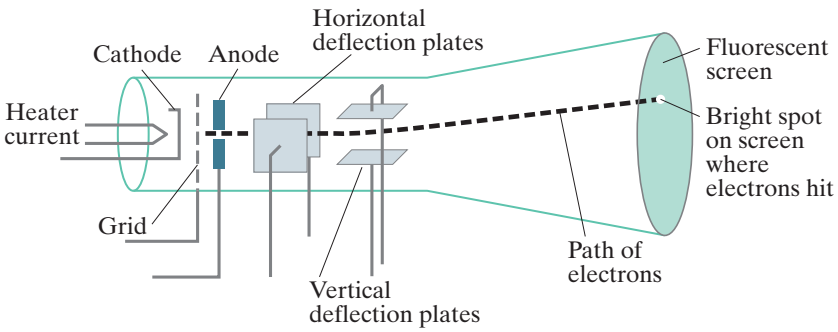
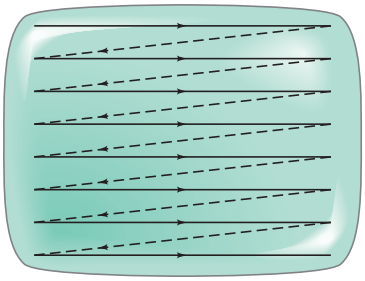


FIGURE 23–28 A cathode-ray tube. Magnetic deflection coils are commonly used in place of the electric deflection plates shown here. The relative positions of the elements have been exaggerated for clarity.

The inside of the tube face on the right (the screen) is coated with a fluorescent material that glows at the spot where the electrons hit. Voltage applied across the horizontal and vertical deflection plates, Fig. 23–28, can be varied to deflect the electron beam to different spots on the screen.

In TV and computer monitors, the CRT electron beam sweeps over the screen in the manner ("raster") shown in Fig. 23–29 by carefully synchronized voltages applied to the deflection plates (more commonly by magnetic deflection coils—Chapter 27). During each horizontal sweep of the electron beam, the **grid** electrode (Fig. 23–28) receives a signal voltage that limits the flow of electrons at each instant during the sweep; the more negative the grid voltage is, the more electrons are repelled and fewer pass through, producing a less bright spot on the screen.

FIGURE 23–29 Electron beam sweeps across a CRT television screen in a succession of horizontal lines, referred to as a **raster**. Each horizontal sweep is made by varying the voltage on the horizontal deflection plates (Fig. 23–28). Then the electron beam is moved down a short distance by a change in voltage on the vertical deflection plates, and the process is repeated.



[†]Vinyl records (33 rpm) and analog audio tape recordings have a background noise ("tape hiss") due to mechanical properties of vinyl grooves and magnetic tape; and the electronics can produce noise such as "hum." *Dolby noise reduction* was invented to reduce tape hiss by increasing the level (voltage) of quiet sounds above about 1 kHz during recording, and then in playback reducing their level (along with the hiss) back down.

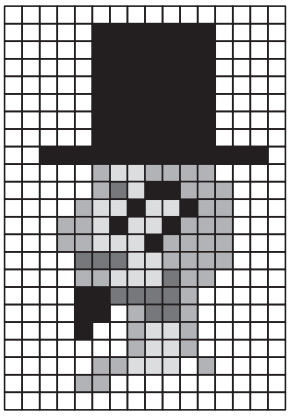


FIGURE 23–30 Example of an image made up of many small squares or *pixels* (picture elements). This one has rather low resolution.

Thus the varying grid voltage is responsible for the brightness of each spot on the screen. At the end of each horizontal sweep of the electron beam, the horizontal deflection voltage changes dramatically to bring the beam back to the opposite side of the screen, and the vertical voltage changes slightly so the beam begins a new horizontal sweep slightly below the previous one. The difference in brightness of the spots on the screen forms the "picture", Fig. 23–30. **Color screens** have red, green, and blue phosphors which glow when struck by the electron beam. The various brightnesses of adjacent red, green, and blue phosphors (so close together we don't distinguish them) produce almost any color. Analog TV for the U.S. provided 480 visible horizontal sweeps[†] to form a complete picture every $\frac{1}{30}$ s. With 30 new frames or pictures every second (25 in countries with 50-Hz line voltage), a "moving picture" is displayed on the TV screen. (Note: commercial movies on film are 24 frames per second.)

*Digital TV and Addressing Pixels

PHYSICS APPLIED
How flat screens work

Today's TV screens and computer monitors, whether flat or curved, contain millions of tiny *picture elements*, or **pixels** which are "addressed" digitally. A standard pixel consists of 3 **subpixels**, a red, a green, and a blue. A close up of a common arrangement of pixels is shown in Fig. 23–31 for an LCD screen. (How liquid crystals work in an LCD screen is described in Section 34–8.) Subpixels are small enough that at normal viewing distances we don't distinguish them and the separate red (R), green (G), and blue (B) subpixels blend to produce almost any color, depending on the relative brightnesses of the three subpixels. Liquid crystals act as filters (R, G, and B) that filter the light from a white **backlight**, usually fluorescent lamps or *light-emitting diodes* (LED, Section 40–9).[‡] The picture you see on the screen depends on the level of brightness of each subpixel (each encoded in 8 bits), as suggested in Fig. 23–30 for a simple black and white picture.

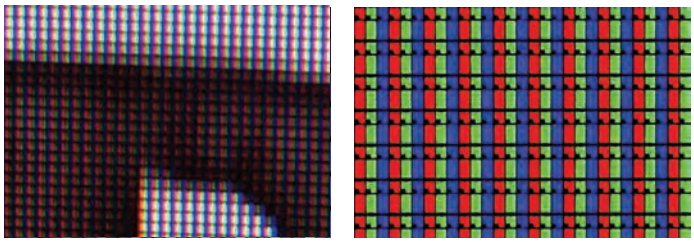


FIGURE 23–31 Close up photographs of a tiny section of two typical LCD screens. You can even make out wires and transistors in the one on the right.

Standard "high definition" (HD) television screens have 1080 horizontal rows of pixels, each row consisting of 1920 pixels across the screen. That is, roughly 2000 pixels (2K) across by 1000 pixels high, for a total of 2 million pixels. Television in the U.S. is transmitted digitally at a rate of 60 Hz—that is, 60 frames or pictures per second (50 Hz in many countries) which makes the "moving picture." To form one frame, each subpixel must have the correct brightness. We now describe one way of doing this. Ultra high definition 4K screens (UHD) have 4 times the number of pixels, 3840×2160; their operation is the same as described below.

The brightness of each LCD subpixel (Section 34–8) depends on the voltage between its front and its back: if this voltage ΔV is zero, that subpixel is at maximum brightness; if ΔV is at its maximum (which might be +5 volts), that subpixel is dark (no color). Giving the correct voltage (to provide the correct brightness) is called **addressing** the subpixel. Typically the front of the subpixel is maintained at a positive voltage, such as +5 V. On the back of the display, the voltage at each subpixel is provided at the intersection of the 1080 horizontal wires (rows) and 1920×3 (colors) $\approx$ 6000 vertical wires (columns).

[†]525 lines in total, but only 480 form the picture; the other 45 lines contain other information such as synchronization. The sweep is **interlaced**: that is, every $\frac{1}{60}$ s every other line is traced, and in the next $\frac{1}{60}$ s, the lines in between are traced.

[‡]LEDs are discussed in Section 40–9. They *emit* light. TVs advertised as LED can be confusing. There are two meanings: an LED TV may mean an LCD screen with an LED backlight (preferable to older fluorescent backlight). Secondly, a true LED TV uses light *emitting* diodes and needs no backlight. LED pixels are somewhat large and are used in very large LED displays such as at stadiums. True LED screens for use at home must be very large because LED pixels are large (you have to view far away so you don't notice the pixels). The best home TV sets have OLED (organic LED) screens. See Section 40–9.

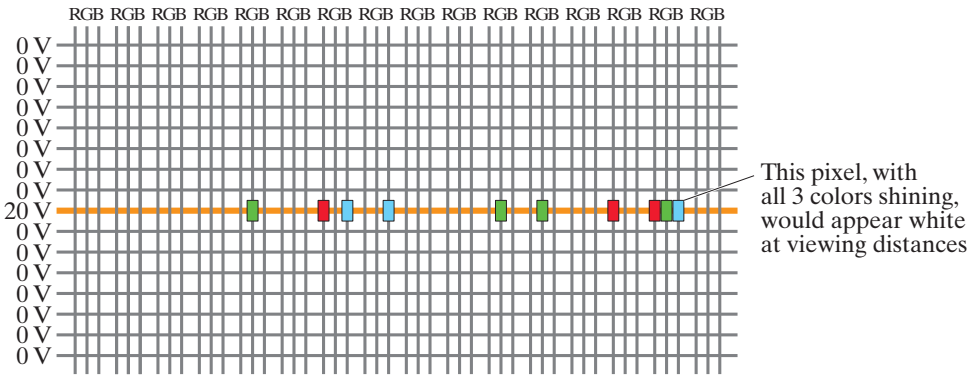


FIGURE 23–32 Array of wires (a matrix) behind all the pixels on an LCD screen. Each intersection of two wires is at a subpixel (red, green, or blue). One horizontal wire is activated at a time (the orange one at the moment shown) meaning it is at a positive voltage (+20 V) which allows that one row of pixels to be addressed at that moment; all other horizontal wires are at 0 V. At this moment, the data stream arrives to all the vertical wires, presenting the needed voltage (between 0 and 5 V) to produce the correct brightness for each of the nearly 6000 subpixels along the activated row.

See Fig. 23–32, which shows the array, or **matrix**, of wires. Each intersection of one vertical and one horizontal wire lies behind one subpixel. Because 60 frames are shown per second (or maybe 120 in some TV sets), the signal voltages applied are brief, like a pulse (see Fig. 23–21b).

The video signal that arrives at the display **activates** only one horizontal wire at a time (the orange line in Fig. 23–32): that one horizontal wire has a voltage (let's say +20 V) whereas all the others are at 0 V. That 20 V is not applied directly to the pixels, but *allows* the vertical wires to apply briefly the proper "signal voltage" to each subpixel along that row (via a transistor, see below). These signal voltages, the **data stream**, are applied to all the vertical wires just as that one horizontal row is activated: they provide the correct brightness for each subpixel in that activated row. A few subpixels are highlighted in Fig. 23–32. Immediately afterward, the other rows are activated, one by one, until the entire frame has been completed (in $\frac{1}{60}$ s). Then a new frame is started. The addressing of subpixels for each row of each frame serves the same purpose as the sweep of the electron beam in a CRT, Fig. 23–29.

*Active Matrix (advanced)

High-definition displays use an **active matrix**, meaning that a tiny **thin-film transistor** (**TFT**) is attached to a corner of the back of each subpixel. (Transistors are discussed in Section 40–10.) One electrode of each TFT, called the "source," is connected to the vertical wire which addresses that subpixel, Fig. 23–33, and the "drain" electrode is connected to the back of the subpixel. The horizontal "activating" wire that serves the subpixel is connected to the transistor's **gate** electrode. The gate's voltage, by attracting charge or not, functions as a switch to connect or disconnect the source voltage to the drain and to the back of the subpixel (its front is fixed at +5 V). The potential difference ΔV across a subpixel, determined by the vertical wire if the gate is "on", effects whether that subpixel will be bright in color ($\Delta V = 0$), black ($\Delta V =$ maximum), or something in between (to produce a possible range of 17 million different colors or shades of color). See Fig. 23–34. All the subpixel TFTs along the one activated horizontal wire (the orange line in Fig. 23–32) will have +20 V at the gate: the TFTs are turned "on," like a switch. That allows electric charge to flow, connecting the vertical wire signal voltage at each TFT source to its drain and to the back of the subpixel, providing the ΔV corresponding to the correct brightness or shade of the color. Thus all subpixels along one row receive the brightness they need for that line of the frame.

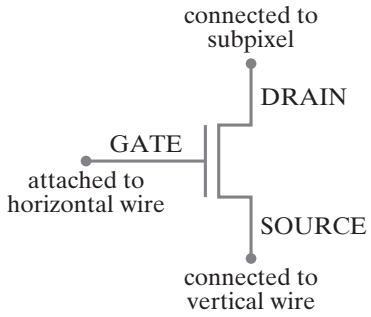


FIGURE 23–33 Thin-film transistor. One is attached to each screen subpixel.

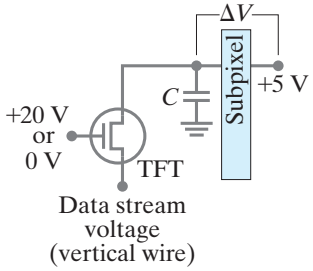


FIGURE 23–34 Circuit diagram for one subpixel. The front of the subpixel is at +5 V. If the TFT gate is at 20 V (horizontal wire activated), the data stream voltage is applied to the back of the subpixel, and determines the brightness of that subpixel. If the gate is at 0 V (horizontal wire not activated), the TFT is "off": no charge passes through, and the capacitance helps maintain ΔV until the subpixel is updated $\frac{1}{60}$ s later.

Within a subpixel's electronics is a capacitance (which stores charge—Chapter 24) that helps maintain the ΔV until that subpixel is **updated** with a new signal for the next frame, $\frac{1}{60}$ s later ($\approx$17 ms). The row below the orange one shown in Fig. 23–32 is activated about 15 μs later $\left[=\left(\frac{1}{60}\text{ s}\right)\left(\frac{1}{1080\text{ lines}}\right)\right]$. The 6000 vertical wires **(data lines)** get their signal voltages (data stream) updated just before each row is activated in order to establish the brightness of each subpixel in that next row. All 1080 rows are activated, one-by-one, within $\frac{1}{60}$ s ($\approx$17 ms) to complete that frame. Then a new frame is started.

[Note: Digital TV in HD is transmitted at about 19 Mb/s as mentioned in Section 23–9. An HD screen has 1080×1920 pixels $\approx 2 \times 10^6$ pixels, and each pixel has 8 bits for each of the 3 colors, for a total of $24 \times 2 \times 10^6 \approx 48$ Mb (megabits); at a frame rate of 60 Hz (60 frames/s), if every pixel was addressed we would need a bit rate of $48\text{ Mb} \times 60\text{s}^{-1} \approx 3000$ Mb/s. But TV is sent at only 19 Mb/s and $3000/19 \approx 150$. So in fact, only about $1/150 \approx 1\%$ of pixels can be addressed for each frame! Huge compression is needed. In fact, if you get close to a TV screen when there is lots of movement on the screen, you will see that it is very blurry.]

The TV set or "box" that receives the digital video signal has to decode the signal in order to send analog voltages to the pixels of the screen, and at just the right time. TV stations in the U.S. are allowed to broadcast HD at 1080×1920 pixels or at 720×1280, or in standard definition (SD) of 480×704 pixels. 4K transmission is up to 57 Mb/s, only 3 times faster but with 4 times the number of pixels. 8K would quadruple pixels again; it may be that many people would not notice much difference between 4K (or 8K) and ordinary HD (2K).

When you read 1080p or 1080i for a TV, the "p" stands for "progressive," meaning an entire frame is made in $\frac{1}{60}$ s as described above. The "i" stands for "interlaced," meaning all the odd rows (half the picture) are done in $\frac{1}{60}$ s and then all the even rows are done in the next $\frac{1}{60}$ s, so a full picture is done at 30 per second or 30 Hz, thus reducing the data (or bit) rate. Analog TV (U.S.) was 480i.



FIGURE 23–35 An electrocardiogram (ECG) trace displayed on a CRT.

*Oscilloscopes

PHYSICS APPLIED
Oscilloscope

An **oscilloscope** is a device for amplifying, measuring, and visually displaying an electrical signal as a function of time on a monitor screen or computer screen. The visible "trace" on the screen, which could be an electrocardiogram of the electric signals from the heart (Fig. 23–35) or a signal from an experiment on nerve conduction, is a plot of the signal voltage (vertically) versus time (horizontally). [In a CRT, the electron beam is swept horizontally at a uniform rate in time by the horizontal deflection plates, Figs. 23–28 and 23–29. The signal to be displayed is applied (after amplification) to the vertical deflection plates.]

*23–11 Electrocardiogram (ECG or EKG)

PHYSICS APPLIED
Electrocardiogram

Each time the heart beats, changes in electrical potential occur on its surface that can be detected using *electrodes* (metal contacts), which are attached to the skin. The changes in potential are small, on the order of millivolts (mV), and must be amplified. They are displayed with a chart recorder on paper, or on a monitor screen, Fig. 23–35. An **electrocardiogram** (ECG or EKG) is the record of the potential changes for a given person's heart. An example is shown in Fig. 23–36.

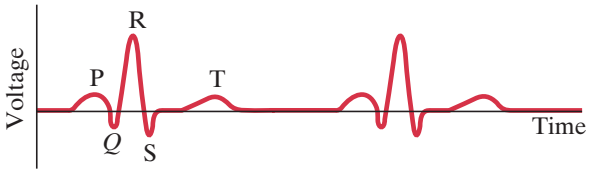


FIGURE 23–36 Typical ECG. Two heart beats are shown.

Both muscle and nerve cells have an electric dipole layer across the cell wall. That is, in the normal situation there is a net positive charge on the exterior surface and a net negative charge on the interior surface, Fig. 23–37a. The amount of charge depends on the size of the cell, but is approximately 10^{-3} C/m^2

of surface. For a cell whose surface area is $10^{-5}\,\text{m}^2$, the total charge on each surface is thus $\approx 10^{-8}\,\text{C}$. Just before the contraction of heart muscles, changes occur in the cell wall, so that positive ions on the exterior of the cell are able to pass through the wall and neutralize charge on the inside, or even make the inside surface slightly positive compared to the exterior. This "depolarization" starts at one end of the cell and progresses toward the opposite end, as indicated by the arrow in Fig. 23–37b, until the whole muscle is depolarized; the muscle then repolarizes to its original state (Fig. 23–37a), all in less than a second. Figure 23–37c shows rough graphs of the potential V as a function of time at the two points P and P′ (on either side of this cell) as the depolarization moves across the cell. The path of depolarization within the heart as a whole is more complicated than described here for one cell, and produces the complex potential difference as a function of time, Fig. 23–36.

It is standard procedure to divide a typical electrocardiogram into regions corresponding to the various deflections (or "waves"), as shown in Fig. 23–36. Each of the deflections corresponds to the activity of a particular part of the heart beat (Fig. 13–44). The P wave corresponds to contraction of the atria. The QRS group corresponds to contraction of the ventricles as the depolarization follows a very complicated path. The T wave corresponds to recovery (repolarization) of the heart in preparation for the next cycle.

The ECG is a powerful tool in identifying heart defects. For example, the right side of the heart enlarges if the right ventricle must push against an abnormally large load (as when blood vessels become hardened or clogged). This problem is readily observed on an ECG, because the S wave becomes very large (negatively). *Infarcts*, which are dead regions of the heart muscle that result from heart attacks, are also detected on an ECG because they reflect the depolarization wave.

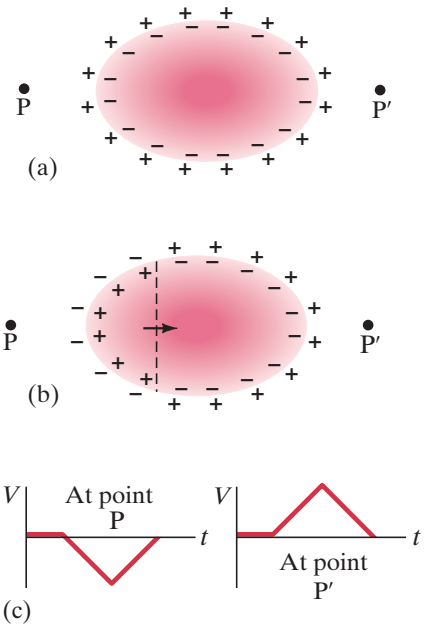


FIGURE 23–37 A heart muscle cell showing (a) charge dipole layer in resting state; (b) depolarization of cell progressing as muscle begins to contract; and (c) potential V at points P and P′ as a function of time.

Summary

Electric potential is defined as electric potential energy per unit charge. That is, the **electric potential difference** between any two points in space is defined as the difference in potential energy of a test charge q placed at those two points, divided by the charge q:

$$V_{\text{ba}} = \frac{U_{\text{b}} - U_{\text{a}}}{q}. \qquad \textbf{(23–2b)}$$

Potential difference is measured in volts $(1\,\text{V} = 1\,\text{J/C})$ and is sometimes referred to as **voltage**.

The change in potential energy of a charge q when it moves through a potential difference V_{ba} is

$$\Delta U = qV_{\text{ba}}. \qquad \textbf{(23–3)}$$

The potential difference V_{ba} between two points, a and b, is given by the relation

$$V_{\text{ba}} = V_{\text{b}} - V_{\text{a}} = -\int_{\text{a}}^{\text{b}} \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}. \qquad \textbf{(23–4a)}$$

Thus V_{ba} can be found in any region where $\vec{\mathbf{E}}$ is known. If the electric field is uniform, the integral is easy: $V_{\text{ba}} = -Ed$, where d is the distance (parallel to the field lines) between the two points.

An **equipotential line** or **surface** is all at the same potential, and is perpendicular to the electric field at all points.

The electric potential due to a single point charge Q, relative to zero potential at infinity, is given by

$$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}. \qquad \textbf{(23–5)}$$

The potential due to any charge distribution can be obtained by summing (or integrating) over the potentials for all the charges.

The potential due to an **electric dipole** drops off as $1/r^2$. The **dipole moment** is $p = Q\ell$, where ℓ is the distance between the two equal but opposite charges of magnitude Q.

When V is known, the components of $\vec{\mathbf{E}}$ can be found from the inverse of Eq. 23–4a, namely

$$E_x = -\frac{\partial V}{\partial x}, \qquad E_y = -\frac{\partial V}{\partial y}, \qquad E_z = -\frac{\partial V}{\partial z}. \qquad \textbf{(23–9)}$$

Digital electronics converts an analog **signal voltage** into an approximate digital voltage based on a **binary code**: each **bit** has two possibilities, 1 or 0 (also "on" or "off"). The binary number 01101 equals 13. A **byte** is 8 bits and provides $2^8 = 256$ voltage levels. **Sampling rate** is the number of voltage measurements done on the analog signal per second. The **bit depth** is the number of digital voltage levels available at each sampling. CDs are 44.1 kHz, 16-bit.

[*TV and computer monitors historically used a **cathode ray tube** (CRT) which accelerates electrons by high voltage, and sweeps them across the screen using magnetic coils or electric deflection plates. **LCD flat screens** contain millions of **pixels**, each with a red, green, and blue **subpixel** whose brightness is addressed via a **matrix** of horizontal and vertical wires using a **digital** (**binary**) code.]

[*An **electrocardiogram** (ECG or EKG) records the potential changes of each heart beat as the cells depolarize and repolarize.]

Questions

1. If two points are at the same potential, does this mean that no net work is done in moving a test charge from one point to the other? Does this imply that no force must be exerted? Explain.
2. If a negative charge is initially at rest in an electric field, will it move toward a region of higher potential or lower potential? What about a positive charge? How does the potential energy of the charge change in each instance? Explain.
3. State clearly the difference (*a*) between electric potential and electric field, (*b*) between electric potential and electric potential energy.
4. An electron is accelerated from rest by a potential difference of 0.20 V. How much greater would its final speed be if it is accelerated with four times as much voltage? Explain.
5. Is there a point along the line joining two equal positive charges where the electric field is zero? Where the electric potential is zero? Explain.
6. Can a particle ever move from a region of low electric potential to one of high potential and yet have its electric potential energy decrease? Explain.
7. If $V = 0$ at a point in space, must $\vec{\mathbf{E}} = 0$? If $\vec{\mathbf{E}} = 0$ at some point, must $V = 0$ at that point? Explain. Give examples for each.
8. Can two equipotential lines cross? Explain.
9. Draw in a few equipotential lines in Figs. 21–35b and c back in Chapter 21, as we did here in Figs. 23–16 and 23–18 for Figs. 21–35d and a.
10. Make a list of rules for and properties of equipotential surfaces or lines. You should be able to find eight distinct rules in the text.
11. When dealing with practical devices, we often take the ground (the Earth) to be 0 V. (*a*) If instead we said the ground was -10 V, how would this affect V and E at other points? (*b*) Does the fact that the Earth carries a net charge affect the choice of V at its surface?
12. A satellite orbits the Earth along a gravitational equipotential line. What shape must the orbit be?
13. A conducting sphere carries a charge Q and a second identical conducting sphere is neutral. The two are initially isolated, but then they are placed in contact. (*a*) What can you say about the potential of each when they are in contact? (*b*) Will charge flow from one to the other? If so, how much? (*c*) If the spheres do not have the same radius, how are your answers to parts a and b altered?
14. Suppose the charged ring of Example 23–9 was not uniformly charged, so that the density of charge was twice as great near the top as near the bottom. Assuming the total charge Q is unchanged, would this affect the potential at point P on the axis (Fig. 23–14)? Would it affect the value of $\vec{\mathbf{E}}$ at that point? Is there a discrepancy here? Explain.
15. Consider a metal conductor in the shape of a football (oblong). If it carries a total charge Q, where would you expect the charge density σ to be greatest, at the ends or along the flatter sides? Explain. [*Hint*: Near the surface of a conductor, $E = \sigma/\epsilon_0$.]
16. If you know V at a point in space, can you calculate $\vec{\mathbf{E}}$ at that point? If you know $\vec{\mathbf{E}}$ at a point can you calculate V at that point? If not, what else must be known in each case?
17. If the electric field $\vec{\mathbf{E}}$ is uniform in a region, what can you infer about the electric potential V? If V is uniform in a region of space, what can you infer about $\vec{\mathbf{E}}$?
18. Is the electric potential energy of two isolated unlike charges positive or negative? What about two like charges? What is the significance of the sign of the potential energy in each case?
19. At a particular location, the electric field points due north. In what direction(s) will the rate of change of potential be (*a*) greatest, (*b*) least, and (*c*) zero?
20. Equipotential lines are spaced 1.00 V apart. Does the distance between the lines in different regions of space tell you anything about the relative strengths of $\vec{\mathbf{E}}$ in those regions? If so, what?

MisConceptual Questions

1. A $+0.2\,\mu$C charge is in a uniform electric field $\vec{\mathbf{E}}$. What happens if that charge is replaced by a $-0.2\,\mu$C charge?
   (*a*) The electric potential changes sign, but the electric potential energy stays the same.
   (*b*) The electric potential stays the same, but the electric potential energy changes sign.
   (*c*) Both the electric potential and electric potential energy change sign.
   (*d*) Both the electric potential and electric potential energy stay the same.
2. Two identical positive charges are placed near each other. At the point halfway between the two charges,
   (*a*) the electric field is zero and the potential is positive.
   (*b*) the electric field is zero and the potential is zero.
   (*c*) the electric field is not zero and the potential is positive.
   (*d*) the electric field is not zero and the potential is zero.
   (*e*) None of these statements is true.
3. Four identical point charges are arranged at the corners of a square [*Hint*: Draw a figure]. The electric field E and potential V at the center of the square are
   (*a*) $E = 0$, $V = 0$.
   (*b*) $E = 0$, $V \neq 0$.
   (*c*) $E \neq 0$, $V \neq 0$.
   (*d*) $E \neq 0$, $V = 0$.
   (*e*) $E = V$ regardless of the value.
4. Which of the following statements is valid?
   (*a*) If the potential at a particular point is zero, the field at that point must be zero.
   (*b*) If the field at a particular point is zero, the potential at that point must be zero.
   (*c*) If the field throughout a particular region is constant, the potential throughout that region must be zero.
   (*d*) If the potential throughout a particular region is constant, the field throughout that region must be zero.

5. A positive charge moves in a direction opposite to that of an electric field. What happens to the energy associated with the charge?
   (*a*) The electric potential energy of the charge increases, and the kinetic energy decreases.
   (*b*) Both the electric potential energy and the kinetic energy of the charge increase.
   (*c*) Both the electric potential energy and the kinetic energy of the charge decrease.
   (*d*) The electric potential energy of the charge decreases, and the kinetic energy increases.

6. A uniform electric field points north. If you move south in this field, the magnitude of the electric potential
   (*a*) decreases.
   (*b*) increases.
   (*c*) stays the same.
   (*d*) cannot be determined—more information is needed.

7. A proton $(Q = +e)$ and an electron $(Q = -e)$ are in a constant electric field created by oppositely charged parallel plates (Fig. 23–1). You release the heavier proton from near the positive plate and the electron from near the negative plate. When they each hit the opposite plate, which has more kinetic energy?
   (*a*) The proton.
   (*b*) The electron.
   (*c*) Both acquire the same kinetic energy.
   (*d*) Neither—there is no change in kinetic energy.
   (*e*) They both acquire the same kinetic energy but with opposite signs.

8. Figure 23–38 shows the electric potential V at five different points (A, B, C, D, E) in a uniform electric field. At which point is the electric potential the largest?
   (*a*) A.
   (*b*) B.
   (*c*) C.
   (*d*) D.
   (*e*) E.
   (*f*) All the same.

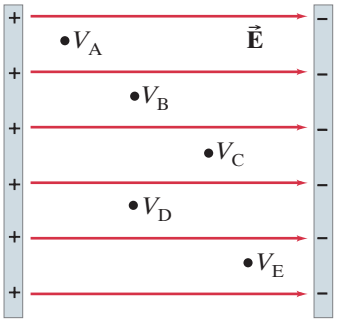


FIGURE 23–38
MisConceptual Question 8.

9. A tiny test charge $+q$ and another of charge $+2q$ are both placed a distance r above a large plate holding charge $+Q$. Which of the following is true? The two test charges
   (*a*) have a different potential and a different potential energy.
   (*b*) have a different potential but the same potential energy.
   (*c*) have the same potential but a different potential energy.
   (*d*) have the same potential and the same potential energy.

10. If it takes an amount of work W to move two $+q$ point charges initially an infinite distance apart to a distance d apart from each other, then how much work should it take to move three $+q$ point charges from infinity to a distance d apart from each other?
   (*a*) $2W$.
   (*b*) $3W$.
   (*c*) $4W$.
   (*d*) $6W$.

11. Which requires the most work—to move a positive charge at point O in a uniform electric field $\vec{\mathbf{E}}$ (Fig. 23–39), from point O to point 1, 2, 3, or 4? All points are the same distance from O.
   (*a*) 1.
   (*b*) 2.
   (*c*) 3.
   (*d*) 4.
   (*e*) All would require the same amount of work.

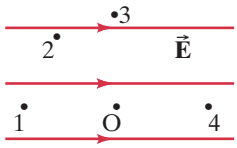


FIGURE 23–39
MisConceptual Question 11.

12. A positive charge moves in the direction of an electric field. Which of the following statements are true?
   (*a*) The potential energy associated with the charge decreases.
   (*b*) The potential energy associated with the charge increases.
   (*c*) The electric field does positive work on the charge.
   (*d*) The electric field does not do any work on the charge.
   (*e*) The electric field does negative work on the charge.

13. Which of the following is a vector?
   (*a*) Electric potential.
   (*b*) Electric potential energy.
   (*c*) Electric field.
   (*d*) Equipotential lines.

Problems

23–1 Electric Potential

1. (I) How much work does the electric field do in moving a $-9.7\,\mu$C charge from ground to a point whose potential is $+65$ V higher?

2. (I) How much work does the electric field do in moving a proton from a point at a potential of $+125$ V to a point at -65 V?

3. (I) What potential difference is needed to stop an electron that has an initial velocity $v = 5.2 \times 10^5$ m/s?

4. (I) An electron acquires 4.65×10^{-16} J of kinetic energy when it is accelerated by an electric field from plate A to plate B. What is the potential difference between the plates, and which plate is at the higher potential?

5. (II) (*a*) An electron accelerates from rest in a uniform electric field $E = 6.0 \times 10^3$ N/C to a speed of 2.0×10^4 m/s. How far did it travel? (*b*) Through what potential difference was the electron accelerated?

6. (II) (*a*) A proton $(Q = +e,\ m = 1.67 \times 10^{-27}\ \text{kg})$ accelerates from rest in a uniform electric field $E = 4.5 \times 10^4$ N/C for 36 cm. What speed does it attain? (*b*) Through what potential difference was the proton accelerated?

7. (II) The work done by an external force to move a $-7.40\,\mu$C charge from point A to point B is 15.0×10^{-4} J. If the charge was started from rest and had 4.82×10^{-4} J of kinetic energy when it reached point B, what must be the potential difference between A and B?

23–2 Potential Related to Electric Field

8. (I) An electric field of 465 V/m is desired between two parallel plates 11.0 mm apart. How large a voltage should be applied?

9. (I) The electric field between two parallel plates connected to a 45-V battery is 1600 V/m. How far apart are the plates?

10. (I) What is the magnitude of the electric field between two parallel plates 4.0 mm apart if the potential difference between them is 85 V?

11. (II) What is the maximum amount of charge that a spherical conductor of radius 5.8 cm can hold in air?

12. (II) What minimum radius must a large conducting sphere (of an electrostatic generating machine) have if it is to be at 45,000 V without discharge into the air? How much charge will it carry?

13. (II) Two parallel plates, connected to a 65-V power supply, are separated by a uniform air gap. How small can the gap be if the air is not to become conducting by exceeding its breakdown value of $E = 3 \times 10^6$ V/m?

14. (II) The electric potential of a very large isolated flat metal plate is V_0. It carries a uniform distribution of charge of surface density σ (C/m^2), or $\sigma/2$ on each surface. Determine V at a distance x from the plate. Consider the point x to be far from the edges and assume x is much smaller than the plate dimensions.

15. (II) The Earth produces an inwardly directed electric field of magnitude 150 V/m near its surface. (*a*) What is the potential of the Earth's surface relative to $V = 0$ at $r = \infty$? (*b*) If the potential of the Earth is chosen to be zero, what is the potential at infinity? (Ignore the fact that positive charge in the ionosphere approximately cancels the Earth's net charge; then later consider how this would affect your answer.)

16. (II) A 32-cm-diameter conducting sphere is charged to 480 V relative to $V = 0$ at $r = \infty$. (*a*) What is the surface charge density σ? (*b*) At what distance will the potential due to the sphere be only 25 V?

17. (II) An insulated spherical conductor of radius r_1 carries a charge Q. A second conducting sphere of radius r_2 and initially uncharged is then connected to the first by a long conducting wire. (*a*) After the connection, what can you say about the electric potential of each sphere? (*b*) How much charge is transferred to the second sphere? Assume the connected spheres are far apart compared to their radii. (Why make this assumption?)

18. (II) Determine the difference in potential between two points that are distances R_a and R_b from a very long ($\gg R_a$ or R_b) straight wire holding a uniform charge per unit length λ.

19. (II) Suppose the end of your finger is charged. (*a*) Estimate the breakdown voltage in air for your finger. (*b*) About what surface charge density would have to be on your finger at this voltage?

20. (II) Estimate the electric field in the membrane wall of a living cell. Assume the wall is 10 nm thick and has a potential of 0.10 V across it.

21. (II) A uniform electric field $\vec{\mathbf{E}} = -6.30\ \text{N/C}\,\hat{\mathbf{i}}$ points in the negative x direction as shown in Fig. 23–40. The x and y coordinates of points A, B, and C are given on the diagram (in meters). Determine the differences in potential (*a*) V_{BA}, (*b*) V_{CB}, and (*c*) V_{CA}.

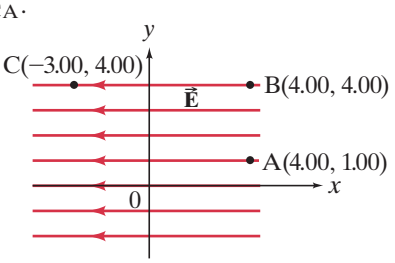


FIGURE 23–40
Problem 21.

22. (III) A hollow spherical conductor, carrying a net charge $+Q$, has inner radius r_1 and outer radius $r_2 = 2r_1$ (Fig. 23–41). At the center of the sphere is a point charge $+Q/2$. (*a*) Write the electric field strength E in all three regions as a function of r. [*Hint*: Use Gauss's law.] Then determine the potential as a function of r, the distance from the center, for (*b*) $r > r_2$, (*c*) $r_1 < r < r_2$, and (*d*) $0 < r < r_1$. (*e*) Plot both V and E as a function of r from $r = 0$ to $r = 2r_2$.

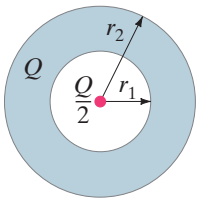


FIGURE 23–41
Problem 22.

23. (III) A very long conducting cylinder (length ℓ) of radius R_0 $(R_0 \ll \ell)$ carries a uniform surface charge density σ (C/m^2). The cylinder is at an electric potential V_0. Determine the potential, at points far from the end, at a distance R from the center of the cylinder for (*a*) $R > R_0$ and (*b*) $R < R_0$. (*c*) Is $V = 0$ at $R = \infty$ (assume $\ell = \infty$)? Explain. [*Hint*: Recall Gauss's law.]

23–3 Potential Due to Point Charges

[Let $V = 0$ at $x = \infty$.]

24. (I) What is the electric potential 15.0 cm from a 3.00 μC point charge?

25. (I) A point charge Q creates an electric potential of +165 V at a distance of 15 cm. What is Q (let $V = 0$ at $r = \infty$)?

26. (II) Two point charges, 3.4 μC and −2.0 μC, are placed 8.0 cm apart on the x axis. At what points along the x axis are (*a*) the electric field zero and (*b*) the potential zero? Let $V = 0$ at $r = \infty$.

27. (II) (*a*) What is the electric potential 2.5×10^{-15} m away from a proton (charge $+e$)? (*b*) What is the electric potential energy of a system that consists of two protons 2.5×10^{-15} m apart—as might occur inside a typical nucleus?

28. (II) Three point charges are arranged at the corners of a square of side ℓ as shown in Fig. 23–42. What is the potential at the fourth corner (point A), taking $V = 0$ at a great distance?

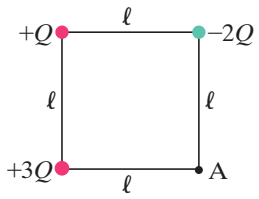


FIGURE 23–42
Problem 28.

29. (II) A +25 μC point charge is placed 6.0 cm from an identical +25 μC point charge. How much work would be required by an external force to move a +0.15 μC test charge from a point midway between them to a point 1.0 cm closer to either of the charges?

30. (II) Point a is 26 cm north of a $-3.8\,\mu C$ point charge, and point b is 36 cm west of the charge (Fig. 23–43). Determine (*a*) $V_b - V_a$, and (*b*) $\vec{E}_b - \vec{E}_a$ (magnitude and direction).

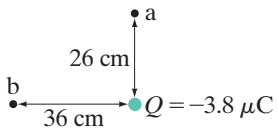


FIGURE 23–43
Problem 30.

31. (II) Two identical $+5.5\,\mu C$ point charges are initially spaced 8.5 cm from each other. If they are released at the same instant from rest, how fast will they be moving when they are very far away from each other? Assume they have identical masses of 1.0 mg.

32. (II) How much voltage must be used to accelerate a proton (radius 1.2×10^{-15} m), starting from rest, so that it has sufficient energy to just "touch" a silicon nucleus? A silicon nucleus has a charge of $+14e$ and its radius is about 3.6×10^{-15} m. Assume the potential is that for point charges.

33. (II) An electron starts from rest 34.5 cm from a fixed point charge with $Q = -0.125$ nC. How fast will the electron be moving when it is very far away?

34. (III) Two equal but opposite charges are separated by a distance d, as shown in Fig. 23–44. Determine a formula for $V_{BA} = V_B - V_A$ for points B and A on the line between the charges situated as shown.

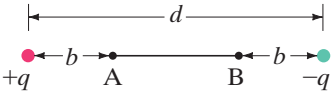


FIGURE 23–44
Problem 34.

23–4 Potential Due to Any Charge Distribution

35. (II) A flat ring of inner radius R_1 and outer radius R_2, Fig. 23–45, carries a uniform surface charge density σ. Determine the electric potential at points along the axis (the x axis). [*Hint*: See Example 23–9.]

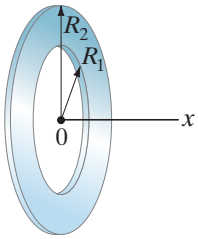


FIGURE 23–45
Problem 35.

36. (II) A total charge Q is uniformly distributed on a thread of length ℓ. The thread forms a semicircle. What is the potential at the full-circle's center? (Assume $V = 0$ at large distances.)

37. (II) A 12.0-cm-radius thin ring carries a uniformly distributed 15.0 μC charge. A small 7.5-g sphere with a charge of 3.6 μC is placed exactly at the center of the ring and given a very small push so it moves along the ring axis ($+x$ axis). How fast will the sphere be moving when it is 2.0 m from the center of the ring (ignore gravity)?

38. (II) A thin rod of length 2ℓ is centered on the x axis as shown in Fig. 23–46. The rod carries a uniformly distributed charge Q. Determine the potential V as a function of y for points along the y axis. Let $V = 0$ at infinity.

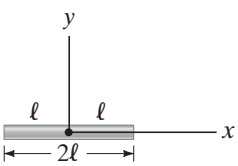


FIGURE 23–46
Problems 38, 39, 40, and 53.

39. (II) Determine the potential $V(x)$ for points along the x axis outside the rod of Fig. 23–46 (Problem 38).

40. (III) The charge on the rod of Fig. 23–46 has a nonuniform linear charge distribution, $\lambda = ax$. Determine the potential V for (*a*) points along the y axis and (*b*) points along the x axis outside the rod.

41. (III) Suppose the flat circular disk of Fig. 23–15 (Example 23–10) has a nonuniform surface charge density that depends on the distance R from the center of the disk: $\sigma = aR^2$. Find the potential $V(x)$ at points along the x axis, relative to $V = 0$ at $x = \infty$.

23–5 Equipotenial Lines and Surfaces

42. (I) Draw a conductor in the oblong shape of a football. This conductor carries a net negative charge, $-Q$. Sketch in a dozen or so electric field lines and equipotential lines surrounding the football.

43. (II) Equipotential surfaces are to be drawn 100 V apart near a very large uniformly charged metal plate carrying a surface charge density $\sigma = 0.75\,\mu C/m^2$. How far apart (in space) are the equipotential surfaces?

44. (II) A metal sphere of radius $r_0 = 0.35$ m carries a charge $Q = 0.50\,\mu C$. Equipotential surfaces are to be drawn for 100-V intervals outside the sphere. Determine the radius r of (*a*) the first, (*b*) the tenth, and (*c*) the 100th equipotential from the surface.

23–6 Dipoles

45. (II) Calculate the electric potential due to a tiny dipole whose dipole moment is 4.8×10^{-30} C · m at a point 4.1×10^{-9} m away if this point is (*a*) along the axis of the dipole nearer the positive charge; (*b*) 45° above the axis but nearer the positive charge; (*c*) 45° above the axis but nearer the negative charge. Let $V = 0$ at $r = \infty$.

46. (III) The dipole moment, considered as a vector, points from the negative to the positive charge. The water molecule, Fig. 23–47, has a dipole moment $\vec{p}$ which can be considered as the vector sum of the two dipole moments $\vec{p}_1$ and $\vec{p}_2$ as shown. The distance between each H and the O is about 0.96×10^{-10} m. The lines joining the center of the O atom with each H atom make an angle of 104° as shown, and the net dipole moment has been measured to be $p = 6.1 \times 10^{-30}$ C · m. (*a*) Determine the effective charge q on each H atom. (*b*) Determine the electric potential, far from the molecule, due to each dipole, $\vec{p}_1$ and $\vec{p}_2$, and show that they sum to

$$V = \frac{1}{4\pi\epsilon_0}\frac{p\cos\theta}{r^2},$$

where p is the magnitude of the net dipole moment, $\vec{p} = \vec{p}_1 + \vec{p}_2$, and V is the total potential due to both $\vec{p}_1$ and $\vec{p}_2$. Take $V = 0$ at $r = \infty$.

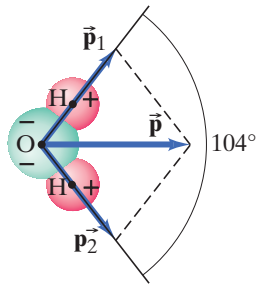


FIGURE 23–47
Problem 46.

23–7 $\vec{E}$ Determined from V

47. (I) Show that the electric field of a single point charge (Eq. 21–4) follows from Eq. 23–5, $V = (1/4\pi\epsilon_0)(Q/r)$.

48. (II) What is the potential gradient just outside the surface of a uranium nucleus $(Q = +92e)$ whose diameter is about 15×10^{-15} m?

49. (II) The electric potential between two parallel conducting plates is given by $V(x) = (8.5\,\text{V/m})x + 4.5\,\text{V}$, with $x = 0$ taken at one of the plates and x positive in the direction toward the other plate. What is the charge density on the plates?

50. (II) The electric potential in a region of space varies as $V = by/(a^2 + y^2)$. Determine $\vec{\mathbf{E}}$.

51. (II) In a certain region of space, the electric potential is given by $V = ay^2 + bxy - cxyz$. Determine the electric field vector, $\vec{\mathbf{E}}$, in this region.

52. (II) A dust particle with mass of 0.050 g and a charge of 2.0×10^{-6} C is in a region of space where the potential is given by $V(x) = (2.0\,\text{V/m}^2)x^2 - (3.0\,\text{V/m}^3)x^3$. If the particle starts at $x = 2.5$ m, what is the initial acceleration of the charge?

53. (III) Use the results of Problems 38 and 39 to determine the electric field due to the uniformly charged rod of Fig. 23–46 for points (*a*) along the *y* axis and (*b*) along the *x* axis.

23–8 Electrostatic Potential Energy; Electron Volt

54. (I) How much work must be done to bring three electrons from a great distance apart to within 1.0×10^{-10} m from one another (at the corners of an equilateral triangle)?

55. (I) What potential difference is needed to give a helium nucleus $(Q = 3.2 \times 10^{-19}\,\text{C})$ 165 keV of kinetic energy?

56. (I) What is the speed of (*a*) a 0.50-keV (kinetic energy) electron and (*b*) a 0.50-keV proton?

57. (II) Many chemical reactions release energy. Suppose that at the beginning of a reaction, an electron and a proton are separated by 0.110 nm, and their final separation is 0.100 nm. How much electric potential energy was lost in this reaction (in units of eV)?

58. (II) Four equal point charges, Q, are fixed at the corners of a square of side b. (*a*) What is their total electrostatic potential energy? (*b*) How much potential energy will a fifth charge, Q, have at the center of the square (relative to $V = 0$ at $r = \infty$)? (*c*) If constrained to remain in that plane, is the fifth charge in stable or unstable equilibrium? (*d*) If a negative $(-Q)$ charge is at the center, is it in stable equilibrium?

59. (II) An electron starting from rest acquires 4.8 keV of kinetic energy in moving from point A to point B. (*a*) How much kinetic energy would a proton acquire, starting from rest at B and moving to point A? (*b*) Determine the ratio of their speeds at the end of their respective trajectories.

60. (III) Determine the total electrostatic potential energy of a conducting sphere of radius r_0 that carries a total charge Q distributed uniformly on its surface.

61. (III) Determine the total electrostatic potential energy of a nonconducting sphere of radius r_0 carrying a total charge Q distributed uniformly throughout its volume.

23–9 Digital

62. (I) Write the decimal number 108 in binary.

63. (I) Write the decimal number 3597 in binary.

64. (I) Write the binary number 01010101 as a decimal number.

65. (I) Write the binary number 1010101010101010 as a decimal number.

66. (II) For each of the following bytes of binary data, state whether an additional parity bit should be a "0" or a "1": (*a*) 01101100, (*b*) 11011011, (*c*) 10010010.

67. (II) (*a*) 16-bit sampling provides how many different possible voltages? (*b*) 24-bit sampling provides how many different possible voltages? (*c*) For color TV, 3 subpixels, each 8 bits, provides a total of how many different colors?

68. (II) Consider a rather coarse 4-bit analog-to-digital conversion where the maximum voltage is 5.0 V. (*a*) What voltage does 1011 represent? (*b*) What is the 4-bit representation for 2.0 V?

69. (II) In the commonly used "ASCII" code, numbers, letters and other symbols are stored as 8-digit binary numbers. The code for the letter A is 00100001, and the next letters are just the next binary numbers. (*a*) What binary numbers are used to store C and T? (*b*) What would be the decimal numbers for ASCII B and H? (*c*) Using the format shown in Fig. 23–22, sketch the waveform of the letters that spell CAT.

70. (II) A few extraterrestrials arrived. They had two hands, but claimed that $3 + 2 = 11$. How many fingers did they have on their two hands? Note that our decimal system (and ten characters: 0, 1, 2, $\cdots$, 9) surely has its origin because we have ten fingers. [*Hint*: 11 is in their system. In our decimal system, the result would be written as 5.]

71. (III) An analog voltage signal can vary from 0 V to 5.00 V, and it is to be converted to an 8-bit binary representation. (*a*) What voltage, to the nearest 0.01 V, would have a binary representation of 01110101? (*b*) What binary number would best represent 3.47 volts?

*23–10 TV and Computer Monitors

***72.** (II) Figure 23–48 is a photograph of a computer screen shot by a camera set at an exposure time of $\frac{1}{4}$ s. During the exposure the cursor arrow was moved around by the mouse, and we see it 15 times. (*a*) Explain why we see the cursor 15 times. (*b*) What is the refresh rate of the screen?



FIGURE 23–48 Problem 72.

***73.** (II) For the greyscale picture in Fig. 23–30, write the line of binary numbers that gives the eleventh row of the picture. Take 000 as white and 100 as black. For the purpose of reading your answer more easily, separate each number by a semi-colon.

***74.** (II) Estimate how many gigabytes of data would be sent for 5.00 minutes of color HD TV in the U.S. if no data compression was used.

*75. (II) (*a*) Calculate the bit-rate that would be required to address all of the 6 million subpixels $(3 \times 1080 \times 1920)$ of an HD TV screen at a (refresh) rate of 60 per second. (*b*) Compare to the present-day digital fixed rate of 19 Mb/s, and explain the difference. (*c*) On average, what % of pixels get refreshed for each frame (refreshed at 60 Hz)? (*d*) If it were necessary to address every pixel for each frame, how many full scans of the screen could be done per second at 19 Mb/s?

*76. (III) (*a*) How many bits of data would it take to convey Fig. 23–30 assuming you had to specify a row number, an element number, and a 5-level greyscale for each pixel? (*b*) Describe a data compression system that would reduce the number of bits required to convey Fig. 23–30. (*c*) How many bits of data would be needed using your data compression scheme? (*d*) Would your system work for any picture with only 4 greytones? Why or why not?

General Problems

77. If the electrons in a single raindrop, 3.5 mm in diameter, could be removed from the Earth (without removing the atomic nuclei), by how much would the potential of the Earth increase?

78. By rubbing a nonconducting material, a charge of 10^{-8} C can readily be produced. If this is done to a sphere of radius 15 cm, estimate the potential produced at the surface. Let $V = 0$ at $r = \infty$.

79. At each corner of a cube of side ℓ there is a point charge $+Q$, Fig. 23–49. (*a*) What is the potential at the center of the cube ($V = 0$ at $r = \infty$)? (*b*) What is the potential at each corner due to the other seven charges? (*c*) What is the total potential energy of this system?

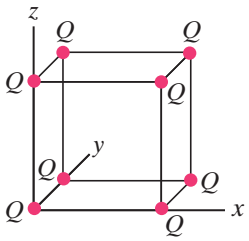


FIGURE 23–49 Problem 79.

80. A $+38\,\mu$C point charge is placed 36 cm from an identical $+38\,\mu$C charge. Then a $-1.8\,\mu$C charge is moved from point A to point B as shown in Fig. 23–50. What is the change in potential energy?

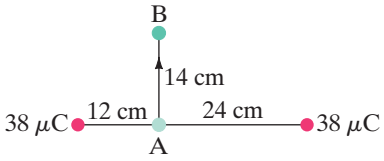


FIGURE 23–50 Problem 80.

81. A proton $(q = +e)$ and an alpha particle $(q = +2e)$ are accelerated by the same voltage V. Which gains the greater kinetic energy, and by what factor?

82. In an older television tube (CRT), electrons are accelerated by thousands of volts through a vacuum. If such a television set is laid on its back, would electrons be able to move upward against the force of gravity? What potential difference, acting over a distance of 2.4 cm, would be needed to balance the downward force of gravity so that an electron would remain stationary? Assume that the electric field is uniform.

83. Two point charges are fixed 4.0 cm apart from each other. Their charges are $Q_1 = Q_2 = 6.5\,\mu$C, and their masses are $m_1 = 2.5$ mg and $m_2 = 3.5$ mg. (*a*) If Q_1 is released from rest, what will be its speed after a very long time? (*b*) If both charges are released from rest at the same time, what will be the speed of Q_1 after a very long time? Ignore the environment.

84. Two charges are placed as shown in Fig. 23–51 with $q_1 = 1.2\,\mu$C and $q_2 = -3.3\,\mu$C. Find the potential difference between points A and B.

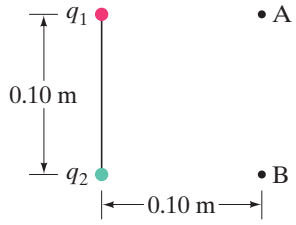


FIGURE 23–51 Problem 84.

85. In a **photocell**, ultraviolet (UV) light provides enough energy to some electrons in barium metal to eject them from the surface at high speed. To measure the maximum energy of the electrons, another plate above the barium surface is kept at a negative enough potential that the emitted electrons are slowed down and stopped, and return to the barium surface. See Fig. 23–52. If the plate voltage is -3.02 V (compared to the barium) when the fastest electrons are stopped, what was the speed of these electrons when they were emitted?

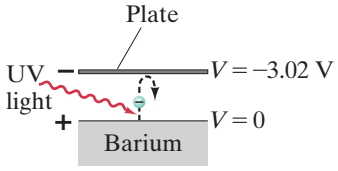


FIGURE 23–52 Problem 85.

86. An electron is accelerated horizontally from rest by a potential difference of 3500 V. It then passes between two horizontal plates 6.5 cm long and 1.3 cm apart that have a potential difference of 250 V (Fig. 23–53). At what angle θ will the electron be traveling after it passes between the plates? Assume the electric field between the plates is uniform.

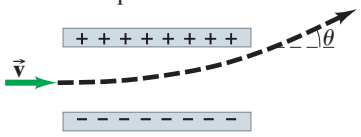


FIGURE 23–53 Problem 86.

87. Three charges are at the corners of an equilateral triangle (side ℓ) as shown in Fig. 23–54. Determine the potential at the midpoint of each of the sides. Let $V = 0$ at $r = \infty$.

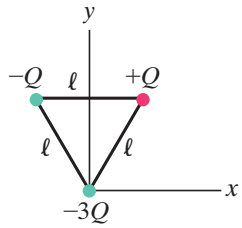


FIGURE 23–54 Problem 87.

88. Near the surface of the Earth there is an electric field of about 150 V/m which points downward. Two identical balls with mass $m = 0.550\,\text{kg}$ are dropped from a height of 2.00 m, but one of the balls is positively charged with $q_1 = 650\,\mu\text{C}$, and the second is negatively charged with $q_2 = -650\,\mu\text{C}$. Use conservation of energy to determine the difference in the speeds of the two balls when they hit the ground. (Neglect air resistance.)

89. A lightning flash transfers 4.0 C of charge and 5.2 MJ of energy to the Earth. (*a*) Across what potential difference did it travel? (*b*) How much water could this boil and vaporize, starting from room temperature? [See also Chapter 19.]

90. Determine the components of the electric field, E_x and E_y, as a function of x and y in the xy plane due to a dipole, Fig. 23–55, starting with Eq. 23–7. Assume $r = (x^2 + y^2)^{\frac{1}{2}} \gg \ell$.

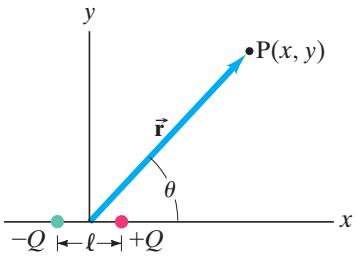


FIGURE 23–55 Problem 90.

91. A nonconducting sphere of radius r_0 carries a total charge Q distributed uniformly throughout its volume. Determine the electric potential (you may want to use Gauss's law) as a function of the distance r from the center of the sphere for (*a*) $r > r_0$ and (*b*) $r < r_0$. Take $V = 0$ at $r = \infty$. (*c*) Plot V versus r and E versus r.

92. The volume charge density ρ_E within a sphere of radius r_0 is distributed according to the following spherically symmetric relation

$$\rho_E(r) = \rho_0\left[1 - \frac{r^2}{r_0^2}\right]$$

where r is measured from the center of the sphere and ρ_0 is a constant. For a point P inside the sphere $(r < r_0)$, determine the electric potential V. Let $V = 0$ at infinity. [*Hint*: Start with Gauss's law.]

93. Binary numbers can be used for addition. The same basic rules apply as for decimal numbers: add in columns, and carry when the total amount in a column is greater than or equal to the value of the column. Consider the decimal numbers 14, 45, and 103. Convert these three numbers to 8-bit binary numbers, add them as binary numbers, and then change that binary sum to a decimal number. Show all of your work. Also show that you get the same result from this process as if you had just added the three numbers in decimal format.

94. Thunderclouds may develop a voltage difference of about 5×10^7 V. Given that an electric field of 3×10^6 V/m is required to produce an electrical spark within a volume of air, estimate the length of a thundercloud *lightning bolt*. [Can you see why, when lightning strikes from a cloud to the ground, the bolt has to propagate as a sequence of steps?]

95. The potential in a region of space is given by $V = B/(x^2 + R^2)^2$ where $B = 120\,\text{V}\cdot\text{m}^4$ and $R = 0.20\,\text{m}$. (*a*) Find V at $x = 0.20\,\text{m}$. (*b*) Find $\vec{\mathbf{E}}$ as a function of x. (*c*) Find $\vec{\mathbf{E}}$ at $x = 0.20\,\text{m}$.

96. A charge $-q_1$ of mass m rests on the y axis at a distance b above the x axis. Two positive charges of magnitude $+q_2$ are fixed on the x axis at $x = +a$ and $x = -a$, respectively (Fig. 23–56). If the $-q_1$ charge is given an initial velocity v_0 in the positive y direction, what is the minimum value of v_0 such that the charge escapes to a point infinitely far away from the two positive charges?

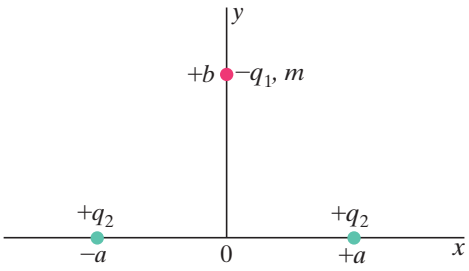


FIGURE 23–56 Problem 96.

97. An electric field greater than about 3×10^6 V/m causes air to break down (electrons are removed from air atoms and then recombine, emitting light). If you shuffle along a carpet and then reach for a doorknob, a spark flies across a gap you estimate to be 1 mm between your finger and the doorknob. Estimate the voltage between your finger and the doorknob. Why is no harm done?

98. A **Geiger counter** is used to detect charged particles emitted by radioactive nuclei. It consists of a thin, positively charged central wire of radius R_a surrounded by a concentric conducting cylinder of radius R_b with an equal negative charge (Fig. 23–57). The charge per unit length on the inner wire is λ (units C/m). The interior space between wire and cylinder is filled with low-pressure inert gas. Charged particles ionize some of these gas atoms; the resulting free electrons are attracted toward the positive central wire. If the radial electric field is strong enough, the freed electrons gain enough energy to ionize other atoms, causing an "avalanche" of electrons to strike the central wire, generating an electric "signal." (*a*) Find the expression for the electric field between the wire and the cylinder, and (*b*) show that the potential difference between R_a and R_b is

$$V_a - V_b = \left(\frac{\lambda}{2\pi\epsilon_0}\right)\ln\left(\frac{R_b}{R_a}\right).$$

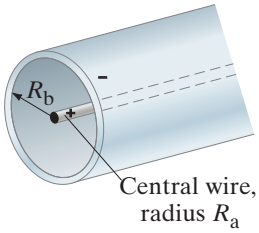


FIGURE 23–57 Problem 98.

99. In the Bohr model of the hydrogen atom, an electron orbits a proton (the nucleus) in a circular orbit of radius 0.53×10^{-10} m. (*a*) What is the electric potential at the electron's orbit due to the proton? (*b*) What is the kinetic energy of the electron? (*c*) What is the total energy of the electron in its orbit? (*d*) What is the *ionization energy*—that is, the energy required to remove the electron from the atom and take it to $r = \infty$, at rest? Express the results of parts (*b*), (*c*), and (*d*) in joules and eV.

100. The **liquid-drop model** of the nucleus suggests that high-energy oscillations of certain nuclei can split ("fission") a large nucleus into two unequal fragments plus a few neutrons. Using this model, consider the case of a uranium nucleus fissioning into two spherical fragments, one with a charge $q_1 = +38e$ and radius $r_1 = 5.5 \times 10^{-15}$ m, the other with $q_2 = +54e$ and $r_2 = 6.2 \times 10^{-15}$ m. Calculate the electric potential energy (MeV) of these fragments, assuming that the charge is uniformly distributed throughout the volume of each spherical nucleus and that their surfaces are initially in contact at rest. The electrons surrounding the nuclei can be neglected. This electric potential energy will then be entirely converted to kinetic energy as the fragments repel each other. How does your predicted kinetic energy of the fragments agree with the observed value associated with uranium fission (approximately 200 MeV total)? [$1\,\text{MeV} = 10^6\,\text{eV}$.]

101. A thin flat disk of radius R_0 carries a total charge Q that is distributed uniformly over its surface. The electric potential at a distance x on the x axis is given by

$$V(x) = \frac{Q}{2\pi\epsilon_0 R_0^2}\left[(x^2 + R_0^2)^{\frac{1}{2}} - x\right].$$

(See Example 23–10.) Show that the electric field at a distance x on the x axis is given by

$$E(x) = \frac{Q}{2\pi\epsilon_0 R_0^2}\left(1 - \frac{x}{(x^2 + R_0^2)^{\frac{1}{2}}}\right).$$

Make graphs of $V(x)$ and $E(x)$ as a function of x/R_0 for $x/R_0 = 0$ to 4. (Do the calculations in steps of 0.1.) Use $Q = 5.0\,\mu\text{C}$ and $R_0 = 10\,\text{cm}$ for the calculation and graphs.

102. A manufacturer claims that a carpet will not generate more than 6.0 kV of static electricity. What magnitude of charge would have to be transferred between a carpet and a shoe for there to be a 6.0-kV potential difference between the shoe and the carpet? Approximate the area of the shoe and assume the shoe and carpet are large sheets of charge separated by a small distance $d = 1.0\,\text{mm}$.

103. A nonconducting sphere of radius r_2 contains a concentric spherical cavity of radius r_1. The material between r_1 and r_2 carries a uniform charge density ρ_E (C/m^3). Determine the electric potential V, relative to $V = 0$ at $r = \infty$, as a function of the distance r from the center for (*a*) $r > r_2$, (*b*) $r_1 < r < r_2$, and (*c*) $0 < r < r_1$. Is V continuous at r_1 and r_2?

104. Figure 23–19 shows contour lines (elevations). Just for fun, assume they are equipotential lines on a flat 2-dimensional surface with the values shown being in volts. Estimate the average magnitude and the direction of the "electric field" (*a*) between Greenstone Lake and the first (lowest) Conness Lake (as you head upstream) and (*b*) on the Conness Glacier. Assume that up is $+y$, right is $+x$, and that Greenstone Lake is about 1.0 km wide at its widest.

105. A **Van de Graaff generator** (Fig. 23–58) can develop a very large potential difference, even millions of volts. Electrons are pulled off the belt by the high voltage pointed electrode (positive) at A, leaving the belt positively charged. (Recall Example 23–5 where we saw that near sharp points the electric field is high and ionization can occur.) The belt carries the positive charge up inside the spherical shell where electrons from the large conducting sphere are attracted over to the pointed conductor at B, leaving the outer surface of the conducting sphere positively charged. As more charge is brought up, the sphere reaches extremely high voltage. Consider a Van de Graaff generator with a sphere of radius 0.20 m. (*a*) What is the electric potential on the surface of the sphere when electrical breakdown occurs $(E = 3 \times 10^6\,\text{V/m})$? Assume $V = 0$ at $r = \infty$. (*b*) What is the charge on the sphere for the potential found in part (*a*)?

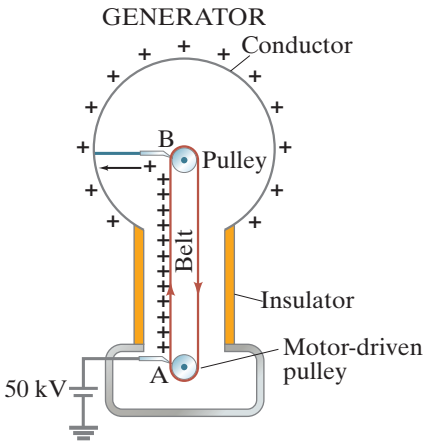
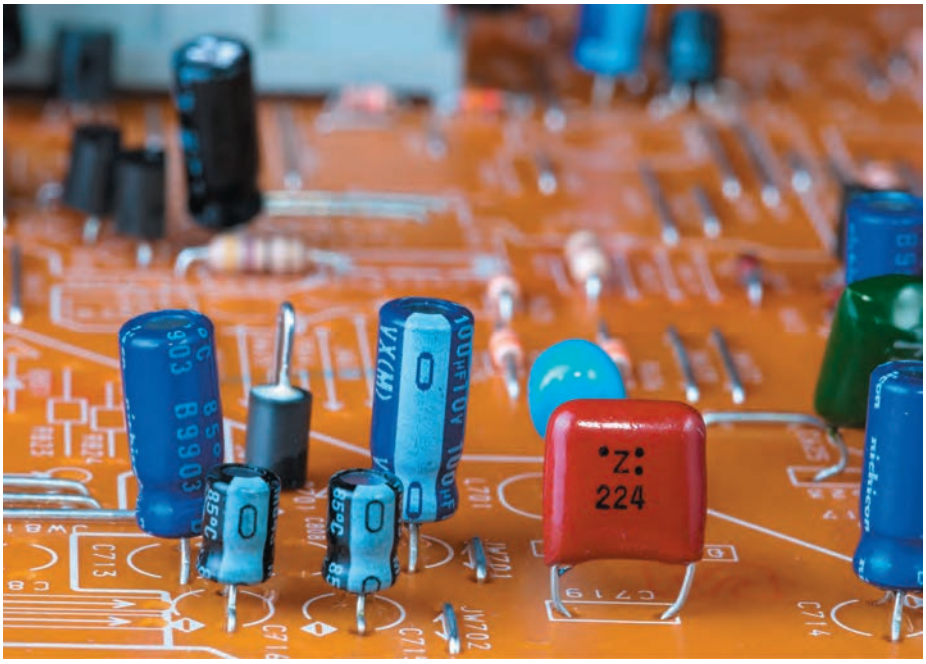


FIGURE 23–58 Problem 105.

ANSWERS TO EXERCISES

A: (*a*) -8.0×10^{-16} J; (*b*) 9.8×10^5 m/s.
B: (*d*).
C: (*d*).
D: 0.72 J.
E: (*c*).
F: (*c*).
G: (*d*).

Capacitors come in a wide range of sizes and shapes, only a few of which are shown here. A capacitor is basically two conductors that do not touch, and which therefore can store charge of opposite sign on its two conductors. Capacitors are used in a wide variety of circuits, for various purposes including storing of electric charge, storage of electric energy, as signal filters, and in timing circuits, as we shall see in this and later Chapters.



CHAPTER 24

Capacitance, Dielectrics, Electric Energy Storage

CONTENTS

24–1 Capacitors
24–2 Determination of Capacitance
24–3 Capacitors in Series and Parallel
24–4 Storage of Electric Energy
24–5 Dielectrics
*24–6 Molecular Description of Dielectrics

CHAPTER-OPENING QUESTION—Guess now!

A fixed potential difference V exists between a pair of close parallel plates carrying opposite charges $+Q$ and $-Q$. Which of the following would not increase the magnitude of charge that you could put on the plates?

(a) Increase the size of the plates.
(b) Move the plates farther apart.
(c) Fill the space between the plates with paper.
(d) Increase the fixed potential difference V.
(e) None of the above.

This Chapter will complete our study of electrostatics. It deals first of all with an important device, the capacitor, which is used in many electronic circuits. We will also discuss electric energy storage and the effects of an insulator, or dielectric, on electric fields and potential differences.

24–1 Capacitors

PHYSICS APPLIED
Uses of capacitors

A **capacitor** is a device that can store electric charge, and normally consists of two conducting objects (usually plates or sheets) placed near each other but not touching. Capacitors (sometimes called **condensers**) are widely used in electronic circuits. They store charge for later use, such as in a camera flash, and as energy backup in devices like computers for power failures. Capacitors also block surges of charge and energy to protect circuits. Very tiny capacitors serve as memory for the "ones" and "zeros" of the binary code in the random access memory (RAM)

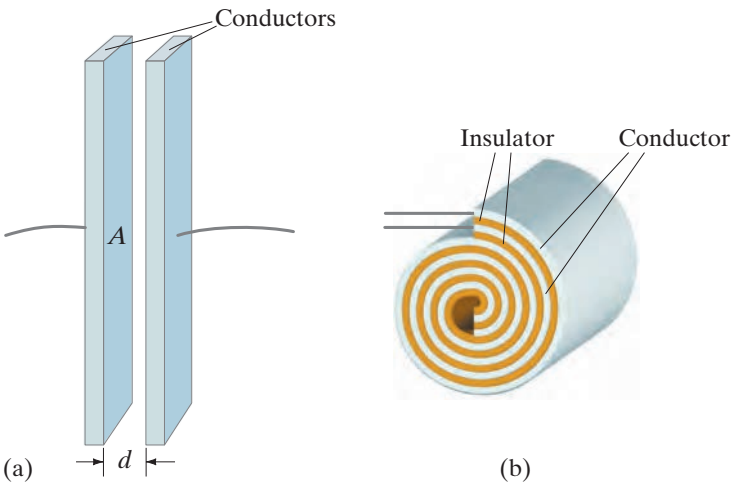
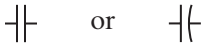


FIGURE 24–1 Capacitors: diagrams of (a) parallel plate, (b) cylindrical (rolled up parallel plate).

of computers and other electronic devices (for example, in Fig. 23–34). Capacitors serve many other applications, some of which we will discuss.

A simple capacitor consists of a pair of parallel plates of area A separated by a small distance d (Fig. 24–1a). Often the two plates are rolled into the form of a cylinder with plastic, paper, or other insulator separating the plates, Fig. 24–1b. In circuit diagrams, the symbol

 or [capacitor symbol]

represents a capacitor. A battery, which is a source of voltage, is indicated by the symbol:

 [battery symbol]

with unequal arms.

If a voltage is applied across a capacitor by connecting the capacitor to a battery with conducting wires as in Fig. 24–2, the two plates quickly become charged: one plate acquires a negative charge, the other an equal amount of positive charge. Each battery terminal and the plate of the capacitor connected to it are at the same potential; hence the full battery voltage appears across the capacitor. For a given capacitor, it is found that the amount of charge Q acquired by each plate is proportional to the magnitude of the potential difference V between the plates:

$$Q = CV. \qquad \textbf{(24–1)}$$

The constant of proportionality, C, in Eq. 24–1 is called the **capacitance** of the capacitor. The unit of capacitance is coulombs per volt, and this unit is called a **farad** (F). Commonly used capacitors have capacitance in the range of 1 pF $(\text{picofarad} = 10^{-12}\,\text{F})$ to $10^{3}\,\mu\text{F}$ $(\text{microfarad} = 10^{-6}\,\text{F})$. The relation, Eq. 24–1, was first suggested by Alessandro Volta (1745–1827, inventor of the electric battery as we discuss in the next Chapter). The capacitance C does not in general depend on Q or V. Its value depends only on the size, shape, and relative position of the two conductors, and also on the material that separates them.

In Eq. 24–1, and from now on, we use simply V (in italics) to represent a *potential difference*, such as that produced by a battery, rather than V_{ba}, ΔV, or $V_b - V_a$, as previously.

Also, be sure not to confuse *italic* letters V and C which stand for the quantities voltage and capacitance, with non-italic V and C which stand for the units volts and coulombs.

EXERCISE A Graphs for charge versus voltage are shown in Fig. 24–3 for three capacitors, A, B, and C. Which has the greatest capacitance?

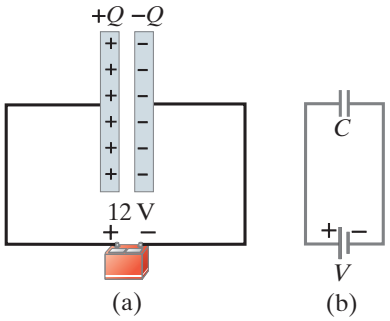
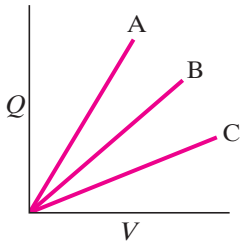


FIGURE 24–2 (a) Parallel-plate capacitor connected to a battery. (b) Same circuit shown using symbols.

FIGURE 24–3 Exercise A.



CAUTION

V = potential difference from here on

24–2 Determination of Capacitance

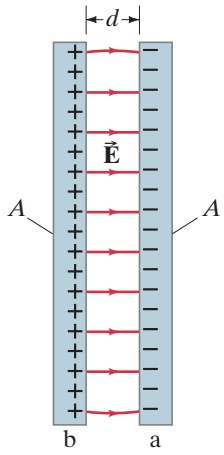


FIGURE 24–4 Parallel-plate capacitor, each of whose plates has area A.

The capacitance of a given capacitor can in principle be determined experimentally directly from Eq. 24–1, by measuring the charge Q on either conductor for a given potential difference V. But for capacitors whose geometry is simple, we can determine C analytically, and in this Section we assume the conductors are separated by a vacuum or air.

First, we determine C for a parallel-plate capacitor, Fig. 24–4. Each plate has area A and the two plates are separated by a distance d. We assume d is small compared to the dimensions of each plate so that the electric field $\vec{\mathbf{E}}$ is uniform between them and we ignore fringing (lines of $\vec{\mathbf{E}}$ not straight) at the edges. We saw earlier (Example 21–14) that the electric field between two closely spaced parallel plates has magnitude $E = \sigma/\epsilon_0$ and its direction is perpendicular to the plates. Since σ is the charge per unit area, $\sigma = Q/A$, then the field between the plates is

$$E = \frac{Q}{\epsilon_0 A}.$$

The relation between electric field and electric potential, as given by Eq. 23–4a, is

$$V = V_{ba} = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}.$$

We can take the line integral along a path antiparallel to the field lines, from plate a in Fig. 24–4 to plate b; then $\theta = 180°$ and $\cos 180° = -1$, so

$$V = V_b - V_a = -\int_a^b E\, d\ell \cos 180° = +\int_a^b E\, d\ell = \frac{Q}{\epsilon_0 A}\int_a^b d\ell = \frac{Qd}{\epsilon_0 A}.$$

This relates Q to V, and from it we can get the capacitance C in terms of the geometry of the plates:

$$C = \frac{Q}{V} = \epsilon_0 \frac{A}{d}. \qquad \text{[parallel-plate capacitor]} \quad \textbf{(24–2)}$$

Note from Eq. 24–2 that the value of C does not depend on Q or V, so Q is predicted to be proportional to V as is found experimentally. Capacitance C is always positive, never negative.

EXAMPLE 24–1 **Capacitor calculations.** (*a*) Calculate the capacitance of a parallel-plate capacitor whose plates are $20\,\text{cm} \times 3.0\,\text{cm}$ and are separated by a 1.0-mm air gap. (*b*) What is the charge on each plate if a 12-V battery is connected across the two plates? (*c*) What is the electric field between the plates? (*d*) Estimate the area of the plates needed to achieve a capacitance of 1 F, assuming the air gap d is 100 times smaller, or 10 microns (1 **micron** = $1\,\mu\text{m} = 10^{-6}\,\text{m}$).

APPROACH The capacitance is found by using Eq. 24–2, $C = \epsilon_0 A/d$. The charge on each plate is obtained from the definition of capacitance, Eq. 24–1, $Q = CV$. The electric field is uniform, so we can use Eq. 23–4b for the magnitude $E = V/d$. In (*d*) we use Eq. 24–2 again.

SOLUTION (*a*) The area $A = (20 \times 10^{-2}\,\text{m})(3.0 \times 10^{-2}\,\text{m}) = 6.0 \times 10^{-3}\,\text{m}^2$. The capacitance C is then

$$C = \epsilon_0 \frac{A}{d} = (8.85 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)\frac{6.0 \times 10^{-3}\,\text{m}^2}{1.0 \times 10^{-3}\,\text{m}} = 53\,\text{pF}.$$

(*b*) The charge on each plate is

$$Q = CV = (53 \times 10^{-12}\,\text{F})(12\,\text{V}) = 6.4 \times 10^{-10}\,\text{C}.$$

(*c*) From Eq. 23–4b for a uniform electric field, the magnitude of E is

$$E = \frac{V}{d} = \frac{12\,\text{V}}{1.0\times10^{-3}\,\text{m}} = 1.2\times10^{4}\,\text{V/m}.$$

(*d*) The plate separation is now $d = 10\,\mu\text{m} = 1.0\times10^{-5}\,\text{m}$ and we want $C = 1\,\text{F}$. We solve for A in Eq. 24–2 and find that we need plates with an area

$$A = \frac{Cd}{\epsilon_0} \approx \frac{(1\,\text{F})(1.0\times10^{-5}\,\text{m})}{(9\times10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)} \approx 10^{6}\,\text{m}^2.$$

NOTE This is the area of a square 10^3 m or 1 km on a side. That is inconveniently large (the size of a small town). Large-capacitance capacitors will not be simple flat parallel plates.

EXERCISE B Two circular plates of radius 5.0 cm are separated by a 0.10-mm air gap. What is the magnitude of the charge on each plate when connected to a 12-V battery?

PHYSICS APPLIED
Very high capacitance; capacitor as power backup

Not long ago, a capacitance greater than a few mF was unusual. Today it is possible to make capacitors that are 1 or 2 F, yet they are just a few cm on a side. Such large capacitance can be used as power backups, for example, in computer memory and electronics where the time and date can be maintained through tiny charge flow. [Capacitors are superior to rechargable batteries for this purpose because they can be recharged more than 10^5 times with no degradation.] Such high-capacitance capacitors can be made of *activated carbon* which has very high porosity, so that the surface area is very large; one tenth of a gram of activated carbon can have a surface area of 100 m^2. Furthermore, the equal and opposite charges exist in an electric "double layer" about 10^{-9} m thick. Thus, the capacitance of 0.1 g of activated carbon, whose internal area can be $10^2\,\text{m}^2$, is equivalent to a parallel-plate capacitor with $C \approx \epsilon_0 A/d = (8.85\times10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)(10^2\,\text{m}^2)/(10^{-9}\,\text{m}) \approx 1\,\text{F}$.

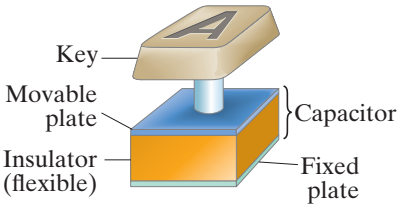


FIGURE 24–5 Key on a computer keyboard. Pressing the key, which is fixed to the movable plate, reduces the capacitor spacing thus increasing the capacitance which can be detected electronically.

PHYSICS APPLIED
Computer key

One type of computer keyboard operates by capacitance. As shown in Fig. 24–5, each key is connected to the upper plate of a capacitor. The upper plate moves down when the key is pressed, reducing the spacing between the capacitor plates, and increasing the capacitance (Eq. 24–2: smaller d, larger C). The *change* in capacitance results in an electric signal that is detected by an electronic circuit.

PHYSICS APPLIED
Condenser microphone

One type of microphone is a **condenser**, or capacitor, **microphone**, diagrammed in Fig. 24–6. The changing air pressure in a sound wave causes one plate of the capacitor C to move back and forth, and that plate is isolated so the charge on it doesn't change. Hence the voltage across the capacitor changes at the same frequency as the sound wave.

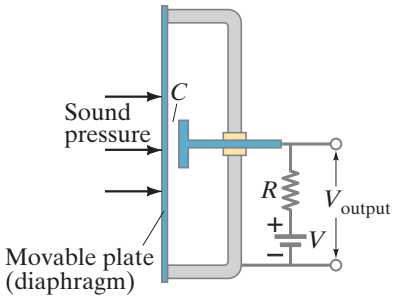
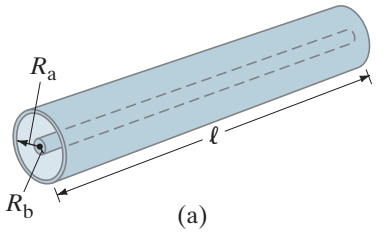
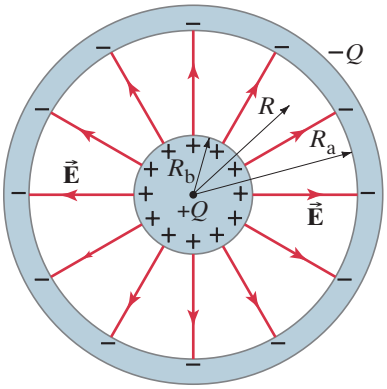


FIGURE 24–6 Diagram of a condenser microphone.

The proportionality, $C \propto A/d$ in Eq. 24–2, is valid for a parallel-plate capacitor even if it is rolled up into a spiral cylinder, as in Fig. 24–1b. However, the constant factor, ϵ_0, must be replaced if an insulator such as paper separates the plates, as is usual (discussed in Section 24–5). For a true cylindrical capacitor—consisting of two long coaxial cylinders—the result is somewhat different as the next Example shows.



(a)



(b)

FIGURE 24–7 (a) Cylindrical capacitor consists of two coaxial cylindrical conductors. (b) The electric field lines are shown in cross-sectional view.

PROBLEM SOLVING
Checking with a limiting case

EXAMPLE 24–2 **Cylindrical capacitor.** A cylindrical capacitor consists of a small cylinder (or wire) of radius R_b surrounded by a larger coaxial cylindrical shell of inner radius R_a, Fig. 24–7a. Both cylinders have length ℓ which we assume is much greater than the separation of the cylinders, $R_a - R_b$, so we can neglect end effects. The capacitor is charged (by connecting it to a battery) so that one cylinder has a charge $+Q$ (say, the inner one) and the other one a charge $-Q$. Determine a formula for the capacitance.

APPROACH To determine the capacitance $C = Q/V$, we need to determine the potential difference V between the cylinders in terms of Q. We can use our earlier result (Example 21–12 or 22–6) that the electric field outside a long wire is directed radially outward and has magnitude $E = (1/2\pi\epsilon_0)(\lambda/R)$, where R is the distance from the central axis and λ is the charge per unit length, Q/ℓ. Then $E = (1/2\pi\epsilon_0)(Q/\ell R)$ for points between the cylinders.

SOLUTION To obtain the potential difference V in terms of Q, we insert this expression for E into Eq. 23–4a, $V = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$, and write the line integral from the outer cylinder to the inner one (so $V > 0$) along a radial line ($d\vec{\boldsymbol{\ell}}$ points inward so $d\ell = -dR$ and $\cos\theta = \cos 180^\circ = -1$):

$$V = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{Q}{2\pi\epsilon_0 \ell}\int_{R_a}^{R_b} \frac{dR}{R}$$
$$= -\frac{Q}{2\pi\epsilon_0 \ell}\ln\frac{R_b}{R_a} = \frac{Q}{2\pi\epsilon_0 \ell}\ln\frac{R_a}{R_b}.$$

Q and V are proportional, and the capacitance C is

$$C = \frac{Q}{V} = \frac{2\pi\epsilon_0 \ell}{\ln(R_a/R_b)}. \qquad \text{[cylindrical capacitor]}$$

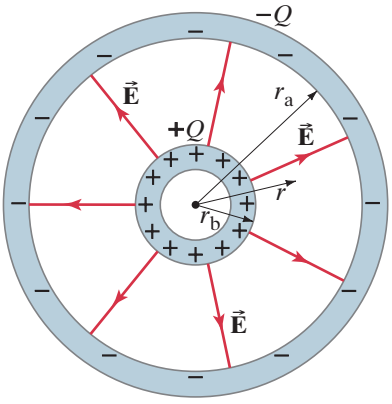
NOTE If the space between cylinders, $R_a - R_b = \Delta R$ is small, we have

$$\ln(R_a/R_b) = \ln[(R_b + \Delta R)/R_b] = \ln[1 + \Delta R/R_b] \approx \Delta R/R_b$$

(see Appendix A–3) so $C \approx 2\pi\epsilon_0 \ell R_b/\Delta R = \epsilon_0 A/\Delta R$ because the area of cylinder b is $A = 2\pi R_b \ell$. This is just Eq. 24–2 for a parallel-plate capacitor ($d = \Delta R$), as we would hope and thus a nice check.

EXERCISE C What is the capacitance per unit length of a cylindrical capacitor with radii $R_a = 2.5$ mm and $R_b = 0.40$ mm? (*a*) 30 pF/m; (*b*) −30 pF/m; (*c*) 56 pF/m; (*d*) −56 pF/m; (*e*) 100 pF/m; (*f*) −100 pF/m.

FIGURE 24–8 Cross section through the center of a spherical capacitor. The thin inner shell has radius r_b and the thin outer shell has radius r_a.



EXAMPLE 24–3 **Spherical capacitor.** A spherical capacitor consists of two thin concentric spherical conducting shells, of radius r_a and r_b as shown in Fig. 24–8. The inner shell carries a uniformly distributed charge Q on its surface, and the outer shell an equal but opposite charge $-Q$. Determine the capacitance of the two shells.

APPROACH In Example 22–3 we used Gauss's law to show that the electric field outside a uniformly charged conducting sphere is $E = Q/4\pi\epsilon_0 r^2$ as if all the charge were concentrated at the center. We then use Eq. 23–4a, $V = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$.

SOLUTION We integrate Eq. 23–4a along a radial line from r_a to r_b to obtain the potential difference between the two conducting shells:

$$V_{ba} = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{Q}{4\pi\epsilon_0}\int_{r_a}^{r_b} \frac{1}{r^2}\,dr$$
$$= \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_b} - \frac{1}{r_a}\right) = \frac{Q}{4\pi\epsilon_0}\left(\frac{r_a - r_b}{r_a r_b}\right).$$

Finally,

$$C = \frac{Q}{V_{ba}} = 4\pi\epsilon_0\left(\frac{r_a r_b}{r_a - r_b}\right).$$

NOTE If the separation $\Delta r = r_a - r_b$ is very small, then $C = 4\pi\epsilon_0 r^2/\Delta r \approx \epsilon_0 A/\Delta r$ (since $A = 4\pi r^2$), which is the parallel-plate formula, Eq. 24–2.

A single isolated conductor can also be said to have a capacitance, C. In this case, C can still be defined as the ratio of the charge to absolute potential V on the conductor (relative to $V = 0$ at $r = \infty$), so that the relation

$$Q = CV$$

remains valid. For example, the potential of a single conducting sphere of radius r_b can be obtained from our results in Example 24–3 by letting r_a become infinitely large. As $r_a \to \infty$, then

$$V = \frac{Q}{4\pi\epsilon_0}\left(\frac{1}{r_b} - \frac{1}{r_a}\right) = \frac{1}{4\pi\epsilon_0}\frac{Q}{r_b};$$

so its capacitance is

$$C = \frac{Q}{V} = 4\pi\epsilon_0 r_b.$$

In practical cases, a single conductor may be near other conductors or the Earth (which can be thought of as the other "plate" of a capacitor), and these will affect the value of the capacitance.

EXAMPLE 24–4 **Capacitance of two long parallel wires.** Estimate the capacitance per unit length of two very long straight parallel wires, each of radius R, carrying uniform charges $+Q$ and $-Q$, and separated by a distance d which is large compared to R ($d \gg R$), Fig. 24–9.

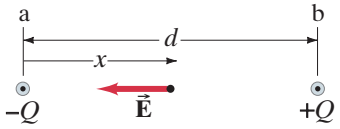


FIGURE 24–9 Example 24–4.

APPROACH We calculate the potential difference between the wires by treating the electric field at any point between them as the superposition of the two fields created by each wire. (The electric field inside each wire conductor is zero.)

SOLUTION The electric field outside of a long straight conductor was found in Examples 21–12 and 22–6 to be radial and given by $E = \lambda/(2\pi\epsilon_0 x)$ where λ is the charge per unit length, $\lambda = Q/\ell$. The total electric field due to both wires at distance x from the left-hand wire in Fig. 24–9 has magnitude

$$E = \frac{\lambda}{2\pi\epsilon_0 x} + \frac{\lambda}{2\pi\epsilon_0(d - x)},$$

and points to the left (from + to −). Now we find the potential difference between the two wires using Eq. 23–4a and integrating along the straight line from the surface of the negative wire to the surface of the positive wire, noting that $\vec{\mathbf{E}}$ and $d\vec{\boldsymbol{\ell}}$ point in opposite directions ($\vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} < 0$):

$$V = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = \left(\frac{\lambda}{2\pi\epsilon_0}\right)\int_R^{d-R}\left[\frac{1}{x} + \frac{1}{(d - x)}\right]dx$$

$$= \left(\frac{\lambda}{2\pi\epsilon_0}\right)[\ln(x) - \ln(d - x)]\Big|_R^{d-R}$$

$$= \left(\frac{\lambda}{2\pi\epsilon_0}\right)[\ln(d - R) - \ln R - \ln R + \ln(d - R)]$$

$$= \left(\frac{\lambda}{\pi\epsilon_0}\right)[\ln(d - R) - \ln(R)] \approx \left(\frac{\lambda}{\pi\epsilon_0}\right)[\ln(d) - \ln(R)]$$

because we are given that $d \gg R$. So

$$V \approx \left(\frac{Q}{\pi\epsilon_0\ell}\right)\left[\ln\left(\frac{d}{R}\right)\right].$$

The capacitance from Eq. 24–1 is $C = Q/V \approx (\pi\epsilon_0\ell)/\ln(d/R)$, so the capacitance per unit length is given approximately by

$$\frac{C}{\ell} \approx \frac{\pi\epsilon_0}{\ln(d/R)}.$$

24–3 Capacitors in Series and Parallel

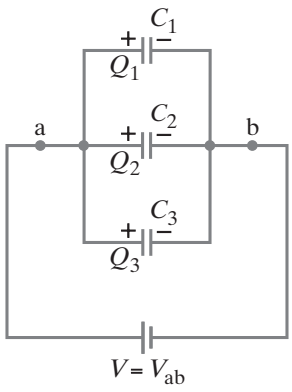


FIGURE 24–10 Capacitors in parallel: $C_{eq} = C_1 + C_2 + C_3$.

Capacitors are found in many electric circuits. By electric circuit we mean a closed path of conductors, usually wires connecting capacitors and/or other devices, in which charge can flow and which includes a source of voltage such as a battery. The battery voltage is usually given the symbol V, which means that V represents a *potential difference*. Capacitors can be connected together in various ways. Two common ways are in *series*, or in *parallel*, and we now discuss both.

A circuit containing three capacitors connected in **parallel** is shown in Fig. 24–10. They are in "parallel" because when a battery of voltage V is connected to points a and b, this same potential difference $V = V_{ab}$ exists across each of the capacitors. That is, since the left-hand plates of all the capacitors are connected by conductors, they all reach the same potential V_a when connected to the battery; and the right-hand plates each reach potential V_b. Each capacitor plate acquires a charge given by $Q_1 = C_1 V$, $Q_2 = C_2 V$, and $Q_3 = C_3 V$. The total charge Q that must leave the battery is then

$$Q = Q_1 + Q_2 + Q_3 = C_1 V + C_2 V + C_3 V.$$

Let us try to find a single equivalent capacitor that will hold the same charge Q at the same voltage $V = V_{ab}$. It will have a capacitance C_{eq} given by

$$Q = C_{eq} V.$$

Combining the two previous equations, we have

$$C_{eq} V = C_1 V + C_2 V + C_3 V = (C_1 + C_2 + C_3) V$$

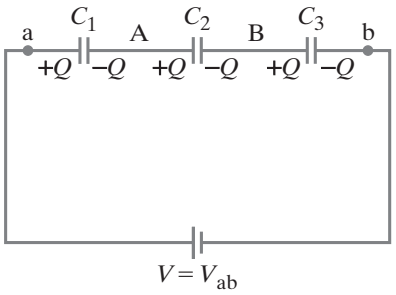
or

Capacitors in parallel

$$C_{eq} = C_1 + C_2 + C_3. \qquad \text{[parallel]} \quad \textbf{(24–3)}$$

The net effect of connecting capacitors in parallel is thus to *increase* the capacitance. This makes sense because we are essentially increasing the area of the plates where charge can accumulate (see, for example, Eq. 24–2).

FIGURE 24–11 Capacitors in series: $\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$.



Capacitors can also be connected in **series**: that is, end to end as shown in Fig. 24–11. A charge $+Q$ flows from the battery to one plate of C_1, and $-Q$ flows to one plate of C_3. The regions A and B between the capacitors were originally neutral, so the net charge there must still be zero. The $+Q$ on the left plate of C_1 attracts a charge of $-Q$ on the opposite plate. Because region A must have a zero net charge, there is $+Q$ on the left plate of C_2. The same applies to the other capacitors, so we see that the charge on each capacitor has the same magnitude Q. A single capacitor that could replace these three in series without affecting the circuit (that is, Q and V the same) would have a capacitance C_{eq} where

$$Q = C_{eq} V.$$

Now the total voltage V across the three capacitors in series must equal the sum of the voltages across each capacitor:

$$V = V_1 + V_2 + V_3.$$

We also have for each capacitor $V_1 = Q/C_1$, $V_2 = Q/C_2$, and $V_3 = Q/C_3$, since all three have the same charge Q. Also, $V = Q/C_{eq}$, so we substitute for V, V_1, V_2, and V_3 into $V = V_1 + V_2 + V_3$ and get

$$\frac{Q}{C_{eq}} = \frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3} = Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}\right)$$

or

Capacitors in series

$$\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}. \qquad \text{[series]} \quad \textbf{(24–4)}$$

Notice that the equivalent capacitance C_{eq} is *smaller* than the smallest contributing capacitance.

Other connections of capacitors can be analyzed similarly using charge conservation, and often simply in terms of series and parallel connections.

EXAMPLE 24–5 **Equivalent capacitance.** Determine the capacitance of a single capacitor that will have the same effect as the combination shown in Fig. 24–12a. Take $C_1 = C_2 = C_3 = C$.

APPROACH First we find the equivalent capacitance of C_2 and C_3 in parallel, and then consider that capacitance in series with C_1 (Fig. 24–12b).

SOLUTION Capacitors C_2 and C_3 are connected in parallel, so they are equivalent to a single capacitor having capacitance

$$C_{23} = C_2 + C_3 = C + C = 2C.$$

This C_{23} is in *series* with C_1, Fig. 24–12b, so the equivalent capacitance of the entire circuit, C_{eq}, is given by

$$\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_{23}} = \frac{1}{C} + \frac{1}{2C} = \frac{3}{2C}.$$

Hence the equivalent capacitance of the entire combination is $C_{eq} = \frac{2}{3}C$, and it is smaller than any of the contributing capacitors, $C_1 = C_2 = C_3 = C$.

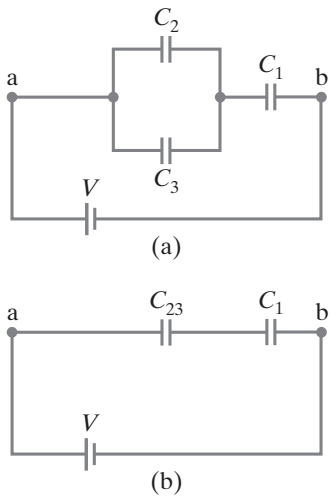


FIGURE 24–12 Example 24–5 and Example 24–6.

EXAMPLE 24–6 **Charge and voltage on capacitors.** Determine the charge on each capacitor in Fig. 24–12a of Example 24–5 and the voltage across each, assuming $C = 3.0\,\mu\text{F}$ and the battery voltage is $V = 4.0\,\text{V}$.

APPROACH We have to work "backward" through Example 24–5. That is, we find the charge Q that leaves the battery, using the equivalent capacitance. Then we find the charge on each separate capacitor and the voltage across each. Each step uses Eq. 24–1, $Q = CV$.

SOLUTION The 4.0-V battery behaves as if it is connected to a capacitance $C_{eq} = \frac{2}{3}C = \frac{2}{3}(3.0\,\mu\text{F}) = 2.0\,\mu\text{F}$. Therefore the charge Q that leaves the battery, by Eq. 24–1, is

$$Q = CV = (2.0\,\mu\text{F})(4.0\,\text{V}) = 8.0\,\mu\text{C}.$$

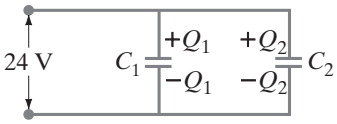
From Fig. 24–12a, this charge arrives at the negative plate of C_1, so $Q_1 = 8.0\,\mu\text{C}$. The charge Q that leaves the positive plate of the battery is split evenly between C_2 and C_3 (symmetry: $C_2 = C_3$) and is $Q_2 = Q_3 = \frac{1}{2}Q = 4.0\,\mu\text{C}$. Next, the voltages across C_2 and C_3 have to be the same. The voltage across each capacitor is obtained using $V = Q/C$. So

$$V_1 = Q_1/C_1 = (8.0\,\mu\text{C})/(3.0\,\mu\text{F}) = 2.7\,\text{V}$$

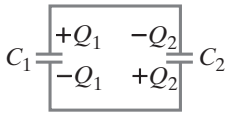
$$V_2 = Q_2/C_2 = (4.0\,\mu\text{C})/(3.0\,\mu\text{F}) = 1.3\,\text{V}$$

$$V_3 = Q_3/C_3 = (4.0\,\mu\text{C})/(3.0\,\mu\text{F}) = 1.3\,\text{V}.$$

EXERCISE D Consider two identical capacitors $C_1 = C_2 = 10\,\mu\text{F}$. What are the smallest and largest capacitances that can be obtained by connecting these in series or parallel combinations? (*a*) $0.2\,\mu\text{F}, 5\,\mu\text{F}$; (*b*) $0.2\,\mu\text{F}, 10\,\mu\text{F}$; (*c*) $0.2\,\mu\text{F}, 20\,\mu\text{F}$; (*d*) $5\,\mu\text{F}, 10\,\mu\text{F}$; (*e*) $5\,\mu\text{F}, 20\,\mu\text{F}$; (*f*) $10\,\mu\text{F}, 20\,\mu\text{F}$.



(a) Initial configuration.



(b) At the instant of reconnection only.



(c) A short time later.

FIGURE 24–13 Example 24–7.

EXAMPLE 24–7 **Capacitors reconnected.** Two capacitors, $C_1 = 2.2\,\mu\text{F}$ and $C_2 = 1.2\,\mu\text{F}$, are connected in parallel to a 24-V source as shown in Fig. 24–13a. After they are charged they are disconnected from the source and from each other, and then reconnected directly to each other with plates of opposite sign connected together (see Fig. 24–13b). Find the charge on each capacitor and the potential across each after equilibrium is established.

APPROACH We find the charge $Q = CV$ on each capacitor initially. Charge is conserved, although rearranged after the switch. The two new voltages will have to be equal.

SOLUTION First we calculate how much charge has been placed on each capacitor after the power source has charged them fully, using Eq. 24–1:

$$Q_1 = C_1 V = (2.2\,\mu\text{F})(24\,\text{V}) = 52.8\,\mu\text{C},$$
$$Q_2 = C_2 V = (1.2\,\mu\text{F})(24\,\text{V}) = 28.8\,\mu\text{C}.$$

Next the capacitors are connected in parallel, Fig. 24–13b, and the potential difference across each must quickly equalize. Thus, the charge cannot remain as shown in Fig. 24–13b, but the charge must rearrange itself so that the upper plates have the same sign of charge, with the lower plates having the opposite charge as shown in Fig. 24–13c. Equation 24–1 applies for each:

$$Q'_1 = C_1 V' \quad \text{and} \quad Q'_2 = C_2 V',$$

where V' is the voltage across each capacitor after the charges have rearranged themselves. We don't know Q'_1, Q'_2, or V', so we need a third equation. This is provided by charge conservation. The charges have rearranged themselves between Figs. 24–13b and c. The total charge on the upper plates in those two Figures must be the same, so we have

$$Q'_1 + Q'_2 = Q_1 - Q_2 = (52.8 - 28.8)\,\mu\text{C} = 24.0\,\mu\text{C}.$$

Combining the last three equations we find:

$$Q'_1 + Q'_2 = C_1 V' + C_2 V' = (C_1 + C_2)\,V'$$

or

$$V' = (Q'_1 + Q'_2)/(C_1 + C_2) = 24.0\,\mu\text{C}/3.4\,\mu\text{F} = 7.06\,\text{V} \approx 7.1\,\text{V}$$
$$Q'_1 = C_1 V' = (2.2\,\mu\text{F})(7.06\,\text{V}) = 15.5\,\mu\text{C} \approx 16\,\mu\text{C}$$
$$Q'_2 = C_2 V' = (1.2\,\mu\text{F})(7.06\,\text{V}) = 8.5\,\mu\text{C}$$

where we have kept only two significant figures in our final answers.

24–4 Storage of Electric Energy

A charged capacitor stores electric energy. The energy stored in a capacitor will be equal to the work done to charge it. The net effect of charging a capacitor is to remove charge from one plate and add it to the other plate. This is what a battery does when it is connected to a capacitor. A capacitor does not become charged instantly. It takes some time (Section 26–5). Initially, when the capacitor is uncharged, no work is required to move the first bit of charge over. When some charge is on each plate, it requires work to add more charge of the same sign because of the electric repulsion. The more charge already on a plate, the more work required to add additional charge. The work needed to add a small amount of charge dq, when a potential difference V is across the plates, is $dW = V\,dq$.

Since $V = q/C$ at any moment (Eq. 24–1), where C is the capacitance, the work needed to store a total charge Q is

$$W = \int_0^Q V\,dq = \frac{1}{C}\int_0^Q q\,dq = \tfrac{1}{2}\frac{Q^2}{C}.$$

Thus we can say that the energy "stored" in a capacitor is

$$U = \tfrac{1}{2}\frac{Q^2}{C}$$

when the capacitor C carries charges $+Q$ and $-Q$ on its two conductors. Since $Q = CV$, where V is the potential difference across the capacitor, we can also write

$$U = \tfrac{1}{2}\frac{Q^2}{C} = \tfrac{1}{2}CV^2 = \tfrac{1}{2}QV. \qquad \textbf{(24–5)}$$

EXAMPLE 24–8 **Energy stored in a capacitor.** A camera flash unit (Fig. 24–14) stores energy in a 660-μF capacitor at 330 V. (*a*) How much electric energy can be stored? (*b*) What is the power output if nearly all this energy is released in 1.0 ms as a flash of light?

PHYSICS APPLIED
Camera flash



FIGURE 24–14 A camera flash unit. The 660-μF capacitor is the black cylinder.

APPROACH We use Eq. 24–5 in the form $U = \frac{1}{2}CV^2$ because we are given C and V.

SOLUTION (*a*) The energy stored is

$$U = \tfrac{1}{2}CV^2 = \tfrac{1}{2}(660\times10^{-6}\,\text{F})(330\,\text{V})^2 = 36\,\text{J}.$$

(*b*) If this energy is released in $\frac{1}{1000}$ of a second, the power output is $P = U/t = (36\,\text{J})/(1.0\times10^{-3}\,\text{s}) = 36{,}000\,\text{W}$.

CONCEPTUAL EXAMPLE 24–9 **Capacitor plate separation increased.** A parallel-plate capacitor carries charge Q and is then disconnected from a battery. The two plates are initially separated by a distance d. Suppose the plates are pulled apart until the separation is $2d$. How has the energy stored in this capacitor changed?

RESPONSE If we increase the plate separation d, we decrease the capacitance according to Eq. 24–2, $C = \epsilon_0 A/d$, by a factor of 2. The charge Q hasn't changed. So according to Eq. 24–5, where we choose the form $U = \frac{1}{2}Q^2/C$ because we know Q is the same and C has been halved, the reduced C means the potential energy stored increases by a factor of $1/(1/2) = 2$.

NOTE We can see why the energy stored increases from a physical point of view: the two plates are charged equal and opposite, so they attract each other. If we pull them apart, we must do work, so we raise their potential energy.

EXAMPLE 24–10 **Repositioning parallel capacitor plates.** The plates of a parallel-plate capacitor have area A, separation x, and are connected to a battery with voltage V. While connected to the battery, the plates are pulled apart until they are separated by $3x$. (*a*) What are the initial and final energies stored in the capacitor? (*b*) How much work is required to pull the plates apart (assume slow constant speed)? (*c*) How much energy is exchanged with the battery?

APPROACH The stored energy when the plates are separated by x is given by Eq. 24–5: $U = \frac{1}{2}CV^2$, where $C = \epsilon_0 A/x$. Unlike Example 24–9, here the capacitor remains connected to the battery so Q can change but V stays the same. Hence charge and energy can flow to or from the battery, and we can not set the work $W = \Delta U$. Instead, the work can be calculated from Eq. 7–7, $W = \int \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}}$.

SOLUTION (*a*) When the separation is x, the capacitance is $C_1 = \epsilon_0 A/x$ and the energy stored is

$$U_1 = \tfrac{1}{2} C_1 V^2 = \tfrac{1}{2}\frac{\epsilon_0 A}{x} V^2.$$

When the separation is $3x$, $C_2 = \epsilon_0 A/3x$ and

$$U_2 = \tfrac{1}{2}\frac{\epsilon_0 A}{3x} V^2.$$

Then

$$\Delta U_{\text{capacitor}} = U_2 - U_1 = \left(\frac{1}{6} - \frac{1}{2}\right)\frac{\epsilon_0 A}{x} V^2 = -\frac{\epsilon_0 A V^2}{3x}.$$

The potential energy decreases as the oppositely charged plates are pulled apart: the plates remain connected to the battery, so V does not change and C decreases; hence some charge leaves each plate $(Q = CV)$, going into the battery, causing U to decrease.

(*b*) The work done in pulling the plates apart is $W = \int_x^{3x} F\, d\ell = \int_x^{3x} QE\, d\ell$, where Q is the charge on one plate at a given moment when the plates are a distance ℓ apart, and E is the field due to the other plate at that instant. You might think we could use $E = V/\ell$ where ℓ is the separation of the plates (Eq. 23–4b). But we want the force on one plate due to the electric field of the other plate only, which is half by *symmetry*; that is, we can say each plate produces half the total electric field, $E = V/2\ell$, to produce the total field $E = (V/2\ell) + (V/2\ell) = V/\ell$. Each plate carries the same charge with opposite sign. The charge Q on each plate at any separation ℓ is given by $Q = CV = (\epsilon_0 A/\ell)V$, so $QE = (\epsilon_0 A/\ell)V(V/2\ell) = \epsilon_0 AV^2/2\ell^2$. Then

$$W = \int_{\ell=x}^{\ell=3x} QE\, d\ell = \frac{\epsilon_0 AV^2}{2}\int_x^{3x}\frac{d\ell}{\ell^2} = -\frac{\epsilon_0 AV^2}{2\ell}\Bigg|_{\ell=x}^{\ell=3x}$$
$$= \frac{\epsilon_0 AV^2}{2}\left(\frac{-1}{3x} + \frac{1}{x}\right) = \frac{\epsilon_0 AV^2}{3x}.$$

As you might expect, the work required to pull these oppositely charged plates apart is positive.

(*c*) Even though the work done is positive, the potential energy stored in the capacitor has decreased, as we saw in part (*a*), which tells us that energy must have gone into the battery (as if charging it). Conservation of energy tells us that the work W done on the system must equal the change in potential energy of the capacitor plus that of the battery (kinetic energy can be assumed to be essentially zero):

$$W = \Delta U_{\text{capacitor}} + \Delta U_{\text{batt}}.$$

Thus the battery experiences a change in energy of

$$\Delta U_{\text{batt}} = W - \Delta U_{\text{capacitor}} = \frac{\epsilon_0 AV^2}{3x} + \frac{\epsilon_0 AV^2}{3x} = \frac{2\epsilon_0 AV^2}{3x}.$$

Thus charge flows back into the battery, raising *its* stored energy. In fact, the battery energy increase is double the work we do.

It is useful to think of the energy stored in a capacitor as being stored in the electric field between the plates. As an example let us calculate the energy stored in a parallel-plate capacitor in terms of the electric field.

We have seen (Fig. 23–1) that the electric field $\vec{\mathbf{E}}$ between two close parallel plates is nearly uniform and its magnitude is related to the potential difference by $V = Ed$ where d is the plate separation (Eq. 23–4b). Also, Eq. 24–2 tells us $C = \epsilon_0 A/d$ for a parallel-plate capacitor. Thus

$$U = \tfrac{1}{2}CV^2 = \tfrac{1}{2}\left(\frac{\epsilon_0 A}{d}\right)(E^2 d^2)$$
$$= \tfrac{1}{2}\epsilon_0 E^2 A d.$$

The quantity Ad is the volume between the plates in which the electric field E exists. If we divide both sides of this equation by the volume, we obtain an expression for the *energy per unit volume* or **energy density**, u:

$$u = \text{energy density} = \frac{\text{energy}}{\text{volume}} = \tfrac{1}{2}\epsilon_0 E^2. \qquad \textbf{(24–6)}$$

The *electric energy stored per unit volume in any region of space is proportional to the square of the electric field* in that region. We derived Eq. 24–6 for the special case of a parallel-plate capacitor. But it can be shown to be true for any region of space where there is an electric field. Note that the units check: for $(\epsilon_0 E^2)$ we have $(\mathrm{C^2/N \cdot m^2})(\mathrm{N/C})^2 = \mathrm{N/m^2} = (\mathrm{N \cdot m})/\mathrm{m^3} = \mathrm{J/m^3}$.

Health Effects

The energy stored in a large capacitance can give you a burn or a shock. One reason you are warned not to touch a circuit, or open up an electronic device, is because capacitors may still be carrying charge even if the external power is turned off.

Capacitors have medical applications. The basis of a **heart defibrillator** is a capacitor charged to a high voltage. A heart attack can be characterized by fast irregular beating of the heart, known as *ventricular* (or *cardiac*) *fibrillation*. The heart then does not pump blood to the rest of the body properly, and if the irregular rhythm lasts for long, death results. A sudden, brief jolt of charge through the heart from a defibrillator can cause complete heart stoppage, sometimes followed by a resumption of normal beating. The defibrillator capacitor is charged to a high voltage, typically a few thousand volts, and is allowed to discharge very rapidly through the heart via a pair of wide contacts known as "pads" or "paddles" that spread out the current over the chest (Fig. 24–15).

PHYSICS APPLIED
Shocks, burns, defibrillators



FIGURE 24–15 Heart defibrillator.

24–5 Dielectrics

In most capacitors there is an insulating sheet of material, such as paper or plastic, called a **dielectric** between the plates (Fig. 24–1b). This serves several purposes. First, dielectrics do not break down (allowing electric charge to flow) as easily as air, so higher voltages can be applied without charge passing across the gap. Second, a dielectric allows the plates to be placed closer together without touching, thus allowing an increased capacitance because the thickness d is smaller in Eq. 24–2. Third, it is found experimentally that if the dielectric fills the space between the two conductors, it increases the capacitance by a factor K, known as the **dielectric constant**. Thus

$$C = KC_0, \qquad \textbf{(24–7)}$$

where C_0 is the capacitance when the space between the two conductors of the capacitor is air (or vacuum), and C is the capacitance when the space is filled with a material whose dielectric constant is K.

The values of the dielectric constant for various materials are given in Table 24–1. Also shown in Table 24–1 is the **dielectric strength**, the maximum electric field before breakdown (charge flow) occurs.

TABLE 24–1
Dielectric Constants (at 20°C)

Material	Dielectric constant K	Dielectric strength (V/m)
Vacuum	1.0000	
Air (1 atm)	1.0006	3×10^6
Paraffin	2.2	10×10^6
Polystyrene	2.6	24×10^6
Vinyl (plastic)	2–4	50×10^6
Paper	3.7	15×10^6
Quartz	4.3	8×10^6
Oil	4	12×10^6
Glass, Pyrex	5	14×10^6
Rubber, neoprene	6.7	20×10^6
Porcelain	6–8	5×10^6
Mica	7	120×10^6
Water (liquid)	80	70×10^6
Strontium titanate	300	8×10^6

For a parallel-plate capacitor (combining Eqs. 24–2 and 24–7),

$$C = K\epsilon_0 \frac{A}{d} \qquad \text{[parallel-plate capacitor]} \quad \textbf{(24–8)}$$

when the space between the plates is completely filled with a dielectric whose dielectric constant is K. (The situation when the dielectric only partially fills the space will be discussed in Example 24–14.)

The quantity $K\epsilon_0$ appears so often in formulas that we define a new quantity

$$\epsilon = K\epsilon_0 \qquad \textbf{(24–9)}$$

called the **permittivity** of a material. Then the capacitance of a parallel-plate capacitor becomes in general

$$C = \epsilon \frac{A}{d}.$$

Note that ϵ_0 represents the permittivity of free space (a vacuum), as in Section 21–5.

CONCEPTUAL EXAMPLE 24–11 **Inserting a dielectric at constant V.** An air-filled capacitor (C_0) consisting of two parallel plates separated by a distance d is connected to a battery of constant voltage $V = V_0$ and acquires a charge Q_0. While it is still connected to the battery, a slab of dielectric material with $K = 3$ is inserted to fill the space between the plates of the capacitor. Will Q increase, decrease, or stay the same?

RESPONSE Since the capacitor remains connected to the battery, the voltage stays constant and equal to the battery voltage V_0. The capacitance C increases from C_0 to KC_0 when the dielectric material is inserted because K in Eq. 24–8 has increased. From the relation $Q = CV$, if V stays constant but C increases, Q must increase to

$$Q = KQ_0.$$

CONCEPTUAL EXAMPLE 24–12 **Inserting a dielectric into an isolated capacitor.** Suppose the air-filled capacitor (C_0) of Example 24–11 is charged (to Q_0) and then disconnected from the battery (V_0). Next that same dielectric is inserted between the plates. Will Q, C, or V change?

RESPONSE The charge Q_0 remains the same—the capacitor is isolated, so there is nowhere for the charge to go. The capacitance, though, *increases* as a result of inserting the dielectric (Eq. 24–8) to $C = KC_0$. The voltage across the capacitor also changes. The voltage *decreases* because, by Eq. 24–1, $Q = Q_0 = CV$, so $V = Q_0/C = Q_0/KC_0 = V_0/K$ where $V_0 = Q_0/C_0$ is the initial voltage. So Q stays constant, C increases, and V decreases by a factor of K:

$$V = \frac{V_0}{K}.$$

EXERCISE E Return to the Chapter-Opening Question, page 692, and answer it again now. Try to explain why you may have answered differently the first time.

The energy density stored in an electric field E (Section 24–4) when a dielectric (K) is present becomes (see Eq. 24–6)

$$u = \tfrac{1}{2}K\epsilon_0 E^2 = \tfrac{1}{2}\epsilon E^2. \qquad [E \text{ in a dielectric}]$$

An electric field is also altered when a dielectric is inserted. When no dielectric is present, the electric field between the plates of a parallel-plate capacitor, for example, is given by Eq. 23–4b:

$$E_0 = \frac{V_0}{d},$$

where V_0 is the potential difference between the plates and d is their separation. If the capacitor is isolated so that the charge remains fixed on the plates when a

dielectric is inserted, filling the space between the plates, the potential difference drops to $V = V_0/K$ (as we saw in Example 24–12). So the electric field inside the dielectric is

$$E = E_D = \frac{V}{d} = \frac{V_0}{Kd}$$

or

$$E_D = \frac{E_0}{K}. \qquad \text{[in a dielectric]} \quad \textbf{(24–10)}$$

The electric field in a dielectric is reduced by a factor equal to the dielectric constant. But the field in a dielectric (or insulator) is not reduced all the way to zero as in a conductor. Equation 24–10 is valid inside a dielectric even if the dielectric's thickness is smaller than the gap between the capacitor plates.

EXAMPLE 24–13 **Dielectric removal.** A parallel-plate capacitor, filled with a dielectric with $K = 3.4$, is connected to a 100-V battery (Fig. 24–16a). After the capacitor is fully charged, the battery is disconnected. The plates have area $A = 4.0\,\text{m}^2$, and are separated by $d = 4.0\,\text{mm}$. (*a*) Find the capacitance, the charge on the capacitor, the electric field strength, and the energy stored in the capacitor. (*b*) The dielectric is then carefully removed, without changing the plate separation nor does any charge leave the capacitor (Fig. 24–16b). Find the new values of capacitance, electric field strength, voltage between the plates, and the energy stored in the capacitor.

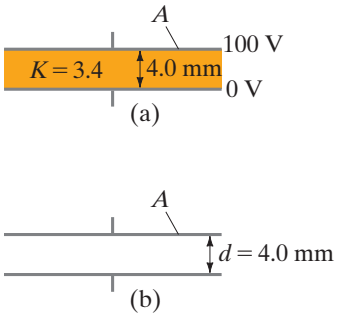


FIGURE 24–16 Example 24–13.

APPROACH We use the expressions for parallel-plate capacitance and electric field with and without a dielectric.

SOLUTION (*a*) First we find the capacitance, with dielectric:

$$C = \frac{K\epsilon_0 A}{d} = \frac{3.4\left(8.85\times10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2\right)\left(4.0\,\text{m}^2\right)}{4.0\times10^{-3}\,\text{m}}$$
$$= 3.0\times10^{-8}\,\text{F} = 30\,\text{nF}.$$

The charge Q on the plates is

$$Q = CV = \left(3.0\times10^{-8}\,\text{F}\right)(100\,\text{V}) = 3.0\times10^{-6}\,\text{C}.$$

The electric field between the plates is

$$E = \frac{V}{d} = \frac{100\,\text{V}}{4.0\times10^{-3}\,\text{m}} = 25\,\text{kV/m}.$$

Finally, the total energy stored in the capacitor is

$$U = \tfrac{1}{2}CV^2 = \tfrac{1}{2}\left(3.0\times10^{-8}\,\text{F}\right)(100\,\text{V})^2 = 1.5\times10^{-4}\,\text{J}.$$

(*b*) The capacitance without dielectric decreases by a factor $K = 3.4$:

$$C_0 = \frac{C}{K} = \frac{\left(3.0\times10^{-8}\,\text{F}\right)}{3.4} = 8.8\times10^{-9}\,\text{F} = 8.8\,\text{nF}.$$

Because the battery has been disconnected, the charge Q cannot go anywhere; it is still Q. When the dielectric is removed, $V = Q/C$ increases by a factor $K = 3.4$ to 340 V, since C decreases. The electric field is

$$E = \frac{V}{d} = \frac{340\,\text{V}}{4.0\times10^{-3}\,\text{m}} = 85\,\text{kV/m},$$

so is 3.4× larger than before. The energy stored is

$$U = \tfrac{1}{2}CV^2 = \tfrac{1}{2}\left(8.8\times10^{-9}\,\text{F}\right)(340\,\text{V})^2 = 5.1\times10^{-4}\,\text{J},$$

also increasing by a factor of $K = 3.4$.

NOTE Where did all this extra energy come from? The energy increased because work had to be done to remove the dielectric. The work required was $W = \left(5.1\times10^{-4}\,\text{J}\right) - \left(1.5\times10^{-4}\,\text{J}\right) = 3.6\times10^{-4}\,\text{J}$. (We will see in the next Section that work is required because of the force of attraction between induced charge on the dielectric and the charges on the plates, Fig. 24–17c.)

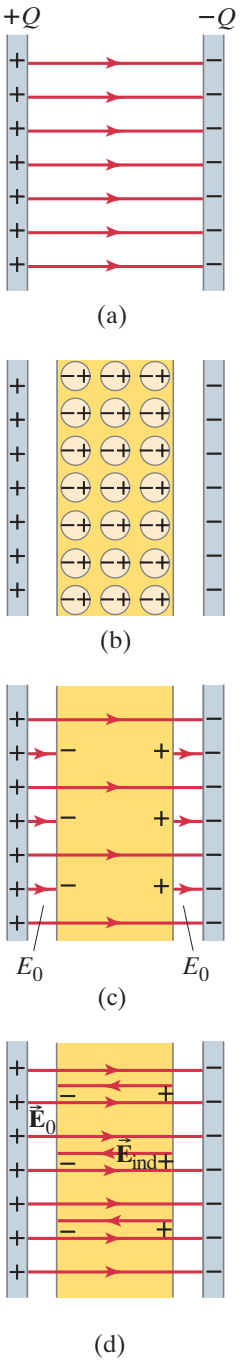


FIGURE 24–17 Molecular view of the effects of a dielectric.

*24–6 Molecular Description of Dielectrics

Let us examine, from the molecular point of view, why the capacitance of a capacitor should be larger when a dielectric is between the plates. A capacitor whose plates are separated by an air gap has capacitance C_0 and a charge $+Q$ on one plate and $-Q$ on the other (Fig. 24–17a). Assume it is isolated (not connected to a battery) so charge cannot flow to or from the plates. The potential difference between the plates, V_0, is given by Eq. 24–1:

$$Q = C_0 V_0,$$

where the subscripts refer to air between the plates. Now we insert a dielectric between the plates (Fig. 24–17b). Because of the electric field between the capacitor plates, the dielectric molecules will tend to become oriented as shown in Fig. 24–17b. If the dielectric molecules are *polar*, the positive end is attracted to the negative plate and vice versa. Even if the dielectric molecules are not polar, electrons within them will tend to move slightly toward the positive capacitor plate, so the effect is the same. The net effect of the aligned dipoles is a net negative charge on the outer edge of the dielectric facing the positive plate, and a net positive charge on the opposite side, as shown in Fig. 24–17c.

Some of the electric field lines, then, do not pass through the dielectric but instead end on charges induced on the surface of the dielectric as shown in Fig. 24–17c. Hence the electric field within the dielectric is less than in air. That is, the electric field in the space between the capacitor plates, assumed filled by the dielectric, has been reduced by some factor K. The voltage across the capacitor is reduced by the same factor K because $V = Ed$ (Eq. 23–4b) and hence, by Eq. 24–1, $Q = CV$, the capacitance C must increase by that same factor K to keep Q constant.

As shown in Fig. 24–17d, the electric field within the dielectric E_{D} can be considered as the vector sum of the electric field $\vec{\mathbf{E}}_0$ due to the "free" charges on the conducting plates, and the field $\vec{\mathbf{E}}_{\text{ind}}$ due to the induced charge on the surfaces of the dielectric. Since these two fields are in opposite directions, the net field within the dielectric, $E_0 - E_{\text{ind}}$, is less than E_0. The precise relationship is given by Eq. 24–10, even if the dielectric does not fill the gap between the plates:

$$E_{\text{D}} = E_0 - E_{\text{ind}} = \frac{E_0}{K},$$

so

$$E_{\text{ind}} = E_0\left(1 - \frac{1}{K}\right).$$

The electric field in air between two parallel plates is related to the surface charge density, σ, by $E = \sigma/\epsilon_0$ (Example 21–14 or 22–8). Thus

$$E_0 = \sigma/\epsilon_0$$

where $\sigma = Q/A$ is the surface charge density on the conductor; Q is the net charge on the conductor and is often called the **free charge** (since charge is free to move in a conductor). Similarly, we define an equivalent induced surface charge density σ_{ind} on the dielectric; then

$$E_{\text{ind}} = \sigma_{\text{ind}}/\epsilon_0$$

where E_{ind} is the electric field due to the induced charge $Q_{\text{ind}} = \sigma_{\text{ind}} A$ on the surface of the dielectric, Fig. 24–17d. Q_{ind} is often called the **bound charge**, since it is on an insulator and is not free to move. Since $E_{\text{ind}} = E_0(1 - 1/K)$ as shown above, we now have

$$\sigma_{\text{ind}} = \sigma\left(1 - \frac{1}{K}\right) \qquad \textbf{(24–11a)}$$

and

$$Q_{\text{ind}} = Q\left(1 - \frac{1}{K}\right). \qquad \textbf{(24–11b)}$$

K is always greater than 1, so we see that the charge induced on the dielectric is always less than the free charge on each of the capacitor plates.

EXAMPLE 24–14 **Dielectric partially fills capacitor.** A parallel-plate capacitor has plates of area $A = 250\,\text{cm}^2$ and separation $d = 2.00\,\text{mm}$. The capacitor is charged to a potential difference $V_0 = 150\,\text{V}$. Then the battery is disconnected (the charge Q on the plates then won't change), and a dielectric sheet $(K = 3.50)$ of the same area A but thickness $\ell = 1.00\,\text{mm}$ is placed between the plates as shown in Fig. 24–18. Determine (*a*) the initial capacitance of the air-filled capacitor, (*b*) the charge on each plate before the dielectric is inserted, (*c*) the charge induced on each face of the dielectric after the dielectric is inserted, (*d*) the electric field in the space between each plate and the dielectric, (*e*) the electric field in the dielectric, (*f*) the potential difference between the plates after the dielectric is added, and (*g*) the capacitance after the dielectric is in place.

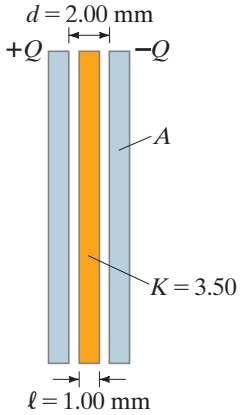


FIGURE 24–18 Example 24–14.

APPROACH We use the expressions for capacitance and charge developed in this Section plus (part *e*) Eq. 23–4a, $V = -\int \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$.

SOLUTION (*a*) Before the dielectric is in place, the capacitance is

$$C_0 = \epsilon_0 \frac{A}{d} = (8.85 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)\left(\frac{2.50 \times 10^{-2}\,\text{m}^2}{2.00 \times 10^{-3}\,\text{m}}\right) = 111\,\text{pF}.$$

(*b*) The charge on each plate is

$$Q = C_0 V_0 = (1.11 \times 10^{-10}\,\text{F})(150\,\text{V}) = 1.66 \times 10^{-8}\,\text{C}.$$

(*c*) Now the dielectric is between the plates. Equations 24–10 and 24–11 are valid even when the dielectric does not fill the gap, so (Eq. 24–11b)

$$Q_{\text{ind}} = Q\left(1 - \frac{1}{K}\right) = (1.66 \times 10^{-8}\,\text{C})\left(1 - \frac{1}{3.50}\right) = 1.19 \times 10^{-8}\,\text{C}.$$

(*d*) The electric field in the gaps between the plates and the dielectric (see Fig. 24–17c) is the same as in the absence of the dielectric since the charge on the plates has not been altered. That is, $E_0 = \sigma/\epsilon_0$ as we just saw on the previous page (also Example 21–14). Thus

$$E_0 = \frac{\sigma}{\epsilon_0} = \frac{Q}{\epsilon_0 A} = \frac{1.66 \times 10^{-8}\,\text{C}}{(8.85 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)(2.50 \times 10^{-2}\,\text{m}^2)} = 7.50 \times 10^4\,\text{V/m}.$$

(*e*) In the dielectric the electric field is (Eq. 24–10)

$$E_{\text{D}} = \frac{E_0}{K} = \frac{7.50 \times 10^4\,\text{V/m}}{3.50} = 2.14 \times 10^4\,\text{V/m}.$$

(*f*) To obtain the potential difference in the presence of the dielectric we use Eq. 23–4a, and integrate from the surface of the negative plate a to the positive plate b along a straight line parallel to the field lines:

$$V = -\int_{\text{a}}^{\text{b}} \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = E_0(d - \ell) + E_{\text{D}}\ell,$$

which can be simplified to

$$\begin{aligned} V &= E_0\left(d - \ell + \frac{\ell}{K}\right) \\ &= (7.50 \times 10^4\,\text{V/m})\left(1.00 \times 10^{-3}\,\text{m} + \frac{1.00 \times 10^{-3}\,\text{m}}{3.50}\right) \\ &= 96.4\,\text{V}. \end{aligned}$$

PROBLEM SOLVING
Voltage across plates a distance d apart with dielectric K of thickness ℓ

(*g*) In the presence of the dielectric, the capacitance is

$$C = \frac{Q}{V} = \frac{1.66 \times 10^{-8}\,\text{C}}{96.4\,\text{V}} = 172\,\text{pF}.$$

EXERCISE F If the dielectric filled the space between the plates $(\ell = d)$, what would be the answers to (*f*) and (*g*)?

Summary

A **capacitor** is a device used to store charge (and electric energy), and consists of two nontouching conductors. The two conductors generally hold equal and opposite charges of magnitude Q. The ratio of this charge Q to the potential difference V between the conductors is called the **capacitance**, C:

$$C = \frac{Q}{V} \quad \text{or} \quad Q = CV. \qquad \textbf{(24–1)}$$

The capacitance of a parallel-plate capacitor is proportional to the area A of each plate and inversely proportional to their separation d:

$$C = \epsilon_0 \frac{A}{d}. \qquad \textbf{(24–2)}$$

When capacitors are connected in **parallel**, the equivalent capacitance is the sum of the individual capacitances:

$$C_{\text{eq}} = C_1 + C_2 + \cdots. \qquad \textbf{(24–3)}$$

When capacitors are connected in **series**, the reciprocal of the equivalent capacitance equals the sum of the reciprocals of the individual capacitances:

$$\frac{1}{C_{\text{eq}}} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots. \qquad \textbf{(24–4)}$$

A charged capacitor stores an amount of electric energy given by

$$U = \tfrac{1}{2}QV = \tfrac{1}{2}CV^2 = \tfrac{1}{2}\frac{Q^2}{C}. \qquad \textbf{(24–5)}$$

This energy can be thought of as stored in the electric field between the plates. In any electric field $\vec{\mathbf{E}}$ in free space the **energy density** u (energy per unit volume) is

$$u = \tfrac{1}{2}\epsilon_0 E^2. \qquad \textbf{(24–6)}$$

The space between the two conductors of a capacitor usually contains a nonconducting material such as air, paper, or plastic. These materials are referred to as **dielectrics**, and the capacitance is proportional to a property of dielectrics called the **dielectric constant**, K (nearly equal to 1 for air). For a parallel-plate capacitor

$$C = K\epsilon_0 \frac{A}{d} = \epsilon \frac{A}{d} \qquad \textbf{(24–8)}$$

where $\epsilon = K\epsilon_0$ is called the **permittivity** of the dielectric material. When a dielectric is present, the energy density is

$$u = \tfrac{1}{2}K\epsilon_0 E^2 = \tfrac{1}{2}\epsilon E^2.$$

Questions

1. Suppose two nearby conductors carry the same negative charge. Can there be a potential difference between them? If so, can the definition of capacitance, $C = Q/V$, be used here? Why or why not?
2. Suppose the separation of plates d in a parallel-plate capacitor is not very small compared to the dimensions of the plates. Would you expect Eq. 24–2 to give an overestimate or underestimate of the true capacitance? Explain.
3. Suppose one of the plates of a parallel-plate capacitor was moved so that the area of overlap was reduced by half, but they are still parallel. How would this affect the capacitance?
4. When a battery is connected to a capacitor, why do the two plates acquire charges of the same magnitude? Will this be true if the two plates are different sizes or shapes?
5. Describe a simple method of measuring ϵ_0 using a capacitor.
6. Suppose three identical capacitors are connected to a battery. Will they store more energy if connected in series or in parallel?
7. A large copper sheet of thickness ℓ is placed between the parallel plates of a capacitor, but does not touch the plates. How will this affect the capacitance?
8. The parallel plates of an isolated capacitor carry opposite charges, Q. If the separation of the plates is increased, is a force required to do so? Is the potential difference changed? What happens to the work done in the pulling process?
9. How does the energy stored in a capacitor change, as the capacitor remains connected to a battery, if the separation of the plates is doubled?
10. If the voltage across a fixed capacitor is doubled, the amount of energy it stores (*a*) doubles; (*b*) is halved; (*c*) is quadrupled; (*d*) is unaffected; (*e*) none of these. Explain.
11. An isolated charged capacitor has horizontal plates. If a thin dielectric is inserted a short way between the plates, Fig. 24–19, will it move left or right when it is released? Explain.



FIGURE 24–19 Question 11.

12. Suppose a battery remains connected to the capacitor in Question 11. What then will happen when the dielectric is released? Explain.
13. How does the energy stored in a capacitor change when a dielectric is inserted if (*a*) the capacitor is isolated so Q does not change; (*b*) the capacitor remains connected to a battery so V does not change? Explain.
14. For dielectrics consisting of polar molecules, how would you expect the dielectric constant to change with temperature?
15. A dielectric is pulled out from between the plates of a capacitor which remains connected to a battery. What changes occur to (*a*) the capacitance, (*b*) the charge on the plates, (*c*) the potential difference, (*d*) the energy stored in the capacitor, and (*e*) the electric field? Explain your answers.
16. We have seen that the capacitance C depends on the size and position of the two conductors, as well as on the dielectric constant K. What then did we mean when we said that C is a constant in Eq. 24–1?
17. What value might we assign to the dielectric constant for a good conductor? Explain.

MisConceptual Questions

1. Which of the following do not affect capacitance?
(*a*) Area of the plates.
(*b*) Separation of the plates.
(*c*) Material between the plates.
(*d*) Charge on the plates.
(*e*) Energy stored in the capacitor.

2. Two uncharged capacitors are connected in series and then to a battery. The battery transfers a charge of $-Q$ to one side of one capacitor as shown in Figure 24–20. What are the charges on the capacitor sides A and B respectively?
(*a*) $0, 0$.
(*b*) $-2Q, +Q$.
(*c*) $-Q, +Q$.
(*d*) $0, -Q$.
(*e*) $+2Q, -2Q$.
(*f*) Not enough information to tell.

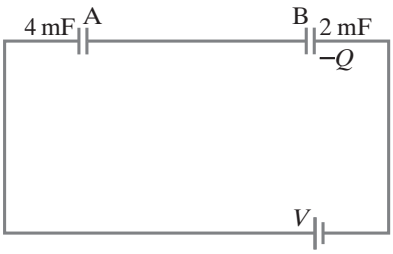


FIGURE 24–20
MisConceptual Question 2.

3. Two flat pieces of metal form a parallel-plate capacitor. The plates carry equal but opposite charges and are not connected to anything. As the plates are moved closer together,
(*a*) both the charge and voltage decrease.
(*b*) both the charge and voltage increase.
(*c*) the charge decreases while the voltage remains constant.
(*d*) the charge remains constant while the voltage decreases.
(*e*) the charge remains constant while the voltage increases.

4. Which of the following would decrease the capacitance of a capacitor?
(*a*) Increasing the spacing between the plates.
(*b*) Increasing the area of the plates.
(*c*) Adding a dielectric between the plates.
(*d*) Increasing the voltage applied across the capacitor.

5. When capacitors are connected in series, the equivalent capacitance is _____ the smallest capacitance; when capacitors are connected in parallel, the equivalent capacitance is _____ the largest capacitance.
(*a*) greater than; equal to
(*b*) greater than; less than
(*c*) less than; greater than
(*d*) equal to; less than
(*e*) equal to; equal to

6. Two different capacitors store the same amount of charge. Which capacitor has a larger potential difference across it?
(*a*) The one with the larger capacitance.
(*b*) The one with the smaller capacitance.
(*c*) Both have the same potential difference.
(*d*) The answer cannot be determined without knowing the actual values involved.

7. The plates of a parallel-plate capacitor are maintained at constant voltage as they are gradually moved closer together, without touching. How is the amount of charge on the plates affected during this process?
(*a*) It quickly becomes 0.
(*b*) It remains constant.
(*c*) It gradually increases.
(*d*) It gradually decreases.

8. The plates of a parallel-plate capacitor are maintained at constant voltage (by a battery) as they are gradually pulled apart. How is the strength of the electric field affected?
(*a*) It gradually increases.
(*b*) It gradually decreases.
(*c*) It becomes zero immediately.
(*d*) It increases very rapidly.
(*e*) It remains constant.

9. A battery establishes a voltage V across a parallel-plate capacitor. After the battery is disconnected, the distance between the plates is doubled without loss of charge. Accordingly, the capacitance _____ and the voltage between the plates _____.
(*a*) increases; decreases.
(*b*) decreases; increases.
(*c*) increases; increases.
(*d*) decreases; decreases.
(*e*) stays the same; stays the same.

10. Adding a dielectric that fills the space between the plates of a capacitor increases all of the following except
(*a*) electric field between the plates.
(*b*) capacitance.
(*c*) charge stored for a given voltage.
(*d*) maximum voltage that can be applied between the plates.
(*e*) the energy stored in a capacitor for a given voltage.

11. As a result of being connected to a voltage V, a capacitor stores energy U. If the voltage is changed to $2V$,
(*a*) the energy is unchanged.
(*b*) the energy is $2U$.
(*c*) the energy is $\frac{1}{2}U$.
(*d*) the energy is $4U$.
(*e*) the energy changes, but not to one of these values.

12. Two unequal capacitors are connected in series and the series combination is connected to a voltage V_0. For this arrangement
(*a*) each capacitor carries the same charge.
(*b*) the charges are different on each capacitor but must add up to the charge supplied by the voltage source.
(*c*) each capacitor has the same voltage and it is the same as V_0.
(*d*) None of these statements is true.

13. When the distance d between the two plates of an isolated charged (Q_0) parallel-plate capacitor is reduced to $d/2$, the energy stored
(*a*) is reduced by $\frac{1}{4}$.
(*b*) is reduced by $\frac{1}{2}$.
(*c*) stays the same.
(*d*) is increased by 2.
(*e*) is increased by 4.

Problems

24–1 Capacitors

1. (I) The two plates of a capacitor hold $+3500\,\mu C$ and $-3500\,\mu C$ of charge, respectively, when the potential difference is 960 V. What is the capacitance?
2. (I) A 6500-pF capacitor holds plus and minus charges of 16.5×10^{-8} C. What is the voltage across the capacitor?
3. (I) How much charge flows from each terminal of a 12.0-V battery when it is connected to a 7.50-μF capacitor?
4. (II) The charge on a capacitor increases by 15 μC when the voltage across it increases from 83 V to 121 V. What is the capacitance of the capacitor?
5. (II) An isolated capacitor C_1 carries a charge Q_0. Its wires are then connected to those of a second capacitor C_2, previously uncharged. What charge will each carry now? What will be the potential difference across each?
6. (II) A 7.7-μF capacitor is charged by a 185-V battery (Fig. 24–21a) and then is disconnected from the battery. When this capacitor (C_1) is then connected (Fig. 24–21b) to a second (initially uncharged) capacitor, C_2, the final voltage on each capacitor is 15 V. What is the value of C_2? [*Hint*: Charge is conserved.]

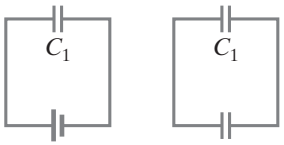


FIGURE 24–21
Problem 6.

7. (II) It takes 24 J of energy to move a 0.30-mC charge from one plate of a 15-μF capacitor to the other. How much charge is on each plate?
8. (II) A 2.70-μF capacitor is charged to 435 V and a 4.00-μF capacitor is charged to 525 V. (*a*) These capacitors are then disconnected from their batteries. Then the positive plates are connected to each other and the negative plates are connected to each other. What will be the potential difference across each capacitor and the charge on each? (*b*) What is the voltage and charge for each capacitor if plates of opposite sign are connected? [*Hint*: Charge is conserved.]

24–2 Determination of Capacitance

9. (I) To make a 0.40-μF capacitor, what area must the plates have if they are to be separated by a 3.4-mm air gap?
10. (I) What is the capacitance per unit length (F/m) of a coaxial cable whose inner conductor has a 1.0-mm diameter and the outer cylindrical sheath has a 5.0-mm diameter? Assume the space between is filled with air.
11. (I) Determine the capacitance of the Earth, assuming it to be a spherical conductor.
12. (II) If a capacitor has opposite 4.2-μC charges on the plates, and an electric field of 1.8 kV/mm is desired between the plates, what must each plate's area be?
13. (II) To get an idea how big a farad is, suppose you want to make a 1-F air-filled parallel-plate capacitor for a circuit you are building. To make it a reasonable size, suppose you limit the plate area to 1.0 cm^2. What would the gap have to be between the plates? Is this practically achievable?
14. (II) Use Gauss's law to show that $\vec{\mathbf{E}} = 0$ inside the inner conductor of a cylindrical capacitor (see Fig. 24–7 and Example 24–2) as well as outside the outer cylinder.
15. (II) An electric field of 8.50×10^5 V/m is desired between two parallel plates, each of area 38.0 cm^2 and separated by 2.45 mm of air. What charge must be on each plate?
16. (II) Dry air will break down if the electric field exceeds about 3.0×10^6 V/m. What amount of charge can be placed on an air-gap capacitor if the area of each plate is 8.2 cm^2?
17. (II) How strong is the electric field between the plates of a 0.80-μF air-gap capacitor if they are 2.0 mm apart and each has a charge of magnitude 84 μC?
18. (III) A large metal sheet of thickness ℓ is placed between, and parallel to, the plates of the parallel-plate capacitor of Fig. 24–4. It does not touch the plates, and extends beyond their edges. (*a*) What is now the net capacitance in terms of A, d, and ℓ? (*b*) If $\ell = 0.40\,d$, by what factor does the capacitance change when the sheet is inserted?
19. (III) Small distances can be measured using a capacitor whose plate separation x is variable. Consider an air-filled parallel-plate capacitor with fixed plate area $A = 25\ mm^2$ separated by a variable distance x. Assume this capacitor is attached to a capacitance-measuring instrument which can measure capacitance C in the range 1.0 pF to 1000.0 pF with an accuracy of $\Delta C = 0.1$ pF. (*a*) If C is measured while x is varied, over what range $(x_{min} \le x \le x_{max})$ can the plate-separation distance (in μm) be determined by this setup? (*b*) Define Δx to be the accuracy (magnitude) to which x can be determined, and determine a formula for Δx. (*c*) Determine the percent accuracy to which x_{min} and x_{max} can be measured.
20. (III) In an **electrostatic air cleaner ("precipitator")**, the strong nonuniform electric field in the central region of a cylindrical capacitor (with outer and inner cylindrical radii R_a and R_b) is used to create ionized air molecules for use in charging dust and soot particles (Fig. 24–22). Under standard atmospheric conditions, if air is subjected to an electric field magnitude that exceeds its dielectric strength $E_S \approx 3.0 \times 10^6$ N/C, air molecules will dissociate into positively charged ions and free electrons. In a precipitator, the region within which air is ionized (the *corona discharge* region) occupies a cylindrical volume of radius R that is typically five times that of the inner cylinder. Assume a particular precipitator is constructed with $R_b = 0.10$ mm and $R_a = 10.0$ cm. In order to create a corona discharge region with radius $R = 5.0\,R_b$, what potential difference V should be applied between the precipitator's inner and outer conducting cylinders? [Besides dissociating air, the charged inner cylinder repels the resulting positive ions from the corona discharge region, where they are put to use in charging dust particles, which are then "collected" on the negatively charged outer cylinder.]

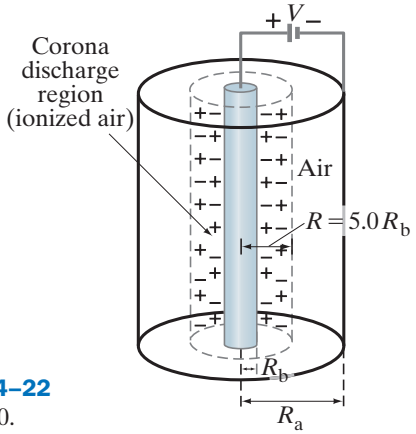


FIGURE 24–22
Problem 20.

24–3 Capacitors in Series and Parallel

21. (I) The capacitance of a portion of a circuit is to be reduced from 3400 pF to 1600 pF. What capacitance can be added to the circuit to produce this effect without removing existing circuit elements? Must any existing connections be broken to accomplish this?

22. (I) (*a*) Six 3.2-μF capacitors are connected in parallel. What is the equivalent capacitance? (*b*) What is their equivalent capacitance if connected in series?

23. (II) Suppose three parallel-plate capacitors, whose plates have areas A_1, A_2, and A_3 and separations d_1, d_2, and d_3, are connected in parallel. Show, using only Eq. 24–2, that Eq. 24–3 is valid.

24. (II) An electric circuit was accidentally constructed using a 5.0-μF capacitor instead of the required 22-μF value. Without removing the 5.0-μF capacitor, what can a technician add to correct this circuit?

25. (II) Three conducting plates, each of area A, are connected as shown in Fig. 24–23. (*a*) Are the two capacitors thus formed connected in series or in parallel? (*b*) Determine C as a function of d_1, d_2, and A. Assume $d_1 + d_2$ is much less than the dimensions of the plates. (*c*) The middle plate can be moved (changing the values of d_1 and d_2), so as to vary the capacitance. What are the minimum and maximum values of the net capacitance?



FIGURE 24–23
Problem 25.

26. (II) Answer the same questions as for Problem 25, for the three plates connected as shown in Fig. 24–24.



FIGURE 24–24
Problem 26.

27. (II) Consider three capacitors, of capacitance 3100 pF, 5800 pF, and 0.0100 μF. What maximum and minimum capacitance can you form from these? How do you make the connection in each case?

28. (II) In Fig. 24–25, suppose $C_1 = C_2 = C_3 = C_4 = C$. (*a*) Determine the equivalent capacitance between points a and b. (*b*) Determine the charge on each capacitor and the potential difference across each in terms of V.

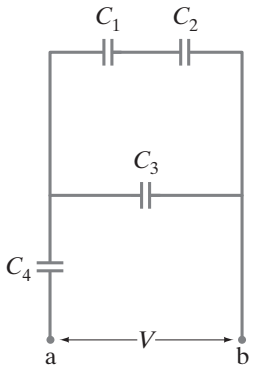


FIGURE 24–25
Problems 28 and 29.

29. (II) Suppose in Fig. 24–25 that $C_1 = C_2 = C_3 = 16.0\,\mu$F and $C_4 = 28.5\,\mu$F. If the charge on C_2 is $Q_2 = 14.8\,\mu$C, determine the charge on each of the other capacitors, the voltage across each capacitor, and the voltage V_{ab} across the entire combination.

30. (II) A 0.50-μF and a 0.80-μF capacitor are connected in series to a 9.0-V battery. Calculate (*a*) the potential difference across each capacitor and (*b*) the charge on each. (*c*) Repeat parts (*a*) and (*b*) assuming the two capacitors are in parallel.

31. (II) The switch S in Fig. 24–26 is connected downward so that capacitor C_2 becomes fully charged by the battery of voltage V_0. If the switch is then connected upward, determine the charge on each capacitor after the switching.

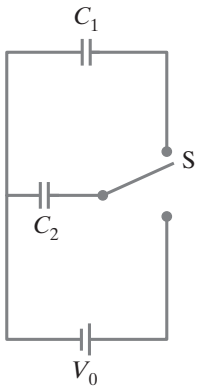


FIGURE 24–26
Problem 31.

32. (II) Given three capacitors, $C_1 = 2.0\,\mu$F, $C_2 = 1.5\,\mu$F, and $C_3 = 3.0\,\mu$F, what arrangement of parallel and series connections with a 12-V battery will give the minimum voltage drop across the 2.0-μF capacitor? What is the minimum voltage drop?

33. (II) A circuit contains a single 250-pF capacitor hooked across a battery. It is desired to store 5.0 times as much energy in a combination of two capacitors by adding a single capacitor to this one. How would you hook it up, and what would its value be?

34. (II) Suppose in Fig. 24–27 that $C_1 = C_3 = 8.0\,\mu$F, $C_2 = C_4 = 16\,\mu$F, and $Q_3 = 21\,\mu$C. Determine (*a*) the charge on each of the other capacitors, (*b*) the voltage across each capacitor, and (*c*) the voltage V_{ba} across the combination.

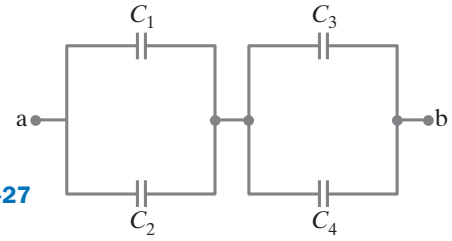


FIGURE 24–27
Problem 34.

35. (II) In the **capacitance bridge** shown in Fig. 24–28, a voltage V_0 is applied and the variable capacitor C_1 is adjusted until there is zero voltage between points a and b as measured on the voltmeter (•–Ⓥ–•). Determine the unknown capacitance C_x if $C_1 = 8.9\,\mu$F and the fixed capacitors have $C_2 = 16.0\,\mu$F and $C_3 = 4.8\,\mu$F. No charge flows through the voltmeter when it reads zero.)

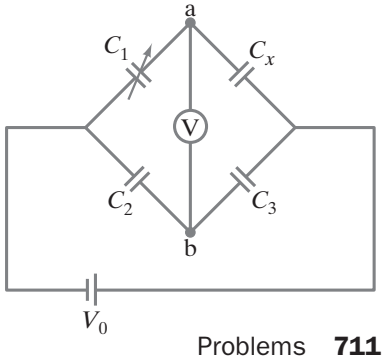


FIGURE 24–28
Problem 35.

36. (II) Two capacitors connected in parallel produce an equivalent capacitance of 32.9 μF, but when connected in series the equivalent capacitance is only 5.5 μF. What is the individual capacitance of each capacitor?

37. (III) Suppose one plate of a parallel-plate capacitor is tilted so it makes a small angle θ with the other plate, as shown in Fig. 24–29. Determine a formula for the capacitance C in terms of A, d, and θ, where A is the area of each plate and θ is small. Assume the plates are square. [*Hint*: Imagine the capacitor as many infinitesimal capacitors in parallel.]

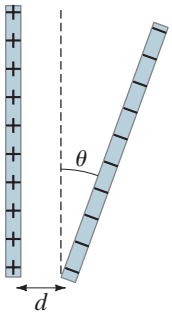


FIGURE 24–29
Problem 37.

24–4 Electric Energy Storage

38. (I) 720 V is applied to a 2800-pF capacitor. How much energy is stored?

39. (I) There is an electric field near the Earth's surface whose intensity is about 150 V/m. How much energy is stored per cubic meter in this field?

40. (I) A **cardiac defibrillator** is used to shock a heart that is beating erratically. A capacitor in this device is charged to 5.0 kV and stores 1200 J of energy. What is its capacitance?

41. (II) How much energy is stored by the electric field between two square plates, 8.0 cm on a side, separated by a 1.5-mm air gap? The charges on the plates are equal and opposite and of magnitude 420 μC.

42. (II) A parallel-plate capacitor has fixed charges $+Q$ and $-Q$. The separation of the plates is then halved. (*a*) By what factor does the energy stored in the electric field change? (*b*) How much work must be done to decrease the separation of the plates from d to $\frac{1}{2}d$? The area of each plate is A.

43. (II) How much energy must a 24-V battery expend to charge a 0.45-μF and a 0.20-μF capacitor fully when they are placed (*a*) in parallel, (*b*) in series? (*c*) How much charge flowed from the battery in each case?

44. (II) (*a*) Suppose the radius R_a of the outer cylinder of a cylindrical capacitor was tripled, but the charge was kept constant. By what factor would the stored energy change? Where would the energy come from? (*b*) Repeat part (*a*), assuming the voltage remains constant.

45. (II) A 3.70-μF capacitor is charged by a 16.0-V battery. It is disconnected from the battery and then connected to an uncharged 5.00-μF capacitor. Determine the total stored energy (*a*) before the two capacitors are connected, and (*b*) after they are connected. (*c*) What is the change in energy?

46. (II) How much work would be required to remove a metal sheet from between the plates of a capacitor (as in Problem 18a), assuming: (*a*) the battery remains connected so the voltage remains constant; (*b*) the battery is disconnected so the charge remains constant?

47. (II) When two uncharged capacitors are connected in parallel and then connected to a battery, the total stored energy is 6.0 times greater than when they are connected in series and then connected to the same battery. What is the ratio of the two capacitances?

48. (II) The electric field near the Earth is about 150 N/C. (*a*) What is the energy density near the Earth? (*b*) Approximately how much electric energy is stored in the Earth's electric field in the first 10 m above the Earth's surface? (*c*) Compare (by a ratio) the answer to part (*b*) with the daily output from a 2000-MW power station.

49. (II) (*a*) Show that each plate of a parallel-plate capacitor exerts a force

$$F = \frac{1}{2}\frac{Q^2}{\epsilon_0 A}$$

on the other, by calculating dW/dx where dW is the work needed to increase the separation by dx. (*b*) Why does using $F = QE$, with E being the electric field between the plates, give the wrong answer?

50. (II) A cylindrical capacitor (Example 24–2) has $R_a = 3.5$ mm and $R_b = 0.50$ mm. The two conductors have a potential difference of 625 V, with the inner conductor at the higher potential. (*a*) Calculate the energy stored in a 1.0-m length of the capacitor. (*b*) What is the linear charge density (C/m) on the capacitor's conductors? (*c*) Calculate the electric field at the surface of the inner conductor.

51. (III) Determine the energy stored per meter of capacitor length, for the capacitor of Problem 50, by integrating the energy density from R_b to R_a.

52. (III) Show that the electrostatic energy stored in the electric field outside an isolated spherical conductor of radius r_0 carrying a net charge Q is

$$U = \frac{1}{8\pi\epsilon_0}\frac{Q^2}{r_0}.$$

Do this in three ways: (*a*) Use Eq. 24–6 for the energy density in an electric field [*Hint*: Consider spherical shells of thickness dr]; (*b*) use Eq. 24–5 together with the capacitance of an isolated sphere (Section 24–2); (*c*) by calculating the work needed to bring all the charge Q up from infinity in infinitesimal bits dq.

24–5 Dielectrics

53. (I) What is the capacitance of a pair of circular plates with a radius of 5.0 cm separated by 2.3 mm of mica?

54. (I) What is the capacitance of two square parallel plates 4.7 cm on a side that are separated by 1.8 mm of paraffin?

55. (II) Suppose the capacitor in Example 24–13 remains connected to the battery as the dielectric is removed. What will be the work required to remove the dielectric in this case?

56. (II) How much energy would be stored in the capacitor of Problem 41 if a mica dielectric is placed between the plates? Assume the mica is 1.5 mm thick (and therefore fills the space between the plates).

57. (II) A 3500-pF air-gap capacitor is connected to an 18-V battery. If a piece of mica fills the space between the plates, how much charge will flow from the battery?

58. (II) Two different dielectrics each fill half the space between the plates of a parallel-plate capacitor as shown in Fig. 24–30. Determine a formula for the capacitance in terms of K_1, K_2, the area A of the plates, and the separation d. [*Hint*: Can you consider this capacitor as two capacitors in series or in parallel?]

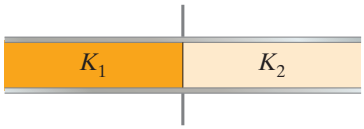


FIGURE 24–30 Problem 58.

59. (II) Two different dielectrics fill the space between the plates of a parallel-plate capacitor as shown in Fig. 24–31. Determine a formula for the capacitance in terms of K_1, K_2, the area A of the plates, and the separation $d_1 = d_2 = d/2$. [*Hint*: Can you consider this capacitor as two capacitors in series or in parallel?]

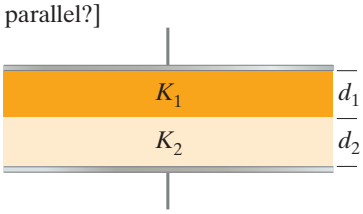


FIGURE 24–31 Problems 59 and 60.

60. (II) Repeat Problem 59 (Fig. 24–31) but assume the separation $d_1 \neq d_2$.

61. (II) Two identical capacitors are connected in parallel and each acquires a charge Q_0 when connected to a source of voltage V_0. The voltage source is disconnected and then a dielectric $(K = 3.6)$ is inserted to fill the space between the plates of one of the capacitors. Determine (*a*) the charge now on each capacitor, and (*b*) the voltage now across each capacitor.

62. (II) An uncharged capacitor is connected to an 18.0-V battery until it is fully charged, after which it is disconnected from the battery. A slab of paraffin is then inserted between the plates. What will now be the voltage across the plates?

63. (II) The electric field between the plates of a paper-separated $(K = 3.75)$ capacitor is 8.24×10^4 V/m. The plates are 1.95 mm apart, and the charge on each is 0.765 μC. Determine the capacitance of this capacitor and the area of each plate.

64. (III) A slab of width d and dielectric constant K is inserted a distance x into the space between the square parallel plates (of side ℓ) of a capacitor as shown in Fig. 24–32. Determine, as a function of x, (*a*) the capacitance, (*b*) the energy stored if the potential difference is V_0, and (*c*) the magnitude and direction of the force exerted on the slab (assume V_0 is constant).

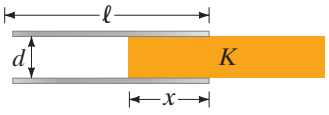


FIGURE 24–32 Problem 64.

65. (III) The quantity of liquid (such as cryogenic liquid nitrogen) available in its storage tank is often monitored by a capacitive level sensor. This sensor is a vertically aligned cylindrical capacitor with outer and inner conductor radii R_a and R_b, whose length ℓ spans the height of the tank. When a nonconducting liquid fills the tank to a height h $(\leq \ell)$ from the tank's bottom, the dielectric in the lower and upper regions between the cylindrical conductors is the liquid (K_{liq}) and its vapor (K_V), respectively (Fig. 24–33). (*a*) Determine a formula for the fraction F of the tank filled by liquid in terms of the level-sensor capacitance C. [*Hint*: Consider the sensor as a combination of two capacitors.] (*b*) By connecting a capacitance-measuring instrument to the level sensor, F can be monitored. Assume the sensor dimensions are $\ell = 2.0$ m, $R_a = 5.0$ mm, and $R_b = 4.5$ mm. For liquid nitrogen $(K_{liq} = 1.4,\ K_V = 1.0)$, what values of C (in pF) will correspond to the tank being completely full and completely empty?

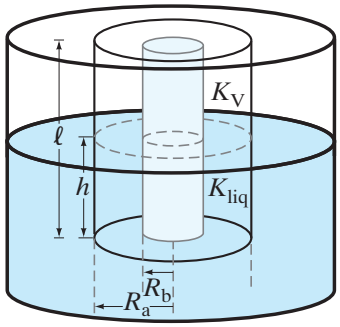


FIGURE 24–33
Problem 65.

*24–6 Molecular Description of Dielectrics

***66.** (II) In Example 24–14 what percent of the stored energy is stored in the electric field in the dielectric?

***67.** (II) Show that the capacitor in Example 24–14 with dielectric inserted can be considered as equivalent to three capacitors in series, and using this assumption show that the same value for the capacitance is obtained as was obtained in part (*g*) of the Example.

***68.** (II) Repeat Example 24–14 assuming the battery remains connected when the dielectric is inserted. Also, what is the free charge on the plates after the dielectric is added (let this be part (*h*) of this Problem)?

***69.** (II) Using Example 24–14 as a model, derive a formula for the capacitance of a parallel-plate capacitor whose plates have area A, separation d, with a dielectric of dielectric constant K and thickness ℓ $(\ell < d)$ placed between the plates.

***70.** (III) The capacitor shown in Fig. 24–34 is connected to an 80.0-V battery. Calculate (and sketch) the electric field everywhere between the capacitor plates. Find both the free charge on each capacitor plate and the induced charge on the faces of the glass dielectric plate.

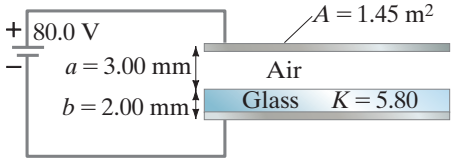


FIGURE 24–34 Problem 70.

General Problems

71. A homemade capacitor is assembled by placing two 9-in. pie pans 6 cm apart and connecting them to the opposite terminals of a 9-V battery. Estimate (*a*) the capacitance, (*b*) the charge on each plate, (*c*) the electric field halfway between the plates, and (*d*) the work done by the battery to charge them. (*e*) Which of the above values change if a dielectric is inserted?

72. (*a*) A general rule for estimating the capacitance C of an isolated conducting sphere with radius r is C (in pF) $\approx r$ (in cm). That is, the numerical value of C in pF is about the same as the numerical value of the sphere's radius in cm. Justify this rule. (*b*) Modeling the human body as a 1-m-radius conducting sphere, use the given rule to estimate your body's capacitance. (*c*) While walking across a carpet, you acquire an excess "static electricity" charge Q and produce a 0.5-cm spark when reaching out to touch a metallic doorknob. The dielectric strength of air is 30 kV/cm. Use this information to estimate Q (in μC).

73. A huge 4.0-F capacitor has enough stored energy to heat 2.4 kg of water from 21°C to 95°C. What is the potential difference across the plates?

74. How does the energy stored in an isolated capacitor change if (*a*) the potential difference is doubled, or (*b*) the separation of the plates is doubled?

75. By what factor does the energy stored in a capacitor change if (*a*) the potential difference is doubled, or (*b*) the charge on each plate is doubled? (*c*) What happens to the energy stored if the separation of the plates is doubled, as the capacitor remains connected to a battery?

76. A **coaxial cable**, Fig. 24–35, consists of an inner cylindrical conducting wire of radius R_b surrounded by a dielectric insulator. Surrounding the dielectric insulator is an outer conductor of radius R_a, which is usually "grounded." (*a*) Determine an expression for the capacitance per unit length of a cable whose insulator has dielectric constant K. (*b*) For a given cable, $R_b = 2.5$ mm and $R_a = 9.0$ mm. The dielectric constant of the dielectric insulator is $K = 3.2$. Suppose that there is a potential of 1.0 kV between the inner conducting wire and the outer conducting sheath. Find the charge, and capacitance, per meter of the cable.

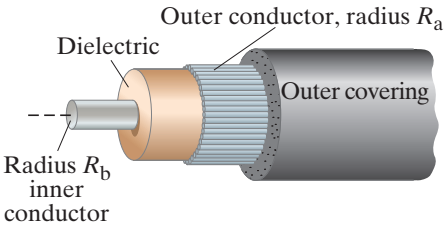


FIGURE 24–35 Problem 76.

77. A 2.1-μF capacitor is fully charged by a 9.0-V battery. The battery is then disconnected. The capacitor is not ideal and the charge slowly leaks out from the plates. The next day, the capacitor has lost half its stored energy. Calculate the amount of charge lost.

78. A capacitor is made from two 1.1-cm-diameter coins separated by a 0.10-mm-thick piece of paper ($K = 3.7$). A 12-V battery is connected to the capacitor. How much charge is on each coin?

79. New compact **ultracapacitors** can have capacitance up to several thousand farads. One application is in providing power for electrical circuits when other sources (such as a battery) are turned off. To get an idea of how much charge can be stored in such a capacitor, assume a 1200-F ultracapacitor is initially charged to 12.0 V by a battery and is then disconnected from the battery. If charge is then drawn off the plates of this capacitor at a rate of 1.0 mC/s to power the backup memory of some electrical device, how long (in days) will it take for the potential difference across this capacitor to drop to 6.0 V?

80. A parallel-plate capacitor is isolated with a charge $\pm Q$ on each plate. If the separation of the plates is halved and a dielectric (constant K) is inserted in place of air, by what factor does the energy storage change? To what do you attribute the change in stored potential energy? How does the new value of the electric field between the plates compare with the original value?

81. A multilayer "film" capacitor has a maximum voltage rating of 100 V and a capacitance of 1.0 μF. It is made from alternating sheets of very thin metal foil connected together, separated by films of polyester dielectric. The sheets are 12.0 mm by 14.0 mm and the total thickness of the capacitor is 6.0 mm (not counting the thickness of the insulator on the outside). The metal foils are actually very thin layers of metal deposited directly on the dielectric, so most of the thickness of the capacitor is due to the dielectric. The dielectric strength of the polyester is about 30×10^6 V/m. Estimate the dielectric constant of the polyester material in the capacitor.

82. A 3.5-μF capacitor is charged by an 11.2-V battery and then is disconnected from the battery. When this capacitor (C_1) is then connected to a second (initially uncharged) capacitor, C_2, the voltage on the first drops to 5.9 V. What is the value of C_2?

83. The power supply for a pulsed nitrogen laser has a 0.080-μF capacitor with a maximum voltage rating of 25 kV. (*a*) Estimate how much energy could be stored in this capacitor. (*b*) If 15% of this stored electrical energy is converted to light energy in a pulse that is 4.0 μs long, what is the power of the laser pulse?

84. A parallel-plate capacitor has square plates 12 cm on a side separated by 0.10 mm of plastic with a dielectric constant of $K = 3.8$. The plates are connected to a battery, causing them to become oppositely charged. Since the oppositely charged plates attract each other, they exert a pressure on the dielectric. If this pressure is 40.0 Pa, what is the battery voltage?

85. A parallel-plate capacitor with plate area 2.0 cm^2 and air-gap separation 0.50 mm is connected to a 12-V battery, and fully charged. The battery is then disconnected. (*a*) What is the charge on the capacitor? (*b*) The plates are now pulled to a separation of 0.85 mm. What is the charge on the capacitor now? (*c*) What is the potential difference across the plates now? (*d*) How much work was required to pull the plates to their new separation?

86. A 155-pF capacitor is connected in series with an unknown capacitor, and as a series combination they are connected to a 25.0-V battery. If the 155-pF capacitor stores 125 pC of charge on its plates, what is the unknown capacitance?

87. The *variable capacitance* of an old radio tuner consists of four plates connected together placed alternately between four other plates, also connected together (Fig. 24–36). Each plate is separated from its neighbor by 1.6 mm of air. One set of plates can move so that the area of overlap of each plate varies from $2.0\,\text{cm}^2$ to $9.0\,\text{cm}^2$. (*a*) Are these seven capacitors connected in series or in parallel? (*b*) Determine the range of capacitance values.

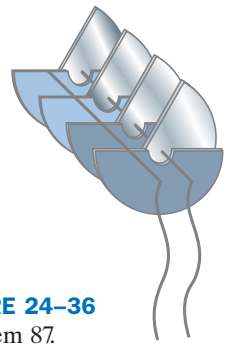


FIGURE 24–36
Problem 87.

88. The long cylindrical capacitor shown in Fig. 24–37 consists of four concentric cylinders, with respective radii R_a, R_b, R_c, and R_d. The cylinders b and c are joined by metal strips. Determine the capacitance per unit length of this arrangement. (Assume equal and opposite charges are placed on the innermost and outermost cylinders.)

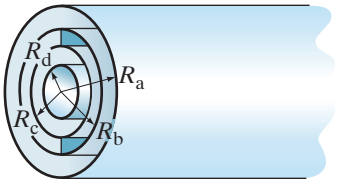


FIGURE 24–37
Problem 88.

89. A parallel-plate capacitor has plate area A, plate separation x, and has a charge Q stored on its plates (Fig. 24–38). (*a*) Determine the work required to double the plate separation to $2x$, assuming the charge remains constant at Q. (*Hint*: See Example 24–10.) (*b*) Show that your answer is consistent with the change in energy stored by the capacitor.

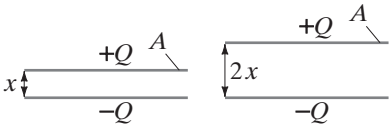


FIGURE 24–38
Problem 89.

90. Let us try to estimate the maximum "static electricity" charge that might result during each walking step across a non-metal floor. Assume the sole of a person's shoe has area $A \approx 150\,\text{cm}^2$, and when the foot is lifted from the ground during each step, the sole acquires an excess charge Q from rubbing contact with the floor. (*a*) Model the sole as a plane conducting surface with Q uniformly distributed across it as the foot is lifted from the ground. Given the dielectric strength of the air between the sole and floor as the foot is lifted is $E_S = 3 \times 10^6\,\text{N/C}$, determine Q_{max}, the maximum possible excess charge that can be transferred to the sole during each step. (*b*) Modeling a person as an isolated conducting sphere of radius $r \approx 1\,\text{m}$, estimate a person's capacitance. (*c*) After lifting the foot from the floor, assume the excess charge Q quickly redistributes itself over the entire surface area of the "spherical" person. Estimate the maximum potential difference that the person can develop with respect to the floor.

91. A parallel-plate capacitor with plate area $A = 2.0\,\text{m}^2$ and plate separation $d = 3.0\,\text{mm}$ is connected to a 45-V battery (Fig. 24–39a). (*a*) Determine the charge on the capacitor, the electric field, the capacitance, and the energy U_0 stored in the capacitor. (*b*) With the capacitor still connected to the battery, a slab of plastic with dielectric constant $K = 3.2$ is placed between the plates of the capacitor, so that the gap is completely filled with the dielectric. What are the new values of charge, electric field, capacitance, and the energy U stored in the capacitor?

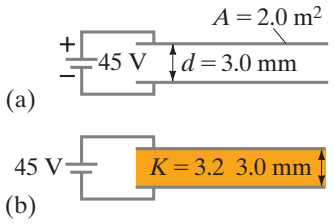


FIGURE 24–39
Problem 91.

92. Consider the use of capacitors as memory cells. A charged capacitor would represent a one and an uncharged capacitor a zero. Suppose these capacitors were fabricated on a silicon chip and each has a capacitance of 30 femto-farads $(1\,\text{fF} = 10^{-15}\,\text{F}.)$ The dielectric filling the space between the parallel plates has dielectric constant $K = 15$ and a dielectric strength of $1.0 \times 10^9\,\text{V/m}$. (*a*) If the operating voltage is 1.5 V, how many electrons would be stored on one of these capacitors when charged? (*b*) If no safety factor is allowed, how thin a dielectric layer could we use for operation at 1.5 V? (*c*) Using the layer thickness from your answer to part (*b*), what would be the area of the capacitor plates?

93. The potential energy stored in a capacitor can be written as either $CV^2/2$ or $Q^2/2C$. In the first case the energy is proportional to C; in the second case the energy is proportional to $1/C$. (*a*) Explain how both of these equations can be correct. (*b*) When might you use the first equation and when might you use the second equation? (*c*) If a paper dielectric is inserted into a parallel-plate capacitor that is attached to a battery (V does not change), by what factor will the energy stored in the capacitor change? (*d*) If a quartz dielectric is inserted into a charged parallel-plate capacitor that is isolated from any battery, by what factor will the energy stored in the capacitor change?

94. A student wearing shoes with thin insulating soles is standing on a grounded metal floor when he puts his hand on the metal case of a high voltage supply. The voltage inside, 6.3 mm from his hand, is 25,000 V. The student's hand and the voltage supply form a capacitor, considering the student as a conductor. Another capacitor is between the floor and the student (his feet). Using reasonable numbers for hand and foot areas, estimate the student's voltage relative to the floor. Assume vinyl-soled shoes 1 cm thick.

95. Capacitors can be used as "electric charge counters." Consider an initially uncharged capacitor of capacitance C with its bottom plate grounded and its top plate connected to a source of electrons. (*a*) If N electrons flow onto the capacitor's top plate, show that the resulting potential difference V across the capacitor is directly proportional to N. (*b*) Assume a voltage-measuring device can accurately resolve voltage changes of about 1 mV. What value of C would be necessary to resolve the arrival of an individual electron? (*c*) Using semiconductor technology, a micron-size capacitor can be constructed with parallel conducting plates separated by an insulator of dielectric constant $K = 3$ and thickness $d = 100\,\text{nm}$. What side length ℓ should the square plates have (in μm)?

96. In the **dynamic random access memory (DRAM)** of a cell phone, each memory cell contains a capacitor for charge storage. Each of these cells represents a single binary-bit value of "1" when its 25-fF capacitor $(1\,\text{fF} = 10^{-15}\,\text{F})$ is charged at 0.6 V, or "0" when uncharged at 0 V. (*a*) When fully charged, how many excess electrons are on a cell capacitor's negative plate? (*b*) After charge has been placed on a cell capacitor's plate, it slowly "leaks" off at a rate of about 0.5 pA. How long does it take for the potential difference across this capacitor to decrease by 2.0% from its fully charged value? (Because of this leakage effect, the charge on a DRAM capacitor is "refreshed" many times per second.) Note: A DRAM cell is shown in Fig. 29–31.

97. In the DRAM computer chip of Problem 96, suppose the two parallel plates of one cell's 25-fF capacitor are separated by a 10-nm-thick insulating material with dielectric constant $K = 25$. (*a*) Determine the area A $(\text{in } \mu\text{m}^2)$ of the cell capacitor's plates. (*b*) If the plate area A accounts for half of the area of each cell, estimate how many megabytes of memory can be placed on a 3.0-cm^2 silicon wafer. (1 byte = 8 bits.)

98. In lightning storms, the potential difference between the Earth and the bottom of the thunderclouds can be as high as 35,000,000 V. The bottoms of thunderclouds are typically 1500 m above the Earth, and can have an area of 110 km^2. Modeling the Earth–cloud system as a huge capacitor, calculate (*a*) the capacitance of the Earth–cloud system, (*b*) the charge stored in the "capacitor," and (*c*) the energy stored in the "capacitor."

99. Suppose it takes 75 kW of power for your car to travel at a constant speed on the highway. (*a*) What is this in horsepower? (*b*) How much energy in joules would it take for your car to travel at highway speed for 4.0 hours? (*c*) Suppose this amount of energy is to be stored in the electric field of a parallel-plate capacitor. If the voltage on the capacitor is to be 850 V, what is the required capacitance? (*d*) If this capacitor were to be made from activated carbon (Section 24–2), the voltage would be limited to no more than 10 V. In this case, how many grams of activated carbon would be required? (*e*) Is this practical?

100. Capacitors can be recharged many thousands of times with little degradation, so consider why electric cars don't use a bank of capacitors rather than a bank of batteries. (*a*) Estimate how large a capacitance would be needed to store the energy equivalent of 10 gallons of gasoline $(\approx 10^9\,\text{J})$, assuming 12 V. (*b*) Physically how large would this capacitor be if it was a conventional parallel-plate capacitor using the best dielectric in Table 24–1? (*c*) If this capacitor was 0.1 mm thick and cut and stacked into 1 m^2 sheets to make it more compact, could it make a practical replacement for the battery of an electric car? (*d*) If the capacitor instead used the activated carbon capacitor material described in Section 24–2, how heavy would the capacitor be?

101. Paper has a dielectric constant $K = 3.7$ and a dielectric strength of $15 \times 10^6\,\text{V/m}$. Suppose that a typical sheet of paper has a thickness of 0.11 mm. You make a "homemade" capacitor by placing a sheet of 21 cm × 14 cm paper between two aluminum foil sheets (Fig. 24–40) of the same size. (*a*) What is the capacitance C_0 of your device? (*b*) About how much charge could you store on your capacitor before it would break down? (*c*) Show in a sketch how you could overlay sheets of paper and aluminum for a parallel combination. (*d*) Suppose you made 100 such capacitors by connecting the edges of alternate sheets of 0.016-mm-thick aluminum foil to form capacitors in parallel. How thick would this single large capacitor of capacitance 100 C_0 be? (*e*) What is the maximum voltage you can apply to this 100-C_0 capacitor without breakdown?

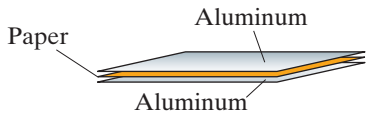


FIGURE 24–40 Problem 101.

ANSWERS TO EXERCISES

A: A.
B: 8.3×10^{-9} C.
C: (*a*).
D: (*e*).
E: (*b*).
F: 42.9 V; 387 pF.



The glow of the wire heating element of an electric heater (in a mountain restaurant) is caused by the electric current passing through it. Electric energy is transformed to thermal energy (via collisions between moving electrons and atoms of the wire), which causes the wire's temperature to become so high that it glows.

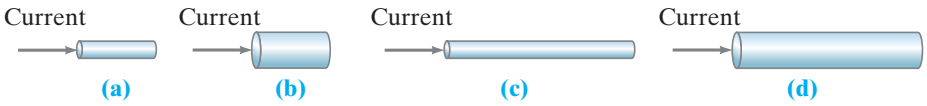
Electric current and electric power in electric circuits are of basic importance in everyday life. We examine both dc and ac in this Chapter, and include the microscopic analysis of electric current.

CHAPTER 25

Electric Current and Resistance

CHAPTER-OPENING QUESTION—Guess now!

The conductors shown below are all made of copper and are at the same temperature. Which conductor would have the greatest resistance to the flow of charge entering from the left? Which would offer the least resistance?



In the previous four Chapters we have been studying static electricity: electric charges at rest. In this Chapter we begin our study of charges in motion, and we call a flow of charge an electric current.

In everyday life we are familiar with electric currents in wires and other conductors. Most practical electrical devices depend on electric current: current through a lightbulb, current in the heating element of a stove, hair dryer, or electric heater, as well as currents in electronic devices. Electric currents can exist in conductors such as wires, but also in semiconductor devices, human cells and their membranes (Section 25–10), and in empty space.

CONTENTS

25 – 1 The Electric Battery
25 – 2 Electric Current
25 – 3 Ohm's Law: Resistance and Resistors
25 – 4 Resistivity
25 – 5 Electric Power
25 – 6 Power in Household Circuits
25 – 7 Alternating Current
25 – 8 Microscopic View of Electric Current
*25 – 9 Superconductivity
*25–10 Electrical Conduction in the Human Nervous System



FIGURE 25–1 Alessandro Volta. In this portrait, Volta demonstrates his battery to Napoleon in 1801.

In electrostatic situations, we saw in Sections 21–9 and 22–3 that the electric field must be zero inside a conductor (if it weren't, the charges would move). But when charges *are moving* in a conductor, an electric field is needed to set charges into motion, and to keep them in motion against resistance in any normal conductor. We can control the flow of charge using electric fields and electric potential (voltage), concepts we have just been discussing. In order to have a current in a wire, a potential difference is needed, which can be provided by a battery.

We first look at electric current from a macroscopic point of view: that is, current as measured in a laboratory. Later in this Chapter we look at currents from a microscopic (theoretical) point of view as a flow of electrons in a wire.

Until the year 1800, the technical development of electricity consisted mainly of producing a static charge by friction. It all changed in 1800 when Alessandro Volta (1745–1827; Fig. 25–1) invented the electric battery, and with it produced the first steady flow of electric charge—that is, a steady electric current.

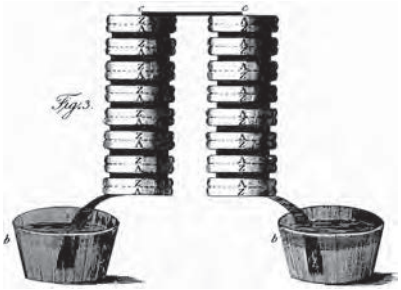
25–1 The Electric Battery

The events that led to the discovery of the battery are interesting. Not only was this an important discovery, but it also gave rise to a famous scientific debate.

In the 1780s, Luigi Galvani (1737–1798), professor at the University of Bologna, carried out a series of experiments on the contraction of a frog's leg muscle by using static electricity. Galvani found that the muscle also contracted when dissimilar metals were inserted into the frog. Galvani believed that the source of the electric charge was in the frog muscle or nerve itself, and that the metal merely transmitted the charge to the proper points. When he published his work in 1791, he termed this charge "animal electricity." Many wondered, including Galvani himself, if he had discovered the long-sought "life-force."

Volta, at the University of Pavia 200 km away, was skeptical of Galvani's results, and came to believe that the source of the electricity was not in the animal itself, but rather in the *contact between the dissimilar metals*. Volta realized that a moist conductor, such as a frog muscle or moisture at the contact point of two dissimilar metals, was necessary in the circuit if it was to be effective. He also saw that the contracting frog muscle was a sensitive instrument for detecting electric "tension" or "electromotive force" (his words for what we now call potential), in fact more sensitive than the best available electroscopes (Section 21–4) that he and others had developed.[†]

FIGURE 25–2 A voltaic battery, from Volta's original publication.



Volta's research found that certain combinations of metals produced a greater effect than others, and, using his measurements, he listed them in order of effectiveness. (This "electrochemical series" is still used by chemists today.) He also found that carbon could be used in place of one of the metals.

Volta then conceived his greatest contribution to science. Between a disc of zinc and one of silver, he placed a piece of cloth or paper soaked in salt solution or dilute acid and piled a battery of such couplings, one on top of another, as shown in Fig. 25–2. This "pile" or "battery" produced a much increased potential difference. Indeed, when strips of metal connected to the two ends of the pile were brought close, a spark was produced. Volta had designed and built the first **electric battery**. He published his discovery in 1800.

Electric Cells and Batteries

A battery produces electricity by transforming chemical energy into electrical energy. Today a great variety of electric cells and batteries are available, from flashlight batteries to the starter batteries of gasoline cars, from those used in cell phones to those for energy storage in electric vehicles.

The simplest batteries contain two plates or rods made of dissimilar metals (one can be carbon) called **electrodes**. The electrodes are immersed in a solution or paste, such as a dilute acid, called the **electrolyte**. Such a device is properly called an **electric cell**, and

[†]Volta's most sensitive electroscope (Section 21–4) measured about 40 V per degree (angle of leaf separation). Nonetheless, he was able to estimate the potential differences produced by combinations of dissimilar metals in contact. For a silver–zinc contact he got about 0.7 V, remarkably close to today's value of 0.78 V.

several cells connected together is a **battery**, although today even a single cell is called a battery. The chemical reactions involved in most electric cells are quite complicated. Here we describe how one very simple cell works, emphasizing the physical aspects.

The cell shown in Fig. 25–3 uses dilute sulfuric acid as the electrolyte. One of the electrodes is made of carbon, the other of zinc. The part of each electrode outside the solution is called the **terminal**, and connections to wires and circuits are made here. The acid tends to dissolve the zinc electrode. Each zinc atom leaves two electrons behind on the electrode and enters the solution as a positive ion. The zinc electrode thus acquires a negative charge. The electrolyte becomes positively charged, and can pull electrons off the carbon electrode. Thus the carbon electrode becomes positively charged. Because there is an opposite charge on the two electrodes, there is a potential difference between the two terminals. The positive electrode is called the **anode**, the negative one the **cathode**.

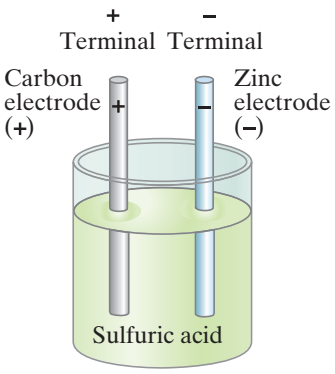


FIGURE 25–3 Simple electric cell.

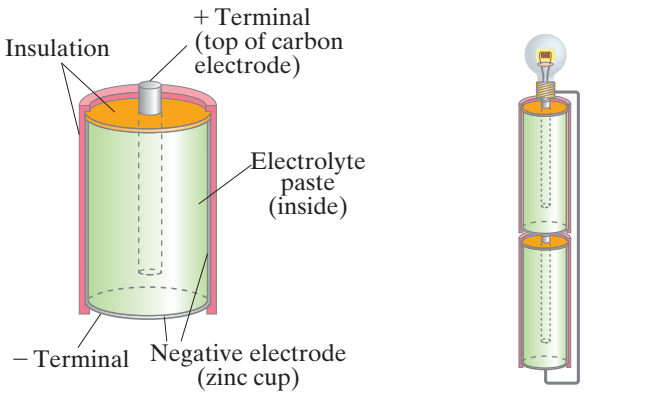
In a cell whose terminals are not connected, only a small amount of the zinc is dissolved, for as the zinc electrode becomes increasingly negative, any new positive zinc ions produced are attracted back to the electrode. Thus, a particular potential difference (or voltage) is maintained between the two terminals. If charge is allowed to flow between the terminals, say, through a wire (or a toaster heating element), then more zinc can be dissolved. After a time, one or the other electrode is used up and the cell becomes "dead."

The voltage that exists between the terminals of a battery depends on what the electrodes are made of and their relative ability to be dissolved or give up electrons.

When two or more cells are connected so that the positive terminal of one is connected to the negative terminal of the next, they are said to be connected in *series* and their voltages add up. Thus, the voltage between the ends of two 1.5-V AA flashlight batteries connected in series is 3.0 V, whereas the six 2-V cells of an ordinary car battery give 12 V. Figure 25–4a shows a diagram of a common "dry cell" or "flashlight battery" used in flashlights and many portable electronic devices, and Fig. 25–4b shows two batteries (AA or AAA) connected in series to a flashlight bulb.

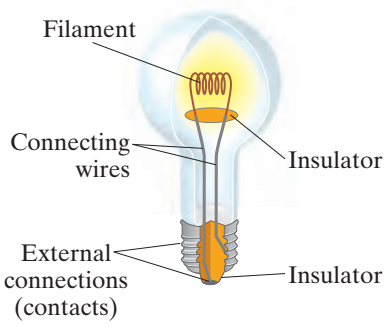
The simple incandescent lightbulb in Fig. 25–4b consists of a thin, coiled wire (filament) inside an evacuated glass bulb, as shown in Fig. 25–5. When charge passes through the filament, it gets very hot ($\approx 2700\,\text{K}$) and glows. Halogen bulbs are incandescent with the filament enclosed in a capsule containing a halogen such as bromine or iodine; they last longer and can glow at 2900 K giving better efficiency. [Halogens, via chemical reactions, restore many of the tungsten atoms that were evaporated from the hot filament, allowing longer life. The higher temperature (typically 2900 K) gives whiter light, closer to what the Sun produces, around 6000 K. Color vs. temperature is discussed in Section 37–1 on blackbody radiation.] Other bulb types, such as fluorescent (Chapter 39) and LED (Chapter 40) work in a more complicated way.

FIGURE 25–4 (a) Diagram of an ordinary dry cell (like a D-cell or AA). The cylindrical zinc cup is covered on the sides by an insulator (pink here); its flat bottom is the negative terminal. (b) Two dry cells (AA type) connected in series. Note that the positive terminal of one cell pushes against the negative terminal of the other.



(a) One D-cell (b) Two AA batteries

FIGURE 25–5 A plain incandescent lightbulb: the fine wire of the filament becomes so hot that it glows. Incandescent *halogen* bulbs enclose the filament in a small quartz tube filled with a halogen gas (bromine or iodine) which allows longer filament life and higher filament temperature for greater efficiency and whiteness.



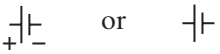
Electric Cars

PHYSICS APPLIED
Electric cars

Considerable research is being done to improve batteries for electric cars and for hybrids (which use both a gasoline internal combustion engine and an electric motor). One type of battery is lithium-ion, in which the anode contains lithium and the cathode is carbon. Electric cars need no gear changes and can develop full torque starting from rest, and so can accelerate quickly and smoothly. The distance an electric car can go between charges of the battery (its "range") is an important parameter because each recharging of an electric car battery may take hours, not minutes like a gas fill-up. Because charging an electric car can draw a large current over a period of several hours, electric power companies need upgraded power grids so they won't fail when many electric cars are being charged at the same time in any given urban area.

25–2 Electric Current

The purpose of a battery is to produce a potential difference, which can then make charges move. When a continuous conducting path, such as a wire and some electrical device, is connected between the terminals of a battery, we have an electric **circuit**, Fig. 25–6a. On any diagram of a circuit, as in Fig. 25–6b, we use the symbol



or [battery symbol]

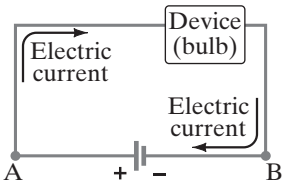
to represent a battery. A device connected to the battery could be a lightbulb, a heater, a radio, or some other device. When such a circuit is formed, charge can move (or flow) through the wires of the circuit, from one terminal of the battery to the other, as long as the conducting path is continuous. Any flow of charge such as this is called an **electric current**.

FIGURE 25–6 (a) A simple electric circuit. (b) Schematic drawing of the same circuit, consisting of a battery, connecting wires (thick gray lines), and a lightbulb or other device.

(a)



(b)



More precisely, the electric current in a wire is defined as the net amount of charge that passes through the wire's full cross section at any point per unit time. Thus, the average current $\bar{I}$ is defined as

$$\bar{I} = \frac{\Delta Q}{\Delta t}, \qquad \textbf{(25–1a)}$$

where ΔQ is the amount of charge that passes through the conductor at any location during the time interval Δt. The instantaneous current is defined by the derivative limit

$$I = \frac{dQ}{dt}. \qquad \textbf{(25–1b)}$$

Electric current is measured in coulombs per second; this is given a special name, the **ampere** (abbreviated amp or A), after the French physicist André Ampère (1775–1836). Thus, $1\ \text{A} = 1\ \text{C/s}$. Smaller units of current are often used, such as the milliampere $(1\ \text{mA} = 10^{-3}\ \text{A})$ and microampere $(1\ \mu\text{A} = 10^{-6}\ \text{A})$.

A current can flow in a circuit only if there is a *continuous* conducting path. We then have a **complete circuit**. If there is a break in the circuit, say, a cut wire, we call it an **open circuit** and no current flows. In any single circuit, with only a single path for current to follow such as in Fig. 25–6b, a steady current at any instant is the same at one point (say, point A) as at any other point (such as B). This follows from the conservation of electric charge: charge doesn't disappear. A battery does not create (or destroy) any net charge, nor does a lightbulb absorb or destroy charge.

CAUTION
A battery does not create charge; a lightbulb does not destroy charge

EXAMPLE 25–1 **Current is flow of charge.** A steady current of 2.5 A exists in a wire for 4.0 min. (*a*) How much total charge passes by a given point in the circuit during those 4.0 min? (*b*) How many electrons would this be?

APPROACH (*a*) Current is flow of charge per unit time, Eqs. 25–1, so the amount of charge passing a point is the product of the current and the time interval. (*b*) To get the number of electrons, we divide the total charge by the charge on one electron.

SOLUTION (*a*) Since the current was 2.5 A, or 2.5 C/s, then in 4.0 min $(= 240\ \text{s})$ the total charge that flowed past a given point in the wire was, from Eq. 25–1a,

$$\Delta Q = I\,\Delta t = (2.5\ \text{C/s})(240\ \text{s}) = 600\ \text{C}.$$

(*b*) The charge on one electron is $1.60 \times 10^{-19}\ \text{C}$, so 600 C would consist of

$$\frac{600\ \text{C}}{1.6 \times 10^{-19}\ \text{C/electron}} = 3.8 \times 10^{21}\ \text{electrons}.$$

EXERCISE A If 1 million electrons per second pass a point in a wire, what is the current?

CONCEPTUAL EXAMPLE 25–2 **How to connect a battery.** What is wrong with each of the schemes shown in Fig. 25–7 for lighting a flashlight bulb with a flashlight battery and a single wire?

RESPONSE (*a*) There is no closed path for charge to flow around. Charges might briefly start to flow from the battery toward the lightbulb, but there they run into a "dead end," and the flow would immediately come to a stop.

(*b*) Now there is a closed path passing to and from the lightbulb; but the wire touches only one battery terminal, so there is no potential difference in the circuit to make the charge move. Neither here, nor in (*a*), does the bulb light up.

(*c*) Nothing is wrong here. This is a complete circuit: charge can flow out from one terminal of the battery, through the wire and the bulb, and into the other terminal. This scheme will light the bulb.

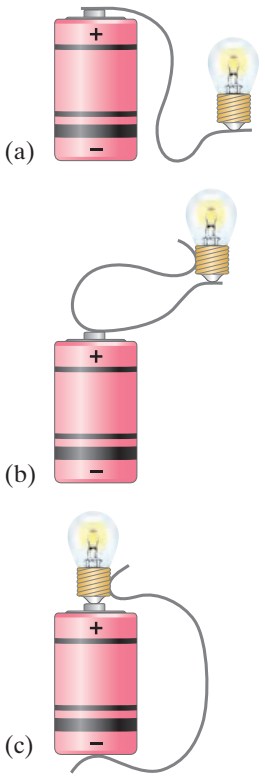


FIGURE 25–7 Example 25–2.

In many real circuits, wires are connected to a common conductor that provides continuity. This common conductor is called **ground**, usually represented as ⏚ or ⏚, and really is connected to the ground for a building or house. In a car, one terminal of the battery is called "ground," but is not connected to the earth itself—it is connected to the frame of the car, as is one connection to each lightbulb and other devices. Thus the car frame is a conductor in each circuit, ensuring a continuous path for charge flow, and is called "ground" for the car's circuits. (The car frame is well insulated from the earth by the rubber tires.)

We saw in Chapter 21 that conductors contain many free electrons. Thus, if a continuous conducting wire is connected to the terminals of a battery, negatively charged electrons flow in the wire. When the wire is first connected, the potential difference between the terminals of the battery sets up an electric field inside the wire[†] and parallel to it. Free electrons at one end of the wire are attracted into the positive terminal, and at the same time other electrons enter the other end of the wire at the negative terminal of the battery. There is a continuous flow of electrons throughout the wire that begins as soon as the wire is connected to *both* terminals.

When the conventions of positive and negative charge were invented two centuries ago, however, it was assumed that positive charge flowed in a wire. For nearly all purposes, positive charge flowing in one direction is exactly equivalent to negative charge flowing in the opposite direction,[‡] as shown in Fig. 25–8.

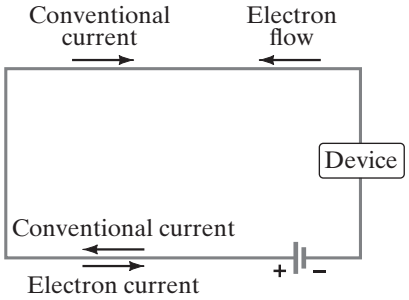


FIGURE 25–8 Conventional current from + to − is equivalent to a negative electron flow from − to +.

Today, we still use the historical convention of positive charge flow when discussing the direction of a current. So when we speak of the current direction in a circuit, we mean the direction *positive* charge would flow. This is sometimes referred to as **conventional current**. When we want to speak of the direction of electron flow, we will specifically state it is the electron current. In liquids and gases, both positive and negative charges (ions) can move.

CAUTION
Distinguish conventional current from electron flow

In practical life, such as rating the total charge available (positive or negative) in a car battery, you may see the unit **ampere-hour** (A · h): from Eq. 25–1a, $\Delta Q = I\,\Delta t$.

EXERCISE B How many coulombs is 1.00 A · h?

[†]This does not contradict what was said in Section 21–9 that in the *static* case, there can be no electric field within a conductor since otherwise the charges would move. Indeed, when there is an electric field in a conductor, charges do move, and we get an electric current.

[‡]An exception is discussed in Section 27–8, the Hall effect.

25–3 Ohm's Law: Resistance and Resistors

To produce an electric current in a circuit, a difference in potential is required. One way of producing a potential difference along a wire is to connect its ends to the opposite terminals of a battery. It was Georg Simon Ohm (1787–1854) who established experimentally that the current in a metal wire is proportional to the potential difference V applied to its two ends:

$$I \propto V.$$

If, for example, we connect a wire to the two terminals of a 6-V battery, the current in the wire will be twice what it would be if the wire were connected to a 3-V battery. It is also found that reversing the sign of the voltage does not affect the magnitude of the current.

Exactly how large the current is in a wire depends not only on the voltage between its ends, but also on the resistance the wire offers to the flow of electrons. We can think of electron flow as being impeded partly because of collisions with the atoms of the wire.[†] We define electrical **resistance** R as the proportionality factor between the voltage V (between the ends of the wire) and the current I (passing through the wire):

OHM'S "LAW"

$$V = IR. \qquad \textbf{(25–2)}$$

Ohm found experimentally that in metal conductors R is a constant independent of V, a result known as **Ohm's law**. Equation 25–2, $V = IR$, is itself sometimes called Ohm's law, but only when referring to materials or devices for which R is a constant independent of V. But R is not a constant for many substances other than metals, nor for devices such as diodes, vacuum tubes, transistors, and so on. Even for metals, R is not constant if the temperature changes much: for an incandescent lightbulb filament the measured resistance is low for small currents, but is much higher at the filament's normal large operating current that puts it at the high temperature needed to make it glow (≈ 3000 K). Thus Ohm's "law" is not a fundamental law of nature, but rather a description of a certain class of materials: metal conductors, whose temperature does not change much. Such materials are said to be "ohmic." Materials or devices that do not follow Ohm's law are said to be *nonohmic*. See Fig. 25–9.

The unit for resistance is called the **ohm** and is abbreviated Ω (Greek capital letter omega). Because $R = V/I$, we see that 1.0 Ω is equivalent to 1.0 V/A.

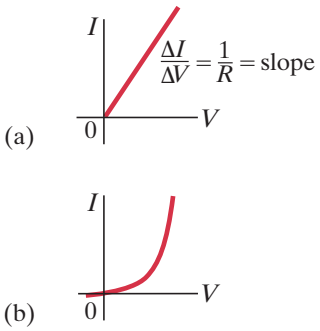
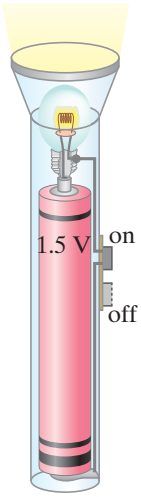


FIGURE 25–9 Graphs of current vs. voltage (a) for a metal conductor which obeys Ohm's law, and (b) for a nonohmic device, in this case a semiconductor diode.

FIGURE 25–10 Flashlight (Example 25–3). Note how the circuit is completed along the side strip. The on-off switch is shown in 2 positions. When "on" it connects the 2 wire contacts.



EXAMPLE 25–3 **Flashlight bulb resistance.** A small flashlight bulb (Fig. 25–10) draws 300 mA from its 1.5-V battery. (*a*) What is the resistance of the bulb? (*b*) If the battery becomes weak and the voltage drops to 1.2 V, how would the current change? Assume the bulb is approximately ohmic.

APPROACH We apply Ohm's law to the bulb, where the voltage applied across it is the battery voltage.

SOLUTION (*a*) We change 300 mA to 0.30 A and use Eq. 25–2:

$$R = \frac{V}{I} = \frac{1.5\,\text{V}}{0.30\,\text{A}} = 5.0\,\Omega.$$

(*b*) If the resistance stays the same, the current would be

$$I = \frac{V}{R} = \frac{1.2\,\text{V}}{5.0\,\Omega} = 0.24\,\text{A} = 240\,\text{mA},$$

or a decrease of 60 mA.

NOTE With the smaller current in part (*b*), the bulb filament's temperature would be lower and the bulb less bright. Also, resistance *does* depend on temperature (Section 25–4), so our calculation is only a rough approximation.

[†]A useful analogy compares the flow of electric charge in a wire to the flow of water in a river, or in a pipe, acted on by gravity. If the river (or pipe) is nearly level, the flow rate is small. But if one end is somewhat higher than the other, the water flow rate—or current—is greater. The greater the difference in height, the greater the current. We saw in Chapter 23 that electric potential is analogous to the height of a cliff for gravity. Just as an increase in height can cause a greater flow of water, so a greater electric potential difference, or voltage, causes a greater electric current. Resistance in a wire is analogous to rocks in a river that retard water flow.

EXERCISE C What resistance should be connected across a 9.0-V battery to make a 10-mA current? (*a*) 9 Ω, (*b*) 0.9 Ω, (*c*) 900 Ω, (*d*) 1.1 Ω, (*e*) 0.11 Ω.

All electric devices, from heaters to lightbulbs to ovens and electronic devices, offer resistance to the flow of current. The filaments of incandescent lightbulbs (Fig. 25–5), toasters, stove burners (Fig. 25–16), and electric heaters are special types of wires whose resistance results in their becoming very hot. Generally, the connecting wires have very low resistance in comparison to the resistance of the wire filaments or coils, so the connecting wires usually have a minimal effect on the magnitude of the current.

In many circuits, particularly in electronic devices, **resistors** are used to control the amount of current. Resistors have resistances ranging from less than an ohm to millions of ohms (see Figs. 25–11 and 25–12). The main types are "wire-wound" resistors which consist of a coil of fine wire, "composition" resistors which are usually made of carbon, resistors made of thin carbon or metal films, and (on tiny integrated circuit "chips") undoped semiconductors.

When we draw a diagram of a circuit, we use the symbol



[resistor symbol]

to indicate a resistance. Wires whose resistance is negligible, however, are shown simply as straight lines. Figure 25–12 and its Table show one way to specify the resistance of a resistor.

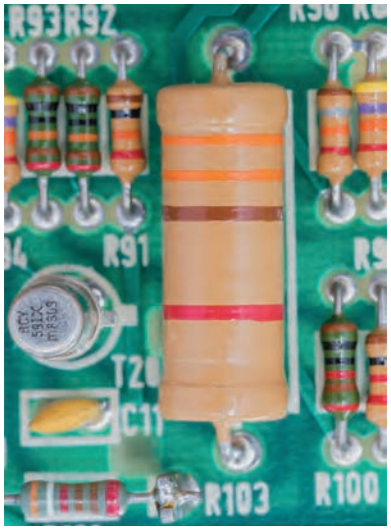


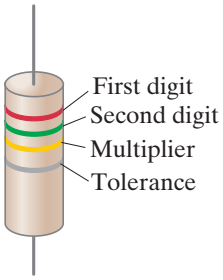
FIGURE 25–11 Photo of resistors with color-coded stripes representing resistance and tolerances. Most are $\frac{1}{4}$ watt (maximum $P = I^2R$ without damage). The large one can handle more current and power, perhaps 4 or 5 W.

Resistor Color Code

Color	Number	Multiplier	Tolerance
Black	0	1	
Brown	1	10^1	1%†
Red	2	10^2	2%†
Orange	3	10^3	
Yellow	4	10^4	
Green	5	10^5	
Blue	6	10^6	
Violet	7	10^7	
Gray	8	10^8	
White	9	10^9	
Gold		10^{-1}	5%
Silver		10^{-2}	10%
No color			20%

†Used on 5-line precision resistors.

FIGURE 25–12 The resistance value of a given resistor is written on the exterior, or may be given as a color code as shown below and in the Table to the left: for a 4-line code, the first two colors represent the first two digits in the value of the resistance, the third color represents the power of ten that it must be multiplied by, and the fourth is the manufactured tolerance. For example, a resistor whose four colors are red, green, yellow, and silver has a resistance of $25 \times 10^4\ \Omega = 250{,}000\ \Omega = 250\ \text{k}\Omega$, plus or minus 10%. [An alternative code is a number such as 104, which means $R = 1.0 \times 10^4\ \Omega$.] A 5 or 6 line color code is a precision resistor with an accuracy of 1% or 2% or even better. For these, the third line is the third significant figure, the fourth line is the multiplier, such as $452 \times 10^4\ \Omega$, and the fifth line is the tolerance. A sixth line represents temperature dependence of its R.



CONCEPTUAL EXAMPLE 25–4 **Current and potential.** Current I enters a resistor R as shown in Fig. 25–13. (*a*) Is the potential higher at point A or at point B? (*b*) Is the current greater at point A or at point B?

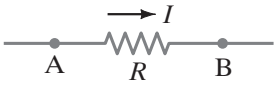


FIGURE 25–13 Example 25–4.

RESPONSE (*a*) Positive charge always flows from + to −, from high potential to low potential. So if current I is conventional (positive) current, point A is at a higher potential than point B.

(*b*) Conservation of charge requires that whatever charge flows into the resistor at point A, an equal amount of charge emerges at point B. Charge or current does not get "used up" by a resistor. So the current is the same at A and B.

An electric potential decrease, as from point A to point B in Example 25–4, is often called a **potential drop** or a **voltage drop**.

Some Helpful Clarifications

Here we briefly summarize some possible misunderstandings and clarifications. Batteries do not put out a constant current. Instead, batteries are intended to maintain a constant potential difference, or very nearly so. (Details in the next Chapter.) Thus a battery should be considered a source of voltage. The voltage is applied *across* a wire or device.

CAUTION
Voltage is applied across *a device; current passes* through *a device*

Electric current passes *through* a wire or device (connected to a battery), and its magnitude depends on that device's resistance. The resistance is a *property* of the wire or device. The voltage, on the other hand, is external to the wire or device, and is applied across the two ends of the wire or device. The current through the device might be called the "response": the current increases if the voltage increases or the resistance decreases, as $I = V/R$.

In a thin wire, the direction of the current is always parallel to the wire at each point, no matter how the wire curves, just like water in a pipe. The direction of conventional (positive) current is from high potential $(+)$ toward lower potential $(-)$.

CAUTION
Current is not consumed

Current and charge do not increase or decrease or get "used up" when going through a wire or other device. The amount of charge that goes in at one end comes out at the other end.

25–4 Resistivity

It is found experimentally that the resistance R of any wire is directly proportional to its length ℓ and inversely proportional to its cross-sectional area A. That is,

$$R = \rho \frac{\ell}{A}, \tag{25–3}$$

where ρ, the constant of proportionality, is called the **resistivity** and depends on the material used. Typical values of ρ, whose units are $\Omega \cdot \text{m}$ (see Eq. 25–3), are given for various materials in the middle column of Table 25–1, which is divided into the categories *conductors*, *insulators*, and *semiconductors* (Section 21–3). The values depend somewhat on purity, heat treatment, temperature, and other factors. Notice that silver has the lowest resistivity and is thus the best conductor (although it is expensive). Copper is close, and much less expensive, which is why most wires are made of copper. Aluminum, although it has a higher resistivity, is much less dense than copper; it is thus preferable to copper in some situations, such as for transmission lines, because its resistance for the same weight is less than that for copper.

TABLE 25–1 Resistivity and Temperature Coefficients (at 20°C)

Material	Resistivity, ρ ($\Omega \cdot \text{m}$)	Temperature Coefficient, α $(\text{C}°)^{-1}$
Conductors		
Silver	1.59×10^{-8}	0.0061
Copper	1.68×10^{-8}	0.0068
Gold	2.21×10^{-8}	0.0034
Aluminum	2.65×10^{-8}	0.00429
Tungsten	5.28×10^{-8}	0.0045
Iron	9.61×10^{-8}	0.00651
Platinum	10.5×10^{-8}	0.003927
Mercury	96.1×10^{-8}	0.0009
Nichrome (Ni, Fe, Cr alloy)	100×10^{-8}	0.0004
Semiconductors†		
Carbon (graphite)	$(3–60) \times 10^{-5}$	−0.0005
Germanium	$(1–500) \times 10^{-3}$	−0.05
Silicon	0.1–60	−0.07
Insulators		
Glass	10^{9}–10^{12}	
Hard rubber	10^{13}–10^{15}	

†Values depend strongly on the presence of even slight amounts of impurities.

The reciprocal of the resistivity, called the **conductivity** σ, is

$$\sigma = \frac{1}{\rho} \tag{25–4}$$

and has units of $(\Omega \cdot \text{m})^{-1}$.

EXERCISE D Return to the Chapter-Opening Question, page 717, and answer it again now. Try to explain why you may have answered differently the first time.

EXERCISE E A copper wire has a resistance of 10 Ω. What would its resistance be if it had the same diameter but was only half as long? (*a*) 20 Ω, (*b*) 10 Ω, (*c*) 5 Ω, (*d*) 1 Ω, (*e*) none of these.

EXAMPLE 25–5 **Speaker wires.** Suppose you want to connect your home theater or stereo system to remote speakers (Fig. 25–14). (*a*) If each wire must be 20 m long, what diameter copper wire should you use to keep the resistance less than 0.10 Ω per wire? (*b*) If the current to each speaker is 4.0 A, what is the potential difference, or voltage drop, across each wire?

APPROACH We solve Eq. 25–3 to get the area A, from which we can calculate the wire's radius using $A = \pi r^2$. The diameter is $2r$. In (*b*) we can use Ohm's law, $V = IR$.

SOLUTION (*a*) We solve Eq. 25–3 for the area A and find ρ for copper in Table 25–1:

$$A = \rho \frac{\ell}{R} = \frac{(1.68 \times 10^{-8}\ \Omega \cdot \text{m})(20\ \text{m})}{(0.10\ \Omega)} = 3.4 \times 10^{-6}\ \text{m}^2.$$

The cross-sectional area A of a circular wire is $A = \pi r^2$. The radius must then be at least

$$r = \sqrt{\frac{A}{\pi}} = 1.04 \times 10^{-3}\ \text{m} = 1.04\ \text{mm}.$$

The diameter is twice the radius and so must be at least $2r = 2.1$ mm.

(*b*) From $V = IR$ we find that the voltage drop across each wire is

$$V = IR = (4.0\ \text{A})(0.10\ \Omega) = 0.40\ \text{V}.$$

NOTE The voltage drop across the wires reduces the voltage that reaches the speakers from the stereo amplifier, thus reducing the sound level a bit.



FIGURE 25–14 Example 25–5.

CONCEPTUAL EXAMPLE 25–6 **Stretching changes resistance.** Suppose a wire of resistance R could be stretched uniformly until it was twice its original length. What would happen to its resistance? Make the reasonable assumption that the amount of material, and its volume, doesn't change.

RESPONSE If the length ℓ doubles, then the cross-sectional area A is halved, because the volume $(V = A\ell)$ of the wire remains the same. From Eq. 25–3 we see that the resistance would increase by a factor of four $\left(2/\frac{1}{2} = 4\right)$.

EXERCISE F Copper wires in houses typically have a diameter of about 1.5 mm. How long a wire would have a 1.0-Ω resistance?

Temperature Dependence of Resistivity

The resistivity of a material depends somewhat on temperature. The resistance of metals generally increases with temperature. This is not surprising, because at higher temperatures, the atoms are moving more rapidly and are arranged in a less orderly fashion. So they might be expected to interfere more with the flow of electrons.

If the temperature change is not too great, the resistivity of metals usually increases nearly linearly with temperature. That is,

$$\rho_T = \rho_0[1 + \alpha(T - T_0)] \qquad \textbf{(25–5a)}$$

where ρ_0 is the resistivity at some reference temperature T_0 (such as 0°C or 20°C), ρ_T is the resistivity at a higher temperature T, and α is the **temperature coefficient of resistivity**. Values for α are given in Table 25–1. Note that the temperature coefficient for semiconductors can be negative. Why? At higher temperatures, some of the electrons that are normally not free in a semiconductor can become free and contribute to the current. Thus, the resistance of a semiconductor can decrease with an increase in temperature.

PHYSICS APPLIED
Resistance thermometer

EXAMPLE 25–7 **Resistance thermometer.** The variation in electrical resistance with temperature can be used to make precise temperature measurements. Platinum is commonly used since it is relatively free from corrosive effects and has a high melting point. Suppose at 20.0°C the resistance of a platinum resistance thermometer is 164.2 Ω. When placed in a particular solution, the resistance is 187.4 Ω. What is the temperature of this solution?

APPROACH Since the resistance R is directly proportional to the resistivity ρ, we can combine Eq. 25–3 with Eq. 25–5a to find R as a function of temperature T, and then solve that equation for T.

SOLUTION Equation 25–3 tells us $R = \rho\ell/A$, so we multiply Eq. 25–5a by (ℓ/A) to obtain

$$R = R_0[1 + \alpha(T - T_0)]. \qquad \textbf{(25–5b)}$$

Here $R_0 = \rho_0\ell/A$ is the resistance of the wire at $T_0 = 20.0°C$. We solve this equation for T and find (see Table 25–1 for α)

$$T = T_0 + \frac{R - R_0}{\alpha R_0} = 20.0°C + \frac{187.4\ \Omega - 164.2\ \Omega}{(3.927 \times 10^{-3}(C°)^{-1})(164.2\ \Omega)} = 56.0°C.$$

NOTE Resistance thermometers have the advantage that they can be used at very high or low temperatures where gas or liquid thermometers would be useless.

NOTE More convenient for some applications is a **thermistor** (Fig. 25–15), which consists of a metal oxide or semiconductor whose resistance also varies in a repeatable way with temperature. Thermistors can be made quite small and respond very quickly to temperature changes.

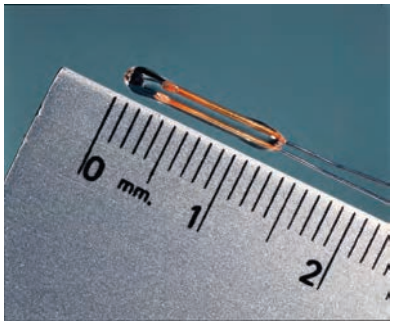


FIGURE 25–15 A thermistor only 13 mm long, shown next to a millimeter ruler for scale.

EXERCISE G The resistance of the tungsten filament of a common incandescent lightbulb is how many times greater at its operating temperature of 2800 K than its resistance at room temperature? (*a*) Less than 1% greater; (*b*) roughly 10% greater; (*c*) about 2 times greater; (*d*) roughly 10 times greater; (*e*) more than 100 times greater.

The value of α in Eqs. 25–5 can itself depend on temperature, so it is important to check the temperature range of validity of any value (say, in a handbook of physical data). If the temperature range is wide, Eq. 25–5a or b is not adequate and terms proportional to the square and cube of the temperature are needed, but these terms are generally very small except when $T - T_0$ is large.

25–5 Electric Power

Electric energy is useful to us because it can be easily transformed into other forms of energy. Motors transform electric energy into mechanical energy, and are examined in Chapter 27.

In other devices such as electric heaters, stoves, toasters, and hair dryers, electric energy is transformed into thermal energy in a wire resistance known as a "heating element." Heating elements (Fig. 25–16) in household appliances have resistances typically of a few ohms to a few hundred ohms.



(a)



(b)

FIGURE 25–16 (a) Hot electric stove burner, and (b) the hot heating element of a toaster, glow because of energy transformed by electric current.

Electric energy is transformed into thermal energy (and light) in such devices, and there are many collisions between the moving electrons and the atoms of the wire. In each collision, part of the electron's kinetic energy is transferred to the atom with which it collides. As a result, the kinetic energy of the wire's atoms increases and hence the temperature of the wire element increases (Section 18–1). The increased thermal energy can be transferred as heat by conduction and convection to the air in a heater or to food in a pan, by radiation to bread in a toaster and to people under an outside restaurant heater in cold weather (big photo page 717), or radiated as light.

To find the power transformed by an electric device, recall that the energy transformed when an infinitesimal charge dq moves through a potential difference V is $dU = V\,dq$ (Eq. 23–3). Let dt be the time required for an amount of charge dq to move through a potential difference V. Then the power P, which is the rate energy is transformed, is

$$P = \frac{dU}{dt} = \frac{dq}{dt}V.$$

The charge that flows per second, dq/dt, is the electric current I. Thus we have

$$P = IV. \qquad \textbf{(25–6)}$$

This general relation gives us the power transformed by any device, where I is the current passing through it and V is the potential difference across it. It also gives the power delivered by a source such as a battery. The SI unit of electric power is the same as for any kind of power, the **watt** $(1\,\text{W} = 1\,\text{J/s})$.

The rate of energy transformation in a resistance R can be written in two other ways, starting with the general relation $P = IV$ and substituting in $V = IR$:

$$P = IV = I(IR) = I^2R \qquad \textbf{(25–7a)}$$

$$P = IV = \left(\frac{V}{R}\right)V = \frac{V^2}{R}. \qquad \textbf{(25–7b)}$$

Equations 25–7a and b apply only to resistors, whereas Eq. 25–6, $P = IV$, is more general and applies to any device.

EXAMPLE 25–8 **Headlights.** Calculate the resistance of a 40-W automobile headlight designed for 12 V (Fig. 25–17).

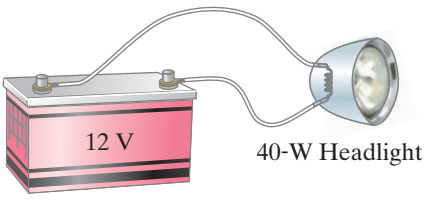
APPROACH We solve for R in Eq. 25–7b which has the given variables.

SOLUTION From Eq. 25–7b,

$$R = \frac{V^2}{P} = \frac{(12\,\text{V})^2}{(40\,\text{W})} = 3.6\,\Omega.$$

NOTE This is the resistance when the bulb is burning brightly at 40 W. When the bulb is cold, the resistance is much lower, as we saw in Eq. 25–5b (also Exercise G). Since the current is high when the resistance is low, lightbulbs burn out most often when first turned on.

FIGURE 25–17 Example 25–8.



PHYSICS APPLIED

Why lightbulbs burn out when first turned on

CAUTION
You pay for energy, which is power × time, not for power

It is energy, not power, that you pay for on your electric bill. Since power is the *rate* energy is transformed, the total energy used by any device is simply its power consumption multiplied by the time it is on. If the power is in watts and the time is in seconds, the energy will be in joules since $1\,\text{W} = 1\,\text{J/s}$. Electric companies usually specify the energy with a much larger unit, the **kilowatt-hour** (kWh). One kWh $= (1000\,\text{W})(3600\,\text{s}) = 3.60 \times 10^6\,\text{J}$.

EXAMPLE 25–9 **Electric heater.** An electric heater draws a steady 15.0 A on a 120-V line. How much power does it require and how much does it cost per month (30 days) if it operates 3.0 h per day and the electric company charges 12 cents per kWh?

APPROACH We use Eq. 25–6, $P = IV$, to find the power. We multiply the power (in kW) by the time (h) used in a month and by the cost per energy unit, \$0.12 per kWh, to get the cost per month.

SOLUTION The power is

$$P = IV = (15.0\,\text{A})(120\,\text{V}) = 1800\,\text{W} = 1.80\,\text{kW}.$$

The time (in hours) the heater is used per month is $(3.0\,\text{h/d})(30\,\text{d}) = 90\,\text{h}$, which at \$ 0.12/kWh would cost $(1.80\,\text{kW})(90\,\text{h})(\$0.12/\text{kWh}) \approx \$19$ just for this heater.

NOTE Household current is actually alternating (ac), but our solution is still valid assuming the given values for V and I are the proper averages (rms) as we discuss in Section 25–7.

PHYSICS APPLIED
Lightning

EXAMPLE 25–10 **ESTIMATE** **Lightning bolt.** Lightning is a spectacular example of electric current in a natural phenomenon (Fig. 25–18). There is much variability to lightning bolts, but a typical event might transfer $10^9\,\text{J}$ of energy across a potential difference of perhaps $5 \times 10^7\,\text{V}$ during a time interval of about 0.2 s. Use this information to estimate (*a*) the total amount of charge transferred between cloud and ground, (*b*) the current in the lightning bolt, and (*c*) the average power delivered over the 0.2 s.



FIGURE 25–18 Example 25–10. A lightning bolt.

APPROACH We estimate the charge Q, recalling that potential energy change equals the potential difference ΔV times the charge Q, Eq. 23–3. We equate ΔU with the energy transferred, $\Delta U \approx 10^9\,\text{J}$. Next, the current I is Q/t (Eq. 25–1a) and the power P is energy/time.

SOLUTION (*a*) From Eq. 23–3, the energy transformed is $\Delta U = Q\,\Delta V$. We solve for Q:

$$Q = \frac{\Delta U}{\Delta V} \approx \frac{10^9\,\text{J}}{5 \times 10^7\,\text{V}} = 20\,\text{coulombs}.$$

(*b*) The current during the 0.2 s is about

$$I = \frac{Q}{t} \approx \frac{20\,\text{C}}{0.2\,\text{s}} = 100\,\text{A}.$$

(*c*) The average power delivered is

$$P = \frac{\text{energy}}{\text{time}} = \frac{10^9\,\text{J}}{0.2\,\text{s}} = 5 \times 10^9\,\text{W} = 5\,\text{GW}.$$

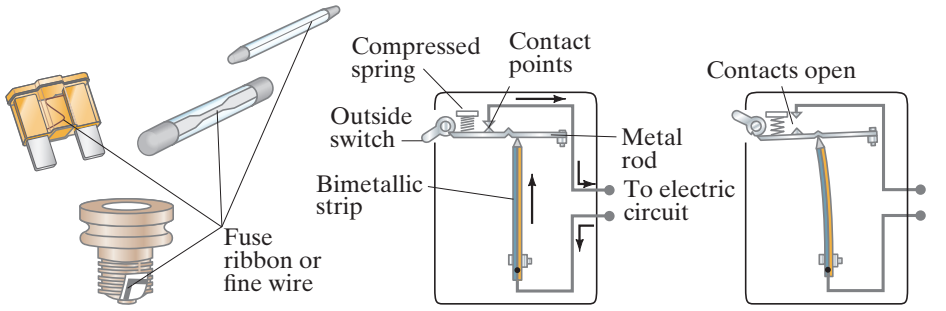
We can also use Eq. 25–6:

$$P = IV = (100\,\text{A})(5 \times 10^7\,\text{V}) = 5\,\text{GW}.$$

NOTE Since most lightning bolts consist of several stages, it is possible that individual parts could carry currents much higher than the 100 A calculated above.

25–6 Power in Household Circuits

The electric wires that carry electricity to lights and other electric appliances have some resistance, although usually it is quite small. Nonetheless, if the current is large enough, the wires will heat up and produce thermal energy at a rate equal to I^2R, where R is the wire's resistance. One possible hazard is that the current-carrying wires in the wall of a building may become so hot as to start a fire. Thicker wires have less resistance (see Eq. 25–3) and thus can carry more current without becoming too hot. When a wire carries more current than is safe, it is said to be "overloaded." To prevent overloading, **fuses** or **circuit breakers** are installed in circuits.



(a) Types of fuses (b) Circuit breaker (closed) (c) Circuit breaker (open)

FIGURE 25–19 (a) Fuses. When the current exceeds a certain value, the metallic ribbon or wire inside melts and the circuit opens. Then the fuse must be replaced. (b) One type of circuit breaker. The electric current passes through a bimetallic strip. When the current exceeds a safe level, the heating of the bimetallic strip causes the strip to bend so far to the left that the notch in the spring-loaded metal rod drops down over the end of the bimetallic strip (c) and the circuit opens at the contact points (one is attached to the rod) and the outside switch is also flipped. When the bimetallic strip cools, it can be reset using the outside switch. Better magnetic-type circuit breakers are discussed in Chapters 28 and 29.

PHYSICS APPLIED
Safety—wires getting hot

They are basically switches (Fig. 25–19) that open the circuit when the current exceeds a safe value. A 20-A fuse or circuit breaker, for example, opens when the current passing through it exceeds 20 A. If a circuit repeatedly burns out a fuse or opens a circuit breaker, and no connected device requires more than 20 A, there are two possibilities: there may be too many devices drawing current in that circuit; or there is a fault somewhere, such as a "short." A short, or "short circuit," means that two wires have touched that should not have (perhaps because the insulation has worn through) so the path of the current is shortened through a path of very low resistance. With reduced resistance, the current becomes very large and can make a wire hot enough to start a fire. Short circuits should be remedied immediately.

Household circuits are designed with the various devices connected so that each receives the standard voltage (usually 120 V in the United States) from the electric company (Fig. 25–20). Circuits with the devices arranged as in Fig. 25–20 are called *parallel circuits*, just as for capacitors, Section 24–3, and as we discuss for resistors in the next Chapter. When a fuse blows or circuit breaker opens, it is important to check the total current being drawn on that circuit, which is the sum of the currents through each device.

EXAMPLE 25–11 **Will a fuse blow?** Determine the total current drawn by all the devices in the circuit of Fig. 25–20.

APPROACH Each device has the same 120-V voltage across it. The current each draws from the source is found from $I = P/V$, Eq. 25–6.

SOLUTION The circuit in Fig. 25–20 draws the following currents: the lightbulb draws $I = P/V = 100\text{ W}/120\text{ V} = 0.8\text{ A}$; the heater draws $1800\text{ W}/120\text{ V} = 15.0\text{ A}$; the stereo or home theater amplifier draws a maximum of $175\text{ W}/120\text{ V} = 1.5\text{ A}$; and the hair dryer draws $1500\text{ W}/120\text{ V} = 12.5\text{ A}$. The total current drawn, if all devices are used at the same time, is

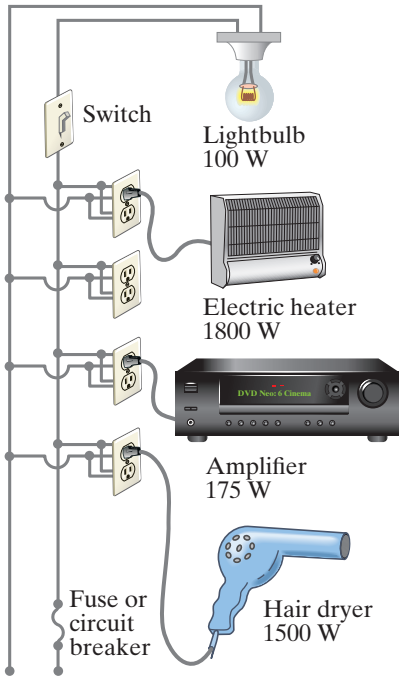
$$0.8\text{ A} + 15.0\text{ A} + 1.5\text{ A} + 12.5\text{ A} = 29.8\text{ A}.$$

NOTE The heater draws as much current as 18 100-W lightbulbs. For safety, the heater should probably be on a circuit by itself.

If the circuit in Fig. 25–20 is designed for a 20-A fuse, the fuse should blow, and we hope it will, to prevent overloaded wires from getting hot enough to start a fire.

PHYSICS APPLIED
Fuses and circuit breakers, and shorts

FIGURE 25–20 Connection of household appliances.



Something will have to be turned off to get this circuit below 20 A. (Houses and apartments usually have several circuits, each with its own fuse or circuit breaker; try moving one of the devices to another circuit.) If the circuit in Fig. 25–20 is designed with heavier wire and a 30-A fuse, the fuse shouldn't blow—if it does, a short may be the problem. (The most likely place for a short is in the cord of one of the devices.) Proper fuse size is selected according to the wire used to supply the current. A properly rated fuse should *never* be replaced by a higher-rated one, even in a car. A fuse blowing or a circuit breaker opening is acting like a switch, making an "open circuit." By an open circuit, we mean that there is no longer a complete conducting path, so no current can flow; it is as if $R = \infty$.

PHYSICS APPLIED
Proper fuses and shorts

PHYSICS APPLIED
Extension cords and possible danger

CONCEPTUAL EXAMPLE 25–12 **A dangerous extension cord.** Your 1800-W portable electric heater is too far from your desk to warm your feet. Its cord is too short, so you plug it into an extension cord rated at 11 A. Why is this dangerous?

RESPONSE 1800 W at 120 V draws a 15-A current. The wires in the extension cord rated at 11 A could become hot enough to melt the insulation and cause a fire.

EXERCISE H How many 60-W 120-V lightbulbs can operate on a 20-A line? (*a*) 2; (*b*) 3; (*c*) 6; (*d*) 20; (*e*) 40.

Resistors used in electronic circuits, like those shown in Figs. 25–11 and 25–12, can also be overheated and destroyed if too much current passes through them. The upper limit is sometimes written on the resistor, but frequently it is not. The physical size of the resistor is a clue that engineers keep track of. The common small ones shown in Fig. 25–11 are probably $\frac{1}{4}$ W.

25–7 Alternating Current

When a battery is connected to a circuit, the current moves steadily in one direction. This is called a **direct current**, or **dc**. Electric generators at electric power plants, however, produce **alternating current**, or **ac**. (Sometimes capital letters are used, **DC** and **AC**.) An alternating current reverses direction many times per second and is commonly sinusoidal, Fig. 25–21. The electrons in a wire first move in one direction and then in the other. The current supplied to homes and businesses by electric companies is ac throughout virtually the entire world. We will discuss and analyze ac circuits in detail in Chapter 30. But because ac circuits are so common in real life, we will discuss some of their basic aspects here.

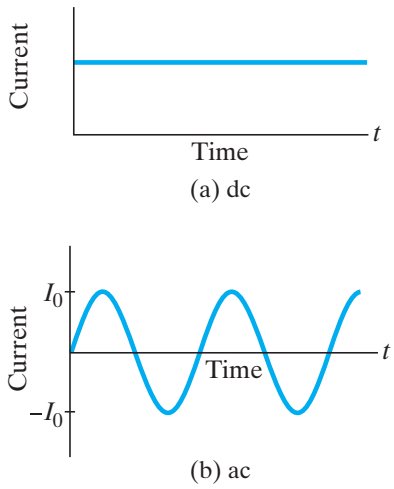


FIGURE 25–21 (a) Direct current, and (b) alternating current, as functions of time.

The voltage produced by an ac electric generator is sinusoidal, as we shall see later. The current it produces is thus sinusoidal (Fig. 25–21b). We can write the voltage as a function of time as

$$V = V_0 \sin 2\pi f t = V_0 \sin \omega t, \qquad \textbf{(25–8a)}$$

where $\omega = 2\pi f$. The potential V oscillates between $+V_0$ and $-V_0$, and V_0 is referred to as the **peak voltage**. The frequency f is the number of complete oscillations made per second, and $\omega = 2\pi f$. In most areas of the United States and Canada, f is 60 Hz (the unit "hertz," as we saw in Chapters 10 and 14, means cycles per second). In many countries, 50 Hz is used.

Equation 25–2, $V = IR$, works also for ac: if a voltage V exists across a resistance R, then the current I through the resistance is

$$I = \frac{V}{R} = \frac{V_0}{R} \sin \omega t = I_0 \sin \omega t. \qquad \textbf{(25–8b)}$$

The quantity $I_0 = V_0/R$ is the **peak current**. The current is considered positive when the electrons flow in one direction and negative when they flow in the opposite direction. It is clear from Fig. 25–21b that an alternating current is as often positive as it is negative. Thus, the average current is zero. This does not mean, however, that no power is needed or that no heat is produced in a resistor. Electrons

do move back and forth, and do produce heat. Indeed, the power transformed in a resistance R at any instant is (Eq. 25–8b)

$$P = I^2R = I_0^2 R \sin^2 \omega t.$$

Because the current is squared, we see that the power is always positive, as graphed in Fig. 25–22. The quantity $\sin^2 \omega t$ varies between 0 and 1; and it is not too difficult to show[†] that its average value is $\frac{1}{2}$, as indicated in Fig. 25–22. Thus, the *average power* transformed, $\overline{P}$, is

$$\overline{P} = \tfrac{1}{2} I_0^2 R.$$

Since power can also be written $P = V^2/R = (V_0^2/R)\sin^2 \omega t$, we also have that the average power is

$$\overline{P} = \tfrac{1}{2}\frac{V_0^2}{R}.$$

The average or mean value of the *square* of the current or voltage is thus what is important for calculating average power: $\overline{I^2} = \frac{1}{2} I_0^2$ and $\overline{V^2} = \frac{1}{2} V_0^2$. The square root of each of these is the **rms** (root-mean-square) value of the current or voltage:

$$I_{\text{rms}} = \sqrt{\overline{I^2}} = \frac{I_0}{\sqrt{2}} = 0.707 I_0, \qquad \textbf{(25–9a)}$$

$$V_{\text{rms}} = \sqrt{\overline{V^2}} = \frac{V_0}{\sqrt{2}} = 0.707 V_0. \qquad \textbf{(25–9b)}$$

The rms values of V and I are sometimes called the *effective values*, as well as rms values. They are useful because they can be substituted directly into the power formulas, Eqs. 25–6 and 25–7, to get the average power:

$$\overline{P} = I_{\text{rms}} V_{\text{rms}} \qquad \textbf{(25–10a)}$$

$$\overline{P} = \tfrac{1}{2} I_0^2 R = I_{\text{rms}}^2 R \qquad \textbf{(25–10b)}$$

$$\overline{P} = \tfrac{1}{2}\frac{V_0^2}{R} = \frac{V_{\text{rms}}^2}{R}. \qquad \textbf{(25–10c)}$$

Thus, a direct current whose values of I and V equal the rms values of I and V for an alternating current will produce the same power. Hence it is usually the rms value of current and voltage that is specified or measured. For example, in the United States and Canada, standard line voltage is 120-V ac. The 120 V is V_{rms}; the peak voltage V_0 is

$$V_0 = \sqrt{2}\, V_{\text{rms}} = 170 \text{ V}.$$

In much of the world (Europe, Australia, Asia) the rms voltage is 240 V, so the peak voltage is 340 V. The line voltage can vary, depending on the total load (resistance); the frequency of 60 Hz or 50 Hz, however, remains extremely steady.

The power specified by any device or appliance is the average power $\overline{P}$. The instantaneous power varies from zero to $2\overline{P}$ as we saw in Fig. 25–22. The glow of an electric heating element looks steady because the changes are fast, 60 Hz or 50 Hz; the heating element cannot cool much during these brief moments when P drops to zero and we don't even see the brightness flicker at these frequencies (at 4 or 5 Hz we might).

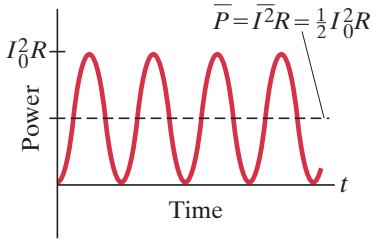


FIGURE 25–22 Power transformed in a resistor in an ac circuit is always positive.

[†]A graph of $\cos^2 \omega t$ versus t is identical to that for $\sin^2 \omega t$ in Fig. 25–22, except that the points are shifted (by $\frac{1}{4}$ cycle) on the time axis. Hence the average value of $\sin^2$ and $\cos^2$, averaged over one or more full cycles, will be the same. From the trigonometric identity $\sin^2\theta + \cos^2\theta = 1$, we can write

$$\overline{(\sin^2 \omega t)} + \overline{(\cos^2 \omega t)} = 2\overline{(\sin^2 \omega t)} = 1.$$

Hence the average value of $\sin^2 \omega t$ is $\frac{1}{2}$.

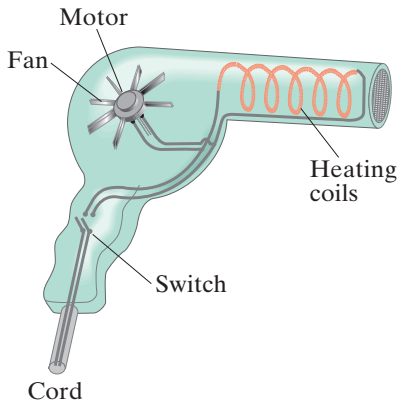


FIGURE 25–23 A hair dryer. Most of the current goes through the heating coils, a pure resistance; a small part goes to the motor to turn the fan. Example 25–13.

EXAMPLE 25–13 **Hair dryer.** (*a*) Calculate the resistance and the peak current in an 1875-W hair dryer (Fig. 25–23) connected to a 120-V ac line. (*b*) What happens if it is connected to a 240-V ac line in Britain?

APPROACH We are given $\overline{P}$ and V_{rms}, so $I_{\text{rms}} = \overline{P}/V_{\text{rms}}$ (Eq. 25–10a or 25–6), and $I_0 = \sqrt{2}\, I_{\text{rms}}$. Then we find R from $V = IR$.

SOLUTION (*a*) We solve Eq. 25–10a for the rms current:

$$I_{\text{rms}} = \frac{\overline{P}}{V_{\text{rms}}} = \frac{1875\ \text{W}}{120\ \text{V}} = 15.6\ \text{A}.$$

Then

$$I_0 = \sqrt{2}\, I_{\text{rms}} = 22.1\ \text{A}.$$

The resistance is

$$R = \frac{V_{\text{rms}}}{I_{\text{rms}}} = \frac{120\ \text{V}}{15.6\ \text{A}} = 7.7\ \Omega.$$

The resistance could equally well be calculated using peak values:

$$R = \frac{V_0}{I_0} = \frac{170\ \text{V}}{22.1\ \text{A}} = 7.7\ \Omega.$$

(*b*) When connected to a 240-V line, more current would flow and the resistance would change with the increased temperature (Section 25–4). But let us make an estimate of the power transformed based on the same 7.7-Ω resistance. The average power would be

$$\overline{P} = \frac{V_{\text{rms}}^2}{R} = \frac{(240\ \text{V})^2}{(7.7\ \Omega)} = 7500\ \text{W}.$$

This is four times the hair dryer's power rating and would undoubtedly melt the heating element or the wire coils of the motor.

EXERCISE I A home sound system sends an average audio power output of 100 W into an 8-Ω loudspeaker (see Fig. 25–14). What are the rms voltage and the rms current fed to the speaker (*a*) at the highest rated power of 100 W, and (*b*) at 1.0 W when the volume is turned down?

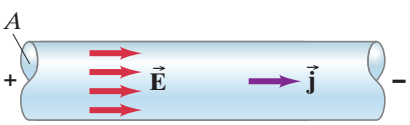
This Section has given a brief introduction to the basic aspects of alternating currents. We will discuss ac circuits in more detail in Chapter 30. In the next chapter, Chapter 26, we will deal with the details of dc circuits only.

25–8 Microscopic View of Electric Current

Up to now in this Chapter we have dealt mainly with a macroscopic view of electric current. We saw, however, that according to atomic theory, the electric current in metal wires is carried by negatively charged electrons, and that in liquid solutions current can also be carried by positive and/or negative ions. Let us now look at this microscopic picture in more detail.

When a potential difference is applied to the two ends of a wire of uniform cross section, the direction of the electric field $\vec{\mathbf{E}}$ is parallel to the walls of the wire (Fig. 25–24). The existence of $\vec{\mathbf{E}}$ within the conducting wire does not contradict our earlier result that $\vec{\mathbf{E}} = 0$ inside a conductor in the electrostatic case, as we are no longer dealing with the static case. Charges are free to move in a conductor, and hence can move under the action of the electric field. If all the charges are at rest, then $\vec{\mathbf{E}}$ must be zero (electrostatics).

FIGURE 25–24 Electric field $\vec{\mathbf{E}}$ in a uniform wire of cross-sectional area A carrying a current I. The current density $j = I/A$.



Current Density and Drift Speed

We now define a new microscopic quantity, the **current density**, $\vec{\mathbf{j}}$. It is defined as the *electric current per unit cross-sectional area* at any point in space. If the current density $\vec{\mathbf{j}}$ in a wire of cross-sectional area A is uniform over the cross section, then j is related to the electric current by

$$j = \frac{I}{A} \qquad \text{or} \qquad I = jA. \tag{25–11}$$

If the current density is not uniform, then the general relation is

$$I = \int \vec{\mathbf{j}} \cdot d\vec{\mathbf{A}}, \tag{25–12}$$

where I is the current through the whole surface and $d\vec{\mathbf{A}}$ is an element of surface area over which the integration is taken. The direction of the current density at any point is the direction that a positive charge would move when placed at that point—that is, the direction of $\vec{\mathbf{j}}$ at any point is generally the same as the direction of $\vec{\mathbf{E}}$, Fig. 25–24. The current density exists for any *point* in space. The current I, on the other hand, refers to a conductor as a whole, and hence is a macroscopic quantity.

The direction of $\vec{\mathbf{j}}$ is chosen to represent the direction of net flow of positive charge. In a metal, it is negatively charged electrons that move, so they move in the direction of $-\vec{\mathbf{j}}$, or $-\vec{\mathbf{E}}$ (to the left in Fig. 25–24). We can imagine the free electrons as moving about randomly at high speeds, bouncing off the metal atoms of the wire (somewhat like the molecules of a gas—Chapter 18). When an electric field exists in the wire, Fig. 25–25, the electrons feel a force and initially begin to accelerate. But they soon reach a more or less steady average speed, known as their **drift speed**, v_d (collisions with atoms in the wire keep them from accelerating further). The drift speed is normally very much smaller than the electrons' average random speed inside the metal wire.

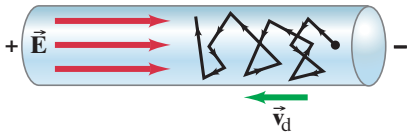


FIGURE 25–25 Black zagged line represents the motion of an electron in a metal wire due to an electric field. The field $\vec{\mathbf{E}}$ gives electrons in random motion a net drift speed v_d. Its direction (the net charge flow) is $\vec{\mathbf{v}}_d$ and is in the opposite direction of $\vec{\mathbf{E}}$ because electrons have a negative charge ($\vec{\mathbf{F}} = q\vec{\mathbf{E}}$).

We can relate the electron drift speed v_d to the macroscopic current I in the wire. In a time Δt, the electrons will travel a distance $\ell = v_d\,\Delta t$ on average. Suppose the wire has cross-sectional area A. Then in time Δt, electrons in a volume $V = A\ell = Av_d\,\Delta t$ will pass through the cross section A of wire, as shown in Fig. 25–26. If there are n free electrons (each of charge $-e$) per unit volume, then the total number of electrons is $N = nV$ (V is volume, not voltage) and the total charge ΔQ that passes through the area A in a time Δt is

$$\begin{aligned}\Delta Q &= (\text{number of charges, } N) \times (\text{charge per particle}) \\ &= (nV)(-e) = -(nAv_d\,\Delta t)(e).\end{aligned}$$

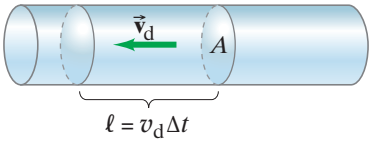


FIGURE 25–26 Electrons in the volume $A\ell$ will all pass through the cross section indicated in a time Δt, where $\ell = v_d\,\Delta t$.

The current I in the wire is thus

$$I = \frac{\Delta Q}{\Delta t} = -neAv_d. \tag{25–13}$$

The current density, $j = I/A$, is

$$j = -nev_d, \tag{25–14}$$

where the minus sign indicates that the direction of (positive) current flow is opposite to the drift speed of electrons.

We can generalize Eq. 25–14 to any type of charge flow, such as flow of ions in an electrolyte (as on page 719). If there are several types of ions (which can include free electrons), each of density n_i (number per unit volume), charge q_i ($q_i = -e$ for electrons) and drift speed v_{di}, then the net current density at any point is

$$j = \sum_i n_i q_i v_{di}. \tag{25–15}$$

The total current I passing through an area A perpendicular to the direction of a uniform j is then

$$I = \sum_i n_i q_i v_{di} A.$$

EXAMPLE 25–14 **Electron speeds in a wire.** A copper wire 3.2 mm in diameter carries a 5.0-A current. Determine (*a*) the current density in the wire (assumed uniform), and (*b*) the drift speed of the free electrons. (*c*) Estimate the rms speed of electrons assuming they behave like an ideal gas at 20°C. Assume that one electron per Cu atom is free to move (the others remain bound to the atom).

APPROACH For (*a*) $j = I/A = I/\pi r^2$. For (*b*) we can apply Eq. 25–14 to find v_d if we can determine the number n of free electrons per unit volume. Since we assume there is one free electron per atom, the density of free electrons, n, is the same as the number of Cu atoms per unit volume. The atomic mass of Cu is 63.5 u (see Periodic Table inside the back cover), so 63.5 g of Cu contains one mole or 6.02×10^{23} free electrons. To find the volume V of this amount of copper, and then $n = N/V$, we use the mass density of copper (Table 13–1), $\rho_D = 8.9 \times 10^3\ \text{kg/m}^3$, where $\rho_D = m/V$. (We use ρ_D to distinguish it here from ρ for resistivity.) In (*c*) we use $K = \frac{3}{2}kT$, Eq. 18–4. (Do not confuse V for volume with V for voltage.)

SOLUTION (*a*) The radius of the wire is $r = \frac{1}{2}(3.2\ \text{mm}) = 1.6 \times 10^{-3}\ \text{m}$, so the current density is (if we assume it is uniform):

$$j = \frac{I}{A} = \frac{I}{\pi r^2} = \frac{5.0\ \text{A}}{\pi(1.6 \times 10^{-3}\ \text{m})^2} = 6.2 \times 10^5\ \text{A/m}^2.$$

(*b*) The number of free electrons per unit volume, $n = N/V$ (where $V = \text{volume} = m/\rho_D$), is

$$n = \frac{N}{V} = \frac{N}{m/\rho_D} = \frac{N(1\ \text{mole})}{m(1\ \text{mole})}\rho_D$$

$$n = \left(\frac{6.02 \times 10^{23}\ \text{electrons}}{63.5 \times 10^{-3}\ \text{kg}}\right)(8.9 \times 10^3\ \text{kg/m}^3) = 8.4 \times 10^{28}\ \text{m}^{-3}.$$

Then, by Eq. 25–14, the drift speed has magnitude

$$v_d = \frac{j}{ne} = \frac{6.2 \times 10^5\ \text{A/m}^2}{(8.4 \times 10^{28}\ \text{m}^{-3})(1.6 \times 10^{-19}\ \text{C})}$$

$$= 4.6 \times 10^{-5}\ \text{m/s} \approx 0.05\ \text{mm/s}.$$

(*c*) If we model the free electrons as an ideal gas (a rather rough approximation), we use Eq. 18–5 to estimate the random rms speed of an electron as it darts around at 20°C = 293 K:

$$v_{\text{rms}} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{3(1.38 \times 10^{-23}\ \text{J/K})(293\ \text{K})}{9.11 \times 10^{-31}\ \text{kg}}} = 1.2 \times 10^5\ \text{m/s}.$$

The drift speed (average speed in the direction of the current) is very much less than the rms thermal speed of the electrons, by a factor of about 10^9.

NOTE The result in (*c*) is an underestimate. Quantum theory calculations, and experiments, give the rms speed in copper to be about $1.6 \times 10^6\ \text{m/s}$.

The drift speed of electrons in a wire is very slow, only about 0.05 mm/s (Example 25–14 above), which means it takes an electron 20×10^3 s, or $5\frac{1}{2}$ h, to travel only 1 m. This is not how fast "electricity travels": when you flip a light switch, the light—even if many meters away—goes on nearly instantaneously. Why? Because electric fields travel essentially at the speed of light $(3 \times 10^8\ \text{m/s})$. We can think of electrons in a wire as being like a pipe full of water: when a little water enters one end of the pipe, some water comes out at the other end almost immediately.

Electric Field Inside a Wire

Equation 25–2, $V = IR$ where V = voltage (not volume), can be written in terms of microscopic quantities as follows. We write the resistance R in terms of the resistivity ρ, Eq. 25–3:

$$R = \rho \frac{\ell}{A}.$$

Then we write V and I as

$$I = jA \qquad \text{and} \qquad V = E\ell.$$

These relations follow from Eqs. 25–11 and 23–4, where we assume the electric field is uniform within the wire and ℓ is the length of the wire (or a portion of the wire) between whose ends the potential difference is V. Thus, from $V = IR$, we have

$$V = IR$$

$$E\ell = (jA)\left(\rho \frac{\ell}{A}\right) = j\rho\ell$$

so

$$j = \frac{1}{\rho} E = \sigma E, \tag{25–16}$$

where $\sigma = 1/\rho$ is the *conductivity* (Eq. 25–4). For a metal conductor, ρ and σ do not depend on V (and hence not on E). Therefore the current density $\vec{\mathbf{j}}$ is proportional to the electrical field $\vec{\mathbf{E}}$ in the conductor. This is the "microscopic" statement of Ohm's law. Equation 25–16, which can be written in vector form as

$$\vec{\mathbf{j}} = \sigma\vec{\mathbf{E}} = \frac{1}{\rho}\vec{\mathbf{E}},$$

is sometimes taken as the definition of conductivity σ and resistivity ρ.

EXAMPLE 25–15 **Electric field inside a wire.** What is the electric field inside the wire of Example 25–14?

APPROACH We use Eq. 25–16 and $\rho = 1.68 \times 10^{-8}\ \Omega \cdot \text{m}$ for copper.

SOLUTION Example 25–14 gives $j = 6.2 \times 10^5\ \text{A/m}^2$, so

$$E = \rho j = (1.68 \times 10^{-8}\ \Omega \cdot \text{m})(6.2 \times 10^5\ \text{A/m}^2) = 1.0 \times 10^{-2}\ \text{V/m}.$$

NOTE For comparison, the electric field between the plates of a capacitor is often much larger; in Example 24–1, for example, E is on the order of $10^4\ \text{V/m}$. Thus we see that only a modest electric field is needed for current flow in practical cases.

*25–9 Superconductivity

At very low temperatures, well below 0°C, the resistivity (Section 25–4) of certain metals and certain compounds or alloys becomes zero as measured by the highest-precision techniques. Materials in such a state are said to be **superconducting**. This phenomenon was first observed by H. K. Onnes (1853–1926) in 1911 when he cooled mercury below 4.2 K (−269°C) and found that the resistance of mercury suddenly dropped to zero. In general, superconductors become superconducting only below a certain **transition temperature** or **critical temperature**, T_C, which until recently was always within a few degrees of absolute zero. Electric current in a ring-shaped superconducting material has been observed to flow for years in the absence of a potential difference, with no measurable decrease. Measurements show that the resistivity ρ of superconductors is less than $4 \times 10^{-25}\ \Omega \cdot \text{m}$, which is over 10^{16} times smaller than that for copper, and is considered to be zero in practice. See Fig. 25–27.

Before 1986 the highest temperature at which a material was found to superconduct was 23 K, which required liquid helium to keep the material cold enough to be superconducting. In 1987, a compound of yttrium, barium, copper, and oxygen (YBCO) was developed that can be superconducting at 90 K.

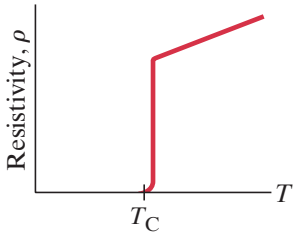


FIGURE 25–27 A superconducting material has zero resistivity when its temperature is below T_C, its "critical temperature." At temperatures above T_C, the resistivity has a "normal" nonzero value and increases with temperature as normal metals do (Eq. 25–5 and Table 25–1).

Since this is above the boiling temperature of liquid nitrogen, 77 K, liquid nitrogen is sufficiently cold to keep the material superconducting. This was an important breakthrough because liquid nitrogen is much more easily and cheaply obtained (N_2 is 80% of the air around us) than is the much rarer (and colder) liquid helium needed for earlier superconductors. Superconductivity at "high" temperatures, such as 160 K, has been reported, and perhaps also at $-23°C$ (250 K), though in fragile compounds and under very high pressure.

To develop high-temperature superconductors (HTS–above 77 K) for use as wires, such as for wires in "superconducting electromagnets" (Section 28–9), is challenging because HTS materials are ceramic; they are very brittle and do not form into flexible wires. The first generation HTS "wires" made use of bismuth-strontium-calcium-copper oxide (BSCCO), which was embedded as tiny filaments or powders into a metal alloy base, which is not resistanceless but has resistance much less than a conventional copper cable. Second-generation HTS apply superconducting powders or a very thin film of YBCO (other rare earths like neodymium also work) onto specially prepared metal tapes; they can be wound into cables, and carry hundreds of amperes. More recently, a compound of magnesium and boron has been made into wires; even though they superconduct at a lower temperature, they are much cheaper to produce.

*25–10 Electrical Conduction in the Human Nervous System

The flow of electric charge in the human nervous system provides us the means for being aware of the world, for communication within the body, and for controlling the body's muscles. Indeed, we could not exist without electricity in our bodies. Although the detailed functioning of the nervous system is complex and still not well understood, we do have a reasonable understanding of how messages are transmitted within the nervous system: they are electrical signals passing along the basic element of the nervous system, the **neuron**.

Neurons are living cells of unusual shape (Fig. 25–28). Attached to the main cell body are several small appendages known as *dendrites* and a long tail called the *axon*. Signals are received by the dendrites and are propagated along the axon. When a signal reaches the nerve endings, it is transmitted to the next neuron or to a muscle at a connection called a **synapse**.

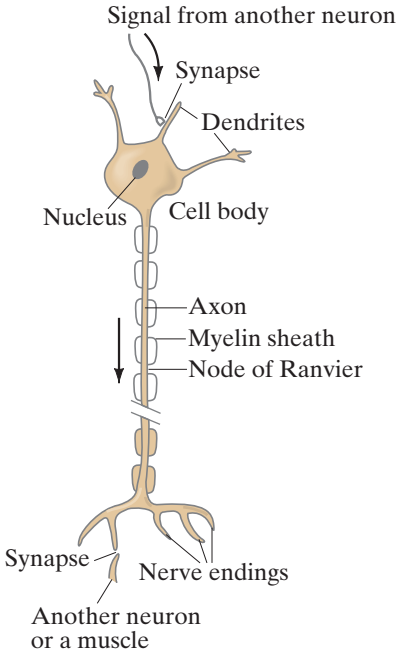


FIGURE 25–28 A simplified sketch of a typical neuron.

A neuron, before transmitting an electrical signal, is in the so-called "resting state." Like nearly all living cells, neurons have a net positive charge on the outer surface of the cell membrane and a negative charge on the inner surface. This difference in charge, or **dipole layer**, means that a potential difference exists across the cell membrane. When a neuron is not transmitting a signal, this **resting potential**, normally stated as

$$V_{\text{inside}} - V_{\text{outside}},$$

is typically -60 mV to -90 mV, depending on the type of organism. The most common ions in a cell are K^+, Na^+, and Cl^-. There are large differences in the concentrations of these ions inside and outside an axon, as indicated by the typical values given in Table 25–2. Other ions are also present, so the fluids both inside and outside the axon are electrically neutral. Because of the differences in concentration, there is a tendency for ions to diffuse across the membrane (see Section 18–7 on diffusion). However, in the resting state the cell membrane prevents any net flow of Na^+ through a mechanism of active transport† of Na^+ ions out of the cell by a particular protein to which Na^+ attach; energy needed comes from ATP.‡ But a cell membrane does allow the flow of Cl^- ions, and less so of K^+ ions, and it is these two ions that produce the dipole charge layer on the membrane. Because there is a greater concentration of K^+ inside the cell than outside, more K^+ ions tend to diffuse outward across the membrane than diffuse inward.

TABLE 25–2
Concentrations of Ions Inside and Outside a Typical Axon

	Concentration inside axon (mol/m^3)	Concentration outside axon (mol/m^3)
K^+	140	5
Na^+	15	140
Cl^-	9	125

†This transport mechanism is sometimes referred to as the **sodium pump**.

‡ATP = adenosine triphosphate. How it works to store energy is discussed in Section 40–2.

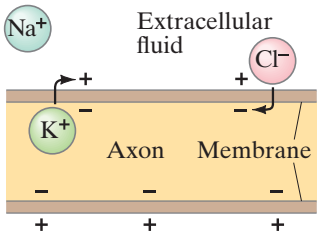
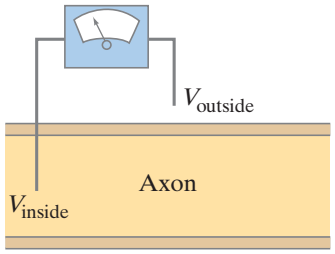


FIGURE 25–29 How a dipole layer of charge forms on a cell membrane.

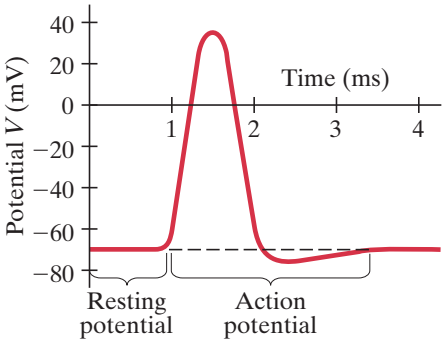
FIGURE 25–30 Measuring the potential difference between the inside and outside of a nerve cell.



A K^+ ion that passes through the membrane becomes attached to the outer surface of the membrane, and leaves behind an equal negative charge that lies on the inner surface of the membrane (Fig. 25–29). The fluids themselves remain neutral. What keeps the ions on the membrane is their attraction for each other across the membrane. Independently, Cl^- ions tend to diffuse *into* the cell since their concentration outside is higher. Both K^+ and Cl^- diffusion tends to charge the interior surface of the membrane negative and the outside positive. As charge accumulates on the membrane surface, it becomes increasingly difficult for more ions to diffuse: K^+ ions trying to move outward, for example, are repelled by the positive charge already there. Equilibrium is reached when the tendency to diffuse because of the concentration difference is just balanced by the electrical potential difference across the membrane. The greater the concentration difference, the greater the potential difference across the membrane (-60 mV to -90 mV).

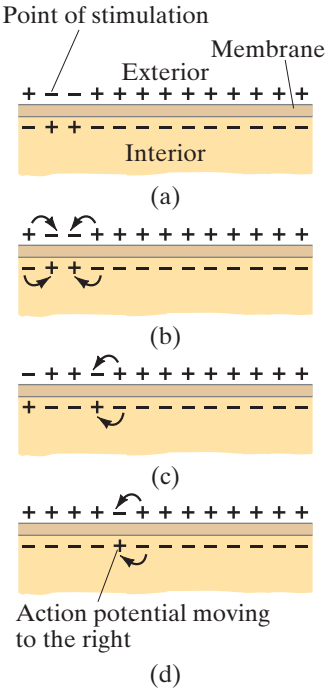
The most important aspect of a neuron is not that it has a resting potential (most cells do), but rather that it can respond to a stimulus and conduct an electrical signal along its length. The stimulus could be thermal (when you touch a hot stove) or chemical (as in taste buds); it could be pressure (as on the skin or at the eardrum), or light (as in the eye); or it could be the electric stimulus of a signal coming from the brain or another neuron. In the laboratory, the stimulus is usually electrical and is applied by a tiny probe at some point on the neuron. If the stimulus exceeds some threshold, a voltage pulse will travel down the axon. This voltage pulse can be detected at a point on the axon using a voltmeter or an oscilloscope connected as in Fig. 25–30. This voltage pulse has the shape shown in Fig. 25–31, and is called an **action potential**. As can be seen, the potential increases from a resting potential of about -70 mV and becomes a positive 30 mV or 40 mV. The action potential lasts for about 1 ms and travels down an axon with a speed of 30 m/s to 150 m/s. When an action potential is stimulated, the nerve is said to have "fired."

FIGURE 25–31 Action potential.



What causes the action potential? At the point where the stimulus occurs, the membrane suddenly alters its permeability, becoming much more permeable to Na^+ than to K^+ and Cl^- ions. Thus, Na^+ ions rush into the cell and the inner surface of the wall becomes positively charged, and the potential difference quickly swings positive ($\approx +30$ mV in Fig. 25–31). Just as suddenly, the membrane returns to its original characteristics; it becomes impermeable to Na^+ and in fact pumps out Na^+ ions. The diffusion of Cl^- and K^+ ions again predominates and the original resting potential is restored (-70 mV in Fig. 25–31).

FIGURE 25–32 Propagation of an action potential along an axon membrane.



What causes the action potential to travel along the axon? The action potential occurs at the point of stimulation, as shown in Fig. 25–32a. The membrane momentarily is positive on the inside and negative on the outside at this point. Nearby charges are attracted toward this region, as shown in Fig. 25–32b. The potential in these adjacent regions then flips, causing an action potential there. Thus, as the membrane returns to normal at the original point, nearby it experiences an action potential, so the action potential moves down the axon (Figs. 25–32c and d).

You may wonder if the number of ions that pass through the membrane would significantly alter the concentrations. The answer is no; and we can show why by treating the axon as a capacitor in the following Example.

EXAMPLE 25–16 ESTIMATE **Capacitance of an axon.** (*a*) Do an order-of-magnitude estimate for the capacitance of an axon 10 cm long of radius 10 μm. The thickness of the membrane is about 10^{-8} m, and the dielectric constant is about 3. (*b*) By what factor does the concentration (number of ions per volume) of Na^+ ions in the cell change as a result of one action potential?

APPROACH We model the membrane of an axon as a cylindrically shaped parallel-plate capacitor, with opposite charges on each side. The separation of the "plates" is the thickness of the membrane, $d \approx 10^{-8}$ m. We first calculate the area of the cylinder and then can use Eq. 24–8, $C = K\epsilon_0 A/d$, to find the capacitance. In (*b*), we use the voltage change during one action potential to find the amount of charge moved across the membrane.

SOLUTION (*a*) The area A is the area of a cylinder of radius r and length ℓ:

$$A = 2\pi r\ell \approx (6.28)(10^{-5}\,\text{m})(0.1\,\text{m}) \approx 6 \times 10^{-6}\,\text{m}^2.$$

From Eq. 24–8, we have

$$C = K\epsilon_0 \frac{A}{d} \approx (3)(8.85 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)\frac{6 \times 10^{-6}\,\text{m}^2}{10^{-8}\,\text{m}} \approx 10^{-8}\,\text{F}.$$

(*b*) Since the voltage changes from -70 mV to about $+30$ mV, the total change is about 100 mV. The amount of charge that moves is then

$$Q = CV \approx (10^{-8}\,\text{F})(0.1\,\text{V}) = 10^{-9}\,\text{C}.$$

Each ion carries a charge $e = 1.6 \times 10^{-19}$ C, so the number of ions that flow per action potential is $Q/e = (10^{-9}\,\text{C})/(1.6 \times 10^{-19}\,\text{C}) \approx 10^{10}$. The volume of our cylindrical axon is

$$V = \pi r^2\ell \approx (3)(10^{-5}\,\text{m})^2(0.1\,\text{m}) = 3 \times 10^{-11}\,\text{m}^3,$$

and the concentration of Na^+ ions inside the cell (Table 25–2) is $15\,\text{mol/m}^3 = 15 \times 6.02 \times 10^{23}\,\text{ions/m}^3 \approx 10^{25}\,\text{ions/m}^3$. Thus, the cell contains $(10^{25}\,\text{ions/m}^3) \times (3 \times 10^{-11}\,\text{m}^3) \approx 3 \times 10^{14}$ Na^+ ions. One action potential, then, will change the concentration of Na^+ ions by about $10^{10}/(3 \times 10^{14}) = \frac{1}{3} \times 10^{-4}$, or 1 part in 30,000. This tiny change would not be measurable.

Thus, even 1000 action potentials will not alter the concentration significantly. The sodium pump does not, therefore, have to remove Na^+ ions quickly after an action potential, but can operate slowly over time to maintain a relatively constant concentration.

Summary

An electric **battery** serves as a source of nearly constant potential difference by transforming chemical energy into electric energy. A simple battery consists of two electrodes made of different metals immersed in a solution or paste known as an **electrolyte**.

Electric current, I, refers to the rate of flow of electric charge and is measured in **amperes** (A): 1 A equals a flow of 1 C/s past a given point.

The direction of **conventional current** is that of positive charge flow. In a wire, it is actually negatively charged electrons that move, so they flow in a direction opposite to what we call conventional current. A positive charge flow in one direction is almost always equivalent to a negative charge flow in the opposite direction. Positive conventional current always flows from a high potential to a low potential.

The **resistance** R of a device is defined by the relation

$$V = IR, \qquad \textbf{(25–2)}$$

where I is the current in the device when a potential difference V is applied across it. For materials such as metals, R is a constant independent of V (thus $I \propto V$), a result known as **Ohm's law**. Thus, the current I coming from a battery of voltage V depends on the resistance R of the circuit connected to it.

Voltage is applied *across* a device or between the ends of a wire. Current passes *through* a wire or device. Resistance is a property *of* the wire or device.

The unit of resistance is the **ohm** (Ω), where $1\ \Omega = 1\ \text{V/A}$. See Table 25–3.

TABLE 25–3 Summary of Units

Current	$1\ \text{A} = 1\ \text{C/s}$
Potential difference	$1\ \text{V} = 1\ \text{J/C}$
Power	$1\ \text{W} = 1\ \text{J/s}$
Resistance	$1\ \Omega = 1\ \text{V/A}$

The resistance R of a wire is inversely proportional to its cross-sectional area A, and directly proportional to its length ℓ and to a property of the material called its resistivity ρ:

$$R = \rho\frac{\ell}{A}. \qquad \textbf{(25–3)}$$

The **resistivity**, ρ, increases with temperature for metals, but for semiconductors it may decrease.

The rate at which energy is transformed in a resistance R from electric to other forms of energy (such as heat and light) is equal to the product of current and voltage. That is, the **power** transformed, measured in watts, is given by

$$P = IV, \qquad \textbf{(25–6)}$$

which for resistors can be written as

$$P = I^2R = \frac{V^2}{R}. \qquad \textbf{(25–7)}$$

The SI unit of power is the **watt** $(1\ \text{W} = 1\ \text{J/s})$.

The total electric energy transformed in any device equals the product of the power and the time during which the device is operated. In SI units, energy is given in joules $(1\ \text{J} = 1\ \text{W}\cdot\text{s})$, but electric companies use a larger unit, the **kilowatt-hour** $(1\ \text{kWh} = 3.6\times10^6\ \text{J})$.

Electric current can be **direct current** (**dc**), in which the current is steady in one direction; or it can be **alternating current** (**ac**), in which the current reverses direction at a particular frequency f, typically 60 Hz. Alternating currents are typically sinusoidal in time,

$$I = I_0\sin\omega t, \qquad \textbf{(25–8)}$$

where $\omega = 2\pi f$, and are produced by an alternating voltage.

The **rms** values of sinusoidally alternating currents and voltages are given by

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}} \quad \text{and} \quad V_{\text{rms}} = \frac{V_0}{\sqrt{2}}, \qquad \textbf{(25–9)}$$

respectively, where I_0 and V_0 are the **peak** values. The power relationship, $P = IV = I^2R = V^2/R$, is valid for the average power in alternating currents when the rms values of V and I are used.

Current density $\vec{\mathbf{j}}$ is the current per cross-sectional area. From a microscopic point of view, the current density is related to the number of charge carriers per unit volume, n, their charge, q, and their **drift speed**, v_{d} by

$$j = nqv_{\text{d}} \qquad \textbf{(25–15)}$$

The electric field within a wire is related to $\vec{\mathbf{j}}$ by $\vec{\mathbf{j}} = \sigma\vec{\mathbf{E}}$ where $\sigma = 1/\rho$ is the **conductivity**.

[*At very low temperatures certain materials become **superconducting**, which means their electrical resistance becomes zero.]

[*The human nervous system operates via electrical conduction: when a nerve "fires," an electrical signal travels as a voltage pulse known as an **action potential**.]

Questions

1. When an electric cell is connected to a circuit, electrons flow away from the negative terminal in the circuit. But within the cell, electrons flow *to* the negative terminal. Explain.
2. When a flashlight is operated, what is being used up: battery current, battery voltage, battery energy, battery power, or battery resistance? Explain.
3. What quantity is measured by a battery rating given in ampere-hours $(\text{A}\cdot\text{h})$? Explain.
4. Can a copper wire and an aluminum wire of the same length have the same resistance? Explain.
5. One terminal of a car battery is said to be connected to "ground." Since it is not really connected to the ground, what is meant by this expression?
6. When you turn on a water faucet, the water usually flows immediately. You don't have to wait for water to flow from the faucet valve to the spout. Why not? Is the same thing true when you connect a wire to the terminals of a battery?
7. The equation $P = V^2/R$ indicates that the power dissipated in a resistor decreases if the resistance is increased, whereas the equation $P = I^2R$ implies the opposite. Is there a contradiction here? Explain.
8. What happens when a lightbulb burns out?
9. If the resistance of a small immersion heater (to heat water for tea or soup, Fig. 25–33) was increased, would it speed up or slow down the heating process? Explain.



FIGURE 25–33 Question 9.

10. If a rectangular solid made of carbon has sides of lengths a, $2a$, and $3a$, to which faces would you connect the wires from a battery so as to obtain (*a*) the least resistance, (*b*) the greatest resistance?

11. Explain why lightbulbs almost always burn out just as they are turned on and not after they have been on for some time.

12. Which draws more current at 110 V, an 18-W lightbulb or a 12-W lightbulb? Which has the higher resistance?

13. Electric power is transferred over large distances at very high voltages. Explain how the high voltage reduces power losses in the transmission lines.

14. A 15-A fuse blows repeatedly. Why is it dangerous to replace this fuse with a 25-A fuse?

15. When electric lights are operated on low-frequency ac (say, 5 Hz), they flicker noticeably. Why?

16. Driven by ac power, the same electrons pass back and forth through your reading lamp over and over again. Explain why the light stays lit instead of going out after the first pass of electrons.

17. The heating element in a toaster is made of Nichrome wire. Immediately after the toaster is turned on, is the current magnitude (I_{rms}) in the wire increasing, decreasing, or staying constant? Explain.

18. Is current used up in a resistor? Explain.

19. Why is it more dangerous to turn on an electric appliance when you are standing outside in bare feet than when you are inside wearing shoes with thick soles?

20. Compare the drift speeds and electric currents in two wires that are geometrically identical and the density of atoms is similar, but the number of free electrons per atom in the material of one wire is twice that in the other.

21. A voltage V is connected across a wire of length ℓ and radius r. How is the electron drift speed affected if (*a*) ℓ is doubled, (*b*) r is doubled, (*c*) V is doubled, assuming in each case that other quantities stay the same?

MisConceptual Questions

1. When connected to a battery, a lightbulb glows brightly. If the battery is reversed and reconnected to the bulb, the bulb will glow
(*a*) brighter.
(*b*) dimmer.
(*c*) with the same brightness.
(*d*) not at all.

2. When a battery is connected to a lightbulb, current flows through the lightbulb and makes it glow. How much current flows through the battery compared with the lightbulb?
(*a*) More.
(*b*) Less.
(*c*) The same amount.
(*d*) No current flows through the battery.

3. Which of the following statements about Ohm's law is true?
(*a*) Ohm's law relates the current through a wire to the voltage across the wire.
(*b*) Ohm's law holds for all materials.
(*c*) Any material that obeys Ohm's law does so independently of temperature.
(*d*) Ohm's law is a fundamental law of physics.
(*e*) Ohm's law is valid for superconductors.

4. Electrons move as a current from a battery to a lightbulb. What happens to the electrons when they reach the lightbulb?
(*a*) The electrons are used up.
(*b*) The electrons stay in the lightbulb.
(*c*) The electrons are emitted as light.
(*d*) Fewer electrons leave the bulb than enter it.
(*e*) None of the above.

5. Where in the circuit of Fig. 25–34 is the current the largest, (*a*), (*b*), (*c*), or (*d*)? Or (*e*) it is the same at all points?

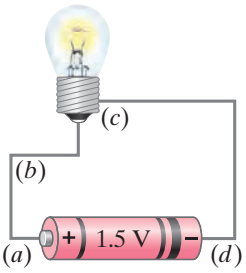


FIGURE 25–34
MisConceptual Question 5.

6. When you double the *voltage* across a certain material or device, you observe that the *current* increases by a factor of 3. What can you conclude?
(*a*) Ohm's law is obeyed, because the current increases when V increases.
(*b*) Ohm's law is not obeyed in this case.
(*c*) Ohm's law is OK if the temperature is over 100°C.

7. When current flows through a resistor,
(*a*) some of the charge is used up by the resistor.
(*b*) some of the current is used up by the resistor.
(*c*) Both (*a*) and (*b*) are true.
(*d*) Neither (*a*) nor (*b*) is true.

8. The unit kilowatt-hour is a measure of
(*a*) the rate at which energy is transformed.
(*b*) power.
(*c*) an amount of energy.
(*d*) the amount of power used per second.

9. Why might a circuit breaker open if you plug too many electrical devices into a single circuit?
(*a*) The voltage becomes too high.
(*b*) The current becomes too high.
(*c*) The resistance becomes too high.
(*d*) A circuit breaker will not "trip" no matter how many electrical devices you plug into the circuit.

10. Nothing happens when birds land on a power line, yet we are warned not to touch a power line with a ladder. What is the difference?
(*a*) Birds have extremely high internal resistance compared to humans.
(*b*) There is little to no voltage drop between a bird's two feet, but there is a significant voltage drop between the top of a ladder touching a power line and the bottom of the ladder on the ground.
(*c*) Dangerous current comes from the ground only.
(*d*) Most birds don't understand the situation.

11. When a light switch is turned on, the light comes on immediately because
(*a*) the electrons coming from the power source move through the initially empty wires very fast.
(*b*) the electrons already in the wire are instantly "pushed" by a voltage difference.
(*c*) the lightbulb may be old with low resistance. It would take longer if the bulb were new and had high resistance.
(*d*) the electricity bill is paid. The electric company can make it take longer when the bill is unpaid.

12. A constant voltage is applied across a circuit. If the resistance in the circuit is doubled, what is the effect on the power dissipated by the circuit?
(*a*) The power dissipated remains constant.
(*b*) The power dissipated is doubled.
(*c*) The power dissipated is reduced by a factor of 4.
(*d*) The power dissipated is reduced by a factor of 2.
(*e*) The power dissipated is quadrupled.

13. Which of the following is not true about alternating current?
(*a*) Alternating current cannot deliver power because the average current and voltage are each zero.
(*b*) $V_0/V_{\text{rms}} = I_0/I_{\text{rms}}$.
(*c*) $V = IR$ holds for voltage across a resistor.
(*d*) rms values are sometimes called effective values.

14. Which of the following is true about the microscopic view of current in a metal?
(*a*) Drift speed is proportional to the current density.
(*b*) For a given voltage and length and thickness of wire, current density is independent of the type of wire.
(*c*) Electrons drift at nearly the speed of light.
(*d*) Electrons drift in a straight line down a wire.
(*e*) Drift speed in a wire is on the order of magnitude of mm/s.

Problems

25–2 and 25–3 Electric Current, Resistance, Ohm's Law

(*Note*: The charge on one electron is 1.60×10^{-19} C.)

1. (I) What voltage will produce 0.25 A of current through a 5400-Ω resistor?

2. (I) What is the resistance of a toaster if 120 V produces a current of 4.2 A?

3. (I) A current of 1.75 A flows in a wire. How many electrons are flowing past any point in the wire per second?

4. (I) How many positive coulombs are stored in a fully-charged 95 ampere-hour car battery?

5. (II) What is the current in amperes if 1200 Na^+ ions flow across a cell membrane in 2.8 μs? The charge on the sodium is the same as on an electron, but positive.

6. (II) A bird stands on a dc electric transmission line carrying 3800 A (Fig. 25–35). The line has 2.5×10^{-5} Ω resistance per meter, and the bird's feet are 4.0 cm apart. What is the potential difference between the bird's feet?



FIGURE 25–35 Problem 6.

7. (II) (*a*) What is the current in the 13-Ω heating element of a 240-V clothes dryer? (*b*) How much charge passes through the element in 15 min? (Assume direct current.)

8. (II) A 12-V battery causes a current of 0.50 A through a resistor. (*a*) What is its resistance, and (*b*) how many joules of energy does the battery lose in a minute?

9. (II) A hair dryer draws 13.5 A when plugged into a 120-V line. (*a*) What is its resistance? (*b*) How much charge passes through it in 8.0 min? (Assume direct current.)

10. (II) A 4.5-V battery is connected to a bulb whose resistance is 2.3 Ω. How many electrons leave the battery per minute?

11. (II) An electric device draws 5.25 A at 240 V. (*a*) If the power company reduces the voltage by 15% because of a "brown out," what will be the current, assuming nothing else changes? (*b*) If the resistance of the device were reduced by 7.5%, what current would be drawn at the lower voltage?

25–4 Resistivity

12. (I) What is the diameter of a 1.00-m length of tungsten wire whose resistance is 0.38 Ω?

13. (I) What is the resistance of a 5.1-m length of copper wire 1.5 mm in diameter?

14. (II) Calculate the ratio of the resistance of 10.0 m of aluminum wire 2.2 mm in diameter, to 27.0 m of copper wire 1.8 mm in diameter.

15. (II) Can a 1.8-mm-diameter copper wire have the same resistance as a tungsten wire of the same length? Give numerical details.

16. (II) A certain copper wire has a resistance of 18.5 Ω. At what point along its length must the wire be cut so that the resistance of one piece is 4.0 times the resistance of the other? What is the resistance of each piece?

17. (II) How much would you have to raise the temperature of a copper wire (originally at 20°C) to increase its resistance by 16%?

18. (II) Two aluminum wires have the same resistance. If one has twice the length of the other, what is the ratio of the diameter of the longer wire to the diameter of the shorter wire?

19. (II) A 100-W halogen incandescent lightbulb may have a resistance of about 12 Ω when cold (20°C). What is its resistance when "on", at 2900 K? Assume an average temperature coefficient of resistivity $\alpha = 0.0045\ (\text{C}^\circ)^{-1}$.

20. (II) Compute the voltage drop along an 18-m length of household no. 14 copper wire (used in 15-A circuits). The wire has diameter 1.628 mm and carries a 12-A current.

21. (II) A rectangular solid made of carbon has sides of lengths 1.0 cm, 2.0 cm, and 4.0 cm, lying along the x, y, and z axes, respectively (Fig. 25–36). Determine the resistance for current that passes through the solid in (*a*) the x direction, (*b*) the y direction, and (*c*) the z direction. Assume the resistivity is $\rho = 3.0 \times 10^{-5}\ \Omega \cdot \text{m}$.

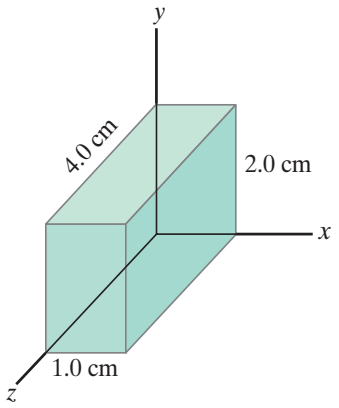


FIGURE 25–36 Problem 21.

22. (II) A length of wire is cut in half and the two lengths are wrapped together side by side to make a thicker wire. How does the resistance of this new combination compare to the resistance of the original wire?

23. (II) A length of aluminum wire is connected to a precision 10.00-V power supply, and a current of 0.4212 A is precisely measured at 23.5°C. The wire is placed in a new environment of unknown temperature where the measured current is 0.3646 A. What is the unknown temperature?

24. (III) For some applications, it is important that the value of a resistance not change with temperature. For example, suppose you made a 3.60-kΩ resistor from a carbon resistor and a Nichrome wire-wound resistor connected together so the total resistance is the sum of their separate resistances. What value should each of these resistors have (at 0°C) so that the combination is temperature independent?

25. (III) Determine a formula for the total resistance of a spherical shell made of material whose conductivity is σ and whose inner and outer radii are r_1 and r_2. Assume the current flows radially outward.

26. (III) The filament of an incandescent lightbulb has a resistance of 12 Ω at 20°C and 140 Ω when hot. (*a*) Calculate the temperature of the filament when it is hot, and take into account the change in length and area of the filament due to thermal expansion (assume tungsten for which the thermal expansion coefficient is $\approx 5.5 \times 10^{-6}\ \text{C}^{\circ -1}$). (*b*) In this temperature range, what is the percentage change in resistance due to thermal expansion, and what is the percentage change in resistance due solely to the change in ρ? Use Eq. 25–5.

27. (III) A hollow cylindrical resistor with inner radius r_1 and outer radius r_2, and length ℓ, is made of a material whose resistivity is ρ (Fig. 25–37). (*a*) Show that the resistance is given by

$$R = \frac{\rho}{2\pi\ell} \ln \frac{r_2}{r_1}$$

for current that flows radially outward. [*Hint*: Divide the resistor into concentric cylindrical shells and integrate.] (*b*) Evaluate the resistance R for such a resistor made of carbon whose inner and outer radii are 1.0 mm and 1.8 mm and whose length is 2.4 cm. (Choose $\rho = 15 \times 10^{-5}\ \Omega \cdot \text{m}$.) (*c*) What is the resistance in part (*b*) for current flowing *parallel* to the axis?

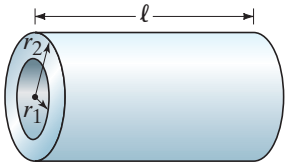


FIGURE 25–37 Problem 27.

25–5 and 25–6 Electric Power

28. (I) What is the maximum power consumption of a 3.0-V portable electronic device that draws a maximum of 240 mA of current?

29. (I) The heating element of an electric oven is designed to produce 3.1 kW of heat when connected to a 240-V source. What must be the resistance of the element?

30. (I) What is the maximum current that can pass through a 4.2-kΩ resistor rated at $\frac{1}{4}$ watt? What happens if the current is greater than this?

31. (I) An electric car uses a 45-kW (160-hp) motor. If the battery pack is designed for 340 V, what current would the motor need to draw from the battery? Neglect any energy losses in getting energy from the battery to the motor.

32. (II) An electric car has a battery that can hold 16 kWh of energy (approximately 6×10^7 J). If the battery is designed to operate at 340 V, how many coulombs of charge would need to leave the battery at 340 V and return at 0 V to equal the stored energy of the battery?

33. (II) A 120-V hair dryer has two settings: 1250 W and 1875 W. (*a*) At which setting do you guess the resistance to be higher? After making a guess, determine the resistance at (*b*) the lower setting, and (*c*) the higher setting.

34. (II) A 110-V fish-tank heater is rated at 130 W. Calculate (*a*) the current through the heater when it is operating, and (*b*) its resistance.

35. (II) You buy a 75-W lightbulb in Europe, where electricity is delivered at 240 V. If you use the bulb in the United States at 120 V (assume its resistance does not change), how bright will it be relative to 75-W 120-V bulbs? [*Hint*: Assume roughly that brightness is proportional to power consumed.]

36. (II) How many kWh of energy does a 650-W toaster use in the morning if it is in operation for a total of 5.0 min? At a cost of 13 euro cents/kWh, estimate how much this would add to your monthly electric energy bill while you were in Europe if you made toast four mornings per week.

37. (II) At $0.12/kWh, what does it cost to leave a 25-W porch light on day and night for a year?

38. (II) What is the total amount of energy stored in a 12-V, 85 A · h car battery when it is fully charged? Give result in J and kWh.

39. (II) A flashlight uses two AA 1.5-V batteries connected in series to provide 3.0 V across the bulb, as in Fig. 25–4b. The bulb draws 135 mA when turned on. (*a*) Calculate the resistance of the bulb and the power dissipated. (*b*) By what factor would the power increase if four AA batteries in series (total 6.0 V) were used with the same bulb? (Neglect heating effects of the filament.) Why shouldn't you try this?

40. (II) An extension cord made of two wires of diameter 0.129 cm (no. 16 copper wire) and of length 2.7 m (9 ft) is connected to an electric heater which draws 16.0 A on a 120-V line. How much power is dissipated in the cord?

41. (II) A power station delivers 750 kW of power at 12,000 V to a factory through wires with total resistance 3.0 Ω. How much less power is wasted if the electricity is delivered at 50,000 V rather than 12,000 V?

42. (II) You want to design a portable electric blanket that runs on a 1.5-V battery. If you use a 0.50-mm-diameter copper wire as the heating element, how long should the wire be if you want to generate 18 W of heating power? What happens if you accidentally connect the blanket to a 9.0-V battery?

43. (II) A traditional incandescent lamp filament may have been lit to a temperature of 2700 K. A contemporary halogen incandescent lamp filament may be at around 2900 K. (*a*) Estimate the percent increase in efficiency of the halogen bulb over the traditional one. [*Hint*: See Section 19–10.] (*b*) To produce the same amount of light as a traditional 100-W bulb, estimate what wattage a halogen bulb should use.

44. (III) A small immersion heater can be used in a car to heat a cup of water for coffee or tea. If the heater can heat 120 mL of water from 25°C to 95°C in 7.0 min, (*a*) approximately how much current does it draw from the car's 12-V battery, and (*b*) what is its resistance? Assume the manufacturer's claim of 85% efficiency.

45. (III) A 1.0-m-long round tungsten wire is to reach a temperature of 3100 K when a current of 18.0 A flows through it. What diameter should the wire be? Assume the wire loses energy only by radiation (emissivity $\epsilon = 1.0$, Section 19–10) and the surrounding temperature is 20°C.

25–7 Alternating Current

46. (I) Calculate the peak current in a 2.5-kΩ resistor connected to a 220-V rms ac source.

47. (I) An ac voltage, whose peak value is 180 V, is across a 350-Ω resistor. What are the rms and peak currents in the resistor?

48. (II) Estimate the resistance of the 120-V_{rms} circuits in your house as seen by the power company, when (*a*) everything electrical is unplugged, and (*b*) a 1500-W heater is on.

49. (II) The peak value of an alternating current in a 1750-W device is 6.1 A. What is the rms voltage across it?

50. (II) An 1800-W arc welder is connected to a 660-V_{rms} ac line. Calculate (*a*) the peak voltage and (*b*) the peak current.

51. (II) An average power output of 150 W is sent into a 4-Ω loudspeaker (see Fig. 25–14). What are the rms voltage and the rms current fed to the speaker (*a*) at the highest rated power of 150 W, and (*b*) at 1.0 W when the volume is turned down?

52. (II) Determine (*a*) the maximum instantaneous power transformed by a 2.7-hp pump connected to a 240-V_{rms} ac power source, and (*b*) the maximum current passing through the pump.

53. (II) For a time-dependent voltage $V(t)$, which is periodic with period T, the rms voltage is defined to be $V_{rms} = \left[\frac{1}{T}\int_0^T V^2\,dt\right]^{\frac{1}{2}}$. Use this definition to determine V_{rms} (in terms of the peak voltage V_0) for (*a*) a sinusoidal voltage, i.e., $V(t) = V_0 \sin(2\pi t/T)$ for $0 \le t \le T$; and (*b*) a positive square-wave voltage, i.e.,

$$V(t) = \begin{cases} V_0 & 0 \le t \le \dfrac{T}{2} \\ 0 & \dfrac{T}{2} \le t \le T. \end{cases}$$

54. (II) A heater coil connected to a 240-V_{rms} ac line has a resistance of 32 Ω. (*a*) What is the average power used? (*b*) What are the maximum and minimum values of the instantaneous power?

25–8 Microscopic View of Electric Current

55. (II) A 0.65-mm-diameter copper wire carries a tiny current of 3.2 μA. Estimate (*a*) the electron drift speed, (*b*) the current density, and (*c*) the electric field in the wire.

56. (II) A 5.30-m length of 2.0-mm-diameter wire carries a 750-mA current when 22.0 mV is applied to its ends. If the drift speed is 1.7×10^{-5} m/s, determine (*a*) the resistance R of the wire, (*b*) the resistivity ρ, (*c*) the current density j, (*d*) the electric field inside the wire, and (*e*) the number n of free electrons per unit volume.

57. (II) At a point high in the Earth's atmosphere, He^{2+} ions in a concentration of $2.8 \times 10^{12}/m^3$ are moving due north at a speed of 2.0×10^6 m/s. Also, a $7.0 \times 10^{11}/m^3$ concentration of O_2^- ions is moving due south at a speed of 6.2×10^6 m/s. Determine the magnitude and direction of the current density $\vec{\mathbf{j}}$ at this point.

*25–10 Nerve Conduction

***58.** (I) What is the magnitude of the electric field across an axon membrane 1.0×10^{-8} m thick if the resting potential is −70 mV?

***59.** (II) A neuron is stimulated with an electric pulse. The action potential is detected at a point 3.40 cm down the axon 0.0052 s later. When the action potential is detected 7.20 cm from the point of stimulation, the time required is 0.0063 s. What is the speed of the electric pulse along the axon? (Why are two measurements needed instead of only one?)

***60.** (III) During an action potential, Na^+ ions move into the cell at a rate of about 3×10^{-7} mol/m² · s. How much power must be produced by the "active Na^+ pumping" system to produce this flow against a +30-mV potential difference? Assume that the axon is 10 cm long and 20 μm in diameter.

General Problems

61. A person accidentally leaves a car with the lights on. If each of the two headlights uses 55 W and each of the two taillights 6 W, for a total of 122 W, how long will a fresh 12-V battery last if it is rated at 75 A · h? Assume the full 12 V appears across each bulb.

62. What is the average current drawn by a 1.0-hp 120-V motor? (1 hp = 746 W.)

63. The **conductance** G of an object is defined as the reciprocal of the resistance R; that is, $G = 1/R$. The unit of conductance is a *mho* $(= \text{ohm}^{-1})$, which is also called the *siemens* (S). What is the conductance (in siemens) of an object that draws 380 mA of current at 3.0 V?

64. The heating element of a 110-V, 1500-W heater is 4.2 m long. If it is made of iron, what must its diameter be?

65. (*a*) A particular household uses a 2.2-kW heater 2.0 h/day ("on" time), four 100-W lightbulbs 6.0 h/day, a 3.0-kW electric stove element for a total of 1.0 h/day, and miscellaneous power amounting to 2.0 kWh/day. If electricity costs \$0.115 per kWh, what will be their monthly bill (30 d)? (*b*) How much coal (which produces 7500 kcal/kg) must be burned by a 35%-efficient power plant to provide the yearly needs of this household?

66. A tungsten filament used in a flashlight bulb operates at 0.20 A and 3.0 V. If its resistance at 20°C is 1.5 Ω, what is the temperature of the filament when the flashlight is on?

67. The current in an electromagnet connected to a 240-V line is 21.5 A. At what rate must cooling water pass over the coils for the water temperature to rise no more than 5.50 C°?

68. A sequence of potential differences V is applied across a wire (diameter = 0.32 mm, length = 11 cm) and the resulting currents I are measured as follows:

V (V)	0.100	0.200	0.300	0.400	0.500
I (mA)	72	142	218	290	357

(*a*) If this wire obeys Ohm's law, graphing I vs. V will result in a straight-line plot. Explain why this is so and determine the theoretical predictions for the straight line's slope and y-intercept. (*b*) Plot I vs. V. Based on this plot, can you conclude that the wire obeys Ohm's law (i.e., did you obtain a straight line with the expected y-intercept, within the values of the significant figures)? If so, determine the wire's resistance R. (*c*) Calculate the wire's resistivity and use Table 25–1 to identify the solid material from which it is composed.

69. For the wire in Fig. 25–38, whose diameter varies uniformly from a to b as shown, suppose a current $I = 2.5$ A enters at a. If $a = 2.5$ mm and $b = 4.0$ mm, what is the current density (assume uniform) at each end?

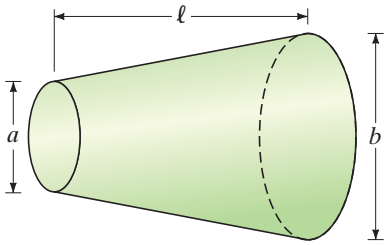


FIGURE 25–38 Problems 69 and 70.

70. The cross section of a portion of wire increases uniformly as shown in Fig. 25–38 so it has the shape of a truncated cone. The diameter at one end is a and at the other it is b, and the total length along the axis is ℓ. If the material has resistivity ρ, determine the resistance R between the two ends in terms of a, b, ℓ, and ρ. Assume that the current flows uniformly through each section, and that the taper is small, i.e., $(b - a) \ll \ell$.

71. Since $1\,\text{kWh} = 3.6 \times 10^6\,\text{J}$, how much mass must be lifted against gravity through 1.0 meter to do the equivalent amount of work?

72. An 1875-W hair dryer is designed for 117 V. (*a*) What will be the percentage change in power output if the voltage drops to 105 V? Assume no change in resistance. (*b*) How would the actual change in resistivity with temperature affect your answer?

73. The wiring in a house must be thick enough so it does not become so hot as to start a fire. What diameter must a copper wire be if it is to carry a maximum current of 35 A and produce no more than 1.5 W of heat per meter of length?

74. A microwave oven running at 72% efficiency delivers 950 W to the interior. Find (*a*) the power drawn from the source, and (*b*) the current drawn. Assume a source voltage of 120 V.

75. A 1.00-Ω wire is stretched uniformly to 1.25 times its original length. What is its resistance now?

76. 220 V is applied to two different conductors made of the same material. One conductor is twice as long and twice the diameter of the second. What is the ratio of the power transformed in the first relative to the second?

77. An electric power plant can produce electricity at a fixed power P, but the plant operator is free to choose the voltage V at which it is produced. This electricity is carried as an electric current I through a transmission line (resistance R) from the plant to the user, where it provides the user with electric power P'. (*a*) Show that the reduction in power $\Delta P = P - P'$ due to transmission losses is given by $\Delta P = P^2R/V^2$. (*b*) In order to reduce power losses during transmission, should the operator choose V to be as large or as small as possible?

78. A 2800-W oven is connected to a 240-V source. (*a*) What is the resistance of the oven's heating elements? (*b*) How long will it take to bring 150 mL of 15°C water to 100°C assuming 65% efficiency? (*c*) How much will this cost at 11 cents/kWh?

79. A proposed electric vehicle makes use of storage batteries as its source of energy. It is powered by 24 batteries, each 12 V, 95 A · h. Assume that the car is driven on level roads at an average speed of 45 km/h, and the average friction force is 440 N. Assume 100% efficiency and neglect energy used for acceleration. No energy is consumed when the vehicle is stopped, since the engine doesn't need to idle. (*a*) Determine the horsepower required. (*b*) After approximately how many kilometers must the batteries be recharged?

80. Determine the resistance of the tungsten filament in a 75-W 120-V incandescent lightbulb (*a*) at its operating temperature of about 2800 K, (*b*) at room temperature.

81. Suppose a current is given by the equation $I = 1.40 \sin 210t$, where I is in amperes and t in seconds. (*a*) What is the frequency? (*b*) What is the rms value of the current? (*c*) If this is the current through a 21.5-Ω resistor, write the equation that describes the voltage as a function of time.

82. A 100-W, 120-V incandescent lightbulb has a resistance of 12 Ω when cold (20°C) and 150 Ω when on (hot). Calculate its power consumption (*a*) at the instant it is turned on, and (*b*) after a few moments when it is hot.

83. A small city requires about 15 MW of power. Suppose that instead of using high-voltage lines to supply the power from a power plant 45 km away, the power is delivered at 120 V. Assuming a two-wire line of 0.50-cm-diameter copper wire, estimate the cost of the energy lost to heat per hour. Assume the cost of electricity is about 12 cents per kWh.

84. In an automobile, the system voltage varies from about 12 V when the car is off to about 13.8 V when the car is on and the car's generator (or "alternator") is in operation, a difference of 15%. By what percentage does the power delivered to the headlights vary as the voltage changes from 12 V to 13.8 V? Assume the headlight resistance remains constant.

85. An air conditioner draws 18 A at 220-V ac. The connecting cord is copper wire with a diameter of 1.62 mm. (*a*) How much total power does the air conditioner draw? (*b*) If the length of the cord (containing two wires) is 3.5 m, how much power is dissipated in the wiring? (*c*) If wire with a diameter of 2.05 mm was used instead, how much power would be dissipated in the wiring? (*d*) Assuming that the air conditioner is run 12 h per day, how much money per month (30 days) would be saved by using the thicker wire? Assume that the cost of electricity is 12 cents per kWh.

86. Lightbulb A is rated at 120 V and 40 W for household applications. Lightbulb B is rated at 12 V and 40 W for automotive applications. (*a*) What is the current through each bulb? (*b*) What is the resistance of each bulb? (*c*) In one hour, how much charge passes through each bulb? (*d*) In one hour, how much energy does each bulb use? (*e*) Which bulb requires larger diameter wires to connect its power source and the bulb?

87. If a wire of resistance R is stretched uniformly so that its length doubles, by what factor does the power dissipated in the wire change, assuming it remains hooked up to the same voltage source? Assume the wire's volume and density remain constant.

88. An electric wheelchair is designed to run on a single 12-V battery rated to provide 85 ampere-hours (85 A · h). (*a*) How much energy is stored in this battery? (*b*) If the wheelchair experiences an average total retarding force (mainly friction) of 260 N, how far can the wheelchair travel on one charge?

89. Battery-powered electricity is very expensive compared with that available from a wall receptacle. Estimate the cost per kWh of (*a*) an alkaline D-cell (cost $1.70) and (*b*) an alkaline AA-cell (cost $1.25). These batteries can provide a continuous current of 25 mA for 820 h and 120 h, respectively, at 1.5 V. (*c*) Compare to the cost of a normal 120-V ac house source at $0.11/kWh.

90. Copper wire of diameter 0.259 cm is used to connect a set of appliances at 120 V, which draw 1250 W of power total. (*a*) What power is wasted in 25.0 m of this wire? (*b*) What is your answer if wire of diameter 0.412 cm is used?

91. A copper pipe has an inside diameter of 3.00 cm and an outside diameter of 5.00 cm (Fig. 25–39). What is the resistance to a dc current of a 12.0-m length of this pipe?

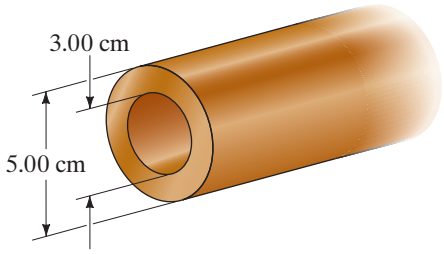


FIGURE 25–39 Problem 91.

***92.** The level of liquid helium (temperature ≈ 4 K) in its storage tank can be monitored using a vertically aligned niobium–titanium (NbTi) wire, whose length ℓ spans the height of the tank. In this level-sensing setup, an electronic circuit maintains a constant electrical current I at all times in the NbTi wire and a voltmeter monitors the voltage V across this wire. The NbTi wire is superconducting ($R = 0$) if below its transition temperature of 10 K, so the portion of the wire immersed in the liquid helium is in the superconducting state, while the portion above the liquid (in helium vapor with temperature above 10 K) is in the normal state. Define $f = x/\ell$ to be the fraction of the tank filled with liquid helium (Fig. 25–40) and V_0 to be the value of the voltage V when the tank is empty ($f = 0$). Determine the relation between f and V (in terms of V_0).

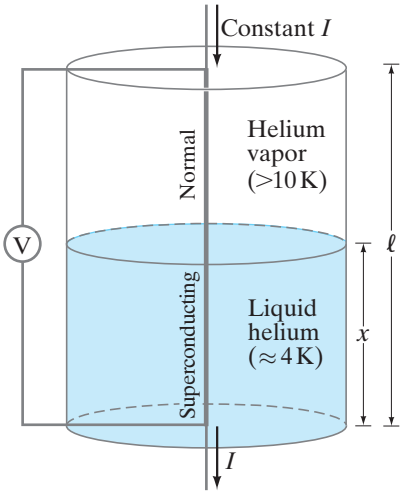


FIGURE 25–40 Problem 92.

93. An electric heater is used to heat a room of volume 65 m^3. Air is brought into the room at 5°C and is completely replaced twice per hour. Heat loss through the walls amounts to approximately 750 kcal/h. If the air is to be maintained at 22°C, what minimum wattage must the heater have? (The specific heat of air is about 0.17 kcal/kg · C°.)

94. You find a small cylindrical resistor that measures 9.00 mm in length and 2.15 mm in diameter, and it has a color code of red, yellow, brown, and gold. What is the resistor made of primarily?

95. Household wiring has sometimes used aluminium instead of copper. (*a*) Using Table 25–1, find the ratio of the resistance of a copper wire to that of an aluminum wire of the same length and diameter. (*b*) Typical copper wire used for home wiring in the U.S. has a diameter of 1.63 mm. What is the resistance of 125 m of this wire? (*c*) What would be the resistance of the same wire if it were made of aluminum? (*d*) How much power would be dissipated in each wire if it carried 18 A of current? (*e*) What should be the diameter of the aluminum wire for it to have the same resistance as the copper wire? (*f*) In Section 25–4, a statement is made about the resistance of copper and aluminum wires of the same weight. Using Table 13–1 for the densities of copper and aluminum, find the resistance of an aluminum wire of the same mass and length as the copper wire in part (*b*). Is the statement true?

96. The LHC accelerator at CERN (Switzerland) is designed to carry a 0.58-A beam of protons $\left(q = 1.6 \times 10^{-19}\,\text{C}\right)$ traveling at very nearly the speed of light $\left(3.0 \times 10^{8}\,\text{m/s}\right)$ around a ring 27 km in circumference. How many protons are in the beam?

97. Estimate how far can an average electron move along one of the connecting wires of a 750-W toaster during an alternating current cycle? The power cord has copper wires of diameter 1.7 mm and is plugged into a 60-Hz 120-V ac outlet. [*Hint*: For sinusoidal motion, Chapter 14, we saw that the maximum distance traveled from equilibrium (amplitude A) is proportional to the maximum (drift) speed (Eq. 14–9a). This maximum drift speed is related to the maximum current (Section 25–8), which is calculated as the first step here; see Chapter 14.]

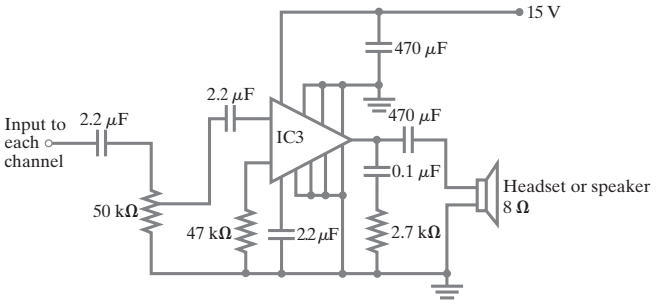
98. Small changes in the length of an object can be measured using a **strain gauge** sensor, which is a wire that when undeformed has length ℓ_0, cross-sectional area A_0, and resistance R_0. This sensor is rigidly affixed to the object's surface, aligning its length in the direction in which length changes are to be measured. As the object deforms, the length of the wire sensor changes by $\Delta\ell$, and the resulting change ΔR in the sensor's resistance is measured. Assuming that as the solid wire is deformed to a length ℓ, its density and volume remain constant (only approximately valid), show that the strain $\left(= \Delta\ell/\ell_0\right)$ of the wire sensor, and thus of the object to which it is attached, is approximately $\Delta R/2R_0$.

ANSWERS TO EXERCISES

A: 1.6×10^{-13} A.
B: 3600 C.
C: (*c*).
D: (*c*), (*b*).
E: (*c*).
F: 105 m.
G: (*d*).
H: (*e*) 40.
I: (*a*) 28 V, 3.5 A; (*b*) 2.8 V, 0.35 A.

This cell phone has an attachment that measures a person's blood sugar level, and plots it over a period of days. All electronic devices contain circuits that are dc, at least in part. The circuit diagram (below right) shows a possible amplifier circuit for an audio output (cell phone ear piece). We have already met two of the circuit elements shown: resistors and capacitors, and we discuss them in circuits in this Chapter. (The large triangle is an amplifier chip containing transistors.) We also discuss how voltmeters and ammeters work, and how measurements affect the quantity being measured.





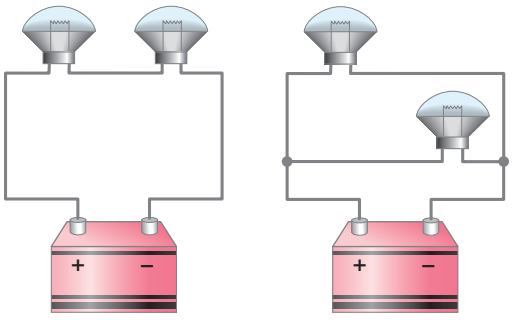
CHAPTER 26

DC Circuits

CHAPTER-OPENING QUESTION—Guess now!

The automobile headlight bulbs shown in the circuits here are identical. The battery connection which produces more light is

(a) circuit 1.
(b) circuit 2.
(c) both the same.
(d) not enough information.



Circuit 1 Circuit 2

Electric circuits are basic parts of all electronic devices from cell phones and TV sets to computers and automobiles. Scientific measurements—whether in physics, biology, or medicine—make use of electric circuits. In Chapter 25 we discussed the basic principles of electric current. Now we apply these principles to analyze dc circuits involving combinations of batteries, resistors, and capacitors. We also study the operation of some useful instruments.†

When we draw a diagram for a circuit, we represent batteries, capacitors, and resistors by the symbols shown in Table 26–1. Wires whose resistance is negligible compared with other resistance in the circuit are drawn as straight lines. A ground symbol (⏚ or ⏚) may mean a real connection to the ground, perhaps via a metal pipe, or it may mean a common connection, such as the frame of a car.

For the most part in this Chapter, except in Section 26–5 on *RC* circuits, we will be interested in circuits operating in their steady state. We won't be looking at a circuit at the moment a change is made in it, such as when a battery or resistor is connected or disconnected, but only when the currents have reached their steady values.

†AC circuits that contain only a voltage source and resistors can be analyzed like the dc circuits in this Chapter. However, ac circuits that contain capacitors and other circuit elements are more complicated, and we discuss them in Chapter 30.

CONTENTS

26–1 EMF and Terminal Voltage
26–2 Resistors in Series and in Parallel
26–3 Kirchhoff's Rules
26–4 EMFs in Series and in Parallel; Charging a Battery
26–5 *RC* Circuits—Resistor and Capacitor in Series
26–6 Electric Hazards and Safety
26–7 Ammeters and Voltmeters—Measurement Affects Quantity Measured

TABLE 26–1 Symbols for Circuit Elements

Symbol	Device
⊣⊢	Battery
⊣⊢ or ⊣(	Capacitor
∿∿	Resistor
——	Wire with negligible resistance
—/—	Switch
⏚ or ⏚	Ground

26–1 EMF and Terminal Voltage

To have current in an electric circuit, we need a device such as a battery or an electric generator that transforms one type of energy (chemical, mechanical, or light, for example) into electric energy. Such a device is called a **source** of **electromotive force**[†] or of **emf**. The *potential difference* between the terminals of such a source, when no current flows to an external circuit, is called the **emf** of the source. The symbol $\mathscr{E}$ is usually used for emf (don't confuse $\mathscr{E}$ with E for electric field), and its unit is volts.

⚠ CAUTION

Why battery voltage isn't perfectly constant

A battery is not a source of constant current—the current out of a battery varies according to the resistance in the circuit. A battery *is*, however, a nearly constant voltage source, but not perfectly constant as we now discuss. For example, if you start a car with the headlights on, you may notice the headlights dim. This happens because the starter draws a large current, and the battery voltage drops below its rated emf as a result. The voltage drop occurs because the chemical reactions in a battery (Section 25–1) cannot supply charge fast enough to maintain the full emf. For one thing, charge must move (within the electrolyte) between the electrodes inside the battery, and there is always some hindrance to completely free flow. Thus, a battery itself has some resistance, which is called its **internal resistance**; it is usually designated r.

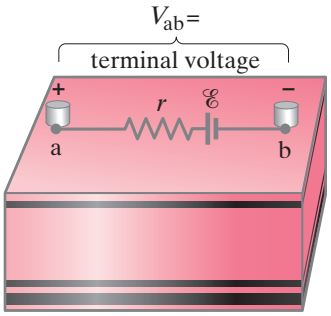


FIGURE 26–1 Diagram for an electric cell or battery.

A real battery is modeled as if it were a perfect emf $\mathscr{E}$ in series with a resistor r, as shown in Fig. 26–1. Since this resistance r is inside the battery, we can never separate it from the battery. The two points a and b in Fig. 26–1 represent the two terminals of the battery. What we measure is the **terminal voltage** $V_{ab} = V_a - V_b$. When no current is drawn from the battery, the terminal voltage equals the emf, which is determined by the chemical reactions in the battery: $V_{ab} = \mathscr{E}$. However, when a current I flows from the battery there is an internal drop in voltage equal to Ir. Thus the terminal voltage (the actual voltage applied to a circuit) is[‡]

$$V_{ab} = \mathscr{E} - Ir. \qquad \text{[current } I \text{ flows from battery]} \quad \textbf{(26–1)}$$

For example, if a 12-V battery has an internal resistance of 0.1 Ω, then when 10 A flows from the battery, the terminal voltage is $12\text{ V} - (10\text{ A})(0.1\ \Omega) = 11\text{ V}$. The internal resistance of a battery is usually small. For example, an ordinary flashlight battery when fresh may have an internal resistance of perhaps 0.20 Ω. (However, as it ages and the electrolyte dries out, the internal resistance increases to many ohms.) Good car batteries have even lower internal resistance.

EXAMPLE 26–1 **Battery with internal resistance.** A 65.0-Ω resistor is connected to the terminals of a battery whose emf is 12.0 V and whose internal resistance is 0.5 Ω, Fig. 26–2. Calculate (*a*) the current in the circuit, (*b*) the terminal voltage of the battery, V_{ab}, and (*c*) the power dissipated in the resistor R and in the battery's internal resistance r.

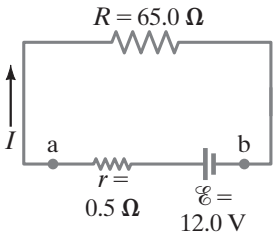


FIGURE 26–2 Example 26–1.

APPROACH We first consider the battery as a whole, which is shown in Fig. 26–2 as an emf $\mathscr{E}$ and internal resistance r between points a and b. Then we apply $V = IR$ to the circuit itself.

[†]The term "electromotive force" is a misnomer—it does not refer to a "force," which is measured in newtons. To avoid confusion, we use the abbreviation, emf.

[‡]When a battery is being charged, a current is forced to pass through it; we then have to write

$$V_{ab} = \mathscr{E} + Ir.$$

See Section 26–4 or Problem 27 and Fig. 26–51.

SOLUTION (*a*) From Eq. 26–1, we have $V_{ab} = \mathcal{E} - Ir$. We apply Ohm's law (Eq. 25–2) to this battery and the resistance R of the circuit: $V_{ab} = IR$. Hence

$$\mathcal{E} - Ir = IR$$

or

$$\mathcal{E} = I(R + r).$$

So

$$I = \frac{\mathcal{E}}{R + r} = \frac{12.0\,\text{V}}{65.0\,\Omega + 0.5\,\Omega} = \frac{12.0\,\text{V}}{65.5\,\Omega} = 0.183\,\text{A}.$$

(*b*) The terminal voltage is

$$V_{ab} = \mathcal{E} - Ir = 12.0\,\text{V} - (0.183\,\text{A})(0.5\,\Omega) = 11.9\,\text{V}.$$

(*c*) The power dissipated in R (Eq. 25–7) is

$$P_R = I^2R = (0.183\,\text{A})^2(65.0\,\Omega) = 2.18\,\text{W},$$

and in the battery's resistance r it is

$$P_r = I^2r = (0.183\,\text{A})^2(0.5\,\Omega) = 0.02\,\text{W}.$$

Unless stated otherwise, we assume the battery's internal resistance is negligible, and the battery voltage given is its terminal voltage, which we will usually write as V rather than V_{ab}. Do not confuse V (italic) for voltage, with V (not italic) for the volt unit.

26–2 Resistors in Series and in Parallel

When two or more resistors are connected end to end along a single path as shown in Fig. 26–3a, they are said to be connected in **series**. The resistors could be simple resistors as were pictured in Fig. 25–11, or they could be lightbulbs (Fig. 26–3b), or heating elements, or other resistive devices. Any charge that passes through R_1 in Fig. 26–3a will also pass through R_2 and then R_3. Hence the same current I passes through each resistor. (If it did not, this would imply either that charge was not conserved, or that charge was accumulating at some point in the circuit, which does not happen in the steady state.)

We let V represent the potential difference (voltage) across all three resistors in Fig. 26–3a. We assume all other resistance in the circuit can be ignored, so V equals the terminal voltage supplied by the battery. We let V_1, V_2, and V_3 be the potential differences across each of the resistors, R_1, R_2, and R_3, respectively. From Ohm's law, $V = IR$, we can write $V_1 = IR_1$, $V_2 = IR_2$, and $V_3 = IR_3$. Because the resistors are connected end to end, energy conservation tells us that the total voltage V is equal to the sum of the voltages[†] across each resistor:

$$V = V_1 + V_2 + V_3 = IR_1 + IR_2 + IR_3. \qquad \text{[series]} \quad \textbf{(26–2)}$$

Now let us determine the equivalent single resistance R_{eq} that would draw the same current I as our combination of three resistors in series; see Fig. 26–3c. Such a single resistance R_{eq} would be related to V by

$$V = IR_{eq}.$$

We equate this expression with Eq. 26–2, $V = I(R_1 + R_2 + R_3)$, and find

$$R_{eq} = R_1 + R_2 + R_3. \qquad \text{[series]} \quad \textbf{(26–3)}$$

When we put several resistances in series, the total or equivalent resistance is the sum of the separate resistances. Sometimes we call it "net resistance." This sum applies to any number of resistances in series.

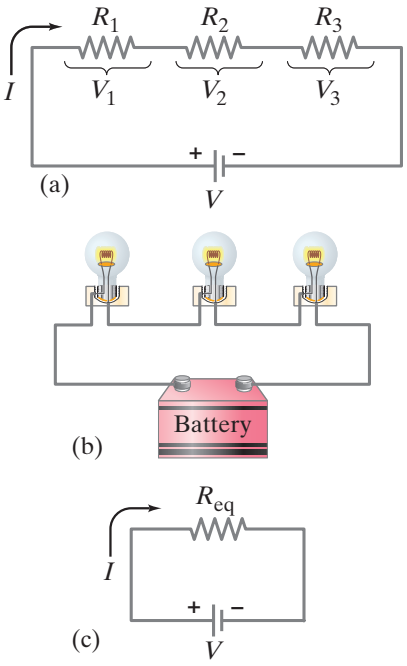


FIGURE 26–3 (a) Resistances connected in series. (b) Resistances could be lightbulbs, or any other type of resistance. (c) Equivalent single resistance R_{eq} that draws the same current: $R_{eq} = R_1 + R_2 + R_3$.

[†]To see in more detail why this is true, note that an electric charge q passing through R_1 loses an amount of potential energy equal to qV_1. In passing through R_2 and R_3, the potential energy U decreases by qV_2 and qV_3, for a total $\Delta U = qV_1 + qV_2 + qV_3$; this sum must equal the energy given to q by the battery, qV, so that energy is conserved. Hence $qV = q(V_1 + V_2 + V_3)$, and so $V = V_1 + V_2 + V_3$, which is Eq. 26–2.

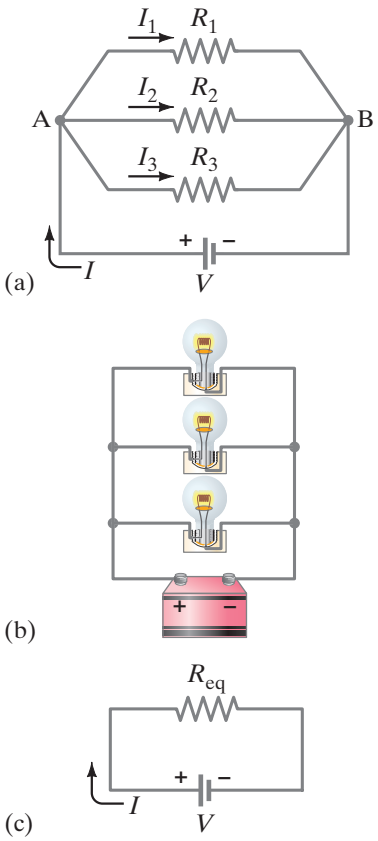


FIGURE 26–4 (a) Resistances connected in parallel. (b) Resistances could be lightbulbs. (c) The equivalent circuit with R_{eq} obtained from Eq. 26–4:

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}.$$

Note that when you add more resistance to the circuit, the current through the circuit will decrease. For example, if a 12-V battery is connected to a 4-Ω resistor, the current will be 3 A. But if the 12-V battery is connected to three 4-Ω resistors in series, the total resistance is 12 Ω and the current through the entire circuit will be only 1 A.

Another way to connect resistors is in **parallel**, so that the current from the source splits into separate branches or paths (Fig. 26–4a and b). Wiring in houses and buildings is arranged so all electric devices are in parallel, as we saw in Chapter 25, Fig. 25–20. With parallel wiring, if you disconnect one device (say, R_1 in Fig. 26–4a), the current to the other devices is not interrupted. Compare to a series circuit, where if one device (say, R_1 in Fig. 26–3a) is disconnected, the current *is* stopped to all the others (for example light bulbs).

In all of our analyses of circuits, we often make use of the fact that any two points in a circuit connected by a wire of negligible resistance are at the same potential.

In a parallel circuit, Fig. 26–4a, the total current I that leaves the battery splits into three separate paths. We let I_1, I_2, and I_3 be the currents through each of the resistors, R_1, R_2, and R_3, respectively. Because *electric charge is conserved*, the current I flowing into junction A (where the different wires or conductors meet in Fig. 26–4a) must equal the current flowing out of the junction. Thus

$$I = I_1 + I_2 + I_3. \qquad \text{[parallel]}$$

When resistors are connected in parallel, each has the same voltage across it. Hence the full voltage of the battery is applied to each resistor in Fig. 26–4a. Applying Ohm's law to each resistor, we have

$$I_1 = \frac{V}{R_1}, \quad I_2 = \frac{V}{R_2}, \quad \text{and} \quad I_3 = \frac{V}{R_3}.$$

Let us now determine what single resistor R_{eq} (Fig. 26–4c) will draw the same current I as these three resistances in parallel. This equivalent resistance R_{eq} must satisfy Ohm's law too:

$$I = \frac{V}{R_{eq}}.$$

We now combine the equations above:

$$I = I_1 + I_2 + I_3,$$

$$\frac{V}{R_{eq}} = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}.$$

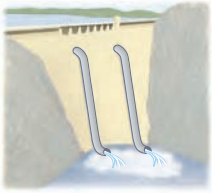
When we divide out the V from each term, we have

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}. \qquad \text{[parallel]} \quad \textbf{(26–4)}$$

For example, suppose you connect two 4-Ω loudspeakers in parallel to a single set of output terminals of an audio amplifier. The equivalent resistance of the two 4-Ω "resistors" in parallel is

$$\frac{1}{R_{eq}} = \frac{1}{4\,\Omega} + \frac{1}{4\,\Omega} = \frac{2}{4\,\Omega} = \frac{1}{2\,\Omega},$$

and so $R_{eq} = 2\,\Omega$. Thus the net (or equivalent) resistance is *less* than each single resistance. This may at first seem surprising. But remember that when you connect resistors in parallel, you are giving the electric current additional paths to follow. Hence the net resistance will be less.[†] (If you have n identical resistors in parallel, $R_{eq} = R/n$.)



[†]An analogy may help. Consider two identical pipes taking in water near the top of a dam and releasing it below as shown in the figure to the left. If both pipes are open, rather than only one, twice as much water will flow through. That is, the net resistance to the flow of water will be reduced by half with two equal pipes open, just as for electrical resistors in parallel. If both pipes are closed, the dam offers infinite resistance to the flow of water, which corresponds in the electrical case to an open circuit—when the path is not continuous and no current flows—so the electrical resistance is infinite.

Equations 26–3 and 26–4 make good sense. Recalling Eq. 25–3 for resistivity, $R = \rho\ell/A$, we see that placing resistors in series effectively increases the length and therefore the resistance; putting resistors in parallel increases the area through which current flows, thus reducing the overall resistance.

Note that whenever a group of resistors is replaced by the equivalent resistance, current and voltage and power in the rest of the circuit are unaffected.

EXERCISE A You have a 10-Ω and a 15-Ω resistor. What is the smallest and largest equivalent resistance that you can make with these two resistors?

CONCEPTUAL EXAMPLE 26–2 **Series or parallel?** (*a*) The lightbulbs in Fig. 26–5 are identical. Which configuration produces more light? (*b*) Which way do you think the headlights of a car are wired? Ignore change of filament resistance R with current.

RESPONSE (*a*) The equivalent resistance of the parallel circuit is found from Eq. 26–4, $1/R_{eq} = 1/R + 1/R = 2/R$. Thus $R_{eq} = R/2$. The parallel combination then has lower resistance $(= R/2)$ than the series combination $(R_{eq} = R + R = 2R)$. There will be more total current in the parallel configuration (2), since $I = V/R_{eq}$ and V is the same for both circuits. The total power transformed, which is related to the light produced, is $P = IV$, so the greater current in (2) means more light is produced. Another simple way to see this result is that the full 12 V is across each bulb in parallel, but only half (6 V) across each bulb in series.

(*b*) Headlights are wired in parallel (2), because if one bulb goes out, the other bulb can stay lit. If they were in series (1), if one bulb burned out, the circuit would be open and no current would flow, so neither bulb would light.

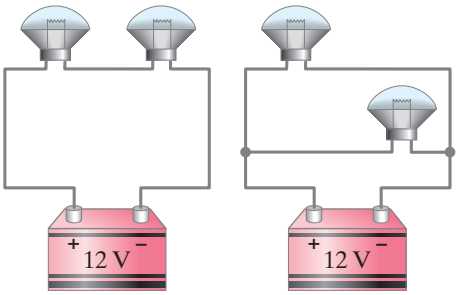


FIGURE 26–5 Example 26–2.

EXERCISE B Return to the Chapter-Opening Question, page 747, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 26–3 **Circuit with series and parallel resistors.** How much current is drawn from the battery shown in Fig. 26–6a?

APPROACH The current I that flows out of the battery all passes through the 400-Ω resistor, but then it splits into I_1 and I_2 passing through the 500-Ω and 700-Ω resistors. The latter two resistors are in parallel with each other. We look for something that we already know how to treat. So let's start by finding the equivalent resistance, R_P, of the parallel resistors, 500 Ω and 700 Ω. Then we can consider this R_P to be in series with the 400-Ω resistor.

SOLUTION The equivalent resistance, R_P, of the 500-Ω and 700-Ω resistors in parallel is

$$\frac{1}{R_P} = \frac{1}{500\ \Omega} + \frac{1}{700\ \Omega} = 0.0020\ \Omega^{-1} + 0.0014\ \Omega^{-1} = 0.0034\ \Omega^{-1}.$$

This is $1/R_P$, so we take the reciprocal to find R_P. It is a common mistake to forget to take this reciprocal. The units of reciprocal ohms, Ω^{-1}, are a reminder. Thus

$$R_P = \frac{1}{0.0034\ \Omega^{-1}} = 290\ \Omega.$$

This 290 Ω is the equivalent resistance of the two parallel resistors, and is in series with the 400-Ω resistor (see equivalent circuit, Fig. 26–6b). The total equivalent resistance R_{eq} is the sum of the 400-Ω and 290-Ω resistances, which are in series:

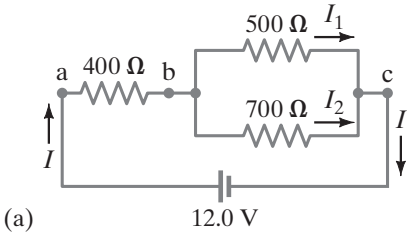
$$R_{eq} = 400\ \Omega + 290\ \Omega = 690\ \Omega.$$

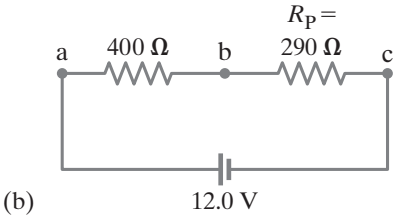
The total current flowing from the battery is then

$$I = \frac{V}{R_{eq}} = \frac{12.0\ \text{V}}{690\ \Omega} = 0.0174\ \text{A} \approx 17\ \text{mA}.$$

NOTE This I is also the current flowing through the 400-Ω resistor, but not through the 500-Ω and 700-Ω resistors (both currents are less—see the next Example).

FIGURE 26–6 (a) Circuit for Example 26–3. (b) Equivalent circuit, showing the equivalent resistance of 290 Ω for the two parallel resistors in (a).





CAUTION
Remember to take the reciprocal

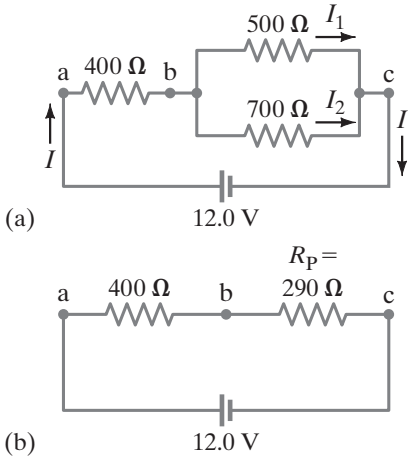


FIGURE 26–6 (Repeated.) (a) Circuit for Examples 26–3 and 26–4. (b) Equivalent circuit, showing the equivalent resistance of 290 Ω for the two parallel resistors in (a).

EXAMPLE 26–4 **Current in one branch.** What is the current I_1 through the 500-Ω resistor in Fig. 26–6a?

APPROACH We need the voltage across the 500-Ω resistor, which is the voltage between points b and c in Fig. 26–6a, and we call it V_{bc}. Once V_{bc} is known, we can apply Ohm's law, $V = IR$, to get the current. First we find the voltage across the 400-Ω resistor, V_{ab}, since we know that 17.4 mA passes through it (Example 26–3).

SOLUTION V_{ab} can be found using $V = IR$:

$$V_{ab} = (0.0174\,\text{A})(400\,\Omega) = 7.0\,\text{V}.$$

The total voltage across the network of resistors is $V_{ac} = 12.0\,\text{V}$, so V_{bc} must be $12.0\,\text{V} - 7.0\,\text{V} = 5.0\,\text{V}$. Ohm's law gives the current I_1 through the 500-Ω resistor:

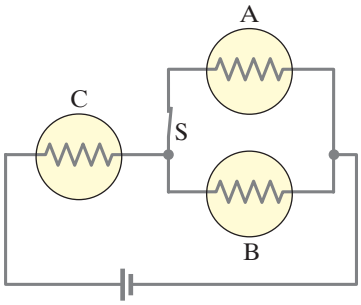
$$I_1 = \frac{5.0\,\text{V}}{500\,\Omega} = 1.0 \times 10^{-2}\,\text{A} = 10\,\text{mA}.$$

This is the answer we wanted. We can also calculate the current I_2 through the 700-Ω resistor since the voltage across it is also 5.0 V:

$$I_2 = \frac{5.0\,\text{V}}{700\,\Omega} = 7\,\text{mA}.$$

NOTE When I_1 combines with I_2 to form the total current I (at point c in Fig. 26–6a), their sum is $10\,\text{mA} + 7\,\text{mA} = 17\,\text{mA}$. This equals the total current I as calculated in Example 26–3, as it should.

FIGURE 26–7 Example 26–5, three identical lightbulbs. Each yellow circle with ⁓ inside represents a lightbulb and its resistance.

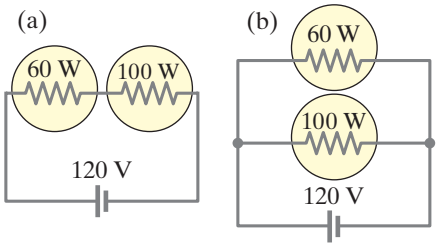


CONCEPTUAL EXAMPLE 26–5 **Bulb brightness in a circuit.** The circuit in Fig. 26–7 has three identical lightbulbs, each of resistance R. (*a*) When switch S is closed, how will the brightness of bulbs A and B compare with that of bulb C? (*b*) What happens when switch S is opened? Use a minimum of mathematics.

RESPONSE (*a*) With switch S closed, the current that passes through bulb C must split into two equal parts when it reaches the junction leading to bulbs A and B (the resistance of bulb A equals that of B). Thus, A and B each receive half of C's current; A and B will be equally bright, but less bright than C ($P = I^2R$).

(*b*) When the switch S is open, no current can flow through bulb A, so it will be dark. Now, the same current passes through bulbs B and C, so B and C will be equally bright. The equivalent resistance of this circuit ($= R + R$) is greater than that of the circuit with the switch closed, so the current leaving the battery is reduced. Thus, bulb C will be dimmer when we open the switch, but bulb B will be brighter because it gets more current when the switch is open (you may want to use some mathematics here).

FIGURE 26–8 Exercise C.



EXERCISE C A 100-W, 120-V lightbulb and a 60-W, 120-V lightbulb are connected in two different ways as shown in Fig. 26–8. In each case, which bulb glows more brightly? Ignore change of filament resistance with current (and temperature).

EXAMPLE 26–6 **ESTIMATE** **A two-speed fan.** One way a multiple-speed ventilation fan for a car can be designed is to put resistors in series with the fan motor. The resistors reduce the current through the motor and make it run more slowly. Suppose the current in the motor is 5.0 A when it is connected directly across a 12-V battery. (*a*) What series resistor should be used to reduce the current to 2.0 A for low-speed operation? (*b*) What power rating should the resistor have?

APPROACH An electric motor in series with a resistor can be treated as two resistors in series. The power comes from $P = IV$.

SOLUTION (*a*) When the motor is connected to 12 V and drawing 5.0 A, its resistance is $R = V/I = (12\text{ V})/(5.0\text{ A}) = 2.4\ \Omega$. We will assume that this is the motor's resistance for all speeds (this is only an approximation). Then, when a current of 2.0 A is flowing, the voltage across the motor is $(2.0\text{ A})(2.4\ \Omega) = 4.8\text{ V}$. The remaining $12.0\text{ V} - 4.8\text{ V} = 7.2\text{ V}$ must appear across the series resistor. When 2.0 A flows through the resistor, its resistance must be $R = (7.2\text{ V})/(2.0\text{ A}) = 3.6\ \Omega$.

(*b*) The power dissipated by the resistor is $P = (7.2\text{ V})(2.0\text{ A}) = 14.4\text{ W}$. To be safe, a power rating of 20 W would be appropriate.

EXAMPLE 26–7 **Analyzing a circuit.** A 9.0-V battery whose internal resistance r is 0.50 Ω is connected in the circuit shown in Fig. 26–9a. (*a*) How much current is drawn from the battery? (*b*) What is the terminal voltage of the battery? (*c*) What is the current in the 6.0-Ω resistor?

APPROACH To find the current out of the battery, we first need to determine the equivalent resistance R_{eq} of the entire circuit, including r, which we do by identifying and isolating simple series or parallel combinations of resistors. Once we find I from Ohm's law, $I = \mathscr{E}/R_{eq}$, we get the terminal voltage using $V_{ab} = \mathscr{E} - Ir$. For (*c*) we apply Ohm's law to the 6.0-Ω resistor.

SOLUTION (*a*) We want to determine the equivalent resistance of the circuit. But where do we start? We note that the 4.0-Ω and 8.0-Ω resistors are in parallel, and so have an equivalent resistance R_{eq1} given by

$$\frac{1}{R_{eq1}} = \frac{1}{8.0\ \Omega} + \frac{1}{4.0\ \Omega} = \frac{3}{8.0\ \Omega};$$

taking the reciprocal, $R_{eq1} = 8.0\ \Omega/3 = 2.7\ \Omega$. This 2.7 Ω is in series with the 6.0-Ω resistor, as shown in the equivalent circuit of Fig. 26–9b. The net resistance of the lower arm of the circuit is then

$$R_{eq2} = 6.0\ \Omega + 2.7\ \Omega = 8.7\ \Omega,$$

as shown in Fig. 26–9c. The equivalent resistance R_{eq3} of the 8.7-Ω and 10.0-Ω resistances in parallel is given by

$$\frac{1}{R_{eq3}} = \frac{1}{10.0\ \Omega} + \frac{1}{8.7\ \Omega} = 0.21\ \Omega^{-1},$$

so $R_{eq3} = (1/0.21\ \Omega^{-1}) = 4.8\ \Omega$. This 4.8 Ω is in series with the 5.0-Ω resistor and the 0.50-Ω internal resistance of the battery (Fig. 26–9d), so the total equivalent resistance R_{eq} of the circuit is $R_{eq} = 4.8\ \Omega + 5.0\ \Omega + 0.50\ \Omega = 10.3\ \Omega$. Hence the current drawn is

$$I = \frac{\mathscr{E}}{R_{eq}} = \frac{9.0\text{ V}}{10.3\ \Omega} = 0.87\text{ A}.$$

(*b*) The terminal voltage of the battery is

$$V_{ab} = \mathscr{E} - Ir = 9.0\text{ V} - (0.87\text{ A})(0.50\ \Omega) = 8.6\text{ V}.$$

(*c*) Now we can work back and get the current in the 6.0-Ω resistor. It must be the same as the current through the 8.7 Ω shown in Fig. 26–9c (why?). The voltage across that 8.7 Ω will be the emf of the battery minus the voltage drops across r and the 5.0-Ω resistor: $V_{8.7} = 9.0\text{ V} - (0.87\text{ A})(0.50\ \Omega + 5.0\ \Omega)$. Applying Ohm's law, we get the current (call it I')

$$I' = \frac{9.0\text{ V} - (0.87\text{ A})(0.50\ \Omega + 5.0\ \Omega)}{8.7\ \Omega} = 0.48\text{ A}.$$

This is the current through the 6.0-Ω resistor.

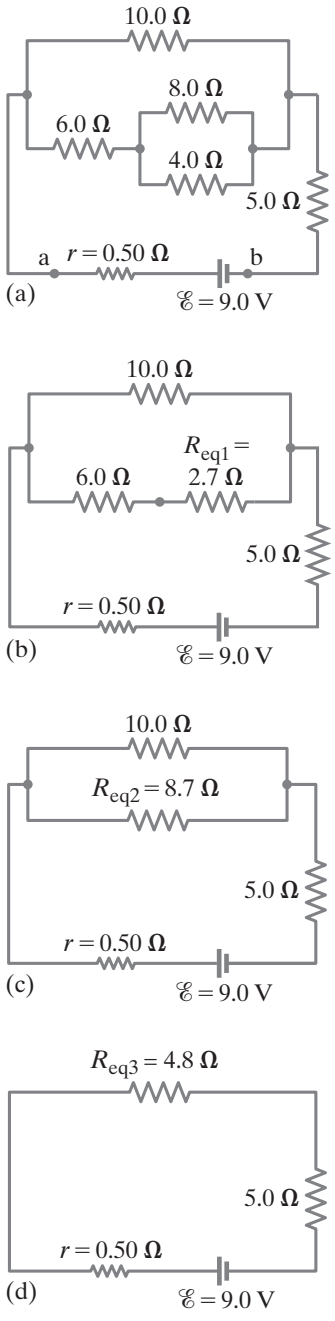


FIGURE 26–9 Circuit for Example 26–7, where r is the internal resistance of the battery.

26–3 Kirchhoff's Rules

In the last few Examples we have been able to find the currents in circuits by combining resistances in series and parallel, and using Ohm's law. This technique can be used for many circuits. However, some circuits are too complicated for that analysis. For example, we cannot find the currents in each part of the circuit shown in Fig. 26–10 simply by combining resistances as we did before.

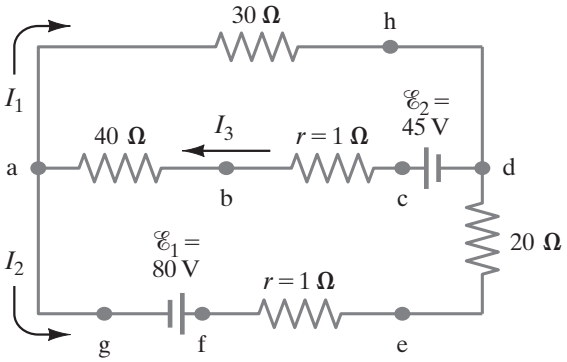


FIGURE 26–10 Currents can be calculated using Kirchhoff's rules.

To deal with complicated circuits, we use Kirchhoff's rules, devised by G. R. Kirchhoff (1824–1887) in the mid-nineteenth century. There are two rules, and they are simply convenient applications of the laws of conservation of charge and energy.

Kirchhoff's first rule or **junction rule** is based on the conservation of electric charge (we already used it to derive the equation for parallel resistors). It states that

Junction rule (conservation of charge)

at any junction point, the sum of all currents entering the junction must equal the sum of all currents leaving the junction.

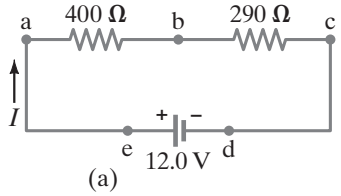
That is, whatever charge goes in must come out. For example, at the junction point a in Fig. 26–10, I_3 is entering whereas I_1 and I_2 are leaving. Thus Kirchhoff's junction rule states that $I_3 = I_1 + I_2$. We already saw an instance of this in the NOTE at the end of Example 26–4.

Kirchhoff's second rule or **loop rule** is based on the conservation of energy. It states that

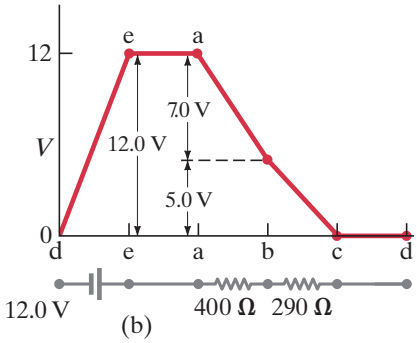
Loop rule (conservation of energy)

the sum of the changes in potential around any closed loop of a circuit must be zero.

FIGURE 26–11 Changes in potential around the circuit in (a) are plotted in (b).



(a)



(b)

To see why this rule should hold, consider a rough analogy with the potential energy of a roller coaster on its track. When the roller coaster starts from the station, it has a particular potential energy. As it is pulled up the first hill, its gravitational potential energy increases and reaches a maximum at the top. As it descends the other side, its potential energy decreases and reaches a local minimum at the bottom of the hill. As the roller coaster continues on its up and down path, its potential energy goes through more changes. But when it arrives back at the starting point, it has exactly as much potential energy as it had when it started at this point. Another way of saying this is that there was as much uphill as there was downhill.

Similar reasoning can be applied to an electric circuit. We will analyze the circuit of Fig. 26–10 shortly, but first we consider the simpler circuit in Fig. 26–11. We have chosen it to be the same as the equivalent circuit of Fig. 26–6b already discussed. The current in this circuit is $I = (12.0\ \text{V})/(690\ \Omega) = 0.0174\ \text{A}$, as we calculated in Example 26–3. (We keep an extra digit in I to reduce rounding errors.) The positive side of the battery, point e in Fig. 26–11a, is at a high potential compared to point d at the negative side of the battery. That is, point e is like the top of a hill for a roller coaster. We follow the current around the circuit starting at any point. We choose to start at point d and follow a small positive test charge completely around this circuit. As we go, we note all changes in potential. When

the test charge returns to point d, the potential will be the same as when we started (total change in potential around the circuit is zero). We plot the changes in potential around the circuit in Fig. 26–11b; point d is arbitrarily taken as zero.

As our positive test charge goes from point d, which is the negative or low potential side of the battery, to point e, which is the positive terminal (high potential side) of the battery, the potential increases by 12.0 V. (This is like the roller coaster being pulled up the first hill.) That is,

$$V_{ed} = +12.0\,\text{V}.$$

When our test charge moves from point e to point a, there is no change in potential because there is no source of emf and negligible resistance in the connecting wires.

Next, as the charge passes through the 400-Ω resistor to get to point b, there is a decrease in potential of $V = IR = (0.0174\,\text{A})\,(400\,\Omega) = 7.0\,\text{V}$. The positive test charge is flowing "downhill" since it is heading toward the negative terminal of the battery, as indicated in the graph of Fig. 26–11b. Because this is a *decrease* in potential, we use a *negative* sign:

$$V_{ba} = V_b - V_a = -7.0\,\text{V}.$$

PROBLEM SOLVING
Be consistent with signs when applying the loop rule

As the charge proceeds from b to c there is another potential decrease (a "voltage drop") of $(0.0174\,\text{A}) \times (290\,\Omega) = 5.0\,\text{V}$, and this too is a decrease in potential:

$$V_{cb} = -5.0\,\text{V}.$$

There is no change in potential as our test charge moves from c to d as we assume negligible resistance in the wires.

The sum of all the changes in potential around the circuit of Fig. 26–11 is

$$+12.0\,\text{V} - 7.0\,\text{V} - 5.0\,\text{V} = 0.$$

This is exactly what Kirchhoff's loop rule said it would be.

PROBLEM SOLVING

Kirchhoff's Rules

1. **Label the current** in each separate branch of the given circuit with a different subscript, such as I_1, I_2, I_3 (see Fig. 26–10). Each current refers to a segment of the circuit between two junctions. Choose the direction of each current, using an arrow. The direction can be chosen arbitrarily: if the current is actually in the opposite direction from what you chose, it will come out with a minus sign in the solution.
2. **Identify the unknowns**. You will need as many independent equations as there are unknowns. You may write down more equations than this, but you will find that some of the equations will be redundant (that is, not be independent in the sense of providing new information). You may use $V = IR$ for each resistor, which sometimes will reduce the number of unknowns.
3. **Apply Kirchhoff's junction rule** at one or more junctions.
4. **Apply Kirchhoff's loop rule** for one or more loops: follow each loop in one direction only. Pay careful attention to subscripts, and to signs:
   (a) For a resistor, apply Ohm's law; the potential difference is negative (a decrease) if your chosen loop direction is the same as the chosen current direction through that resistor. The potential difference is positive (an increase) if your chosen loop direction is *opposite* to the chosen current direction.
   (b) For a battery, the potential difference is positive if your chosen loop direction is from the negative terminal toward the positive terminal; the potential difference is negative if the loop direction is from the positive terminal toward the negative terminal.
5. **Solve the equations** algebraically for the unknowns. Be careful with signs. At the end, check your answers by plugging them into the original equations, or even by using any additional loop or junction rule equations not used previously.

EXAMPLE 26–8 **Using Kirchhoff's rules.** Calculate the currents I_1, I_2, and I_3 in the three branches of the circuit in Fig. 26–12 (which is the same as Fig. 26–10).

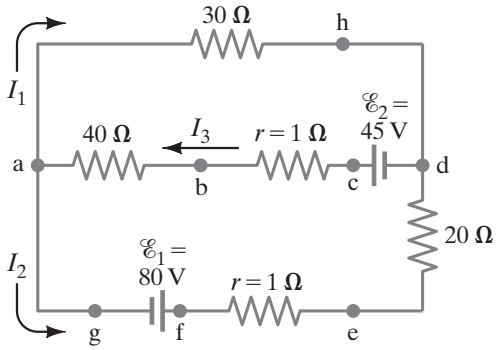


FIGURE 26–12
Currents can be calculated using Kirchhoff's rules.

APPROACH and SOLUTION

PROBLEM SOLVING
Choose current directions arbitrarily

1. **Label the currents** and their directions. Figure 26–12 uses the labels I_1, I_2, and I_3 for the current in the three separate branches. Since (positive) current tends to move away from the positive terminal of a battery, we choose I_2 and I_3 to have the directions shown in Fig. 26–12. The direction of I_1 is not obvious in advance, so we arbitrarily chose the direction indicated. If the current actually flows in the opposite direction, our answer for I_1 will have a negative sign.
2. **Identify the unknowns**. We have three unknowns (I_1, I_2, and I_3) and therefore we need three equations, which we get by applying Kirchhoff's junction and loop rules.
3. **Junction rule**: We apply Kirchhoff's junction rule to the currents at point a, where I_3 enters and I_2 and I_1 leave:
$$I_3 = I_1 + I_2. \qquad \textbf{(i)}$$
This same equation holds at point d, so we get no new information by writing an equation for point d.
4. **Loop rule**: We apply Kirchhoff's loop rule to two different closed loops. First we apply it to the upper loop ahdcba. We start (and end) at point a. From a to h we have a potential decrease $V_{ha} = -(I_1)(30\ \Omega)$. From h to d there is no change, but from d to c the potential increases by 45 V: that is, $V_{cd} = +45$ V. From c to a the potential decreases through the two resistances by an amount $V_{ac} = -(I_3)(40\ \Omega + 1\ \Omega) = -(41\ \Omega)I_3$. Thus we have $V_{ha} + V_{cd} + V_{ac} = 0$, or
$$-30I_1 + 45 - 41I_3 = 0, \qquad \textbf{(ii)}$$
where we have omitted the units (volts and amps) so we can more easily see the algebra. For our second loop, we take the outer loop ahdefga. (We could have chosen the lower loop abcdefga instead.) Again we start at point a, and going to point h we have $V_{ha} = -(I_1)(30\ \Omega)$. Next, $V_{dh} = 0$. But when we take our positive test charge from d to e, it actually is going uphill, against the current—or at least against the *assumed* direction of the current, which is what counts in this calculation. Thus $V_{ed} = +I_2(20\ \Omega)$ has a *positive* sign. Similarly, $V_{fe} = +I_2(1\ \Omega)$. From f to g there is a decrease in potential of 80 V because we go from the high potential terminal of the battery to the low. Thus $V_{gf} = -80$ V. Finally, $V_{ag} = 0$, and the sum of the potential changes around this loop is
$$-30I_1 + (20 + 1)I_2 - 80 = 0. \qquad \textbf{(iii)}$$
Our major work is done. The rest is algebra.

PROBLEM SOLVING
Be consistent with signs when applying the loop rule

5. **Solve the equations**. We have three equations—labeled (i), (ii), and (iii)—and three unknowns. From Eq. (iii) we have
$$I_2 = \frac{80 + 30I_1}{21} = 3.8 + 1.4I_1. \qquad \textbf{(iv)}$$

From Eq. (ii) we have

$$I_3 = \frac{45 - 30I_1}{41} = 1.1 - 0.73I_1. \qquad \textbf{(v)}$$

We substitute Eqs. (iv) and (v) into Eq. (i):

$$I_1 = I_3 - I_2 = 1.1 - 0.73I_1 - 3.8 - 1.4I_1.$$

We solve for I_1, collecting terms:

$$3.1I_1 = -2.7$$
$$I_1 = -0.87\ \text{A}.$$

The negative sign indicates that the direction of I_1 is actually opposite to that initially assumed and shown in Fig. 26–12. The answer automatically comes out in amperes because our voltages and resistances were in the SI units of volts and ohms. From Eq. (iv) we have

$$I_2 = 3.8 + 1.4I_1 = 3.8 + 1.4(-0.87) = 2.6\ \text{A},$$

and from Eq. (v)

$$I_3 = 1.1 - 0.73I_1 = 1.1 - 0.73(-0.87) = 1.7\ \text{A}.$$

This completes the solution.

PROBLEM SOLVING
I_1 is in the opposite direction from that assumed in Fig. 26–12

NOTE The unknowns in different situations are not necessarily currents. It might be that the currents are given and we have to solve for unknown resistance or voltage. The variables are then different, but the technique is the same.

EXERCISE D Write the Kirchhoff equation for the lower loop abcdefga of Fig. 26–12 and Example 26–8, and show (assuming the currents calculated in this Example) that the potentials add to zero for this lower loop.

26–4 EMFs in Series and in Parallel; Charging a Battery

When two or more sources of emf, such as batteries, are arranged in series as in Fig. 26–13a, the total voltage is the algebraic sum of their respective voltages. On the other hand, when a 20-V and a 12-V battery are connected oppositely, as shown in Fig. 26–13b, the net voltage V_{ca} is 8 V (ignoring voltage drop across internal resistances). That is, a positive test charge moved from a to b gains in potential by 20 V, but when it passes from b to c it drops by 12 V. So the net change is $20\ \text{V} - 12\ \text{V} = 8\ \text{V}$. You might think that connecting batteries in reverse like this would be wasteful. For most purposes that would be true. But such a reverse arrangement is precisely how a **battery charger** works. In Fig. 26–13b, the 20-V source is charging up the 12-V battery. Because of its greater voltage, the 20-V source is forcing charge back into the 12-V battery: electrons are being forced into its negative terminal and removed from its positive terminal.

An automobile alternator keeps the car battery charged in the same way. A voltmeter placed across the terminals of a (12-V) car battery with the engine running fairly fast can tell you whether or not the alternator is charging the battery. If it is, the voltmeter reads 13 or 14 V. If the battery is not being charged, the voltage will be 12 V, or less if the battery is discharging. Some batteries, including car batteries, can be recharged, but other batteries (ordinary inexpensive AA, AAA, C, D) are not rechargeable because the chemical reactions cannot be reversed. In such cases, the arrangement of Fig. 26–13b would simply waste energy.

Sources of emf can also be arranged in parallel, Fig. 26–13c, which—if the emfs are the same—can provide more energy when large currents are needed. Each of the cells in parallel has to produce only a fraction of the total current, so the energy loss due to internal resistance is less than for a single cell; and the batteries will go dead less quickly.

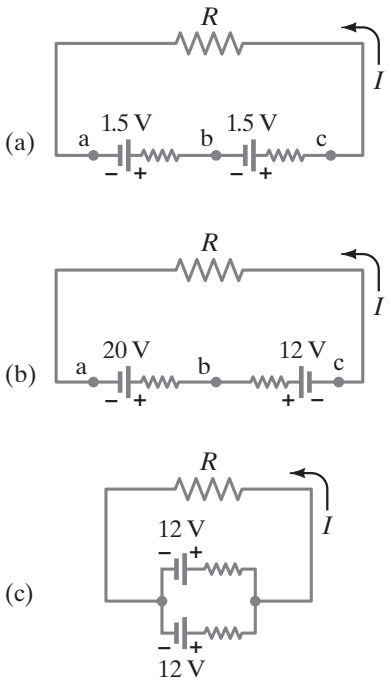
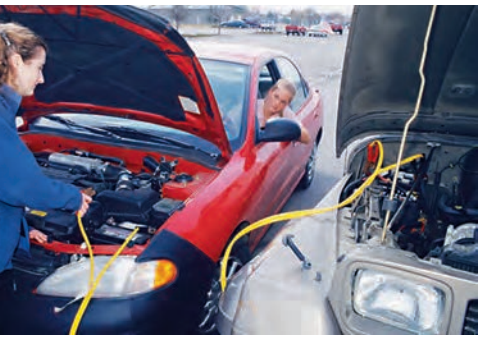


FIGURE 26–13 Batteries in series, (a) and (b), and in parallel (c).



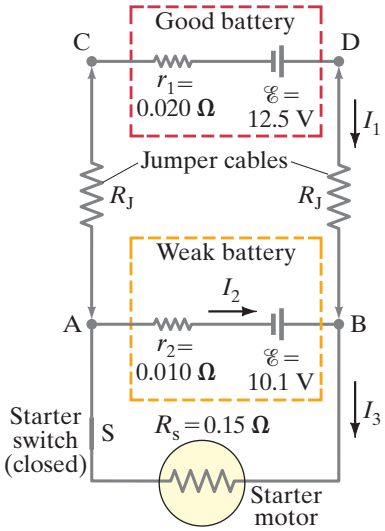


FIGURE 26–14 Example 26–9, a jump start.

EXAMPLE 26–9 **Jump starting a car.** A good car battery is being used to jump start a car with a weak battery. The good battery has an emf of 12.5 V and internal resistance $r_1 = 0.020\ \Omega$. Suppose the weak battery has an emf of 10.1 V and internal resistance $r_2 = 0.10\ \Omega$. Each copper jumper cable is 3.0 m long and 0.50 cm in diameter, and can be attached as shown in Fig. 26–14. Assume the starter motor can be represented as a resistor $R_s = 0.15\ \Omega$. Determine the current through the starter motor (*a*) if only the weak battery is connected to it, and (*b*) if the good battery is also connected, as shown in Fig. 26–14.

APPROACH We apply Kirchhoff's rules, but in (*b*) we will first need to determine the resistance of the jumper cables using their dimensions and the resistivity ($\rho = 1.68 \times 10^{-8}\ \Omega \cdot \text{m}$ for copper) as discussed in Section 25–4.

SOLUTION (*a*) The circuit with only the weak battery and no jumper cables is simple: an emf of 10.1 V connected to two resistances in series, $r_2 + R_s = 0.10\ \Omega + 0.15\ \Omega = 0.25\ \Omega$. Hence the current is $I = V/R = (10.1\ \text{V})/(0.25\ \Omega) = 40\ \text{A}$.

(*b*) We need to find the resistance of the jumper cables that connect the good battery to the weak one. From Eq. 25–3, each has resistance

$$R_J = \frac{\rho \ell}{A} = \frac{(1.68 \times 10^{-8}\ \Omega \cdot \text{m})(3.0\ \text{m})}{(\pi)(0.25 \times 10^{-2}\ \text{m})^2} = 0.0026\ \Omega.$$

Kirchhoff's loop rule for the full outside loop of Fig. 26–14 gives

$$12.5\ \text{V} - I_1(2R_J + r_1) - I_3 R_s = 0$$
$$12.5\ \text{V} - I_1(0.025\ \Omega) - I_3(0.15\ \Omega) = 0 \qquad \textbf{(i)}$$

since $(2R_J + r) = (0.0052\ \Omega + 0.020\ \Omega) = 0.025\ \Omega$. The loop rule for the lower loop, including the weak battery and the starter, gives

$$10.1\ \text{V} - I_3(0.15\ \Omega) - I_2(0.10\ \Omega) = 0. \qquad \textbf{(ii)}$$

The junction rule at point B gives

$$I_1 + I_2 = I_3. \qquad \textbf{(iii)}$$

We have three equations in three unknowns. From Eq. (iii),

$$I_1 = I_3 - I_2$$

and we substitute this into Eq. (i):

$$12.5\ \text{V} - (I_3 - I_2)(0.025\ \Omega) - I_3(0.15\ \Omega) = 0,$$
$$12.5\ \text{V} - I_3(0.175\ \Omega) + I_2(0.025\ \Omega) = 0.$$

We solve for I_2 in Eq. (ii) and insert that into this last equation:

$$12.5\ \text{V} - I_3(0.175\ \Omega) + \left(\frac{10.1\ \text{V} - I_3(0.15\ \Omega)}{0.10\ \Omega}\right)(0.025\ \Omega) = 0$$

or

$$I_3 = \frac{12.5\ \text{V} + 2.5\ \text{V}}{(0.175\ \Omega + 0.0375\ \Omega)} = 71\ \text{A},$$

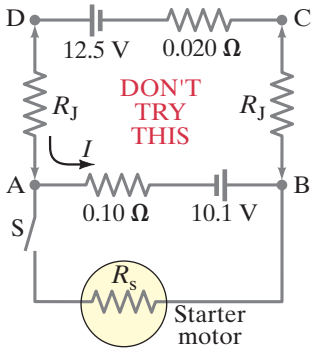
quite a bit better than in part (*a*).

The other currents are $I_2 = -5\ \text{A}$ and $I_1 = 76\ \text{A}$. Note that $I_2 = -5\ \text{A}$ is in the opposite direction from what we assumed in Fig. 26–14. The terminal voltage of the weak 10.1-V battery when being charged is

$$V_{BA} = 10.1\ \text{V} - (-5\ \text{A})(0.10\ \Omega) = 10.6\ \text{V}.$$

NOTE The circuit in Fig. 26–14, without the starter motor, is how a battery can be charged. The stronger battery pushes charge back into the weaker battery.

FIGURE 26–15 Exercise E.



EXERCISE E If the jumper cables of Example 26–9 are mistakenly connected in reverse, the positive terminal of each battery would be connected to the negative terminal of the other battery (Fig. 26–15). What would be the current I even before the starter motor is engaged (the switch S in Fig. 26–15 is open)? Why could this cause the batteries to explode?

Safety When Jump Starting

Before jump starting a car's weak battery, be sure both batteries are 12 V and check the polarity of both batteries. The following (cautious) procedure applies if the negative (−) terminal is ground (attached by a cable to the metal car frame and motor), and the "hot" terminal is positive (+) on both batteries, as is the case for most modern cars. The + terminal is usually marked by a red color, often a red cover. The safest procedure is to

1. connect the hot (+) terminal of the weak battery to the hot terminal of the good battery (using the cable with red clamps). Spread apart the handles of each clamp to squeeze the contacts tightly.
2. Then connect the black cable, first to the ground terminal of the good battery, and the other end to a clean exposed metal part (i.e., at ground) on the car with the weak battery. (This last connection should preferably be not too close to the battery, which in rare cases might leak H_2 gas that could ignite at the spark that may accompany the final connection.) This is safer than connecting directly to the ground terminal.
3. When you are ready to start the disabled car, it helps to have the good car running (to keep its battery fully charged). As soon as the disabled car starts, immediately detach the cables in the exact *reverse* order (ground cable first).

⚠ CAUTION
Connect ground cable last
Disconnect ground cable first

In the photo of Fig. 26–14, the above procedure is *not* being followed. Note the safety error: with ground terminals incorrectly connected first, if the red clamp (+12 V) touches a metal part (= ground), even if dropped by the person, a short circuit with damaging high electric current can occur (hundreds of amps).

⚠ CAUTION
An unsafe procedure

26–5 *RC* Circuits—Resistor and Capacitor in Series

Capacitors and resistors are often found together in a circuit. Such ***RC* circuits** are common in everyday life. They are used to control the speed of a car's windshield wipers and the timing of traffic lights; they are used in camera flashes, in heart pacemakers, and in many other electronic devices. Up to now in this Chapter we have been dealing with steady currents that do not change in time. The current in *RC* circuits *does* vary in time. In *RC* circuits, we are not so interested in the final "steady state" voltage and charge on the capacitor, but rather in how these variables change in time.

Capacitor Charging

A simple *RC* circuit is shown in Fig. 26–16a. When the switch S is closed, current immediately begins to flow through the circuit. Electrons will flow out from the negative terminal of the battery, through the resistor R, and accumulate on the upper plate of the capacitor. And electrons will flow into the positive terminal of the battery, leaving a positive charge on the other plate of the capacitor. As charge accumulates on the capacitor, the potential difference across it increases $(V_C = Q/C)$, and the current is reduced because of this opposing voltage on the capacitor. Eventually the voltage across the capacitor equals the emf of the battery, $\mathscr{E}$. There is then no further current flow and no potential difference across the resistor. Potential difference V_C across the capacitor, which is proportional to the charge on it $(V_C = Q/C,$ Eq. 24–1$)$, thus increases in time as shown in Fig. 26–16b. The mathematical form of this curve—that is, V_C as a function of time—can be derived using conservation of energy (or Kirchhoff's loop rule). The emf $\mathscr{E}$ of the battery will equal the sum of the voltage drops across the resistor (IR) and the capacitor (Q/C):

$$\mathscr{E} = IR + \frac{Q}{C}. \qquad \textbf{(26–5)}$$

The resistance R includes all resistance in the circuit, including the internal resistance of the battery; I is the current at all points in the circuit at any instant, and Q is the charge on the capacitor at that same instant. Although $\mathscr{E}$, R, and C are constants, both Q and I are functions of time. The rate at which charge flows through the resistor $(I = dQ/dt)$ is equal to the rate at which charge accumulates on the capacitor. Thus we can write

$$\mathscr{E} = R\frac{dQ}{dt} + \frac{1}{C}Q.$$

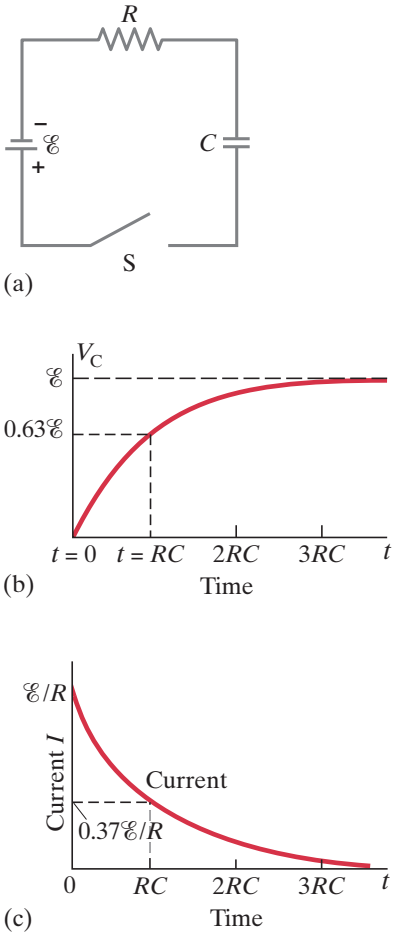


FIGURE 26–16 After the switch S closes in the *RC* circuit shown in (a), the voltage across the capacitor increases with time as shown in (b), and the current through the resistor decreases with time as shown in (c).

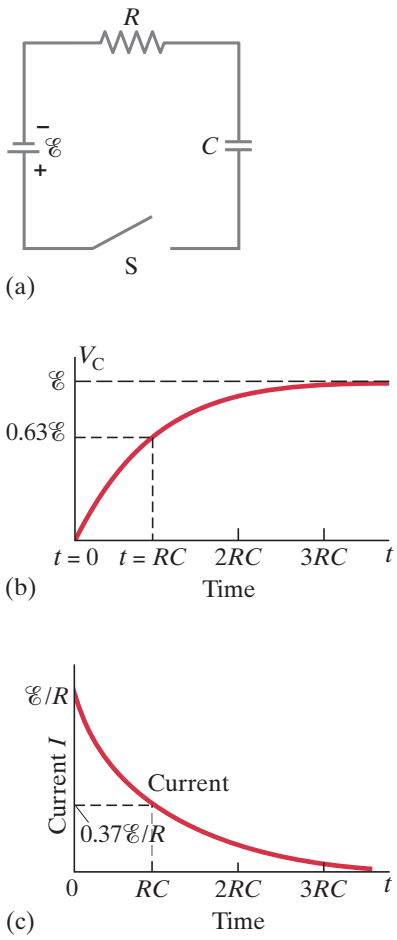


FIGURE 26–16 (Repeated.) After the switch S closes in the *RC* circuit shown in (a), the voltage across the capacitor increases with time as shown in (b), and the current through the resistor decreases with time as shown in (c).

This equation, $\mathscr{E} = R(dQ/dt) + Q/C$, can be solved by rearranging it:

$$\frac{dQ}{C\mathscr{E} - Q} = \frac{dt}{RC}.$$

We now integrate from $t = 0$, when $Q = 0$, to time t when a charge Q is on the capacitor:

$$\int_0^Q \frac{dQ}{C\mathscr{E} - Q} = \frac{1}{RC}\int_0^t dt$$

$$-\ln(C\mathscr{E} - Q)\Big|_0^Q = \frac{t}{RC}\Big|_0^t$$

$$-\ln(C\mathscr{E} - Q) - (-\ln C\mathscr{E}) = \frac{t}{RC}$$

$$\ln(C\mathscr{E} - Q) - \ln(C\mathscr{E}) = -\frac{t}{RC}$$

$$\ln\left(1 - \frac{Q}{C\mathscr{E}}\right) = -\frac{t}{RC}.$$

We take the exponential[†] of both sides

$$1 - \frac{Q}{C\mathscr{E}} = e^{-t/RC}$$

or

$$Q = C\mathscr{E}(1 - e^{-t/RC}) = Q_0(1 - e^{-t/RC}), \qquad \textbf{(26–6a)}$$

where $Q_0 = C\mathscr{E}$ represents the maximum charge on the capacitor. [Careful: Q_0 here is not Q at $t = 0$.] The potential difference across the capacitor is $V_C = Q/C$, so

$$V_C = \mathscr{E}(1 - e^{-t/RC}). \qquad \textbf{(26–6b)}$$

This Eq. 26–6b corresponds to the graph of Fig. 26–16b.

From Eqs. 26–6 we see that the charge Q on the capacitor, and the voltage V_C across it, increase from zero at $t = 0$ to maximum values $Q_0 = C\mathscr{E}$ and $V_C = \mathscr{E}$ after a very long time. The quantity RC that appears in the exponent of Eqs. 26–6a and b is called the **time constant** τ of the circuit:

$$\tau = RC. \qquad \textbf{(26–7)}$$

It represents the time[‡] required for the capacitor to reach $(1 - e^{-1}) = 0.63$ or 63% of its full charge and voltage $(e^{-1} = 1/2.718 = 0.37)$. Thus the product RC is a measure of how quickly the capacitor becomes charged. In a circuit, for example, where $R = 200\,\text{k}\Omega$ and $C = 3.0\,\mu\text{F}$, the time constant is $(2.0 \times 10^5\,\Omega)(3.0 \times 10^{-6}\,\text{F}) = 0.60\,\text{s}$. If the resistance is smaller, the time constant is smaller and the capacitor becomes charged more quickly. This makes sense, because a lower resistance will retard the flow of charge less. All circuits contain some resistance (if only the small amount in the connecting wires), so a capacitor never can be charged instantaneously when connected to a battery.

From Eqs. 26–6, it appears that Q and V_C never quite reach their maximum values within a finite time. However, they reach 86% of maximum in $2RC$, 95% in $3RC$, 98% in $4RC$, and so on. We say that Q and V_C approach their maximum values "asymptotically." For example, if $R = 20\,\text{k}\Omega$ and $C = 0.30\,\mu\text{F}$, the time constant is $(2.0 \times 10^4\,\Omega)(3.0 \times 10^{-7}\,\text{F}) = 6.0 \times 10^{-3}\,\text{s}$. So the capacitor is more than 98% charged in less than $\frac{1}{40}$ of a second.

The current I through the circuit of Fig. 26–16a at any time t can be obtained by differentiating Eq. 26–6a:

$$I = \frac{dQ}{dt} = \frac{\mathscr{E}}{R}e^{-t/RC}, \qquad \textbf{(26–8)}$$

which we call an **exponential decay**. Thus, at $t = 0$, just when the switch of the circuit in Fig. 26–16a is closed, the current is largest because there is no charge on the capacitor to impede it. That is, $I = I_0 = \mathscr{E}/R$, as expected for a circuit containing only a resistor. As charge builds up on the capacitor, the current decreases exponentially in time with a time constant equal to RC, as shown in Fig. 26–16c.

[†]The exponential constant e, known as the base for natural logarithms, has the value $e = 2.718\cdots$. See Appendix A–7. Do not confuse this e with e for the charge on the electron.

[‡]The units of RC are $\Omega \cdot \text{F} = (\text{V/A})(\text{C/V}) = \text{C/(C/s)} = \text{s}$.

The time constant RC represents the time required for the current to drop to $1/e \approx 0.37$ of its initial value.

EXAMPLE 26–10 ***RC* circuit, with emf.** The capacitance in the circuit of Fig. 26–16a is $C = 0.30\,\mu\text{F}$, the total resistance is $R = 20\,\text{k}\Omega$, and the battery emf is 12 V. Determine (*a*) the time constant, (*b*) the maximum charge the capacitor could acquire, (*c*) the time it takes for the charge to reach 99% of this value, (*d*) the current I when the charge Q on the capacitor is half its maximum value, (*e*) the maximum current in the circuit, and (*f*) the charge Q on the capacitor when the current I is 0.20 of its maximum value.

APPROACH We use Fig. 26–16 and Eqs. 26–5, 6, 7, and 8.

SOLUTION (*a*) The time constant is $RC = (2.0 \times 10^4\,\Omega)(3.0 \times 10^{-7}\,\text{F}) = 6.0 \times 10^{-3}\,\text{s} = 6.0\,\text{ms}$.

(*b*) The maximum charge on the capacitor would occur when no further current flows, so $Q = Q_0 = C\mathscr{E} = (3.0 \times 10^{-7}\,\text{F})(12\,\text{V}) = 3.6\,\mu\text{C}$.

(*c*) In Eq. 26–6a, we set $Q = 0.99C\mathscr{E}$:

$$0.99C\mathscr{E} = C\mathscr{E}(1 - e^{-t/RC}),$$

or

$$e^{-t/RC} = 1 - 0.99 = 0.01.$$

We take the natural logarithm of both sides recalling that $\ln e^x = x$ (Appendix A–7):

$$\frac{t}{RC} = -\ln(0.01) = 4.6$$

so

$$t = 4.6RC = (4.6)(6.0 \times 10^{-3}\,\text{s}) = 28 \times 10^{-3}\,\text{s}$$

or 28 ms (less than $\frac{1}{30}$ s).

(*d*) From part (*b*) the maximum charge on the capacitor is $3.6\,\mu\text{C}$. When the charge is half this value, $1.8\,\mu\text{C}$, the current I in the circuit can be found using the original differential equation, or Eq. 26–5:

$$I = \frac{1}{R}\left(\mathscr{E} - \frac{Q}{C}\right) = \frac{1}{2.0 \times 10^4\,\Omega}\left(12\,\text{V} - \frac{1.8 \times 10^{-6}\,\text{C}}{0.30 \times 10^{-6}\,\text{F}}\right) = 300\,\mu\text{A}.$$

(*e*) The current is a maximum at $t = 0$ (the moment when the switch is closed) and there is no charge yet on the capacitor $(Q = 0)$. So the maximum current is

$$I_0 = \frac{\mathscr{E}}{R} = \frac{12\,\text{V}}{2.0 \times 10^4\,\Omega} = 600\,\mu\text{A}.$$

(*f*) Again using Eq. 26–5, now with $I = 0.20I_0 = 120\,\mu\text{A}$, we have

$$Q = C(\mathscr{E} - IR)$$
$$= (3.0 \times 10^{-7}\,\text{F})[12\,\text{V} - (1.2 \times 10^{-4}\,\text{A})(2.0 \times 10^4\,\Omega)] = 2.9\,\mu\text{C}.$$

Capacitor Discharging

The circuit just discussed involved the *charging* of a capacitor by a battery through a resistance. Now let us look at another situation: a capacitor is already charged to a voltage V_0 and charge Q_0, and it is then allowed to *discharge* through a resistance R as shown in Fig. 26–17a. In this case there is no battery. When the switch S is closed, charge begins to flow through resistor R from one side of the capacitor toward the other side, until the capacitor is fully discharged. The voltage across the resistor at any instant equals that across the capacitor:

$$IR = \frac{Q}{C}.$$

The rate at which charge leaves the capacitor equals the negative of the current in the resistor, $I = -dQ/dt$, because the capacitor is discharging (Q is decreasing).

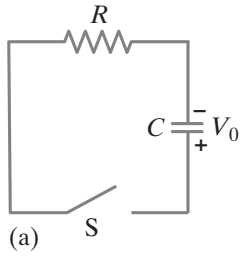
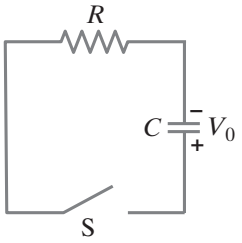
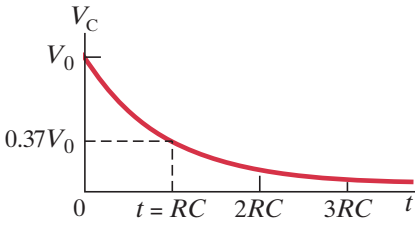


FIGURE 26–17a A simple *RC* circuit where the capacitor is initially charged and there is no battery. (Part (b) is on next page.)



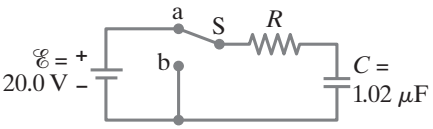
(a)



(b)

FIGURE 26–17 For the *RC* circuit shown in (a), the voltage V_C across the capacitor decreases with time, as shown in (b), after the switch S is closed at $t = 0$. The charge on the capacitor follows the same curve because $Q \propto V_C$.

FIGURE 26–18 Example 26–11.



So, with $I = -dQ/dt$, we write the equation $IR = Q/C$ as

$$-\frac{dQ}{dt}R = \frac{Q}{C}.$$

We rearrange this to

$$\frac{dQ}{Q} = -\frac{dt}{RC}$$

and integrate it from $t = 0$ when the charge on the capacitor is Q_0, to some time t later when the charge is Q:

$$\ln\left(Q/Q_0\right) = -\left(t/RC\right)$$

or

$$Q = Q_0 e^{-t/RC}. \qquad \textbf{(26–9a)}$$

The voltage across the capacitor $(V_C = Q/C)$ as a function of time is

$$V_C = V_0 e^{-t/RC}, \qquad \textbf{(26–9b)}$$

where the initial voltage $V_0 = Q_0/C$. Thus the charge on the capacitor, and the voltage across it, decrease exponentially in time with a time constant RC. This is shown in Fig. 26–17b. The current is

$$I = -\frac{dQ}{dt} = \frac{Q_0}{RC}e^{-t/RC} = I_0 e^{-t/RC}, \qquad \textbf{(26–10)}$$

and it too is seen to decrease exponentially in time with the same time constant RC. The charge on the capacitor, the voltage across it, and the current in the resistor all decrease to 37% of their original value in one time constant $t = \tau = RC$.

EXERCISE F In 10 time constants, the charge on the capacitor in Fig. 26–17 will be about (*a*) $Q_0/20{,}000$, (*b*) $Q_0/5000$, (*c*) $Q_0/1000$, (*d*) $Q_0/10$, (*e*) $Q_0/3$?

EXAMPLE 26–11 **Discharging *RC* circuit.** In the *RC* circuit shown in Fig. 26–18, the battery has fully charged the capacitor, so $Q_0 = C\mathscr{E}$. Then at $t = 0$ the switch is thrown from position a to b. The battery emf is 20.0 V, and the capacitance $C = 1.02\,\mu\text{F}$. The current I is observed to decrease to 0.50 of its initial value in 40 μs. (*a*) What is the value of Q, the charge on the capacitor, at $t = 0$? (*b*) What is the value of R? (*c*) What is Q at $t = 60\,\mu$s?

APPROACH At $t = 0$, the capacitor has charge $Q_0 = C\mathscr{E}$, and then the battery is removed from the circuit and the capacitor begins discharging through the resistor, as in Fig. 26–17. At any time t later (Eq. 26–9a) we have

$$Q = Q_0 e^{-t/RC} = C\mathscr{E}e^{-t/RC}.$$

SOLUTION (*a*) At $t = 0$, the charge on the capacitor is

$$Q_0 = C\mathscr{E} = \left(1.02\times10^{-6}\,\text{F}\right)(20.0\,\text{V}) = 2.04\times10^{-5}\,\text{C} = 20.4\,\mu\text{C}.$$

(*b*) To find R, we are given that at $t = 40\,\mu$s, $I = 0.50I_0$. Hence (Eq. 26–10):

$$0.50I_0 = I_0 e^{-t/RC}.$$

Taking natural logs on both sides $(\ln 0.50 = -0.693)$:

$$0.693 = \frac{t}{RC}$$

so

$$R = \frac{t}{(0.693)C} = \frac{\left(40\times10^{-6}\,\text{s}\right)}{(0.693)\left(1.02\times10^{-6}\,\text{F}\right)} = 57\,\Omega.$$

(*c*) At $t = 60\,\mu$s,

$$Q = Q_0 e^{-t/RC} = (20.4\times10^{-6}\,\text{C})e^{-\frac{60\times10^{-6}\,\text{s}}{(57\,\Omega)(1.02\times10^{-6}\,\text{F})}} = 7.3\,\mu\text{C}.$$

EXAMPLE 26–12 **How quickly does voltage drop.** A charged capacitor, $C = 35\,\mu\text{F}$, is connected to a resistance $R = 120\,\Omega$ as in Fig. 26–17a. How much time will elapse until the voltage falls to 10% of its original (maximum) value?

APPROACH The voltage across the capacitor decreases according to $V_C = V_0 e^{-t/RC}$, Eq. 26–9b. We set $V_C = 0.10V_0$ (10% of V_0), but first we need to calculate $\tau = RC$.

SOLUTION The time constant for this circuit is given by

$$\tau = RC = (120\,\Omega)(35\times10^{-6}\,\text{F}) = 4.2\times10^{-3}\,\text{s}.$$

After a time t the voltage across the capacitor will be

$$V_C = V_0 e^{-t/RC}.$$

We want to know the time t for which $V_C = 0.10V_0$. We substitute into the above equation

$$0.10V_0 = V_0 e^{-t/RC}$$

so

$$e^{-t/RC} = 0.10.$$

The inverse operation to the exponential e is the natural log, ln. Thus

$$\ln\left(e^{-t/RC}\right) = -\frac{t}{RC} = \ln 0.10 = -2.3.$$

Solving for t, we find the elapsed time is

$$t = 2.3(RC) = (2.3)\left(4.2\times10^{-3}\,\text{s}\right) = 9.7\times10^{-3}\,\text{s} = 9.7\,\text{ms}.$$

NOTE We can find the time for any specified voltage across a capacitor by using $t = RC\ln\left(V_0/V_C\right)$.

CONCEPTUAL EXAMPLE 26–13 **Bulb in *RC* circuit.** In the circuit of Fig. 26–19, the capacitor is originally uncharged. Describe the behavior of the lightbulb from the instant switch S is closed until a long time later.

RESPONSE When the switch is first closed, the current in the circuit is high and the lightbulb burns brightly. As the capacitor charges, the voltage across the capacitor increases, causing the current to be reduced, and the lightbulb dims. As the potential difference across the capacitor approaches the same voltage as the battery, the current decreases toward zero and the lightbulb goes out.

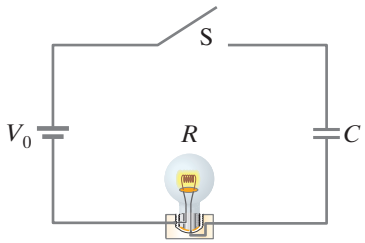


FIGURE 26–19 Example 26–13.

*Medical and Other Applications of *RC* Circuits

The charging and discharging in an *RC* circuit can be used to produce voltage pulses at a regular frequency. The charge on the capacitor increases to a particular voltage, and then discharges. One way of initiating the discharge of the capacitor is by the use of a gas-filled tube which has an electrical breakdown when the voltage across it reaches a certain value V_0. After the discharge is finished, the tube no longer conducts current and the recharging process repeats itself, starting at a lower voltage V_0'. Figure 26–20 shows a possible circuit, and the **sawtooth voltage** it produces.

PHYSICS APPLIED
Sawtooth voltage

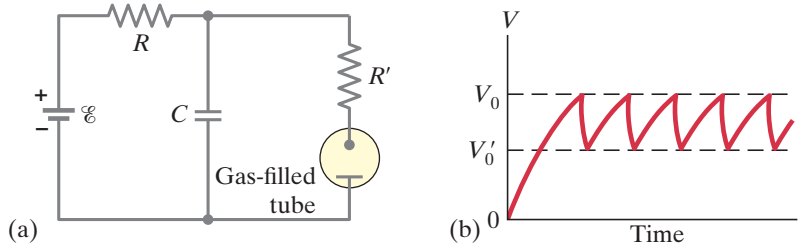


FIGURE 26–20 (a) An *RC* circuit, coupled with a gas-filled tube as a switch, can produce (b) a repeating "sawtooth" voltage. (In (b) note the initial increase in voltage is like Fig. 26–16b.)

A simple blinking light can be an application of a sawtooth oscillator circuit. Here the emf is supplied by a battery; the neon bulb flashes on at a rate of perhaps 1 cycle per second. The main component of a "flasher unit" is a moderately large capacitor.

PHYSICS APPLIED
Blinking flashers

The intermittent windshield wipers of a car can also use an *RC* circuit. The *RC* time constant, which can be changed using a multi-positioned switch for different values of *R* with fixed *C*, determines the rate at which the wipers come on.

PHYSICS APPLIED
Windshield wipers on "intermittent"

EXERCISE G A typical turn signal flashes perhaps twice per second, so its time constant is on the order of 0.5 s. Estimate the resistance in the circuit, assuming a moderate capacitor of $C = 1\,\mu\text{F}$.

PHYSICS APPLIED
Heart pacemaker

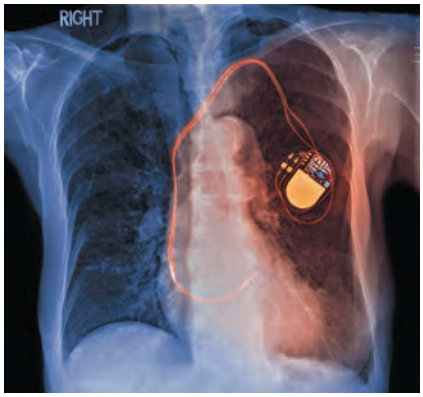


FIGURE 26–21 Electronic battery-powered pacemaker can be seen on the rib cage in this X-ray (color added).

PHYSICS APPLIED
Dangers of electricity

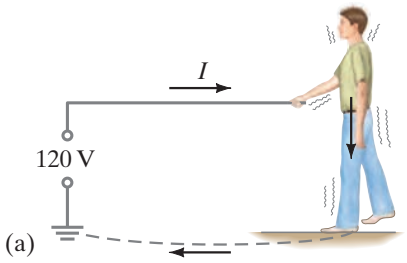
An important medical use of an *RC* circuit is the electronic heart pacemaker, which can make a stopped heart start beating again by applying an electric stimulus through electrodes attached to the chest. The stimulus can be repeated at the normal heartbeat rate if necessary. The heart itself contains *pacemaker* cells, which send out tiny electric pulses at a rate of 60 to 80 per minute. These signals induce the start of each heartbeat. In some forms of heart disease, the natural pacemaker fails to function properly, and the heart loses its beat. Such patients use *electronic pacemakers* which produce a regular voltage pulse that starts and controls the frequency of the heartbeat. The electrodes are implanted in or near the heart (Fig. 26–21), and the circuit contains a capacitor and a resistor. The charge on the capacitor increases to a certain point and then discharges a pulse to the heart. Then it starts charging again. The pulsing rate depends on the time constant *RC*.

26–6 Electric Hazards and Safety

Excess electric current can overheat wires in buildings and cause fires, as discussed in Section 25–6. Electric current can also damage the human body or even be fatal. Electric current through the human body can cause damage in two ways: (1) heating tissue and causing burns; (2) stimulating nerves and muscles, and we feel a "shock." The severity of a shock depends on the magnitude of the current, how long it acts, and through what part of the body it passes. A current passing through vital organs such as the heart or brain is especially damaging.

A current of about 1 mA or more can be felt and may cause pain. Currents above 10 mA cause severe contraction of the muscles, and a person may not be able to let go of the source of the current (say, a faulty appliance or wire). Death from paralysis of the respiratory system can occur. Artificial respiration can sometimes revive a victim. If a current above about 80 to 100 mA passes across the torso, so that a portion passes through the heart for more than a second or two, the heart muscles will begin to contract irregularly and blood will not be properly pumped. This condition is called **ventricular fibrillation**. If it lasts for long, death results. Strangely enough, if the current is much larger, on the order of 1 A, death by heart failure may be less likely,[†] but such currents can cause serious burns if concentrated through a small area of the body.

FIGURE 26–22 You can receive a shock when the circuit is completed.



(a)



(b)

It is current that harms, even though it is voltage that drives the current. The seriousness of an electric shock depends on the current and thus on the applied voltage and the effective resistance of the body. Living tissue has low resistance because the fluid of cells contains ions that can conduct quite well. However, the outer layer of skin, when dry, offers high resistance and is thus protective. The effective resistance between two points on opposite sides of the body when the skin is dry is on the order of 10^4 to $10^6\ \Omega$. But when the skin is wet, the resistance may be $10^3\ \Omega$ or less. A person who is barefoot or wearing thin-soled shoes will be in good contact with the ground, and touching a 120-V line with a wet hand can result in a current

$$I = \frac{120\ \text{V}}{1000\ \Omega} = 120\ \text{mA}.$$

As we saw, this could be lethal.

You can get a shock by becoming part of a complete circuit. Figure 26–22 shows two ways the circuit might be completed when you accidentally touch a "hot" electric wire—"hot" meaning a high potential relative to ground such as 120 V (normal U.S. household voltage) or 240 V (many other countries). The other wire of building wiring is connected to ground—either by a wire connected to a buried conductor, or via a metal water pipe into the ground. The current in Fig. 26–22a passes from the high voltage wire through you to ground through your bare feet, and back along the ground (a fair conductor) to the ground terminal of the source. If you stand on a good insulator—thick rubber-soled shoes or a

[†]Larger currents apparently bring the entire heart to a standstill. Upon release of the current, the heart returns to its normal rhythm. This may not happen when fibrillation occurs because, once started, it can be hard to stop. Fibrillation may also occur as a result of a heart attack or during heart surgery. A device known as a *defibrillator* (described in Section 24–4) can apply a brief high current to the heart, causing complete heart stoppage which is often followed by resumption of normal beating.

dry wood floor—there will be much more resistance in the circuit and much less current through you. If you stand with bare feet on the ground, or in a bathtub, there is lethal danger because the resistance is much less and the current greater. In a bathtub (or swimming pool), not only are you wet, which reduces your resistance, but the water is in contact with the drain pipe which, if metal, leads to the ground. It is strongly recommended that you not touch anything electrical when wet or in bare feet. The use of non-metal pipes is important and protective.

In Fig. 26–22b, a person touches a "hot" wire in a faulty device with one hand, and the other hand touches a sink faucet (connected to ground via the pipe or even by water in a non-metal pipe). The current is particularly dangerous because it passes across the chest, through the heart and lungs. A useful rule: if one hand is touching something electrical, keep your other hand in your back pocket (don't use it!), and wear thick rubber-soled shoes. Also remove metal jewelry, especially rings (your finger is usually moist under a ring).

CAUTION
Keep one hand in your pocket when the other touches electricity

PHYSICS APPLIED
Grounding and shocks

You can come into contact with a hot wire by touching a bare wire whose insulation has worn off, or from a bare wire inside an appliance when you're tinkering with it. (Always unplug an electrical device before investigating its insides!)† Also, a wire inside a device may break or lose its insulation and come in contact with the case. If the case is metal, it will conduct electricity. A person could then suffer a severe shock merely by touching the case, as shown in Fig. 26–23b.

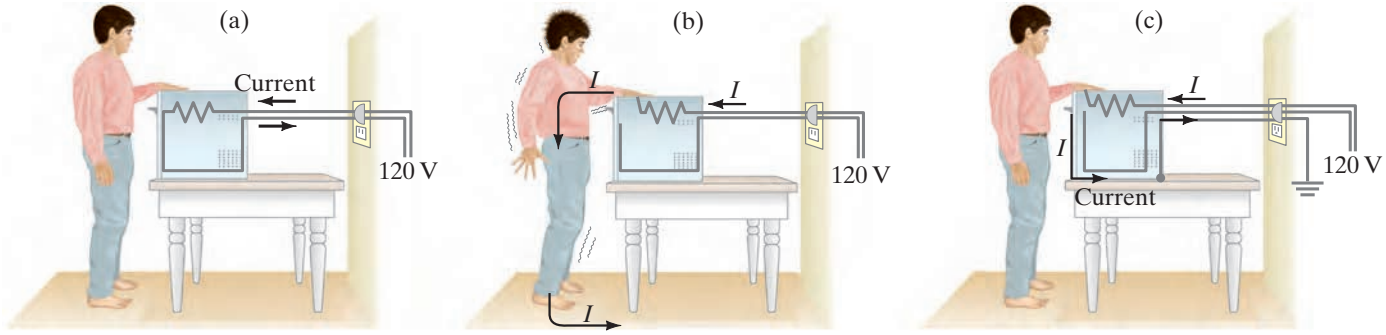


FIGURE 26–23 (a) An electric oven operating normally with a 2-prong plug. (b) A short to a metal case which is ungrounded, causing a shock. (c) A short to the case which is grounded by a 3-prong plug; almost no current goes through the person.

To prevent an accident, metal cases are supposed to be connected directly to ground by a separate ground wire. Then if a "hot" wire touches the grounded case, a short circuit to ground immediately occurs internally, as shown in Fig. 26–23c, and most of the current passes through the low-resistance ground wire rather than through the person. Furthermore, the high current should open a fuse or circuit breaker. Grounding a metal case is done by a separate ground wire connected to the third (round) prong of a 3-prong plug. Never cut off the third prong of a plug—it could save your life. A three-prong plug, and an adapter, are shown in Figs. 26–24a and b.

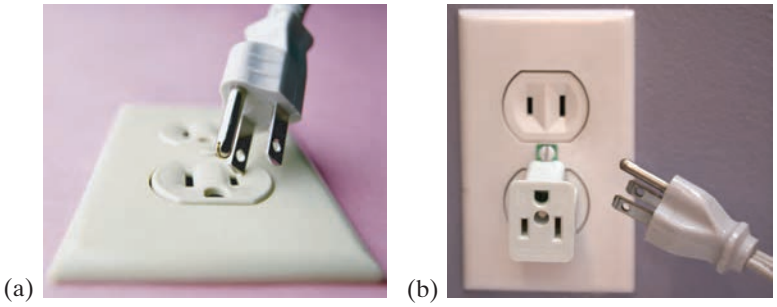


FIGURE 26–24 (a) A 3-prong plug, and (b) an adapter (white) for old-fashioned 2-prong outlets—be sure to screw down the ground tab (green color in photo).

Safe Wiring

Why is a third wire needed? The 120 V is carried by the other two wires—one **hot** (120 V ac), the other **neutral** ($\approx$ 0 V), which is itself grounded. The third "dedicated" ground wire with the round prong may seem redundant. But it is protection for two reasons: (1) It protects against internal wiring that may have been done incorrectly or is faulty as discussed above, Fig. 26–23.

CAUTION
Necessity of third (ground) wire

†Even then you can get a bad shock from a capacitor that hasn't been discharged until you touch it.

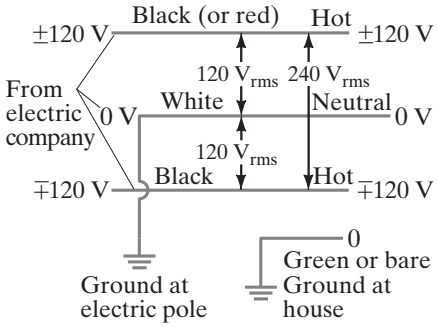


FIGURE 26–25 A polarized 2-prong plug.

CAUTION

Black wire may be either ground or hot. Beware!

FIGURE 26–26 Four wires entering a typical house (at the right). The color codes for wires are not always as shown here—be careful!



PHYSICS APPLIED

Hospital patients with implanted electrodes are at great risk

PHYSICS APPLIED

CAUTION: Do not touch a downed electric wire—run away, one foot on the ground at a time

(2) The *neutral* wire (≈ 0 V) carries the full normal current ("return" current from the hot 120 V) and it does have resistance—so there can be a voltage drop along the neutral wire, normally small; but if connections are poor or corroded, or the plug is loose, the resistance could be large enough that you might feel that voltage if you touched the neutral wire some distance from its grounding point.

Some electrical devices come with only two wires, and the plug's two prongs are of different widths; the plug can be inserted only one way into the outlet so that the intended neutral (wider prong) in the device is connected to neutral in the wiring (Fig. 26–25). For example, the screw threads of a lightbulb are meant to be connected to neutral (and the base contact to hot), to avoid shocks when changing a bulb in a possibly protruding socket. Devices with 2-prong plugs do *not* have their cases grounded; they are supposed to have double electric insulation (or have a nonmetal case). Take extra care anyway.

The insulation on a wire may be color coded. Hand-held meters (next Section) may have red (hot) and black (ground) lead wires. But in a U.S. house, the hot wire is often black (though it may be red), whereas white is neutral and green (or bare) is the dedicated ground, Fig. 26–26. But beware: these color codes cannot always be trusted.

[In the U.S., three wires normally enter a house: two *hot* wires at 120 V each (which add together to 240 V for appliances or devices that run on 240 V) plus the grounded *neutral* (carrying return current for the two hot wires). See Fig. 26–26. The "dedicated" *ground* wire (non-current carrying) is a fourth wire that does not come from the electric company but enters the house from a nearby heavy stake in the ground or a buried metal pipe. The two hot wires can feed separate 120-V circuits in the house, so each 120-V circuit inside the house has only three wires, including the dedicated ground.]

Normal circuit breakers (Sections 25–6 and 28–9) protect equipment and buildings from overload and fires. They protect humans only in some circumstances, such as the very high currents that result from a short, if they respond quickly enough. *Ground fault circuit interrupters* (GFCI or GFI), described in Section 29–9, are designed to protect people from the much lower currents (10 mA to 100 mA) that are lethal but would not throw a 15-A circuit breaker or blow a 20-A fuse. GFCI are our best protection, and should be installed in electric outlets if not already there, especially in bathrooms and kitchens.

Another danger is **leakage current**, by which we mean a current along an unintended path. Leakage currents are often "capacitively coupled." For example, a wire in a lamp or toaster forms a capacitor with the metal case; charges moving in one conductor attract or repel charge in the other, so there is a current. Typical electrical codes limit leakage currents to 1 mA for any device, which is usually harmless. It could be dangerous, however, to a hospital patient with implanted electrodes, due to the absence of the protective skin layer and because the current can pass directly through the heart. Although 100 mA may be needed to cause heart fibrillation when entering through the hands and spreading out through the body (very little of it actually passing through the heart), but as little as 0.02 mA can cause fibrillation when passing directly to the heart. Thus, a "wired" patient is in considerable danger from leakage current even from as simple an act as touching a lamp.

Finally, don't touch a downed power line (lethal!) or even get near it. A hot power line is at thousands of volts. Furthermore a huge current can flow along the ground from the point where the high-voltage wire touches the ground. This current is great enough that the voltage between your two feet could be large and dangerous. Tip: stand on one foot, or run so only one foot touches the ground at a time.

It is current that harms, but voltage drives the current. It is sometimes said that 30 V is the threshhold for serious danger. But even a 12-V car battery (which can supply large currents) can cause nasty burns and shock. Always be attentive with electricity: and keep one hand in your back pocket.

26–7 Ammeters and Voltmeters—Measurement Affects Quantity Measured

Measurement is a fundamental part of physics, and is not as simple as you might think. Measuring instruments can not be taken for granted; their results are not perfect and often need to be interpreted. As an illustration of measurement "theory" we examine here how electrical quantities are measured using meters. We also examine how meters affect the quantity they attempt to measure.

An **ammeter** measures current (amps), and a **voltmeter** measures potential difference or voltage. Each can be either: (1) an *analog* meter, which displays numerical values by the position of a pointer that can move across a scale (Fig. 26–27a); or (2) a *digital* meter, which displays the numerical value in numbers (Fig. 26–27b).We first examine how analog meters work.

Ammeters and Voltmeters (Analog)

An analog ammeter or voltmeter, in which the reading is by a pointer on a scale (Fig. 26–27a), uses a **galvanometer**. A galvanometer works on the principle of the force between a magnetic field and a current-carrying coil of wire; it is straightforward to understand and will be discussed in Chapter 27. For now, we only need to know that the deflection of the needle of a galvanometer is proportional to the current flowing through it. The **full-scale current sensitivity** of a galvanometer, I_m, is the electric current needed to make the needle deflect full scale, typically about 50 μA.

A galvanometer whose sensitivity I_m is 50 μA can measure currents from about 1 μA (currents smaller than this would be hard to read on the scale) up to 50 μA. To measure larger currents, a resistor is placed in parallel with the galvanometer. An analog **ammeter**, represented by the symbol •—Ⓐ—•, consists of a galvanometer (•—Ⓖ—•) in parallel with a resistor called the **shunt resistor**, as shown in Fig. 26–28. ("Shunt" is a synonym for "in parallel.") The shunt resistance is R_{sh}, and the resistance of the galvanometer coil is r. The value of R_{sh} is chosen according to the full-scale deflection desired; R_{sh} is normally very small—giving an ammeter a very small net resistance—so most of the current passes through R_{sh} and very little ($\lesssim 50\,\mu$A) passes through the galvanometer to deflect the needle.



(a)



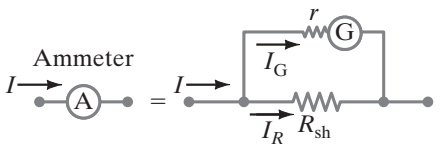
(b)

FIGURE 26–27 (a) An analog multimeter. (b) An electronic digital meter measuring voltage at a circuit breaker.

PHYSICS APPLIED

Ammeters use shunt resistor in parallel

FIGURE 26–28 An ammeter is a galvanometer in parallel with a (shunt) resistor with low resistance, R_{sh}.



EXAMPLE 26–14 **Ammeter design.** Design an ammeter to read 1.0 A at full scale using a galvanometer with a full-scale sensitivity of 50 μA and a resistance $r = 30\,\Omega$. Check if the scale is linear.

APPROACH Only $50\,\mu\text{A}\ (= I_G = 0.000050\text{ A})$ of the 1.0-A current passes through the galvanometer to give full-scale deflection. The rest of the current $(I_R = 0.999950\text{ A})$ passes through the small shunt resistor, R_{sh}, Fig. 26–28. The potential difference across the galvanometer equals that across the shunt resistor (they are in parallel). We apply Ohm's law to find R_{sh}.

SOLUTION Because $I = I_G + I_R$, when $I = 1.0$ A flows into the meter, we want I_R through the shunt resistor to be $I_R = 0.999950$ A. The potential difference across the shunt is the same as across the galvanometer, so

$$I_R R_{sh} = I_G r.$$

Then

$$R_{sh} = \frac{I_G r}{I_R} = \frac{(5.0 \times 10^{-5}\text{ A})(30\,\Omega)}{(0.999950\text{ A})} = 1.5 \times 10^{-3}\,\Omega,$$

or 0.0015 Ω. The shunt resistor must thus have a *very* low resistance and most of the current passes through it.

Because $I_G = I_R(R_{sh}/r)$ and (R_{sh}/r) is constant, we see that the scale is linear (because the needle deflection is proportional to I_G). If the current $I \approx I_R$ into the meter is half of full scale, 0.50 A, the current to the galvanometer will be $I_G = I_R(R_{sh}/r) = (0.50\text{ A})(1.5 \times 10^{-3}\,\Omega)/(30\,\Omega) = 25\,\mu$A, which would make the needle deflect halfway, as it should.

PHYSICS APPLIED
Voltmeters use series resistor

An analog **voltmeter** (•–Ⓥ–•) consists of a galvanometer and a resistor R_{ser} connected in series, Fig. 26–29. R_{ser} is usually large, giving a voltmeter a high internal resistance.



FIGURE 26–29 A voltmeter is a galvanometer in series with a resistor with high resistance, R_{ser}.

EXAMPLE 26–15 **Voltmeter design.** Using a galvanometer with internal resistance $r = 30\ \Omega$ and full-scale current sensitivity of $50\ \mu A$, design a voltmeter that reads from 0 to 15 V. Is the scale linear?

APPROACH When a potential difference of 15 V exists across the terminals of our voltmeter, we want $50\ \mu A$ to be passing through it so as to give a full-scale deflection.

SOLUTION From Ohm's law, $V = IR$, we have (Fig. 26–29)

$$15\ \text{V} = (50\ \mu\text{A})(r + R_{ser}),$$

so

$$R_{ser} = \frac{15\ \text{V}}{5.0 \times 10^{-5}\ \text{A}} - r = 300\ \text{k}\Omega - 30\ \Omega \approx 300\ \text{k}\Omega.$$

Notice that $r = 30\ \Omega$ is so small compared to the value of R_{ser} that it doesn't influence the calculation significantly. The scale will again be linear: if the voltage to be measured is 6.0 V, the current passing through the voltmeter will be $(6.0\ \text{V})/(3.0 \times 10^5\ \Omega) = 2.0 \times 10^{-5}\ \text{A}$, or $20\ \mu A$. This will produce two-fifths of full-scale deflection, as required $(6.0\ \text{V}/15.0\ \text{V} = 2/5)$.

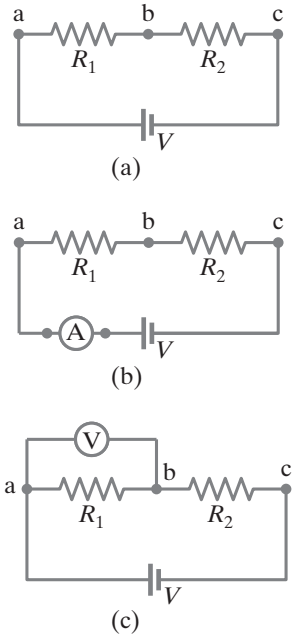


FIGURE 26–30 Measuring current and voltage.

How to Connect Meters

Suppose you wish to determine the current I in the circuit shown in Fig. 26–30a, and the voltage V across the resistor R_1. How exactly are ammeters and voltmeters connected to the circuit being measured?

Because an ammeter is used to measure the current flowing in the circuit, it must be inserted directly into the circuit, in series with the other elements, as shown in Fig. 26–30b. The smaller its internal resistance, the less it affects the circuit.

A voltmeter is connected "externally," in parallel with the circuit element across which the voltage is to be measured. It measures the potential difference between two points. Its two wire "leads" (connecting wires) are connected to the two points, as shown in Fig. 26–30c, where the voltage across R_1 is being measured.The larger its internal resistance ($R_{ser} + r$, Fig. 26–29), the less it affects the circuit being measured. Digital meters, discussed shortly, are connected in the same way and generally have less effect on the circuit.

Effects of Meter Resistance

PHYSICS APPLIED
Correcting for meter resistance

It is important to know the sensitivity of a meter, for in many cases the resistance of the meter can seriously affect your results. Consider the following Example.

EXAMPLE 26–16 **Voltage reading vs. true voltage.** An electronic circuit has two 15-kΩ resistors, R_1 and R_2, connected in series, Fig. 26–31a. The battery voltage is 8.0 V and it has negligible internal resistance. A voltmeter has resistance of 50 kΩ on the 5.0-V scale. What voltage does the meter read when connected across R_1, Fig. 26–31b, and what error is caused by the meter's finite resistance?

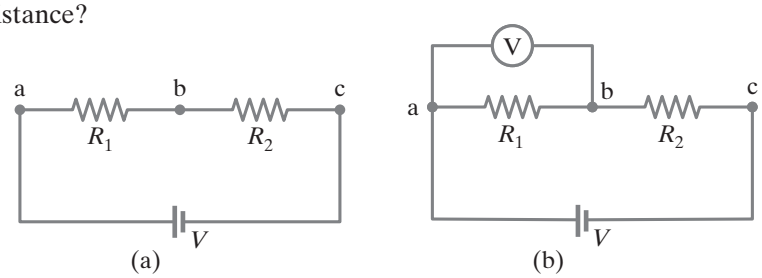


FIGURE 26–31 Example 26–16.

APPROACH The meter acts as a resistor in parallel with R_1. We use parallel and series resistor analyses and Ohm's law to find currents and voltages.

SOLUTION The voltmeter resistance of 50,000 Ω is in parallel with $R_1 = 15\,\text{k}\Omega$, Fig. 26–31b. The net resistance R_{eq} of these two is

$$\frac{1}{R_{\text{eq}}} = \frac{1}{50\,\text{k}\Omega} + \frac{1}{15\,\text{k}\Omega} = \frac{13}{150\,\text{k}\Omega};$$

so $R_{\text{eq}} = 11.5\,\text{k}\Omega$. This $R_{\text{eq}} = 11.5\,\text{k}\Omega$ is in series with $R_2 = 15\,\text{k}\Omega$, so the total resistance is now $26.5\,\text{k}\Omega$ (not the original $30\,\text{k}\Omega$). Hence the current from the battery is

$$I = \frac{8.0\,\text{V}}{26.5\,\text{k}\Omega} = 3.0 \times 10^{-4}\,\text{A} = 0.30\,\text{mA}.$$

Then the voltage drop across R_1, which is the same as that across the voltmeter, is $(3.0 \times 10^{-4}\,\text{A})(11.5 \times 10^{3}\,\Omega) = 3.5\,\text{V}$. [The voltage drop across R_2 is $(3.0 \times 10^{-4}\,\text{A})(15 \times 10^{3}\,\Omega) = 4.5\,\text{V}$, for a total of 8.0 V.] If we assume the meter is accurate, it reads 3.5 V. In the original circuit, without the meter, $R_1 = R_2$ so the voltage across R_1 is half that of the battery, or 4.0 V. Thus the voltmeter, because its internal resistance is not very large (a cheap meter), gives a low reading. It is off by 0.5 V, or more than 10%.

NOTE Often the **sensitivity** of a voltmeter is specified on its face as, for example, 10,000 Ω/V. Then on a 5.0-V scale, the voltmeter would have a resistance given by $(5.0\,\text{V})(10{,}000\,\Omega/\text{V}) = 50{,}000\,\Omega$. The meter's resistance depends on the scale used. Set on a 100-V scale, that meter would have an internal resistance of 1 MΩ.

Example 26–16 illustrates how seriously a meter can affect a circuit and give a misleading reading. If the resistance of a voltmeter is much higher than the resistance of the circuit, however, it will have little effect and its readings can be more accurate, at least to the manufactured precision of the meter, which for analog meters is typically 3% to 4% of full-scale deflection. Even an ammeter can interfere with a circuit, but the effect is minimal if its resistance is much less than that of the circuit as a whole. For both voltmeters and ammeters, the more sensitive the galvanometer, the less effect it will have on the circuit. [A 50,000-Ω/V meter is far better than a 1000-Ω/V meter.]

Whenever we make a measurement on a circuit, to some degree we affect that circuit (Example 26–16). This is true for other types of measurement as well: when we make a measurement on a system, we affect that system in some way. On a temperature measurement, for example, the thermometer can exchange heat with the system, thus altering the temperature it is measuring. It is important to be able to make needed corrections, as we saw in Example 26–16.

⚠ CAUTION
Measurements affect the quantity being measured

Other Meters

The meters described above are for direct current. A dc meter can be modified to measure ac (alternating current, Section 25–7) with the addition of diodes (Chapter 40), which allow current to flow in one direction only. An ac meter can be calibrated to read rms or peak values.

Voltmeters and ammeters can have several series or shunt resistors to offer a choice of range. **Multimeters** can measure voltage, current, and resistance. Sometimes a multimeter is called a VOM (Volt-Ohm-Meter or Volt-Ohm-Milliammeter).

An **ohmmeter** measures resistance, and must contain a battery of known voltage connected in series to a resistor (R_{ser}) and to an ammeter which contains a shunt R_{sh} (Fig. 26–32). The resistor whose resistance is to be measured completes the circuit, and must not be connected in a circuit containing a voltage source. The needle deflection of the meter is inversely proportional to the resistance.

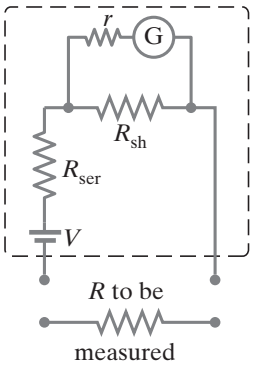


FIGURE 26–32 An ohmmeter.

The scale calibration depends on the value of its series resistor, which is changeable in a multimeter. Because an ohmmeter sends a current through the device whose resistance is to be measured, it should not be used on very delicate devices that could be damaged by the current.

Digital Meters

Digital meters (see Fig. 26–27b) are used in the same way as analog meters: they are inserted directly into the circuit, in series, to measure current (Fig. 26–30b), and connected "outside," in parallel with the circuit, to measure voltage (Fig. 26–30c).

The internal construction of digital meters is very different from analog meters. First, digital meters do not use a galvanometer, but rather semiconductor devices (Chapter 40). The electronic circuitry and digital readout are more sensitive than a galvanometer, and have less effect on the circuit to be measured. When we measure dc voltages, a digital voltmeter's resistance is very high, commonly on the order of 10 MΩ to 100 MΩ $(10^7$–$10^8\ \Omega)$, and doesn't change significantly when different voltage scales are selected. A 100-MΩ digital meter draws very little current when connected across even a 1-MΩ (one megaohm) resistance.

The precision of digital meters is exceptional, often one part in 10^4 $(=0.01\%)$ or better. This precision is not the same as accuracy, however. A precise meter of internal resistance $10^8\ \Omega$ will not give accurate results if used to measure a voltage across a 10^8-Ω resistor—in which case it is necessary to do a calculation like that in Example 26–16.

Summary

A device that transforms another type of energy into electrical energy is called a **source** of **emf**. A battery behaves like a source of emf in series with an **internal resistance**. The emf is the potential difference determined by the chemical reactions in the battery and equals the terminal voltage when no current is drawn. When a current is drawn, the voltage at the battery's terminals is less than its emf by an amount equal to the potential decrease Ir across the internal resistance.

When resistances are connected in **series** (end to end in a single linear path), the equivalent resistance is the sum of the individual resistances:

$$R_{\text{eq}} = R_1 + R_2 + \cdots . \qquad \textbf{(26–3)}$$

In a series combination, R_{eq} is greater than any component resistance.

When resistors are connected in **parallel**, it is the reciprocals that add up:

$$\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots . \qquad \textbf{(26–4)}$$

In a parallel connection, the net resistance is less than any of the individual resistances, even the smallest one.

Kirchhoff's rules are useful in determining the currents and voltages in circuits. Kirchhoff's **junction rule** is based on conservation of electric charge and states that the sum of all currents entering any junction equals the sum of all currents leaving that junction. The second, or **loop rule**, is based on conservation of energy and states that the algebraic sum of the changes in electric potential around any closed path of the circuit must be zero.

When an ***RC* circuit** containing a resistor R in series with a capacitance C is connected to a dc source of emf, the voltage across the capacitor rises gradually in time characterized by an exponential of the form $(1 - e^{-t/RC})$, where the **time constant**

$$\tau = RC \qquad \textbf{(26–7)}$$

is the time it takes for the voltage to reach 63% of its maximum value.

A capacitor discharging through a resistor is characterized by the same time constant: in a time $\tau = RC$, the voltage across the capacitor drops to 37% of its initial value. The charge on the capacitor, and the voltage across it, decrease as $e^{-t/RC}$.

Electric shocks are caused by current passing through the body. To avoid shocks, your body must not become part of a complete circuit by allowing different parts of your body to touch objects at different potentials. Commonly, shocks are caused by one part of your body touching ground $(V = 0)$ and another part touching a nonzero electric potential.

An **ammeter** measures current. An analog ammeter consists of a galvanometer and a parallel **shunt resistor** that carries most of the current. An analog **voltmeter** consists of a galvanometer and a series resistor. An ammeter is inserted *into* the circuit whose current is to be measured. A voltmeter is external, being connected in parallel to the element whose voltage is to be measured. Digital voltmeters have greater internal resistance and so affect the circuit to be measured less than do analog meters.

Questions

1. Explain why birds can sit on power lines safely, even though the wires have no insulation around them, whereas leaning a metal ladder up against a power line is extremely dangerous.
2. Discuss the advantages and disadvantages of Christmas tree lights connected in parallel versus being connected in series.
3. If all you have is a 120-V line, would it be possible to light several 6-V lamps without burning them out? How?
4. Two lightbulbs of resistance R_1 and R_2 $(R_2 > R_1)$ and a battery are all connected in series. Which bulb is brighter? What if they are connected in parallel? Explain.
5. Household outlets are often double outlets. Are these connected in series or parallel? How do you know?
6. With two identical lightbulbs and two identical batteries, explain how and why you would arrange the bulbs and batteries in a circuit to get the maximum possible total power to the lightbulbs. (Ignore internal resistance of batteries.)
7. If two identical resistors are connected in series to a battery, does the battery have to supply more power or less power than when only one of the resistors is connected? Explain.
8. You have a single 60-W bulb lit in your room. How does the overall resistance of your room's electric circuit change when you turn on an additional 100-W bulb? Explain.
9. Compare and discuss the formulas for resistors and for capacitors (Section 24–3) when connected in series and in parallel.
10. When applying Kirchhoff's loop rule (such as in Fig. 26–33), does the sign (or direction) of a battery's emf depend on the direction of current through the battery? What about the terminal voltage?

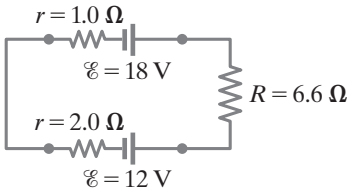


FIGURE 26–33
Question 10.

11. Different lamps might have batteries connected in either of the two arrangements shown in Fig. 26–34. What would be the advantages of each scheme?

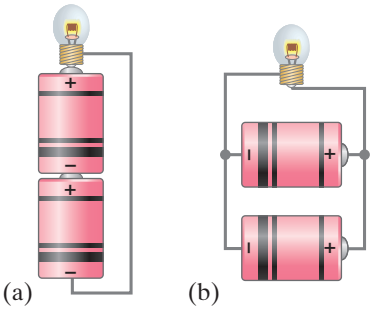


FIGURE 26–34
Question 11.

12. For what use are batteries connected in series? For what use are they connected in parallel? Does it matter if the batteries are nearly identical or not in either case?
13. Can the terminal voltage of a battery ever exceed its emf? Explain.
14. Explain in detail how you could measure the internal resistance of a battery.
15. In an *RC* circuit, current flows from a battery until the capacitor is completely charged. Is the total energy supplied by the battery equal to the total energy stored by the capacitor? If not, where does the extra energy go?
16. Given the circuit shown in Fig. 26–35, use the words "increases," "decreases," or "stays the same" to complete the following statements:
    (*a*) If R_7 increases, the potential difference between A and E _____. Assume no resistance in Ⓐ and ℰ.
    (*b*) If R_7 increases, the potential difference between A and E _____. Assume Ⓐ and ℰ have resistance.
    (*c*) If R_7 increases, the voltage drop across R_4 _____.
    (*d*) If R_2 decreases, the current through R_1 _____.
    (*e*) If R_2 decreases, the current through R_6 _____.
    (*f*) If R_2 decreases, the current through R_3 _____.
    (*g*) If R_5 increases, the voltage drop across R_2 _____.
    (*h*) If R_5 increases, the voltage drop across R_4 _____.
    (*i*) If R_2, R_5, and R_7 increase, ℰ $(r = 0)$ _____.

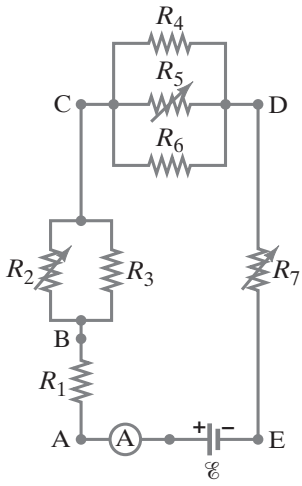


FIGURE 26–35
Question 16. R_2, R_5, and R_7 are **variable resistors** (you can change their resistance); they are, given the symbol ⟋.

17. Design a circuit in which two different switches of the type shown in Fig. 26–36 can be used to operate the same lightbulb from opposite sides of a room.

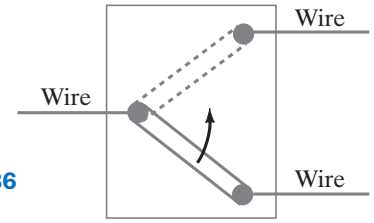


FIGURE 26–36
Question 17.

18. Why is it more dangerous to turn on an electric appliance when you are standing outside in bare feet than when you are inside wearing shoes with thick soles?
19. What is the main difference between an analog voltmeter and an analog ammeter?
20. What would happen if you mistakenly used an ammeter where you needed to use a voltmeter?
21. Explain why an ideal ammeter would have zero resistance and an ideal voltmeter infinite resistance.
22. A voltmeter connected across a resistor always reads *less* than the actual voltage (i.e., when the meter is not present). Explain.
23. A small battery-operated flashlight requires a single 1.5-V battery. The bulb is barely glowing. But when you take the battery out and check it with a digital voltmeter, it registers 1.5 V. How would you explain this?

MisConceptual Questions

1. In which circuits shown in Fig. 26–37 are resistors connected in series?

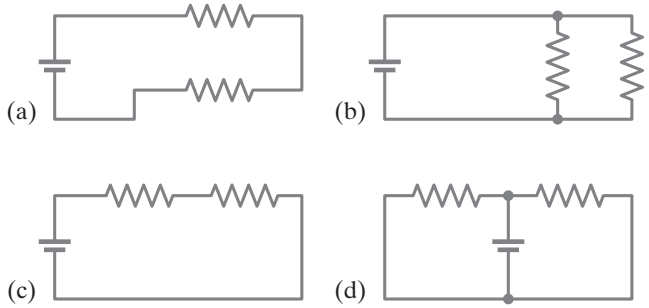


FIGURE 26–37 MisConceptual Question 1.

2. Which resistors in Fig. 26–38 are connected in parallel?
(*a*) All three.
(*b*) R_1 and R_2.
(*c*) R_2 and R_3.
(*d*) R_1 and R_3.
(*e*) None of the above.

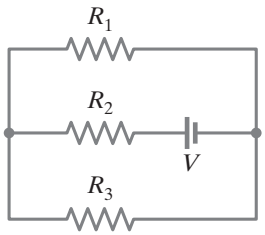


FIGURE 26–38 MisConceptual Question 2.

3. A 10,000-Ω resistor is placed in series with a 100-Ω resistor. The current in the 10,000-Ω resistor is 10 A. If the resistors are swapped, how much current flows through the 100-Ω resistor?
(*a*) >10 A.
(*b*) <10 A.
(*c*) 10 A.
(*d*) Need more information about the circuit.

4. Figure 26–39 shows three identical bulbs in a circuit. What happens to the brightness of bulb A if you replace bulb B with just a connecting wire?
(*a*) Bulb A gets brighter.
(*b*) Bulb A gets dimmer.
(*c*) Bulb A's brightness does not change.
(*d*) Bulb A goes out.

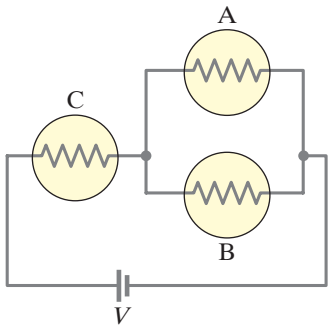


FIGURE 26–39 MisConceptual Question 4.

5. When the switch shown in Fig. 26–40 is closed, what will happen to the voltage across resistor R_4? It will
(*a*) increase.
(*b*) decrease.
(*c*) stay the same.

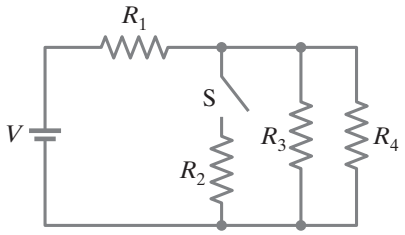


FIGURE 26–40 MisConceptual Questions 5 and 6.

6. When the switch shown in Fig. 26–40 is closed, what will happen to the voltage across resistor R_1? It will
(*a*) increase.
(*b*) decrease.
(*c*) stay the same.

7. As a capacitor is being charged in an *RC* circuit, the current flowing through the resistor is
(*a*) increasing.
(*b*) decreasing.
(*c*) constant.
(*d*) zero.

8. For the circuit shown in Fig. 26–41, what happens when the switch S is closed?
(*a*) Nothing. Current cannot flow through the capacitor.
(*b*) The capacitor immediately charges up to the battery emf.
(*c*) The capacitor eventually charges up to the full battery emf at a rate determined by *R* and *C*.
(*d*) The capacitor charges up to a fraction of the battery emf determined by *R* and *C*.
(*e*) The capacitor charges up to a fraction of the battery emf determined by *R* only.

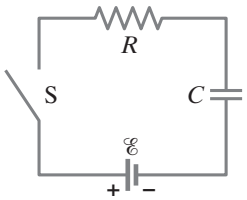


FIGURE 26–41 MisConceptual Question 8.

9. Which resistor shown in Fig. 26–42 has the greatest current going through it? Assume they all have the same resistance *R*.
(*a*) R_1.
(*b*) R_1 and R_2.
(*c*) R_3 and R_4.
(*d*) R_5.
(*e*) All of them the same.

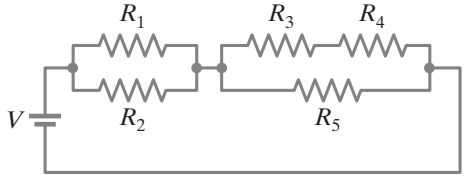


FIGURE 26–42 MisConceptual Question 9.

10. The capacitor in the circuit shown in Fig. 26–43 has already been charged to an initial value Q_0. When the switch is closed, it discharges through the resistor. It takes 2.0 seconds for the charge to drop to $\frac{1}{2}Q_0$. How long does it take to drop to $\frac{1}{4}Q_0$?
(*a*) 3.0 seconds.
(*b*) 4.0 seconds.
(*c*) Between 2.0 and 3.0 seconds.
(*d*) Between 3.0 and 4.0 seconds.
(*e*) More than 4.0 seconds.

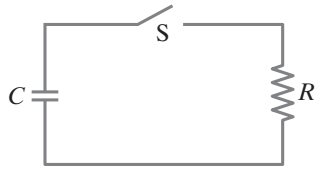


FIGURE 26–43 MisConceptual Question 10.

11. A resistor and a capacitor are used in series to control the timing in the circuit of a heart pacemaker. To design a pacemaker that can double the heart rate when the patient is exercising, which statement below is true? The capacitor
(*a*) needs to discharge faster, so the resistance should be decreased.
(*b*) needs to discharge faster, so the resistance should be increased.
(*c*) needs to discharge slower, so the resistance should be decreased.
(*d*) needs to discharge slower, so the resistance should be increased.
(*e*) does not affect the timing, regardless of the resistance.

12. Why is an appliance cord with a three-prong plug safer than one with two prongs?
(*a*) The 120 V from the outlet is split among three wires, so it isn't as high a voltage as when it is only split between two wires.
(*b*) Three prongs fasten more securely to the wall outlet.
(*c*) The third prong grounds the case, so the case cannot reach a harmful voltage.
(*d*) The third prong acts as a ground wire, so the electrons have an easier time leaving the appliance. As a result, fewer electrons build up in the appliance.
(*e*) The third prong controls the capacitance of the appliance, so it can't build up a high voltage.

13. If ammeters and voltmeters are not to significantly alter the quantities they are measuring,
(*a*) the resistance of an ammeter and a voltmeter should be much higher than that of the circuit element being measured.
(*b*) the resistance of an ammeter should be much lower, and the resistance of a voltmeter should be much higher, than those of the circuit being measured.
(*c*) the resistance of an ammeter should be much higher, and the resistance of a voltmeter should be much lower, than those of the circuit being measured.
(*d*) the resistance of an ammeter and a voltmeter should be much lower than that of the circuit being measured.
(*e*) None of the above.

Problems

26–1 Emf and Terminal Voltage

1. (I) Calculate the terminal voltage for a battery with an internal resistance of 0.900 Ω and an emf of 6.00 V when the battery is connected in series with (*a*) a 61.0-Ω resistor, and (*b*) a 610-Ω resistor.

2. (I) Four 1.50-V cells are connected in series to a 12-Ω lightbulb. If the resulting current is 0.47 A, what is the internal resistance of each cell, assuming they are identical and neglecting the resistance of the wires?

3. (II) A 1.5-V dry cell can be tested by connecting it to a low-resistance ammeter. It should be able to supply at least 25 A. What is the internal resistance of the cell in this case, assuming it is much greater than that of the ammeter?

4. (II) What is the internal resistance of a 12.0-V car battery whose terminal voltage drops to 8.2 V when the starter motor draws 95 A? What is the resistance of the starter?

26–2 Resistors in Series and Parallel

[In these Problems neglect the internal resistance of a battery unless the Problem refers to it.]

5. (I) A 720-Ω and an 1800-Ω resistor are connected in series with a 12-V battery. What is the voltage across the 1800-Ω resistor?

6. (I) Three 45-Ω lightbulbs and three 65-Ω lightbulbs are connected in series. (*a*) What is the total resistance of the circuit? (*b*) What is the total resistance if all six are wired in parallel?

7. (II) How many 10-Ω resistors must be connected in series to give an equivalent resistance to five 100-Ω resistors connected in parallel?

8. (II) Suppose that you have a 520-Ω, a 760-Ω, and a 1.20-kΩ resistor. What is (*a*) the maximum, and (*b*) the minimum resistance you can obtain by combining these?

9. (II) A 75-W, 120-V bulb is connected in parallel with a 25-W, 120-V bulb. What is the net resistance?

10. (II) Three 2.20-kΩ resistors can be connected together in four different ways, making combinations of series and/or parallel circuits. What are these four ways, and what is the net resistance in each case?

11. (II) A battery with an emf of 12.0 V shows a terminal voltage of 11.8 V when operating in a circuit with two lightbulbs, each rated at 3.0 W (at 12.0 V), which are connected in parallel. What is the battery's internal resistance?

12. (II) Eight identical bulbs are connected in series across a 120-V line. (*a*) What is the voltage across each bulb? (*b*) If the current is 0.65 A, what is the resistance of each bulb, and what is the power dissipated in each?

13. (II) Eight bulbs are connected in parallel to a 120-V source by two long leads of total resistance 1.4 Ω. If 210 mA flows through each bulb, what is the resistance of each, and what fraction of the total power is wasted in the leads?

14. (II) A close inspection of an electric circuit reveals that a 480-Ω resistor was inadvertently soldered in the place where a 330-Ω resistor is needed. How can this be fixed without removing anything from the existing circuit?

15. (II) Ten 7.0-W Christmas tree lights are connected in series to each other and to a 120-V source. What is the resistance of each bulb?

16. (II) Determine (*a*) the equivalent resistance of the circuit shown in Fig. 26–44, (*b*) the voltage across each resistor, and (*c*) the current through each resistor.

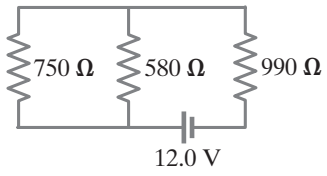


FIGURE 26–44
Problem 16.

17. (II) (*a*) Determine the equivalent resistance of the "ladder" of equal 215-Ω resistors shown in Fig. 26–45. In other words, what resistance would an ohmmeter read if connected between points A and B? (*b*) What is the current through each of the three resistors on the left if a 50.0-V battery is connected between points A and B?

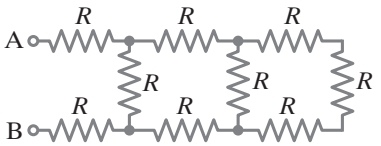


FIGURE 26–45
Problem 17.

18. (II) What is the net resistance of the circuit connected to the battery in Fig. 26–46?

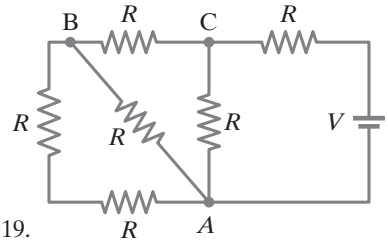


FIGURE 26–46
Problems 18 and 19.

19. (II) Calculate the current through each resistor in Fig. 26–46 if each resistance $R = 4.25\,\text{k}\Omega$ and $V = 12.0\,\text{V}$. What is the potential difference between points A and B?

20. (II) A power supply has a fixed output voltage of 12.0 V, but you need $V_T = 3.0\,\text{V}$ output for an experiment. (*a*) Using the **voltage divider** shown in Fig. 26–47, what should R_2 be if R_1 is 16.5 Ω? (*b*) What will the terminal voltage V_T be if you connect a load to the 3.0-V output, assuming the load has a resistance of 7.0 Ω?

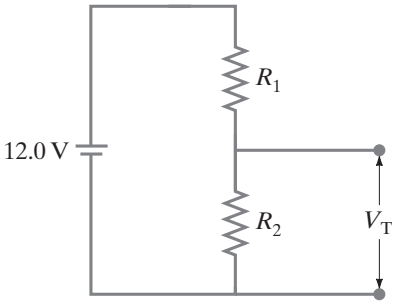


FIGURE 26–47
Problem 20.

21. (II) Suppose that you have a 9.0-V battery and wish to apply a voltage of only 3.5 V. Given an unlimited supply of 1.0-Ω resistors, how could you connect them to make a "voltage divider" that produces a 3.5-V output for a 9.0-V input?

22. (III) Two resistors when connected in series to a 120-V line use one-fourth the power that is used when they are connected in parallel. If one resistor is 4.3 kΩ, what is the resistance of the other?

23. (III) A 2.5-kΩ and a 3.7-kΩ resistor are connected in parallel; this combination is connected in series with a 1.4-kΩ resistor. If each resistor is rated at 1.0 W (maximum without overheating), what is the maximum voltage that can be applied across the whole network?

24. (III) Three equal resistors (R) are connected to a battery as shown in Fig. 26–48. Qualitatively, what happens to (*a*) the voltage drop across each of these resistors, (*b*) the current flow through each, and (*c*) the terminal voltage of the battery, when the switch S is opened, after having been closed for a long time? (*d*) If the emf of the battery is 9.0 V, what is its terminal voltage when the switch is closed if the internal resistance r is 0.40 Ω and $R = 5.50\,\Omega$? (*e*) What is the terminal voltage when the switch is open?

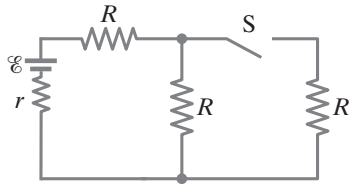


FIGURE 26–48
Problem 24.

25. (III) Consider the network of resistors shown in Fig. 26–49. Answer qualitatively: (*a*) What happens to the voltage across each resistor when the switch S is closed? (*b*) What happens to the current through each when the switch is closed? (*c*) What happens to the power output of the battery when the switch is closed? (*d*) Let $R_1 = R_2 = R_3 = R_4 = 135\,\Omega$ and $V = 22.0\,\text{V}$. Determine the current through each resistor before and after closing the switch. Are your qualitative predictions confirmed?

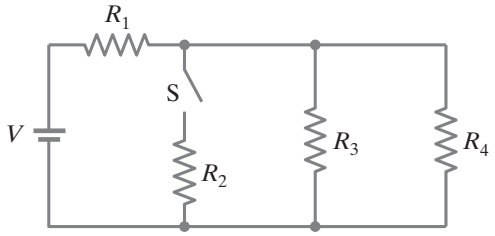


FIGURE 26–49
Problem 25.

26. (III) You are designing a wire resistance heater to heat an enclosed container of gas. For the apparatus to function properly, this heater must transfer heat to the gas at a very constant rate. While in operation, the resistance of the heater will always be close to the value $R = R_0$, but may fluctuate slightly causing its resistance to vary a small amount ΔR ($\ll R_0$). To maintain the heater at constant power, you design the circuit shown in Fig. 26–50, which includes two resistors, each of resistance R'. Determine the value for R' so that the heater power P will remain constant even if its resistance R fluctuates by a small amount. [*Hint*: If $\Delta R \ll R_0$, then $\Delta P \approx \Delta R \left.\frac{dP}{dR}\right|_{R=R_0}$.]

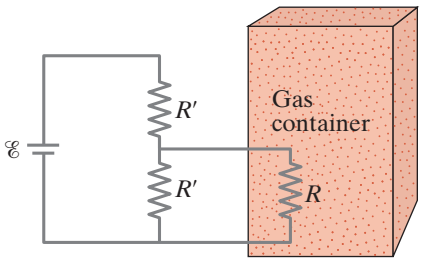


FIGURE 26–50
Problem 26.

26–3 Kirchhoff's Rules

27. (I) Calculate the current in the circuit of Fig. 26–51, and show that the sum of all the voltage changes around the circuit is zero.

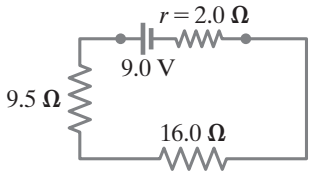


FIGURE 26–51
Problem 27.

28. (II) Determine the terminal voltage of each battery in Fig. 26–52.

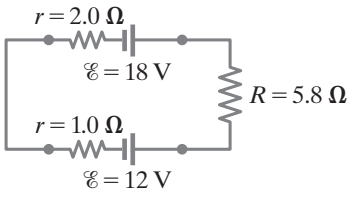


FIGURE 26–52
Problem 28.

29. (II) For the circuit shown in Fig. 26–53, find the potential difference between points a and b. Each resistor has $R = 180\ \Omega$ and each battery is 1.5 V.

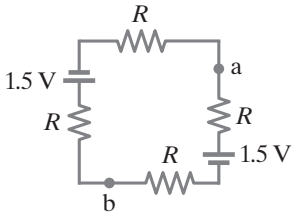


FIGURE 26–53
Problem 29.

30. (II) Calculate the magnitude and direction of the currents in each resistor of Fig. 26–54.

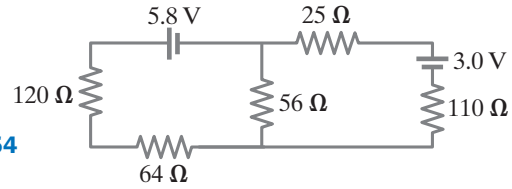


FIGURE 26–54
Problem 30.

31. (II) (*a*) What is the potential difference between points a and d in Fig. 26–55 (similar to Fig. 26–12, Example 26–8), and (*b*) what is the terminal voltage of each battery?

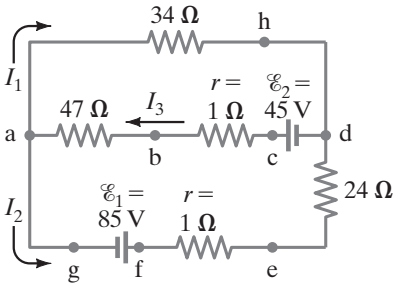


FIGURE 26–55
Problem 31.

32. (II) Determine the magnitudes and directions of the currents through R_1 and R_2 in Fig. 26–56.

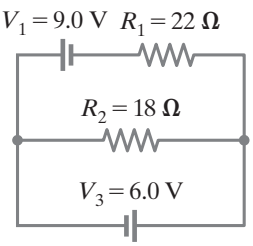


FIGURE 26–56
Problems 32 and 33.

33. (II) Repeat Problem 32, now assuming that each battery has an internal resistance $r = 1.2\ \Omega$.

34. (II) Determine the magnitudes and directions of the currents in each resistor shown in Fig. 26–57. The batteries have emfs of $\mathscr{E}_1 = 9.0$ V and $\mathscr{E}_2 = 12.0$ V and the resistors have values of $R_1 = 25\ \Omega$, $R_2 = 48\ \Omega$, and $R_3 = 35\ \Omega$. (*a*) Ignore internal resistance of the batteries. (*b*) Assume each battery has internal resistance $r = 1.0\ \Omega$.

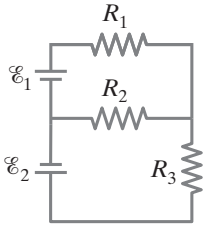


FIGURE 26–57
Problem 34.

35. (II) A voltage V is applied to n identical resistors connected in parallel. If the resistors are instead all connected in series with the applied voltage, show that the power transformed is decreased by a factor n^2.

36. (III) (*a*) Determine the currents I_1, I_2, and I_3 in Fig. 26–58. Assume the internal resistance of each battery is $r = 1.0\ \Omega$. (*b*) What is the terminal voltage of the 6.0-V battery?

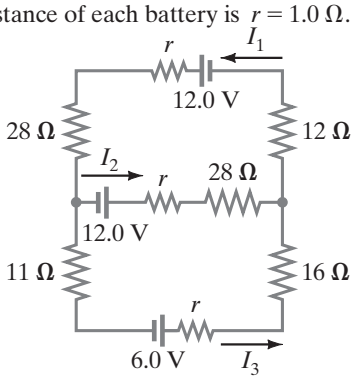


FIGURE 26–58
Problems 36 and 37.

37. (III) What would the current I_1 be in Fig. 26–58 if the 12-Ω resistor is shorted out (resistance = 0)? Let $r = 1.0\ \Omega$.

38. (III) Determine the current through each of the resistors in Fig. 26–59.

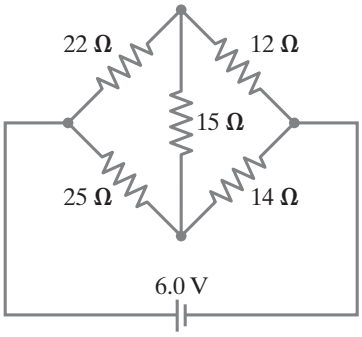


FIGURE 26–59
Problems 38 and 39.

39. (III) If the 25-Ω resistor in Fig. 26–59 is shorted out (resistance = 0), what then would be the current through the 15-Ω resistor?

40. (III) Twelve resistors, each of resistance R, are connected at the corners of a cube as shown in Fig. 26–60. Determine the equivalent resistance (*a*) between points a and b, the ends of a side; (*b*) between points a and c, the ends of a face diagonal; (*c*) between points a and d, the ends of the volume diagonal. [*Hint*: Apply an emf between the 2 points in each case and determine currents; use symmetry at junctions.]

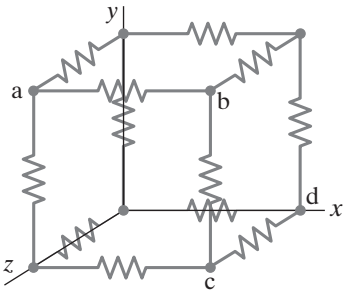


FIGURE 26–60
Problem 40.

41. (III) Determine the net resistance in Fig. 26–61 (*a*) between points a and c, and (*b*) between points a and b. Assume $R' \neq R$. [*Hint*: Apply an emf between the two points in each case and determine currents; use symmetry at junctions.]

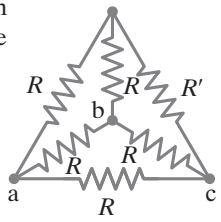


FIGURE 26–61
Problem 41.

42. (III) (*a*) A network of five equal resistors R is connected to a battery $\mathscr{E}$ as shown in Fig. 26–62. Determine the current I that flows out of the battery. (*b*) Use the value determined for I to find the single resistor R_{eq} that is equivalent to the five-resistor network.

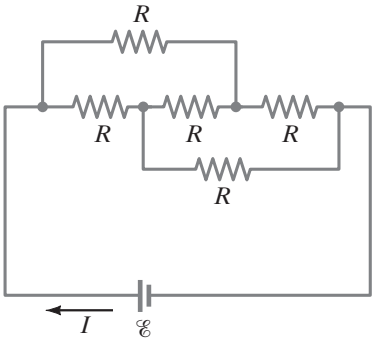


FIGURE 26–62
Problem 42.

26–4 Emfs Combined, Battery Charging

43. (II) Suppose two batteries, with unequal emfs of 2.00 V and 3.00 V, are connected as shown in Fig. 26–63. If each internal resistance is $r = 0.350\ \Omega$, and $R = 4.00\ \Omega$, what is the voltage across the resistor R?

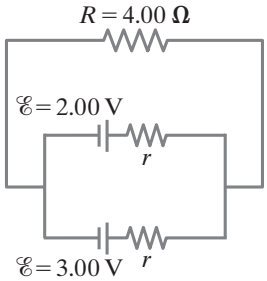


FIGURE 26–63
Problem 43.

26–5 *RC* Circuits

44. (I) Estimate the value of resistances needed to make a variable timer for intermittent windshield wipers: one wipe every 15 s, 8 s, 4 s, 2 s, 1 s. Assume the capacitor used is on the order of 1 μF. See Fig. 26–64.



FIGURE 26–64 Problem 44.

45. (II) In Fig. 26–65 (same as Fig. 26–16a), the total resistance is 15.0 kΩ, and the battery's emf is 24.0 V. If the time constant is measured to be 24.0 μs, calculate (*a*) the total capacitance of the circuit and (*b*) the time it takes for the voltage across the resistor to reach 14.0 V after the switch is closed.

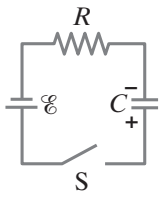


FIGURE 26–65
Problems 45 and 46.

46. (II) How long does it take for the energy stored in a capacitor in a series *RC* circuit (Fig. 26–65) to reach 75% of its maximum value? Express answer in terms of the time constant $\tau = RC$.

47. (II) Two 3.8-μF capacitors, two 2.2-kΩ resistors, and a 12.0-V source are connected in series. Starting from the uncharged state, how long does it take for the current to drop from its initial value to 1.35 mA?

48. (II) A parallel-plate capacitor is filled with a dielectric of dielectric constant K and high resistivity ρ (it conducts very slightly). This capacitor can be modeled as a pure capacitance C in parallel with a resistance R. Assume a battery places a charge $+Q$ and $-Q$ on the capacitor's opposing plates and is then disconnected. Show that the capacitor discharges with a time constant $\tau = K\epsilon_0\rho$ (known as the *dielectric relaxation time*). Evaluate τ if the dielectric is glass with $\rho = 1.0 \times 10^{12}\ \Omega \cdot \text{m}$ and $K = 5.0$.

49. (II) The *RC* circuit of Fig. 26–66 (same as Fig. 26–17a) has $R = 8.7\ \text{k}\Omega$ and $C = 3.0\ \mu\text{F}$. The capacitor is at voltage V_0 at $t = 0$, when the switch is closed. How long does it take the capacitor to discharge to 0.25% of its initial voltage?

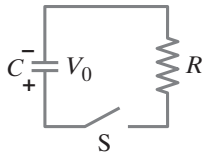


FIGURE 26–66
Problem 49.

50. (II) Consider the circuit shown in Fig. 26–67, where all resistors have the same resistance R. At $t = 0$, with the capacitor C uncharged, the switch is closed. (*a*) At $t = 0$, the three currents can be determined by analyzing a simpler, but equivalent, circuit. Identify this simpler circuit and use it to find the values of I_1, I_2, and I_3 at $t = 0$. (*b*) At $t = \infty$, the currents can be determined by analyzing a simpler, equivalent circuit. Identify this simpler circuit and implement it in finding the values of I_1, I_2, and I_3 at $t = \infty$. (*c*) At $t = \infty$, what is the potential difference across the capacitor?

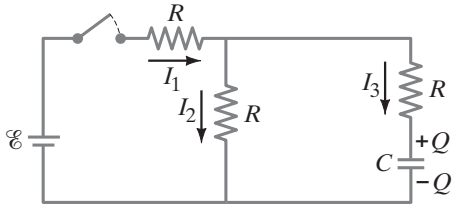


FIGURE 26–67 Problem 50.

51. (II) In the circuit shown in Fig. 26–68, switch S is closed at time $t = 0$. (*a*) After the capacitor is fully charged, what is the voltage across it? How much charge is on it? (*b*) Switch S is now opened. How long does it now take for the capacitor to discharge until it has only 2.0% of its initial charge?

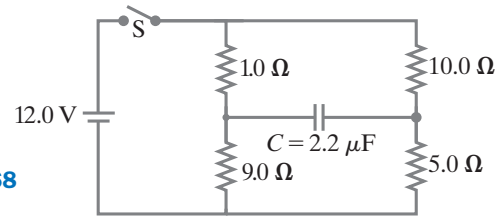


FIGURE 26–68
Problem 51.

52. (III) Determine the time constant for charging the capacitor in the circuit of Fig. 26–69. [*Hint*: Use Kirchhoff's rules.] (*b*) What is the maximum charge on the capacitor?

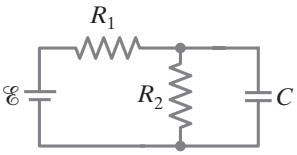


FIGURE 26–69
Problem 52.

53. (III) Two resistors and two uncharged capacitors are arranged as shown in Fig. 26–70. Then a potential difference of 24 V is applied across the combination as shown. (*a*) What is the potential at point a with switch S open? (Let $V = 0$ at the negative terminal of the source.) (*b*) What is the potential at point b with the switch open? (*c*) When the switch is closed, what is the final potential of point b? (*d*) How much charge flows through the switch S after it is closed?

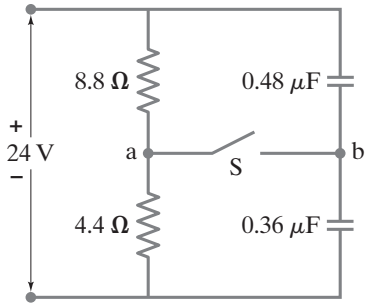


FIGURE 26–70 Problems 53 and 54.

54. (III) Suppose the switch S in Fig. 26–70 is closed. What is the time constant (or time constants) for charging the capacitors after the 24 V is applied?

26–7 Ammeters and Voltmeters

55. (I) An ammeter has a sensitivity of 35,000 Ω/V. What current in the galvanometer produces full-scale deflection?

56. (I) What is the resistance of a voltmeter on the 250-V scale if the meter sensitivity is 35,000 Ω/V?

57. (II) A galvanometer has a sensitivity of 45 kΩ/V and internal resistance 20.0 Ω. How could you make this into (*a*) an ammeter that reads 1.0 A full scale, or (*b*) a voltmeter reading 1.00 V full scale?

58. (II) A galvanometer has an internal resistance of 32 Ω and deflects full scale for a 48-μA current. Describe how to use this galvanometer to make (*a*) an ammeter to read currents up to 25 A, and (*b*) a voltmeter to give a full scale deflection of 250 V.

59. (II) A particular *digital meter* is based on an electronic module that has an internal resistance of 120 MΩ and a *full-scale sensitivity* of 250 mV. Two resistors connected as shown in Fig. 26–71 can be used to change the voltage range. Assume $R_1 = 12$ MΩ. Find the value of R_2 that will result in a voltmeter with a full-scale range of 40 V.

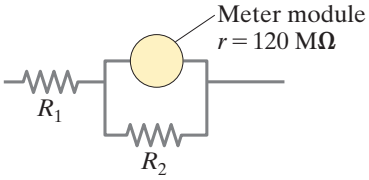


FIGURE 26–71
Problem 59.

60. (II) A milliammeter reads 25 mA full scale. It consists of a 0.20-Ω resistor in parallel with a 33-Ω galvanometer. How can you change this ammeter to a voltmeter giving a full scale reading of 25 V without taking the ammeter apart? What will be the sensitivity (Ω/V) of your voltmeter?

61. (II) A 35-V battery of negligible internal resistance is connected to a 44-kΩ and a 27-kΩ resistor in series. What reading will a voltmeter, of internal resistance 85 kΩ, give when used to measure the voltage across each resistor? What is the percent inaccuracy due to meter resistance for each case?

62. (II) An ammeter whose internal resistance is 53 Ω reads 5.25 mA when connected in a circuit containing a battery and two resistors in series whose values are 760 Ω and 480 Ω. What is the actual current when the ammeter is absent?

63. (II) A battery with $\mathscr{E} = 12.0$ V and internal resistance $r = 1.0$ Ω is connected to two 7.5-kΩ resistors in series. An ammeter of internal resistance 0.50 Ω measures the current, and at the same time a voltmeter with internal resistance 18 kΩ measures the voltage across one of the 7.5-kΩ resistors in the circuit. What do the ammeter and voltmeter read? What is the % "error" from the current and voltage *without* the meters?

64. (II) A 12.0-V battery (assume the internal resistance = 0) is connected to two resistors in series. A voltmeter whose internal resistance is 18.0 kΩ measures 5.5 V and 4.0 V, respectively, when connected across each of the resistors in turn. What is the resistance of each resistor?

65. (III) Two 8.4-kΩ resistors are placed in series and connected to a battery. A voltmeter of sensitivity 1000 Ω/V is on the 3.0-V scale and reads 2.3 V when placed across either resistor. What is the emf of the battery? (Ignore its internal resistance.)

66. (III) When the resistor R in Fig. 26–72 is 45 Ω, the high-resistance voltmeter reads 9.7 V. When R is replaced by a 14.0-Ω resistor, the voltmeter reading drops to 8.1 V. What are the emf and internal resistance of the battery?

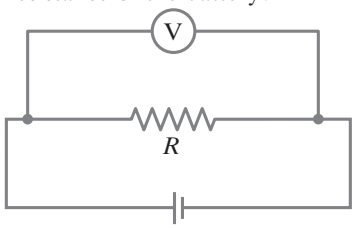


FIGURE 26–72
Problem 66.

67. (III) What internal resistance should the voltmeter of Example 26–16 have to be in error by less than 5%?

General Problems

68. Suppose that you wish to apply a 0.25-V potential difference between two points on the human body. The resistance between those two points is about 1800 Ω, and you only have a 1.5-V battery. How can you connect up one or more resistors to produce the desired voltage? This is another example of a *voltage divider* (Problem 20).

69. Suppose you want to run some apparatus that is 45 m from an electric outlet. Each of the wires connecting your apparatus to the 120-V source has a resistance per unit length of 0.0065 Ω/m. If your apparatus draws 3.0 A, what will be the voltage drop across the connecting wires and what voltage will be applied to your apparatus?

70. For the circuit shown in Fig. 26–17a, show that the decrease in energy stored in the capacitor from $t = 0$ until one time constant has elapsed equals the energy dissipated as heat in the resistor.

71. A **Wheatstone bridge** is a type of "bridge circuit" used to make measurements of resistance. The unknown resistance to be measured, R_x, is placed in the circuit with accurately known resistances R_1, R_2, and R_3 (Fig. 26–73). One of these, R_3, is a variable resistor which is adjusted so that when the switch is closed momentarily, the ammeter Ⓐ shows zero current flow. (*a*) Determine R_x in terms of R_1, R_2, and R_3. (*b*) If a Wheatstone bridge is "balanced" when $R_1 = 685\ \Omega$, $R_2 = 972\ \Omega$, and $R_3 = 78.6\ \Omega$, what is the value of the unknown resistance?

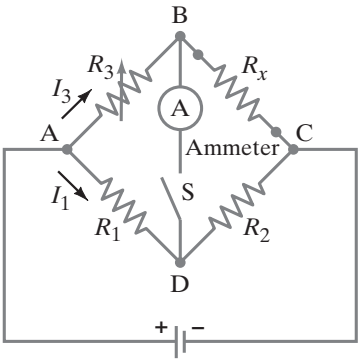


FIGURE 26–73 Problems 71 and 72. Wheatstone bridge.

72. An unknown length of platinum wire 1.22 mm in diameter is placed as the unknown resistance in a Wheatstone bridge (see Problem 71, Fig. 26–73). Arms 1 and 2 have resistance of 38.0 Ω and 29.2 Ω, respectively. Balance is achieved when R_3 is 3.26 Ω. How long is the platinum wire?

73. The internal resistance of a 1.35-V mercury cell is 0.030 Ω, whereas that of a 1.5-V dry cell is 0.35 Ω. Explain why three mercury cells can more effectively power a 2.5-W hearing aid that requires 4.0 V than can three dry cells.

74. How many $\frac{1}{2}$-W resistors, each of the same resistance, must be used to produce an equivalent 3.2-kΩ, 4.5-W resistor? What is the resistance of each, and how must they be connected? Do not exceed $P = \frac{1}{2}$ W in each resistor. (See page 730.)

75. A **solar cell**, 3.0 cm square, has an output of 350 mA at 0.80 V when exposed to full sunlight. A solar panel that delivers close to 1.3 A of current at an emf of 120 V to an external load is needed. How many cells will you need to create the panel? How big a panel will you need, and how should you connect the cells to one another? How can you optimize the output of your solar panel?

76. A battery produces 40.8 V when 7.10 A is drawn from it, and 47.3 V when 2.80 A is drawn. What are the emf and internal resistance of the battery?

77. The current through the 4.0-kΩ resistor in Fig. 26–74 is 2.85 mA. What is the terminal voltage V_{ba} of the "unknown" battery? (There are two answers. Why?)

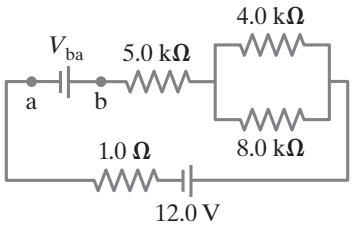


FIGURE 26–74 Problem 77.

78. In the circuit shown in Fig. 26–75, the 33-Ω resistor dissipates 0.80 W. What is the battery voltage?

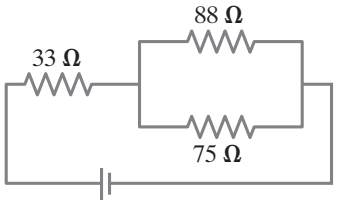


FIGURE 26–75 Problem 78.

79. The current through the 20-Ω resistor in Fig. 26–76 does not change whether the two switches S_1 and S_2 are both open or both closed. Use this clue to determine the value of the unknown resistance R.

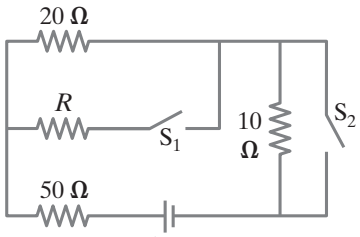


FIGURE 26–76 Problem 79.

80. A **three-way lightbulb** using wire filaments can produce 50 W, 100 W, or 150 W, at 120 V. Such a bulb contains two filaments that can be connected to the 120 V individually or in parallel (Fig. 26–77). (*a*) Describe how the connections to the two filaments are made to give each of the three wattages. (*b*) What must be the resistance of each filament?



FIGURE 26–77 Problem 80.

81. A 10.0-m length of wire consists of 5.0 m of copper followed by 5.0 m of aluminum, both of diameter 1.4 mm. A voltage difference of 75 mV is placed across the composite wire. (*a*) What is the total resistance (sum) of the two wires? (*b*) What is the current through the wire? (*c*) What are the voltages across the aluminum part and across the copper part?

82. Electrocardiographs are often connected as shown in Fig. 26–78. The lead wires to the legs are said to be capacitively coupled. A time constant of 3.0 s is typical and allows rapid changes in potential to be recorded accurately. If $C = 3.0\,\mu\text{F}$, what value must R have? [*Hint*: Consider each leg as a separate circuit.]

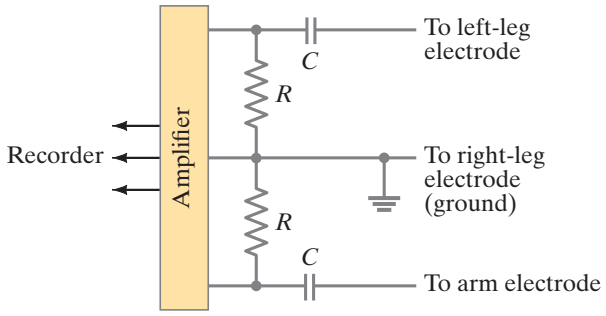


FIGURE 26–78 Problem 82.

83. Electricity can be a hazard in hospitals, particularly to patients who are connected to electrodes, such as an ECG. Suppose that the motor of a motorized bed shorts out to the bed frame, and the bed frame's connection to a ground has broken (or was not there in the first place). If a nurse touches the bed and the patient at the same time, the nurse becomes a conductor and a complete circuit can be made through the patient to ground through the ECG apparatus. This is shown schematically in Fig. 26–79. Calculate the current through the patient.

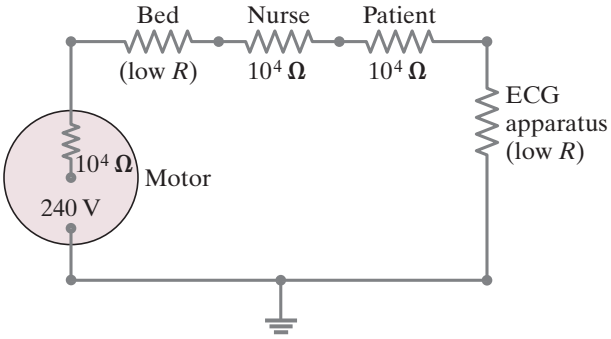


FIGURE 26–79 Problem 83.

84. (*a*) What is the equivalent resistance of the circuit shown in Fig. 26–80? (*b*) What is the current in the 18-Ω resistor? (*c*) What is the current in the 12-Ω resistor? (*d*) What is the power dissipation in the 4.5-Ω resistor?

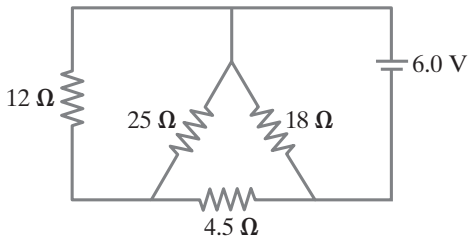


FIGURE 26–80 Problem 84.

85. A flashlight bulb rated at 2.0 W and 3.0 V is operated by a 9.0-V battery. To light the bulb at its rated voltage and power, a resistor R is connected in series as shown in Fig. 26–81. What value should the resistor have?

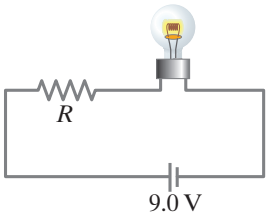


FIGURE 26–81 Problem 85.

86. Some light-dimmer switches use a variable resistor as shown in Fig. 26–82. The slide moves from position $x = 0$ to $x = 1$, and the resistance up to slide position x is proportional to x (the total resistance is $R_{\text{var}} = 150\ \Omega$ at $x = 1$). What is the power expended in the lightbulb if (*a*) $x = 1.00$, (*b*) $x = 0.70$, (*c*) $x = 0.40$?

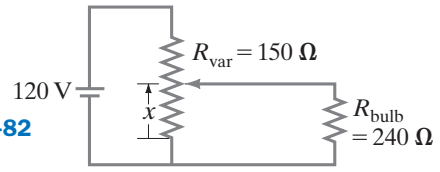


FIGURE 26–82 Problem 86.

87. A heart pacemaker is designed to operate at 72 beats/min using a 5.2-μF capacitor in a simple RC circuit. What value of resistance should be used if the pacemaker is to fire (capacitor discharge) when the voltage reaches 75% of maximum and then drops to 0 V (72 times a minute)?

88. Suppose that a person's body resistance is 950 Ω (moist skin). (*a*) What current passes through the body when the person accidentally is connected to 120 V? (*b*) If there is an alternative path to ground whose resistance is 25 Ω, what then is the current through the body? (*c*) If the voltage source can produce at most 1.5 A, how much current passes through the person in case (*b*)?

89. A **potentiometer** is a device to precisely measure potential differences or emf, using a "null" technique. In the simple potentiometer circuit shown in Fig. 26–83, R' represents the total resistance of the resistor from A to B (which could be a long uniform "slide" wire), whereas R represents the resistance of only the part from A to the movable contact at C. When the unknown emf to be measured, $\mathscr{E}_x$, is placed into the circuit as shown, the movable contact C is moved until the galvanometer G gives a null reading (i.e., zero) when the switch S is closed. The resistance between A and C for this situation we call R_x. Next, a standard emf, $\mathscr{E}_s$, which is known precisely, is inserted into the circuit in place of $\mathscr{E}_x$ and again the contact C is moved until zero current flows through the galvanometer when the switch S is closed. The resistance between A and C now is called R_s. Show that the unknown emf is given by

$$\mathscr{E}_x = \left(\frac{R_x}{R_s}\right)\mathscr{E}_s$$

where R_x, R_s, and $\mathscr{E}_s$ are all precisely known. The *working battery* (at top of circuit diagram), as well as the standard cell $\mathscr{E}_s$, are assumed to be fresh and to give a constant voltage.

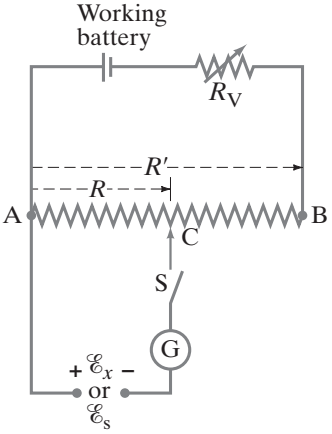


FIGURE 26–83 Potentiometer circuit. Problem 89.

90. Electronic devices often use an *RC* circuit to protect against *power outages* as shown in Fig. 26–84. (*a*) If the protector circuit is supposed to keep the supply voltage at least 75% of full voltage for as long as 0.20 s, how big a resistance R is needed? The capacitor is 7.6 μF. Assume the attached "electronics" draws negligible current. (*b*) Between which two terminals should the device be connected, a and b, b and c, or a and c?

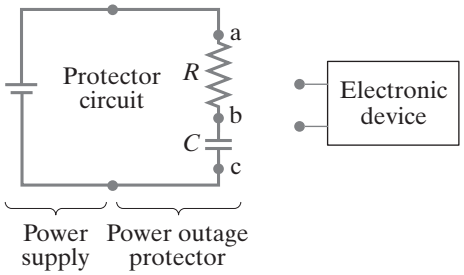


FIGURE 26–84
Problem 90.

91. (*a*) A voltmeter and an ammeter can be connected as shown in Fig. 26–85a to measure a resistance R. If V is the voltmeter reading, and I is the ammeter reading, the value of R will not

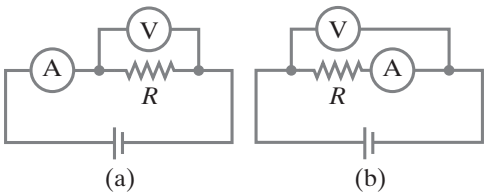


FIGURE 26–85 Problem 91.

quite be V/I (as in Ohm's law) because some current goes through the voltmeter. Show that the actual value of R is

$$\frac{1}{R} = \frac{I}{V} - \frac{1}{R_V},$$

where R_V is the voltmeter resistance. Note that $R \approx V/I$ if $R_V \gg R$. (*b*) A voltmeter and an ammeter can also be connected as shown in Fig. 26–85b to measure a resistance R. Show in this case that

$$R = \frac{V}{I} - R_A,$$

where V and I are the voltmeter and ammeter readings and R_A is the resistance of the ammeter. Note that $R \approx V/I$ if $R_A \ll R$.

92. Figure 26–86 shows the circuit for a simple **sawtooth oscillator**. At time $t = 0$, its switch S is closed. The neon bulb has initially infinite resistance until the voltage across it reaches 90.0 V, and then it begins to conduct with very little resistance (essentially zero). It stops conducting (its resistance becomes essentially infinite) when the voltage drops down to 65.0 V. (*a*) At what time t_1 does the neon bulb reach 90.0 V and start conducting? (*b*) At what time t_2 does the bulb reach 90.0 V for a second time and again become conducting? (*c*) Sketch the sawtooth waveform between $t = 0$ and $t = 0.70$ s.

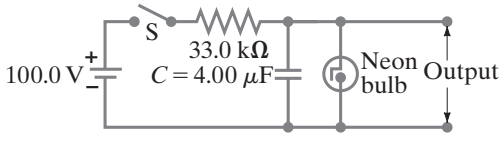


FIGURE 26–86 Problem 92.

93. For the circuit shown in Fig. 26–87, determine (*a*) the current through the 16-V battery and (*b*) the potential difference between points a and b, $V_a - V_b$.

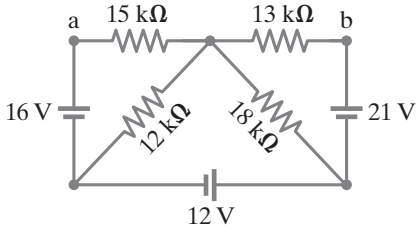


FIGURE 26–87
Problem 93.

94. The performance of the starter circuit in a car can be significantly degraded by a small amount of *corrosion* on a *battery terminal*. Figure 26–88a depicts a properly functioning circuit with a battery (12.5-V emf, 0.02-Ω internal resistance) attached via corrosion-free cables to a starter motor of resistance $R_s = 0.15\ \Omega$. Sometime later, corrosion between a battery terminal and a starter cable introduces an extra series resistance of only $R_C = 0.10\ \Omega$ into the circuit as suggested in Fig. 26–88b. Let P_0 be the power delivered to the starter in the circuit free of corrosion, and let P be the power delivered to the starter with corrosion. Determine the ratio P/P_0.

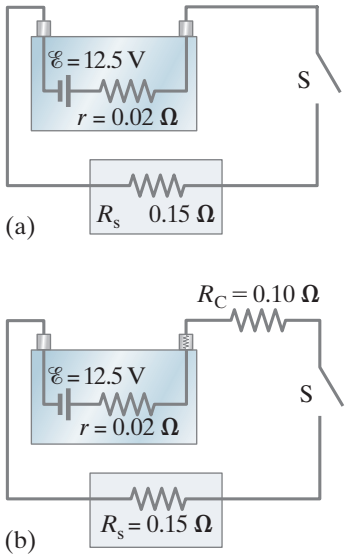


FIGURE 26–88
Problem 94.

95. The circuit shown in Fig. 26–89 is a primitive 4-bit **digital-to-analog converter (DAC)**. In this circuit, to represent each digit (2^n) of a binary number, a "1" has the n^{th} switch closed whereas zero ("0") has the switch open. For example, 0010 is represented by closing switch $n = 1$, while all other switches are open. Show that the voltage V across the 1.0-Ω resistor for the binary numbers 0001, 0010, 0100, and 1010 (which represent 1, 2, 4, 10) follows the pattern that you expect for a 4-bit DAC.

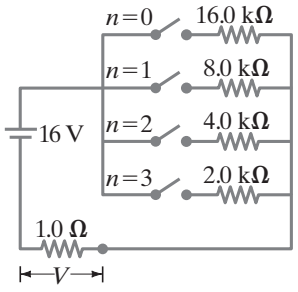


FIGURE 26–89
Problem 95.

96. *Cardiac defibrillators* are discussed in Section 24–4. (*a*) Choose a value for the resistance so that the 1.4-μF capacitor can be charged to 3100 V in 2.0 seconds. Assume that this 3100 V is 95% of the full source voltage. (*b*) The effective resistance of the human body is given in Section 26–6. If the defibrillator discharges with a time constant of 12 ms, what is the effective capacitance of the human body?

97. Determine the current in each resistor of the circuit shown in Fig. 26–90.

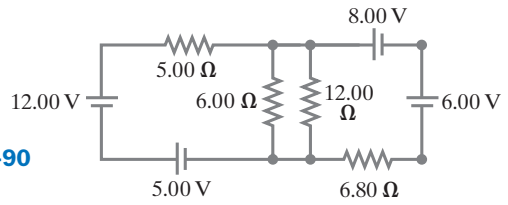


FIGURE 26–90
Problem 97.

98. The circuit shown in Fig. 26–91 uses a neon-filled tube as in Fig. 26–20a. This neon lamp has a threshold voltage V_0 for conduction, because no current flows until the neon gas in the tube is ionized by a sufficiently strong electric field. Once the threshold voltage V_0 is exceeded, the lamp has negligible resistance. The capacitor stores electrical energy, which can be released to flash the lamp. Assume that $C = 0.150\,\mu\text{F}$, $R = 2.15 \times 10^6\,\Omega$, $V_0 = 90.0\,\text{V}$, and $\mathcal{E} = 105\,\text{V}$. (*a*) Assuming the circuit is hooked up to the emf at time $t = 0$, at what time will the light first flash? (*b*) If the value of R is increased, will the time you found in part (*a*) increase or decrease? (*c*) The flashing of the lamp is very brief. Why? (*d*) Explain what happens after the lamp flashes for the first time (see Fig. 26–20 and the text that refers to it).

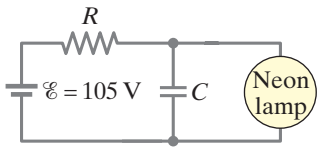


FIGURE 26–91
Problem 98.

99. Measurements made on circuits that contain large resistances can be confusing. Consider a circuit powered by a battery $\mathcal{E} = 15.000\,\text{V}$ with a 10.00-MΩ resistor in series with an unknown resistor R. As shown in Fig. 26–92, a particular voltmeter reads $V_1 = 366\,\text{mV}$ when connected across the 10.00-MΩ resistor, and this meter reads $V_2 = 7.317\,\text{V}$ when connected across R. Determine the value of R. [*Hint*: Define R_V as the voltmeter's internal resistance.]

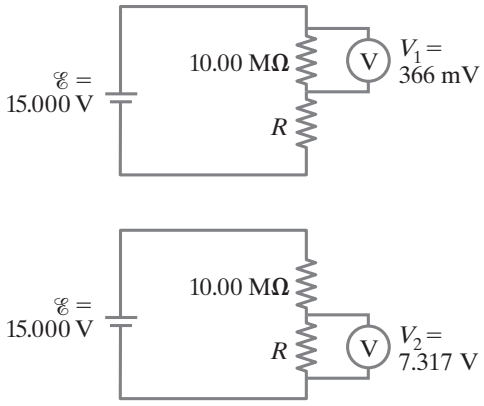


FIGURE 26–92 Problem 99.

100. A typical voltmeter has an internal resistance of 10 MΩ and can only measure potential differences of up to several hundred volts. Figure 26–93 shows the design of a probe to measure a very high voltage V using a voltmeter. If you want the voltmeter to read 50 V when $V = 50\,\text{kV}$, what value R should be used in this probe?

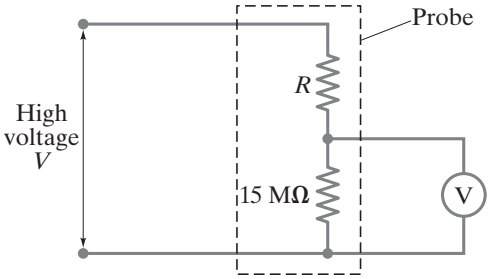


FIGURE 26–93 Problem 100.

101. Consider two unequal resistors, of resistance R_1 and R_2, that are connected either in series or in parallel. Fill in the Table below assuming the electric potential on the low-voltage end of the combination is V_A volts and the potential at the high-voltage end of the combination is V_B volts. First draw diagrams.

Property	Resistors are in Series	Resistors are in Parallel
Equivalent resistance		
Current through equivalent resistance		
Voltage across equivalent resistance		
Voltage across the pair of resistors		
Voltage across each resistor	$V_1 =$ $V_2 =$	$V_1 =$ $V_2 =$
Voltage at a point between the resistors		Not applicable
Current through each resistor	$I_1 =$ $I_2 =$	$I_1 =$ $I_2 =$

ANSWERS TO EXERCISES

A: 6 Ω and 25 Ω.
B: (*b*).
C: (*a*) 60-W bulb; (*b*) 100-W bulb. [Can you explain why? In (*a*), note that the resistance R is higher for the 60-W bulb $(P = V^2/R)$ and recall $P = I^2R$.]
D: $41I_3 - 45 + 21I_2 - 80 = 0$.
E: 180 A; this high current through the batteries could cause them to become very hot (and dangerous—possibly exploding): the power dissipated in the weak battery would be $P = I^2r = (180\,\text{A})^2(0.10\,\Omega) = 3200\,\text{W}$!
F: (*a*).
G: $\approx 500\,\text{k}\Omega$.

Magnets produce magnetic fields, but so do electric currents. Compass needles are magnets, and they align along the direction of any magnetic field present. Here, the compasses show the presence (and direction) of a magnetic field created by a current-carrying wire. We shall see in this Chapter how magnetic field is defined, and how magnetic fields exert forces on electric currents and on charged particles. We also discuss useful applications of the interaction between magnetic fields and electric currents and moving electric charges, such as motors and loudspeakers.



CHAPTER 27

Magnetism

CONTENTS

27–1 Magnets and Magnetic Fields
27–2 Electric Currents Produce Magnetic Fields
27–3 Force on an Electric Current in a Magnetic Field; Definition of $\vec{\mathbf{B}}$
27–4 Force on an Electric Charge Moving in a Magnetic Field
27–5 Torque on a Current Loop; Magnetic Dipole Moment
27–6 Applications: Motors, Loudspeakers, Galvanometers
27–7 Discovery and Properties of the Electron
27–8 The Hall Effect
27–9 Mass Spectrometer

CHAPTER-OPENING QUESTION—Guess now!

Which of the following can experience a force when placed in the magnetic field of a magnet?

(a) An electric charge at rest.
(b) A moving electric charge.
(c) An electric current in a wire.
(d) Another magnet.

The history of magnetism began thousands of years ago when, in a region of Asia Minor known as Magnesia, rocks were found that could attract each other. These rocks were called "magnets" after their place of discovery.

Not until the nineteenth century, however, was it realized that magnetism and electricity are closely related. A crucial discovery was that electric currents produce magnetic effects (we will say "magnetic fields") like magnets do. All kinds of practical devices depend on magnetism, from compasses to motors, loudspeakers, earbuds, computer memory, electric generators, and even magnetic resonance imaging (discussed in Chapter 42).

27–1 Magnets and Magnetic Fields

You probably have observed a magnet attract paper clips, nails, and other objects made of iron, as in Fig. 27–1. Any magnet, whether it is in the shape of a bar or a horseshoe, has two ends or faces, called **poles**, which is where the magnetic effect is strongest. If a bar magnet is suspended from a fine thread, it is found that one pole of the magnet will always point toward the north. It is not known for sure when this fact was discovered, but it is known that the Chinese were making use of it as an aid to navigation by the eleventh century and perhaps earlier. This is the principle

of a compass. A compass needle is simply a bar magnet which is supported at its center of gravity so that it can rotate freely. The pole of a freely suspended magnet that points toward geographic north is called the **north pole** of the magnet. The other pole points toward the south and is called the **south pole**.

It is a familiar observation that when two magnets are brought near one another, each exerts a force on the other. The force can be either attractive or repulsive and can be felt even when the magnets don't touch. If the north pole of one bar magnet is brought near the north pole of a second magnet, the force is repulsive. Similarly, if the south poles are brought close, the force is repulsive. But when the north pole of one magnet is brought near the south pole of another magnet, the force is attractive. These results are shown in Fig. 27–2, and are reminiscent of the forces between electric charges: like poles repel, and unlike poles attract. But *do not confuse magnetic poles with electric charge*. They are very different. One important difference is that a positive or negative electric charge can easily be isolated. But an isolated single magnetic pole has never been observed. If a bar magnet is cut in half, you do not obtain isolated north and south poles. Instead, two new magnets are produced, Fig. 27–3, each with north (N) and south (S) poles. If the cutting operation is repeated, more magnets are produced, each with a north and a south pole. Physicists have searched for isolated single magnetic poles (monopoles), but no **magnetic monopole** has ever been observed.

Besides iron, a few other materials, such as cobalt, nickel, gadolinium, and some of their oxides and alloys, show strong magnetic effects. They are said to be **ferromagnetic** (from the Latin word *ferrum* for iron). Other materials show some slight magnetic effect, but it is very weak and can be detected only with delicate instruments. We will look in more detail at ferromagnetism in Section 28–8.

In Chapter 21, we used the concept of an electric field surrounding an electric charge. In a similar way, we can picture a **magnetic field** surrounding a magnet. The force one magnet exerts on another magnet can then be described as the interaction between one magnet and the magnetic field of the other. Just as we drew electric field lines, we can also draw **magnetic field lines**. They can be drawn, as for electric field lines, so that

1. the direction of the magnetic field is tangent to a field line at any point, and
2. the number of lines per unit area is proportional to the strength of the magnetic field.

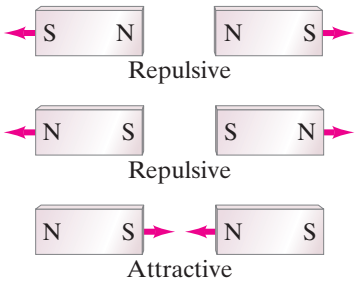
The *direction* of the magnetic field at a given location can be *defined* as the direction that the north pole of a compass needle would point if placed at that location. (We will give a more precise definition of magnetic field shortly.) Figure 27–4a shows how thin iron filings (acting like tiny magnets) reveal the magnetic field lines by lining up like the compass needles. The magnetic field determined in this way for the field surrounding a bar magnet is shown in Fig. 27–4b. Notice that because of our definition, the lines always point out from the north pole and in toward the south pole of a magnet (the north pole of a magnetic compass needle is attracted to the south pole of the magnet).

Magnetic field lines continue inside a magnet, as indicated in Fig. 27–4b. Indeed, given the lack of single magnetic poles, magnetic field lines always form closed loops, unlike electric field lines that begin on positive charges and end on negative charges.



FIGURE 27–1 A horseshoe magnet attracts pins made of iron.

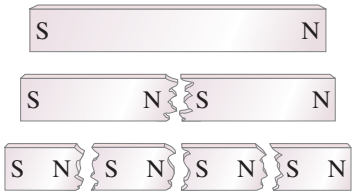
FIGURE 27–2 Like poles of two magnets repel; unlike poles attract.



CAUTION

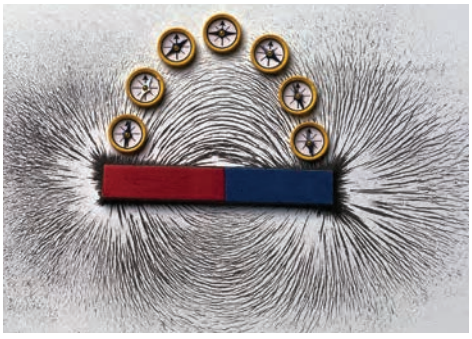
Magnets do not attract all metals

FIGURE 27–3 If you split a magnet, you won't get isolated north and south poles; instead, two new magnets are produced, each with a north and a south pole.

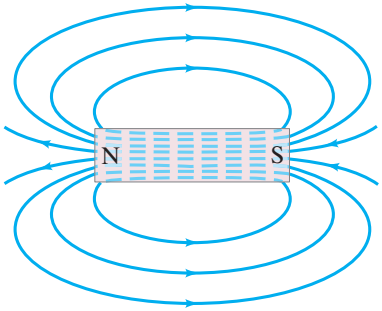


CAUTION

Magnetic field lines form closed loops, unlike electric field lines



(a)



(b)

FIGURE 27–4 (a) Visualizing magnetic field lines around a bar magnet, using iron filings and compass needles. The red end of the bar magnet is its north pole. The N pole of a nearby compass needle points away from the north pole of the magnet. (b) Diagram of magnetic field lines for a bar magnet.

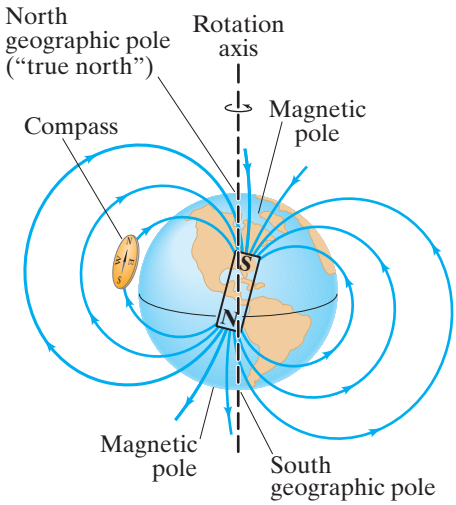


FIGURE 27–5 The Earth acts like a huge magnet. But its magnetic poles are not at the geographic poles (on the Earth's rotation axis).

Earth's Magnetic Field

The Earth's magnetic field is shown in Fig. 27–5, and is thought to be produced by electric currents in the Earth's molten iron outer core. The pattern of field lines is almost as though there were an imaginary bar magnet inside the Earth. Since the north pole (N) of a compass needle points north, the Earth's **magnetic pole** which is in the geographic north is magnetically a south pole, as indicated in Fig. 27–5 by the S on the schematic bar magnet inside the Earth. Remember that the north pole of one magnet is attracted to the south pole of another magnet. Nonetheless, Earth's pole in the north is still often called the "north magnetic pole," or "geomagnetic north," simply because it is in the north. Similarly, the Earth's southern magnetic pole, which is near the geographic south pole, is magnetically a north pole (N). The Earth's magnetic poles do not coincide with the **geographic poles**, which are on the Earth's axis of rotation. The north magnetic pole, for example, is in the Canadian Arctic, now on the order of 1000 km[†] from the geographic north pole, or **true north**. This difference must be taken into account for accurate use of a compass (Fig. 27–6). The angular difference between the direction of a compass needle (which points along the magnetic field lines) at any location and true (geographical) north is called the **magnetic declination**. In the U.S. it varies from 0° to about 20°, depending on location.

Notice in Fig. 27–5 that the Earth's magnetic field at most locations is not tangent to the Earth's surface. The angle that the Earth's magnetic field makes with the horizontal at any point is referred to as the **angle of dip**, or the "inclination." It is 67° at New York, for example, and 55° at Miami.

EXERCISE A Does the Earth's magnetic field have a greater magnitude near the poles or near the equator? [How can you tell using the field lines in Fig. 27–5?]

PHYSICS APPLIED
Use of a compass

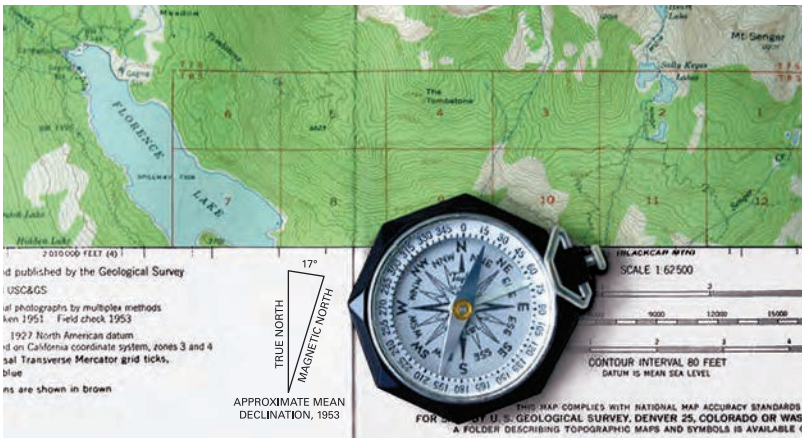
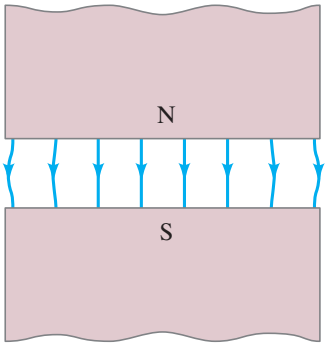


FIGURE 27–6 Using a map and compass in the wilderness. First you align the compass case so the needle points away from true north (N) exactly the number of degrees of declination stated on the map (for this topographic map, it is 17° as shown just to the left of the compass). Then align the map with true north, as shown, *not* with the compass needle. [This is an old map (1953) of a part of California; on newer maps of this area the declination is only 13°, telling us the position of magnetic north has moved—see footnote below.]

Uniform Magnetic Field

The simplest magnetic field is one that is uniform—it doesn't change in magnitude or direction from one point to another. A perfectly uniform field over a large area is not easy to produce. But the field between two flat parallel pole pieces of a magnet is nearly uniform if the area of the pole faces is large compared to their separation, as shown in Fig. 27–7. At the edges, the field "fringes" out somewhat: the magnetic field lines are no longer quite parallel and uniform. The parallel evenly spaced field lines in the central region of the gap indicate that the field is uniform at points not too near the edges, much like the electric field between two parallel plates (Fig. 23–16).

FIGURE 27–7 Magnetic field between two wide poles of a magnet is nearly uniform, except near the edges.



[†]Earth's north magnetic pole has been moving over time, on the order of 10 km per year in recent decades. Magnetism in rocks solidified at various times in the past (age determined by radioactive dating—see Section 41–10) suggests that Earth's magnetic poles have not only moved significantly over geologic time, but have also reversed direction 400 times over the last 330 million years. Also note that a compass gives a false reading if you are standing on rock containing magnetized iron ore (as you move around, the compass needle is inconsistent).



(a) (b) (c) (d)

FIGURE 27–8 (a) Deflection of compass needles near a current-carrying wire, showing the presence and direction of the magnetic field. (b) Iron filings also align along the direction of the magnetic field lines near a straight current-carrying wire. (c) Diagram of the magnetic field lines around an electric current in a straight wire. (d) Right-hand-rule-1 for remembering the direction of the magnetic field: when the thumb points in the direction of the conventional current, the fingers wrapped around the wire point in the direction of the magnetic field. ($\vec{\mathbf{B}}$ is the symbol for magnetic field.)

27–2 Electric Currents Produce Magnetic Fields

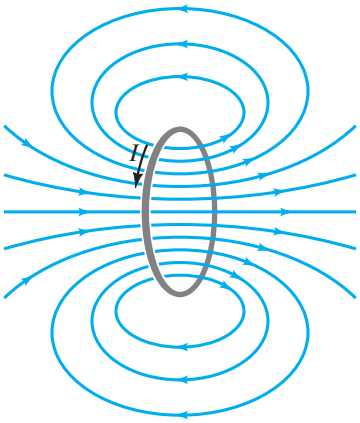
During the eighteenth century, many scientists sought to find a connection between electricity and magnetism. A stationary electric charge and a magnet were shown to have no influence on each other. But in 1820, Hans Christian Oersted (1777–1851) found that when a compass is placed near a wire, the compass needle deflects if (and only if) the wire carries an electric current. As we have seen, a compass needle is deflected by a magnetic field. So Oersted's experiment showed that **an electric current produces a magnetic field**. He had found a connection between electricity and magnetism.

A compass needle placed near a straight section of current-carrying wire experiences a force, causing the needle to align tangent to a circle around the wire, Fig. 27–8a. Thus, the magnetic field lines produced by a current in a straight wire are in the form of circles with the wire at their center, Figs. 27–8b and c. The direction of these lines is indicated by the north pole of the compasses in Fig. 27–8a. There is a simple way to remember the direction of the magnetic field lines in this case. It is called a **right-hand rule**: grasp the wire with your right hand so that your thumb points in the direction of the conventional (positive) current; then your fingers will encircle the wire in the direction of the magnetic field, Fig. 27–8d.

Right-Hand-Rule-1:
Magnetic field direction produced by electric current

The magnetic field lines due to a circular loop of current-carrying wire can be determined in a similar way by placing a compass at various locations near the loop. The result is shown in Fig. 27–9. Again the right-hand rule can be used, as shown in Fig. 27–10. Unlike the uniform field shown in Fig. 27–7, the magnetic fields shown in Figs. 27–8 and 27–9 are *not* uniform—the fields are different in magnitude and direction at different locations.

FIGURE 27–9 Magnetic field lines due to a circular loop of wire.



EXERCISE B A straight wire carries a current directly toward you. In what direction are the magnetic field lines surrounding the wire?

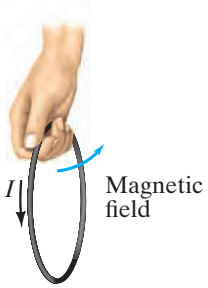


FIGURE 27–10 Right-hand rule for determining the direction of the magnetic field relative to the current in a loop of wire.

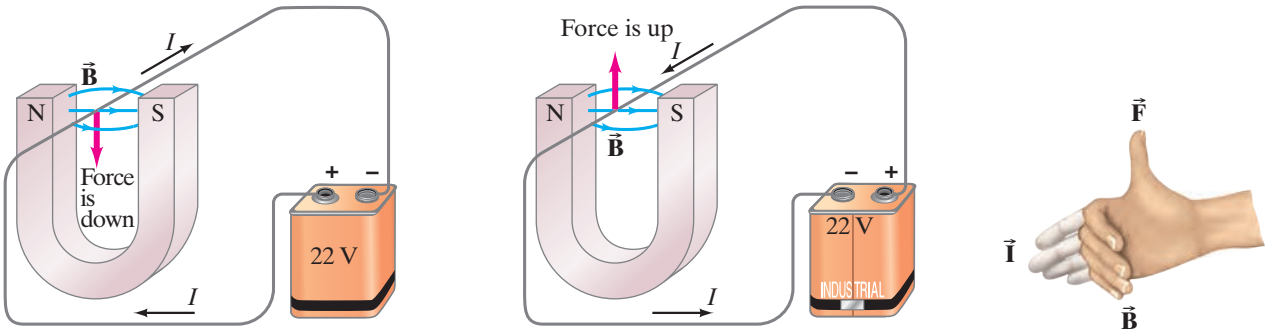


FIGURE 27–11 (a) Force on a current-carrying wire placed in a magnetic field $\vec{\mathbf{B}}$; (b) same, but current reversed; (c) right-hand rule for setup in (b), with current $\vec{\mathbf{I}}$ shown as if a vector with direction. We call this right-hand-rule-2. (RHR-1 is shown in Fig. 27–8d.)

27–3 Force on an Electric Current in a Magnetic Field; Definition of $\vec{\mathbf{B}}$

In Section 27–2 we saw that an electric current exerts a force on a magnet, such as a compass needle. By Newton's third law, we might expect the reverse to be true as well: we should expect that *a magnet exerts a force on a current-carrying wire*. Experiments indeed confirm this effect, and it too was first observed by Oersted.

Suppose a straight wire is placed in the magnetic field between the poles of a horseshoe magnet as shown in Fig. 27–11, where the vector symbol $\vec{\mathbf{B}}$ represents the magnitude and direction of the magnetic field. When a current flows in the wire, experiment shows that a force is exerted on the wire. But this force is *not* toward one or the other pole of the magnet. Instead, the force is directed at right angles to the magnetic field direction, downward in Fig. 27–11a. If the current is reversed in direction, the force is in the opposite direction, upward as shown in Fig. 27–11b. Experiments show that *the direction of the force is always perpendicular to the direction of the current and also perpendicular to the direction of the magnetic field*, $\vec{\mathbf{B}}$.

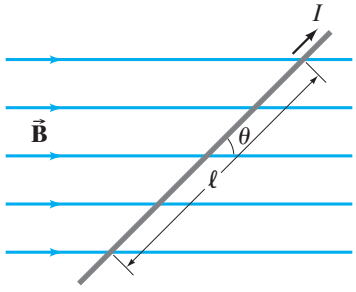
Right-Hand-Rule-2:
Force on current exerted by $\vec{\mathbf{B}}$

The direction of the force is given by another **right-hand rule**, as illustrated in Fig. 27–11c. Orient your right hand until your outstretched fingers can point in the direction of the conventional current I, and when you bend your fingers they point in the direction of the magnetic field lines, $\vec{\mathbf{B}}$. Then your outstretched thumb will point in the direction of the force $\vec{\mathbf{F}}$ on the wire. We call this right-hand-rule-2 (RHR-2) to distinguish it from that shown in Fig. 27–8d (RHR-1).

This right-hand rule describes the direction of the force. What about the *magnitude* of the force on the wire? It is found experimentally that the magnitude of the force is directly proportional to the current I in the wire, to the magnetic field B (assumed uniform), and to the length ℓ of wire exposed to the magnetic field. The force also depends on the angle θ between the current direction and the magnetic field (Fig. 27–12), being proportional to $\sin\theta$. Thus, the force on a wire carrying a current I with length ℓ in a uniform magnetic field B is given by

$$F \propto I\ell B \sin\theta.$$

FIGURE 27–12 Current-carrying wire in a magnetic field. Force on the wire is directed into the page.



When the current is perpendicular to the field lines ($\theta = 90°$ and $\sin 90° = 1$), the force is strongest. When the wire is parallel to the magnetic field lines ($\theta = 0°$), there is no force at all.

Up to now we have not defined the magnetic field strength precisely. In fact, the magnetic field B can be conveniently defined in terms of the above proportion so that the proportionality constant is precisely 1. Thus we have

$$F = I\ell B \sin\theta. \qquad \textbf{(27–1)}$$

If the current's direction is perpendicular to the field $\vec{\mathbf{B}}$ ($\theta = 90°$), then the force is

$$F_{\text{max}} = I\ell B. \qquad [\text{current} \perp \vec{\mathbf{B}}] \qquad \textbf{(27–2)}$$

(If B is not uniform, then B in Eqs. 27–1 and 27–2 could be an average of the field over the length ℓ of the wire. Or calculus can be used, as we will see shortly.)

The magnitude of $\vec{\mathbf{B}}$ can be defined using Eq. 27–2 as $B = F_{\text{max}}/I\ell$, where F_{max} is the magnitude of the force on a straight length ℓ of wire carrying a current I when the wire is perpendicular to $\vec{\mathbf{B}}$.

The relation between the force $\vec{\mathbf{F}}$ on a wire carrying current I, and the magnetic field $\vec{\mathbf{B}}$ that causes the force, can be written as a vector equation. To do so, we recall that the direction of $\vec{\mathbf{F}}$ is given by right-hand-rule-2 (Fig. 27–11c), and the magnitude by Eq. 27–1. This is consistent with the definition of the vector cross product (see Section 11–2), so we can write

$$\vec{\mathbf{F}} = I\vec{\boldsymbol{\ell}} \times \vec{\mathbf{B}}; \tag{27–3}$$

here, $\vec{\boldsymbol{\ell}}$ is a vector whose magnitude is the length of the wire and its direction is along the wire (assumed straight) in the direction of the conventional (positive) current.

Equation 27–3 applies if the magnetic field is uniform and the wire is straight. If $\vec{\mathbf{B}}$ is not uniform, or if the wire does not everywhere make the same angle θ with $\vec{\mathbf{B}}$, then Eq. 27–3 can be written more generally as

$$d\vec{\mathbf{F}} = I\, d\vec{\boldsymbol{\ell}} \times \vec{\mathbf{B}}, \tag{27–4}$$

where $d\vec{\mathbf{F}}$ is the infinitesimal force acting on a differential length $d\vec{\boldsymbol{\ell}}$ of the wire. The total force on the wire is then found by integrating.

Equation 27–4 can serve (just as well as Eq. 27–2 or 27–3) as a practical *definition* of $\vec{\mathbf{B}}$. An equivalent way to define $\vec{\mathbf{B}}$, in terms of the force on a moving electric charge, is discussed in the next Section.

Definition of magnetic field

EXERCISE C A wire carrying current I is perpendicular to a uniform magnetic field of strength B. Assuming a fixed length of wire, which of the following changes will reduce the force on the wire by a factor of 2? (*a*) Decrease the angle from 90° to 45°; (*b*) decrease the angle from 90° to 30°; (*c*) decrease the current in the wire to $I/2$; (*d*) decrease the magnetic field to $B/2$; (*e*) none of these will do it.

The SI unit for magnetic field B is the **tesla** (T). From Eq. 27–1, 2, 3, or 4, we see that $1\,\text{T} = 1\,\text{N/A}\cdot\text{m}$. An older name for the tesla is the "weber per meter squared" $(1\,\text{Wb/m}^2 = 1\,\text{T})$.[†] Another unit sometimes used to specify magnetic field is a cgs unit, the **gauss** (G): $1\,\text{G} = 10^{-4}\,\text{T}$. A field given in gauss should always be changed to teslas before using with other SI units. To get a "feel" for these units, we note that the magnetic field of the Earth at Earth's surface is about $\frac{1}{2}$ G or $0.5 \times 10^{-4}\,\text{T}$. On the other hand, typical strong electromagnets can produce fields on the order of 2 T, and some specialized magnets can produce fields over 30 T.

EXAMPLE 27–1 **Magnetic force on a current-carrying wire.** A wire carrying a steady (dc) 30-A current has a length $\ell = 12\,\text{cm}$ between the pole faces of a magnet. The wire is at an angle $\theta = 60°$ to the field (Fig. 27–13). The magnetic field is approximately uniform at 0.90 T. We ignore the field beyond the pole pieces. Determine the magnitude and direction of the force on the wire.

APPROACH We use Eq. 27–1, $F = I\ell B \sin\theta$.

SOLUTION The force F on the 12-cm length of wire within the uniform field B is

$$F = I\ell B \sin\theta = (30\,\text{A})(0.12\,\text{m})(0.90\,\text{T})(\sin 60°) = 2.8\,\text{N}.$$

We use right-hand-rule-2 to find the direction of $\vec{\mathbf{F}}$. Hold your right hand flat, pointing your fingers in the direction of the current. Then bend your fingers (maybe needing to rotate your hand) so they point along $\vec{\mathbf{B}}$, Fig. 27–13. Your thumb then points into the page, which is thus the direction of the force F.

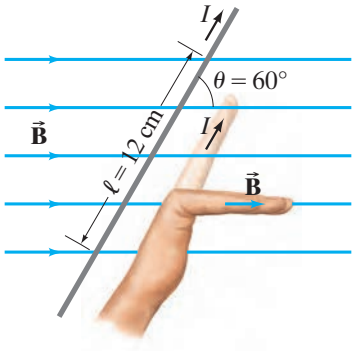


FIGURE 27–13 Example 27–1. For right-hand-rule-2, the thumb points into the page. See Fig. 27–11c.

EXERCISE D A straight power line carries 30 A and is perpendicular to the Earth's magnetic field of $0.50 \times 10^{-4}\,\text{T}$. What magnitude force is exerted on 100 m of this power line?

[†]The "weber" is discussed in Section 29–2. For now, no worries.

On a diagram, when we want to represent an electric current or a magnetic field that is pointing *out* of the page (toward us) or *into* the page, we use $\odot$ or $\times$, respectively. The $\odot$ is meant to resemble the tip of an arrow pointing directly toward the reader, whereas the $\times$ or $\otimes$ resembles the tail of an arrow pointing away. See Fig. 27–14.

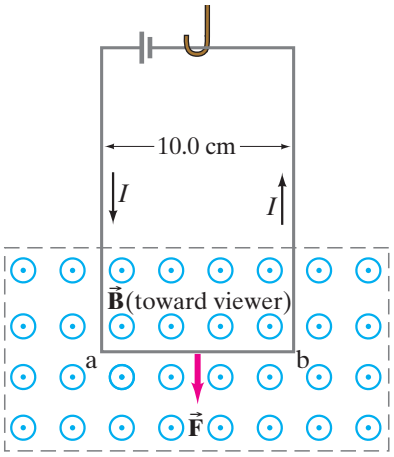


FIGURE 27–14 Measuring a magnetic field $\vec{\mathbf{B}}$. Example 27–2.

EXAMPLE 27–2 **Measuring a magnetic field.** A rectangular loop of wire hangs vertically as shown in Fig. 27–14. A magnetic field $\vec{\mathbf{B}}$ is directed horizontally, perpendicular to the plane of the loop, and points out of the page as represented by the symbol $\odot$. The magnetic field $\vec{\mathbf{B}}$ is very nearly uniform along the horizontal portion of wire ab (length $\ell = 10.0\,\text{cm}$) which is near the center of the gap of a large magnet producing the field. The top portion of the wire loop is out of the field. The loop hangs from a balance (reads 0 when $B = 0$) which measures a downward magnetic force of $F = 3.48 \times 10^{-2}\,\text{N}$ when the wire carries a current $I = 0.245\,\text{A}$. What is the magnitude of the magnetic field B?

APPROACH Three straight sections of the wire loop are in the magnetic field: a horizontal section and two vertical sections. We apply Eq. 27–1 to each section and use the right-hand rule.

SOLUTION Using right-hand-rule-2 (page 786), we see that the magnetic force on the left vertical section of wire points to the left, and the force on the vertical section on the right points to the right. These two forces are equal and in opposite directions and so add up to zero. Hence, the net magnetic force on the loop is that on the horizontal section ab, whose length is $\ell = 0.100\,\text{m}$. The angle θ between $\vec{\mathbf{B}}$ and the wire is $\theta = 90°$, so $\sin\theta = 1$. Thus Eq. 27–1 gives

$$B = \frac{F}{I\ell} = \frac{3.48 \times 10^{-2}\,\text{N}}{(0.245\,\text{A})\,(0.100\,\text{m})} = 1.42\,\text{T}.$$

NOTE This technique can be a precise means of determining magnetic field strength.

EXAMPLE 27–3 **Magnetic force on a semicircular wire.** A rigid wire, carrying a current I, consists of a semicircle of radius R and two straight portions as shown in Fig. 27–15. The wire lies in a plane perpendicular to a uniform magnetic field $\vec{\mathbf{B}}_0$. Note choice of x and y axis. The straight portions each have length ℓ within the field. Determine the net force on the wire due to the magnetic field $\vec{\mathbf{B}}_0$.

APPROACH The forces on the two straight sections are equal $(= I\ell B_0)$ and in opposite directions, so they cancel. Hence the net force is that on the semicircular portion.

SOLUTION We divide the semicircle into short lengths $d\ell = R\,d\phi$ as indicated in Fig. 27–15, and use Eq. 27–4, $d\vec{\mathbf{F}} = I\,d\vec{\boldsymbol{\ell}} \times \vec{\mathbf{B}}$, to find

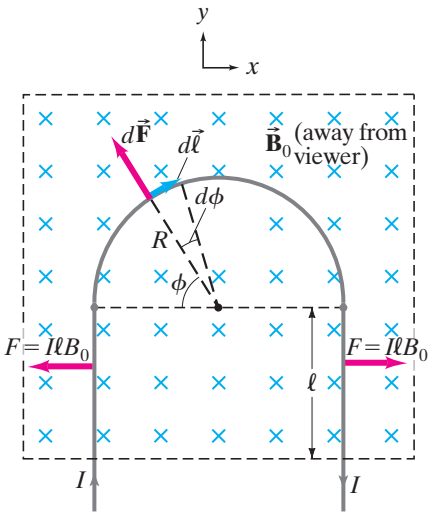
$$dF = IB_0 R\,d\phi,$$

where dF is the force on the length $d\ell = R\,d\phi$, and the angle between $d\vec{\boldsymbol{\ell}}$ and $\vec{\mathbf{B}}_0$ all along the path is 90° (so $\sin\theta = 1$ in the cross product). The x component of the force $d\vec{\mathbf{F}}$ on the segment $d\vec{\boldsymbol{\ell}}$ shown, and the x component of $d\vec{\mathbf{F}}$ for a symmetrically located $d\vec{\boldsymbol{\ell}}$ on the other side of the semicircle, will cancel each other. Thus for the entire semicircle there will be no x component of force. Hence we need be concerned only with the y components, each equal to $dF\sin\phi$, and the total force will have magnitude

$$F = \int_0^\pi dF\sin\phi = IB_0 R\int_0^\pi \sin\phi\,d\phi = -IB_0 R\cos\phi\Big|_0^\pi = 2IB_0 R,$$

with direction vertically upward along the y axis in Fig. 27–15.

FIGURE 27–15 Example 27–3. A magnetic field, $\vec{\mathbf{B}}_0$, is uniform within the dashed area, and points *into* the page (symbols $\times$).



27–4 Force on an Electric Charge Moving in a Magnetic Field

We have seen that a current-carrying wire experiences a force when placed in a magnetic field. Since a current in a wire consists of moving electric charges, we might expect that freely moving charged particles (not in a wire) would also experience a force when passing through a magnetic field. Free electric charges may not be

as easy to produce in the lab as a current in a wire, but it can be done, and experiments do show that moving electric charges experience a force in a magnetic field.

From what we have already discussed, we can predict the force on a single electric charge moving in a magnetic field $\vec{\mathbf{B}}$. If N such particles of charge q pass by a given point in time t, they constitute a current $I = Nq/t$. We let t be the time for a charge q to travel a distance ℓ in a magnetic field $\vec{\mathbf{B}}$; then $\vec{\boldsymbol{\ell}} = \vec{\mathbf{v}}t$ where $\vec{\mathbf{v}}$ is the velocity of the particle. Thus, the force on these N particles is, by Eq. 27–3, $\vec{\mathbf{F}} = I\vec{\boldsymbol{\ell}} \times \vec{\mathbf{B}} = (Nq/t)(\vec{\mathbf{v}}t) \times \vec{\mathbf{B}} = Nq\vec{\mathbf{v}} \times \vec{\mathbf{B}}$. The force on *one* of the N particles is then

$$\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}. \qquad \textbf{(27–5a)}$$

Alternate definition of magnetic field

This basic and important result can be considered as an alternative way of *defining* the magnetic field $\vec{\mathbf{B}}$, in place of Eq. 27–4 or 27–3. The magnitude of the force in Eq. 27–5a is

$$F = qvB \sin\theta. \qquad \textbf{(27–5b)}$$

This gives the magnitude of the force on a particle of charge q moving with velocity $\vec{\mathbf{v}}$ at a point where the magnetic field has magnitude B. The angle between $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$ is θ. The force is greatest when the particle moves perpendicular to $\vec{\mathbf{B}}$ ($\theta = 90°$):

$$F_{\text{max}} = qvB. \qquad [\vec{\mathbf{v}} \perp \vec{\mathbf{B}}]$$

Right-Hand-Rule-3:
Force on moving charge exerted by $\vec{\mathbf{B}}$

The force is *zero* if the particle moves *parallel* to the field lines ($\theta = 0°$ or $180°$). The *direction* of the force is perpendicular to the magnetic field $\vec{\mathbf{B}}$ and to the velocity $\vec{\mathbf{v}}$ of the particle. It is given again by a **right-hand rule** (as for any cross product), which we call right-hand-rule-3: you orient your right hand so that your outstretched fingers point along the direction of the particle's velocity ($\vec{\mathbf{v}}$), and when you bend your fingers they must point along the direction of $\vec{\mathbf{B}}$. Then your thumb will point in the direction of the force. This is true only for *positively* charged particles, and will be "up" for the positive particle shown in Fig. 27–16. For negatively charged particles, the force is in exactly the opposite direction, "down" in Fig. 27–16. You could also use an equivalent "left-hand" rule for negative charges and get the correct result.

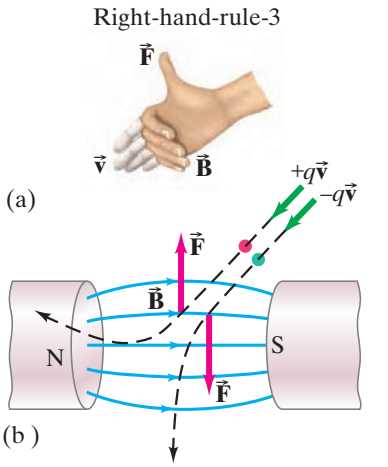


FIGURE 27–16 (a) Right-hand-rule-3 is shown for the force on a positive charge, $+q$. (b) Force on charged particles due to a magnetic field is perpendicular to the magnetic field direction. If $\vec{\mathbf{v}}$ is horizontal, then $\vec{\mathbf{F}}$ is vertical.

CONCEPTUAL EXAMPLE 27–4 **Negative charge near a magnet.** A negative charge $-Q$ is placed at rest near a magnet. Will the charge begin to move? Will it feel a force? What if the charge were positive, $+Q$?

RESPONSE No to all questions. A charge at rest has velocity equal to zero. Magnetic fields exert a force only on moving electric charges (Eqs. 27–5).

EXERCISE E Return to the Chapter-Opening Question, page 782, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 27–5 **Magnetic force on a proton.** A magnetic field exerts a force of 8.0×10^{-14} N toward the west on a proton moving vertically upward (on this page) at a speed of 5.0×10^{6} m/s (Fig. 27–17a). When moving horizontally in a northerly direction, the force on the proton is zero (Fig. 27–17b). Determine the magnitude and direction of the magnetic field in this region. (The charge on a proton is $q = +e = 1.6 \times 10^{-19}$ C.)

APPROACH Since the force on the proton is zero when moving north, the field must be in a north–south direction. To produce a force to the west when the proton moves up along this page, right-hand-rule-3 tells us that $\vec{\mathbf{B}}$ must point toward the north. (Your thumb points west and the outstretched fingers of your right hand point upward only when your bent fingers point north.) The magnitude of $\vec{\mathbf{B}}$ is found using Eq. 27–5b.

SOLUTION Equation 27–5b with $\theta = 90°$ gives

$$B = \frac{F}{qv} = \frac{8.0 \times 10^{-14}\,\text{N}}{(1.6 \times 10^{-19}\,\text{C})(5.0 \times 10^{6}\,\text{m/s})} = 0.10\,\text{T}.$$

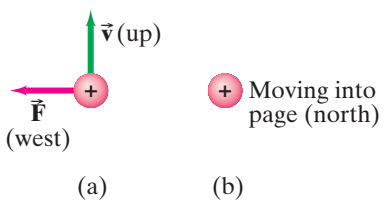


FIGURE 27–17 Example 27–5.

EXERCISE F What is the force on the proton of Example 27–5 if it heads horizontally south?

EXAMPLE 27–6 **ESTIMATE** **Magnetic force on ions during a nerve pulse.** Estimate the magnitude of the magnetic force due to the Earth's magnetic field on ions crossing a cell membrane during an action potential (Section 25–10). Assume the speed of the ions is 10^{-2} m/s.

APPROACH Using $F = qvB$, set the magnetic field of the Earth to be roughly $B \approx 10^{-4}$ T, and the charge $q \approx e \approx 10^{-19}$ C.

SOLUTION $F = qvB \approx (10^{-19}\,\text{C})(10^{-2}\,\text{m/s})(10^{-4}\,\text{T}) = 10^{-25}$ N.

NOTE This is an extremely small force. Yet it is thought that migrating animals do somehow detect the Earth's magnetic field, and this is an area of active research.

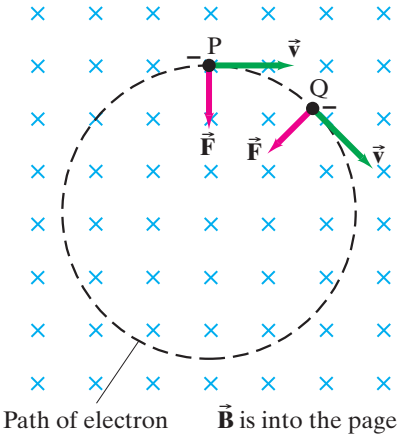


FIGURE 27–18 Force exerted by a uniform magnetic field on a moving charged particle (in this case, an electron) produces a circular path.

The path of a charged particle moving in a plane perpendicular to a uniform magnetic field is a circle as we shall now show. In Fig. 27–18 the magnetic field is directed *into* the paper, as represented by ×'s. An electron at point P is moving to the right, and the force on it at this point is toward the bottom of the page as shown (use right-hand-rule-3 and reverse the direction for negative charge). The electron is thus deflected toward the page bottom. A moment later, when it reaches point Q, the force is still perpendicular to the velocity and is in the direction shown. Because the force is always perpendicular to $\vec{\mathbf{v}}$, the magnitude of $\vec{\mathbf{v}}$ does not change—the electron moves at constant speed. We saw in Chapter 5 that if the force on a particle is always perpendicular to its velocity $\vec{\mathbf{v}}$, the particle moves in a circle and has a centripetal acceleration of magnitude $a = v^2/r$ (Eq. 5–1). Thus a charged particle moves in a circular path with a centripetal acceleration of constant magnitude in a uniform magnetic field (see Fig. 27–19). The electron moves clockwise in Fig. 27–18. A positive particle in this field would experience a force in the opposite direction and would thus move counterclockwise.

EXAMPLE 27–7 **Electron's path in a uniform magnetic field.** An electron travels at 1.5×10^7 m/s in a plane perpendicular to a uniform 0.010-T magnetic field. Describe its path quantitatively. Ignore gravity (= very small in comparison).

APPROACH The electron moves at speed v in a curved path and so must have a centripetal acceleration $a = v^2/r$ (Eq. 5–1). We find the radius of curvature using Newton's second law $F_{\text{net}} = ma$. The force is given by Eq. 27–5b with $\sin\theta = 1$: $F = qvB$.

SOLUTION We insert F and a into Newton's second law:

$$\Sigma F = ma$$

$$qvB = \frac{mv^2}{r}.$$

We solve for r and find

$$r = \frac{mv}{qB}.$$

Since $\vec{\mathbf{F}}$ is perpendicular to $\vec{\mathbf{v}}$, the magnitude of $\vec{\mathbf{v}}$ doesn't change. From this equation we see that if $\vec{\mathbf{B}}$ = constant, then r = constant, and the curve must be a circle as we claimed above. See Fig. 27–19. To get r we put in the numbers:

$$r = \frac{(9.1 \times 10^{-31}\,\text{kg})(1.5 \times 10^7\,\text{m/s})}{(1.6 \times 10^{-19}\,\text{C})(0.010\,\text{T})} = 0.85 \times 10^{-2}\,\text{m} = 8.5\,\text{mm}.$$

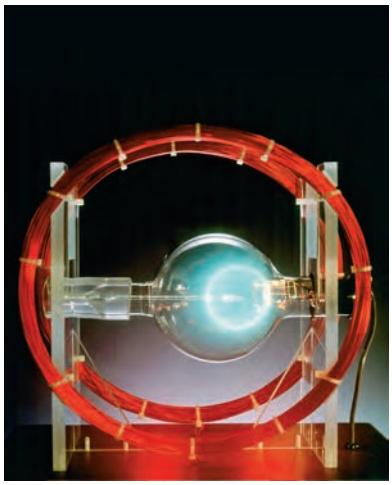


FIGURE 27–19 The white ring inside the glass tube is the glow of a beam of electrons that ionize the gas molecules inside the tube. The red coils of current-carrying wire produce a nearly uniform magnetic field, illustrating the circular path of charged particles in a uniform magnetic field.

EXERCISE G If the magnetic field B is larger, is the radius larger or smaller?

The time T required for a particle of charge q moving with constant speed v to make one circular revolution in a uniform magnetic field $\vec{\mathbf{B}}$ ($\perp \vec{\mathbf{v}}$) is $T = 2\pi r/v$, where $2\pi r$ is the circumference of its circular path (speed = distance $(2\pi r)$/time T). From Example 27–7, $r = mv/qB$, so

$$T = \frac{2\pi m}{qB}.$$

Since T is the period for one revolution, the frequency is

$$f = \frac{1}{T} = \frac{qB}{2\pi m}. \qquad \textbf{(27–6)}$$

This is often called the **cyclotron frequency** of a particle in a magnetic field because this is the frequency at which particles revolve in a cyclotron (see Problem 69).

CONCEPTUAL EXAMPLE 27–8 **Stopping charged particles.** An electric charge q moving in an electric field $\vec{\mathbf{E}}$ can be decelerated to a stop if the force $\vec{\mathbf{F}} = q\vec{\mathbf{E}}$ (Eq. 21–5) acts in the direction opposite to the charge's velocity. Can a magnetic field be used to stop a charged particle?

RESPONSE No, because the force is always *perpendicular* to the velocity of the particle and thus can only change the direction but not the magnitude of its velocity. Also the magnetic force cannot do work on the particle (force and displacement are perpendicular, Eq. 7–3) and so cannot change the kinetic energy of the particle, Eq. 7–11.

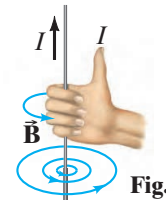
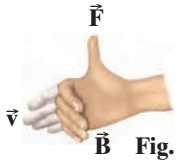
PROBLEM SOLVING

Magnetic Fields

Magnetic fields are somewhat analogous to the electric fields of Chapter 21, but there are several important differences to recall:

1. The force experienced by a charged particle moving in a magnetic field is *perpendicular* to the direction of the magnetic field (and to the direction of the velocity of the particle), whereas the force exerted by an electric field is *parallel* to the direction of the field (and independent of the velocity of the particle).
2. The *right-hand rule*, in its different forms, is intended to help you determine the direction of magnetic field, and the force a field exerts, and/or the directions of electric current or charged particle velocity. The right-hand rules (Table 27–1) are designed to deal with the "perpendicular" nature of these quantities.

TABLE 27–1 Summary of Right-hand Rules (= RHR)

Physical Situation	Example	How to Orient Right Hand	Result
1. Magnetic field produced by current (RHR-1)	 Fig. 27–8d	Wrap fingers around wire with thumb pointing in direction of current I	Fingers curl in direction of $\vec{\mathbf{B}}$
2. Force on electric current I due to magnetic field (RHR-2)	 Fig. 27–11c	Fingers first point straight along current I, then bend along magnetic field $\vec{\mathbf{B}}$	Thumb points in direction of the force $\vec{\mathbf{F}}$
3. Force on a positive electric charge due to magnetic field (RHR-3) (Use left hand for a negative charge)	 Fig. 27–16	Fingers first point along particle's velocity $\vec{\mathbf{v}}$, then along $\vec{\mathbf{B}}$	Thumb points in direction of the force $\vec{\mathbf{F}}$

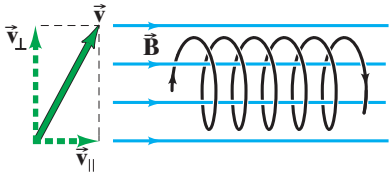
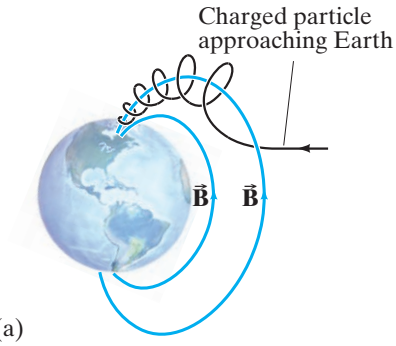


FIGURE 27–20 Example 27–9.

FIGURE 27–21 (a) Diagram showing a charged particle that approaches the Earth and is "captured" by the magnetic field of the Earth. Such particles follow the field lines toward the poles as shown. (b) Photo of aurora borealis.

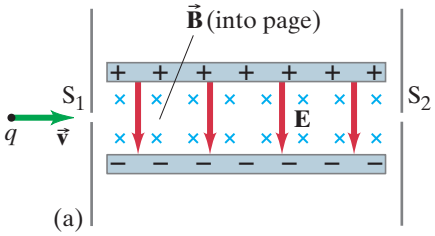


(a)

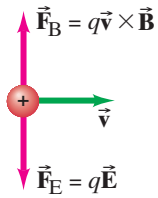


(b)

FIGURE 27–22 (below) A velocity selector: if $v = E/B$, the particles passing through S_1 make it through S_2.



(a)



(b)

CONCEPTUAL EXAMPLE 27–9 **A helical path.** What is the path of a charged particle in a uniform magnetic field if its velocity is *not* perpendicular to the magnetic field?

RESPONSE The velocity vector can be broken down into components parallel and perpendicular to the field. The velocity component parallel to the field lines experiences no force $(\theta = 0)$, so this component remains constant. The velocity component perpendicular to the field results in circular motion about the field lines. Putting these two motions together produces a helical (spiral) motion around the field lines as shown in Fig. 27–20.

EXERCISE H What is the sign of the charge in Fig. 27–20? How would you modify the drawing if the charge had the opposite sign?

*Aurora Borealis

Charged ions approach the Earth from the Sun (the "solar wind") and enter the atmosphere mainly near the poles, sometimes causing a phenomenon called the **aurora borealis** or "northern lights" in northern latitudes. To see why, consider Example 27–9 and Fig. 27–21 (see also Fig. 27–20). In Fig. 27–21a we imagine a stream of charged particles approaching the Earth. The velocity component *perpendicular* to the field for each particle becomes a circular orbit around the field lines, whereas the velocity component *parallel* to the field carries the particle along the field lines toward the poles. As a particle approaches the Earth's North Pole, the magnetic field is stronger and the radius of the helical path becomes smaller (see Example 27–7, $r \propto 1/B$).

A high concentration of high-speed charged particles ionizes the air, and as the electrons recombine with atoms, light is emitted (Chapter 37) which is the aurora. Auroras are especially spectacular during periods of high sunspot activity when more charged particles are emitted and more come toward Earth.

Lorentz Equation

If a particle of charge q moves with velocity $\vec{\mathbf{v}}$ in the presence of both a magnetic field $\vec{\mathbf{B}}$ and an electric field $\vec{\mathbf{E}}$, it will feel a force

$$\vec{\mathbf{F}} = q(\vec{\mathbf{E}} + \vec{\mathbf{v}} \times \vec{\mathbf{B}}) \qquad \textbf{(27–7)}$$

where we have made use of Eqs. 21–5 and 27–5a. Equation 27–7 is often called the **Lorentz equation** and is considered one of the basic equations in physics.

CONCEPTUAL EXAMPLE 27–10 **Velocity selector, or filter: Crossed $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields.** Some electronic devices and experiments need a beam of charged particles all moving at nearly the same velocity. This can be achieved using both a uniform electric field and a uniform magnetic field, arranged so they are at right angles to each other. As shown in Fig. 27–22a, particles of charge q pass through slit S_1 and enter the region where $\vec{\mathbf{B}}$ points into the page and $\vec{\mathbf{E}}$ points down from the positive plate toward the negative plate. If the particles enter with different velocities, show how this device "selects" a particular velocity, and determine what this velocity is.

RESPONSE After passing through slit S_1, each particle is subject to two forces as shown in Fig. 27–22b. If q is positive, the magnetic force is upwards and the electric force downwards. (Vice versa if q is negative.) The exit slit, S_2, is assumed to be directly in line with S_1 and the particles have velocity $\vec{\mathbf{v}}$. Depending on the magnitude of $\vec{\mathbf{v}}$, some particles will be bent upwards and some downwards. The only ones to make it through the slit S_2 will be those for which the net force is zero: $\Sigma F = qvB - qE = 0$. Hence this device selects particles whose velocity is

$$v = \frac{E}{B}. \qquad \textbf{(27–8)}$$

This result does not depend on the sign of the charge q.

EXERCISE I A particle in a velocity selector as diagrammed in Fig. 27–22 hits below the exit hole, S_2. This means that the particle (*a*) is going faster than the selected speed; (*b*) is going slower than the selected speed; (*c*) answer *a* is true if $q > 0$, *b* is true if $q < 0$; (*d*) answer *a* is true if $q < 0$, *b* is true if $q > 0$.

27–5 Torque on a Current Loop; Magnetic Dipole Moment

When an electric current flows in a closed loop of wire placed in an external magnetic field, as shown in Fig. 27–23, the magnetic force on the current can produce a torque. This is the principle behind a number of important practical devices, including motors and analog voltmeters and ammeters, which we discuss in the next Section. The interaction between a current and a magnetic field is important in other areas as well, including atomic physics.

Current flows through the rectangular loop in Fig. 27–23a, whose face we assume is parallel to $\vec{\mathbf{B}}$. $\vec{\mathbf{B}}$ exerts no force and no torque on the horizontal segments of wire because they are parallel to the field and $\sin\theta = 0$ in Eq. 27–1. But the magnetic field does exert a force on each of the vertical sections of wire as shown, $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ (see also top view, Fig. 27–23b). By right-hand-rule-2 (Fig. 27–11c or Table 27–1) the direction of the force on the upward current on the left is in the opposite direction from the equal magnitude force $\vec{\mathbf{F}}_2$ on the downward current on the right. These forces give rise to a net torque that acts to rotate the loop about its vertical axis.

Let us calculate the magnitude of this torque. From Eq. 27–2 $(\text{current} \perp \vec{\mathbf{B}})$, the force $F = IaB$, where *a* is the length of the vertical arm of the loop (Fig. 27–23a). The lever arm for each force is $b/2$, where *b* is the width of the loop and the "axis" is at the midpoint. The torques around this axis produced by $\vec{\mathbf{F}}_1$ and $\vec{\mathbf{F}}_2$ act in the same direction (Fig. 27–23b), so the total torque is the sum of the two torques:

$$\tau = IaB\frac{b}{2} + IaB\frac{b}{2} = IabB = IAB,$$

where $A = ab$ is the area of the loop. If we have a coil that consists of *N* loops of wire, the current is then *NI*, so the torque becomes

$$\tau = NIAB.$$

If this coil makes an angle θ with the magnetic field, as shown in Fig. 27–23c, the forces are unchanged, but each lever arm is reduced from $\frac{1}{2}b$ to $\frac{1}{2}b\sin\theta$. Note that the angle θ is taken to be the angle between $\vec{\mathbf{B}}$ and the perpendicular to the face of the loop or coil, Fig. 27–23c. So the torque becomes

$$\tau = NIAB\sin\theta. \qquad \textbf{(27–9)}$$

This formula, derived here for a rectangular coil, is valid for any shape of flat coil.

The quantity *NIA* is called the **magnetic dipole moment** of the coil and is considered a vector:

$$\vec{\boldsymbol{\mu}} = NI\vec{\mathbf{A}}. \qquad \textbf{(27–10)}$$

The *direction* of $\vec{\mathbf{A}}$ (and therefore of $\vec{\boldsymbol{\mu}}$) is defined as *perpendicular* to the plane of the coil (the greenish-blue arrow in Fig. 27–23c) and consistent with another right-hand rule (cup your right hand so your fingers wrap around the loop in the direction of current flow, then your thumb points in the direction of $\vec{\boldsymbol{\mu}}$ and $\vec{\mathbf{A}}$). With this definition of $\vec{\boldsymbol{\mu}}$, we can rewrite Eq. 27–9 in vector form:

$$\vec{\boldsymbol{\tau}} = NI\vec{\mathbf{A}} \times \vec{\mathbf{B}}$$

or

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\mathbf{B}}, \qquad \textbf{(27–11)}$$

which specifies both the magnitude and direction for the torque vector $\vec{\boldsymbol{\tau}}$.

Equation 27–11 has the same form as Eq. 21–9b for an electric dipole (with electric dipole moment $\vec{\mathbf{p}}$) in an electric field $\vec{\mathbf{E}}$, which is $\vec{\boldsymbol{\tau}} = \vec{\mathbf{p}} \times \vec{\mathbf{E}}$. And just as an electric dipole has potential energy given by $U = -\vec{\mathbf{p}} \cdot \vec{\mathbf{E}}$ when in an electric field, we expect a similar form for a magnetic dipole in a magnetic field. In order to rotate a current loop (Fig. 27–23) so as to increase θ, we must do work against the torque due to the magnetic field.

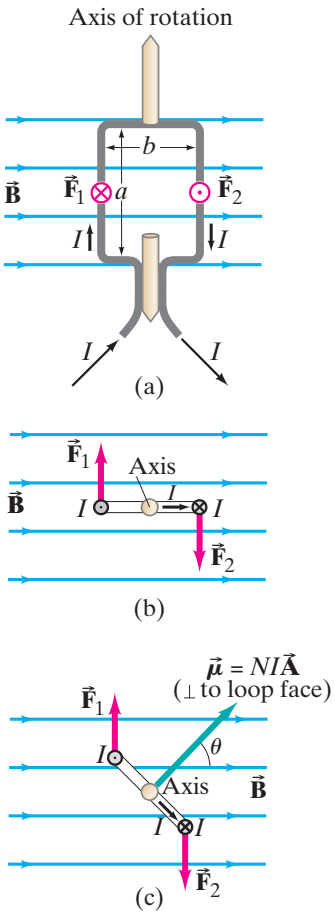


FIGURE 27–23 Calculating the torque on a current loop in a magnetic field $\vec{\mathbf{B}}$. (a) Loop face parallel to $\vec{\mathbf{B}}$ field lines; (b) top view; (c) loop makes an angle to $\vec{\mathbf{B}}$, reducing the torque since the lever arm is reduced. The vector $\vec{\boldsymbol{\mu}}$ is the "magnetic moment."

Hence the potential energy depends on angle (see Eq. 10–22, the work-energy principle for rotational motion) as

$$U = \int \tau \, d\theta = \int NIAB \sin\theta \, d\theta = -\mu B \cos\theta + C.$$

If we choose $U = 0$ at $\theta = \pi/2$, then the arbitrary constant C is zero and the potential energy is

$$U = -\mu B \cos\theta = -\vec{\boldsymbol{\mu}} \cdot \vec{\mathbf{B}}, \qquad \textbf{(27–12)}$$

as expected (compare Eq. 21–10). Bar magnets and compass needles, as well as current loops, can be considered as magnetic dipoles. Note the striking similarities of the fields produced by a bar magnet and a current loop, Figs. 27–4b and 27–9.

EXAMPLE 27–11 **Torque on a coil.** A circular coil of wire has a diameter of 20.0 cm and contains 10 loops. The current in each loop is 3.00 A, and the coil is placed in a 2.00-T external magnetic field. Determine the maximum and minimum torque exerted on the coil by the field.

APPROACH Equation 27–9 is valid for any shape of coil, including circular loops. Maximum and minimum torque are determined by the angle θ the coil makes with the magnetic field.

SOLUTION The area of one loop of the coil is

$$A = \pi r^2 = \pi (0.100\,\text{m})^2 = 3.14 \times 10^{-2}\,\text{m}^2.$$

The maximum torque occurs when the coil's face is parallel to the magnetic field, so $\theta = 90°$ in Fig. 27–23c, and $\sin\theta = 1$ in Eq. 27–9:

$$\tau = NIAB \sin\theta = (10)(3.00\,\text{A})(3.14 \times 10^{-2}\,\text{m}^2)(2.00\,\text{T})(1) = 1.88\,\text{N}\cdot\text{m}.$$

The minimum torque occurs if $\sin\theta = 0$, for which $\theta = 0°$, and then $\tau = 0$ from Eq. 27–9.

NOTE If the coil is free to turn, it will rotate toward the orientation with $\theta = 0°$.

EXAMPLE 27–12 **Magnetic moment of a hydrogen atom.** Determine the magnetic dipole moment of an electron orbiting the proton of a hydrogen atom at a given instant, assuming (in the simplistic "planetary" or Bohr model) the smallest circular orbit which has a radius 0.529×10^{-10} m. [This is a very rough picture of atomic structure, but nonetheless gives an accurate result.]

APPROACH We start by setting the electrostatic force on the electron due to the proton equal to $ma = mv^2/r$ since the electron's acceleration is centripetal.

SOLUTION The electron is held in its orbit by the coulomb force, so Newton's second law, $F = ma$, gives

$$\frac{e^2}{4\pi\epsilon_0 r^2} = \frac{mv^2}{r};$$

so

$$v = \sqrt{\frac{e^2}{4\pi\epsilon_0 mr}} = \sqrt{\frac{(1/4\pi\epsilon_0)e^2}{mr}}$$

$$= \sqrt{\frac{(8.99 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(1.60 \times 10^{-19}\,\text{C})^2}{(9.11 \times 10^{-31}\,\text{kg})(0.529 \times 10^{-10}\,\text{m})}} = 2.19 \times 10^6\,\text{m/s}.$$

Since current is the electric charge that passes a given point per unit time, the revolving electron is equivalent to a current

$$I = \frac{e}{T} = \frac{ev}{2\pi r},$$

where $T = 2\pi r/v$ is the time required for one orbit (speed $= 2\pi r/T$). Since the area of the orbit is $A = \pi r^2$, the magnetic dipole moment is

$$\mu = IA = \frac{ev}{2\pi r}(\pi r^2) = \tfrac{1}{2} evr$$

$$= \tfrac{1}{2}(1.60 \times 10^{-19}\,\text{C})(2.19 \times 10^6\,\text{m/s})(0.529 \times 10^{-10}\,\text{m}) = 9.27 \times 10^{-24}\,\text{A}\cdot\text{m}^2,$$

or $9.27 \times 10^{-24}\,\text{J/T}$.

27–6 Applications: Motors, Loudspeakers, Galvanometers

There are many practical applications of the forces related to magnetism. Among the most common are motors and loudspeakers. First we look at the galvanometer, which is the easiest to explain, and which you find on the instrument panels of automobiles and other devices whose readout is via a pointer or needle.

Galvanometer

The basic component of analog meters (those with pointer and dial), including analog ammeters, voltmeters, and ohmmeters, as well as gauges on car dashboards, is a galvanometer. We discussed in Section 26–7 how these meters are designed. Now we can examine how the crucial element works, the galvanometer. A **galvanometer** consists of a coil with N loops of wire (with attached pointer) suspended in the magnetic field of a permanent magnet, as shown in Fig. 27–24 (only one loop of wire is shown).When current flows through the loop of wire, the magnetic field B exerts a torque τ on the loop, as given by Eq. 27–9,

$$\tau = NIAB \sin\theta.$$

This torque is opposed by a spring which exerts a torque τ_s approximately proportional to the angle ϕ through which it is turned (Hooke's law). That is,

$$\tau_s = k\phi,$$

where k is the stiffness constant of the spring. The coil and attached pointer rotate to the angle where the torques balance. When the needle is in equilibrium at rest, the torques have equal magnitude: $k\phi = NIAB \sin\theta$, so

$$\phi = \frac{NIAB \sin\theta}{k}.$$

The deflection of the pointer, ϕ, is directly proportional to the current I flowing in the coil, but also depends on the angle θ the coil makes with $\vec{\mathbf{B}}$. For a useful meter we need ϕ to depend only on the current I, independent of θ. To solve this problem, magnets with curved pole pieces are used and the galvanometer coil is wrapped around a cylindrical iron core as shown in Fig. 27–25. The iron tends to concentrate the magnetic field lines so that $\vec{\mathbf{B}}$ always points parallel to the face of the coil at the wire outside the core. The force is then always perpendicular to the face of the coil, and the torque will not vary with angle. Thus ϕ will be proportional to I, as required for a useful meter.

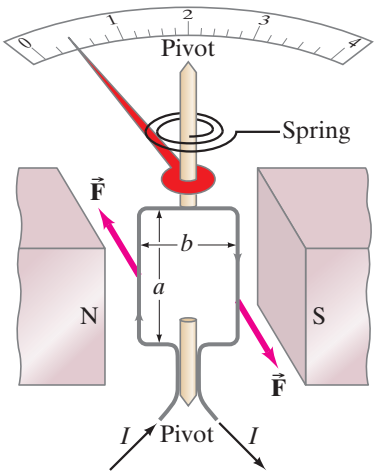
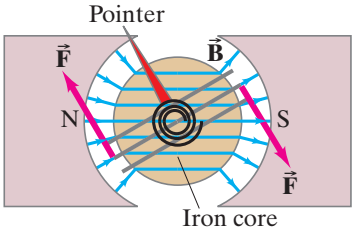


FIGURE 27–24 Galvanometer. Only one loop of wire is shown.

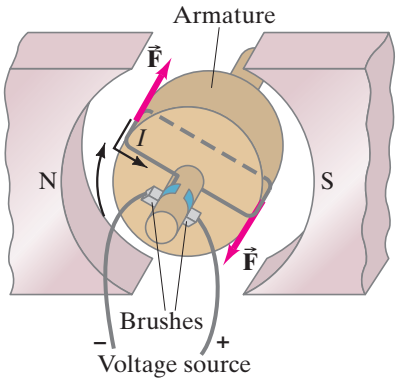
FIGURE 27–25 Galvanometer coil (3 loops shown) wrapped on an iron core.



Electric Motors

An **electric motor** converts electric energy into (rotational) mechanical energy. A motor works on the same principle as a galvanometer (a torque is exerted on a current-carrying coil of wire in a magnetic field) except that the coil must turn continuously in one direction for a useful motor. The coil is mounted on an iron cylinder called the **rotor** or **armature**, Fig. 27–26 (only one loop of wire is shown). The armature is mounted on a shaft or axle. When the armature is in the position shown in Fig. 27–26, the magnetic field exerts forces on the current in the loop as shown (perpendicular to $\vec{\mathbf{B}}$ and to the electric current direction). However, when the coil, which is rotating clockwise in Fig. 27–26, passes beyond the vertical position, the forces would then act to return it back toward the vertical if the current remained the same. But if the current could be reversed at that critical moment, the forces would reverse, and the coil would continue rotating in the same direction. Thus, alternation of the current is necessary if a motor is to turn continuously in one direction.

FIGURE 27–26 Diagram of a simple dc motor. (Magnetic field lines are as shown in Fig. 27–25.)



PHYSICS APPLIED
DC motor

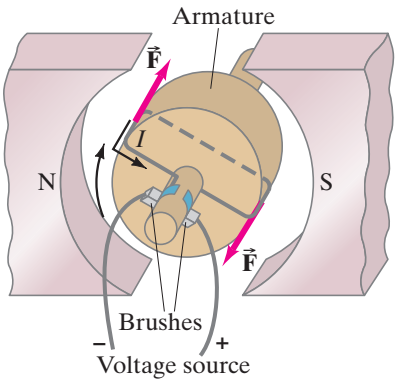


FIGURE 27–26 (Repeated.) Simple dc motor.

This can be achieved in a **dc motor** with the use of **commutators** and **brushes**. As shown in Fig. 27–27, input current passes through stationary brushes that rub against the conducting commutators mounted on the motor shaft. At every half revolution, each commutator changes its connection over to the other brush. Thus the current in the coil reverses every half revolution as required for continuous rotation.

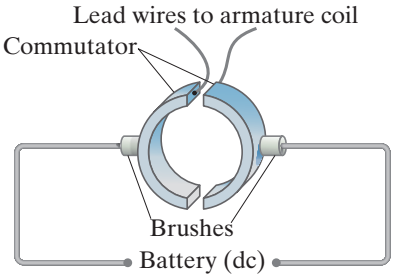
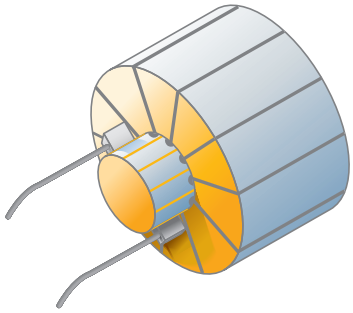


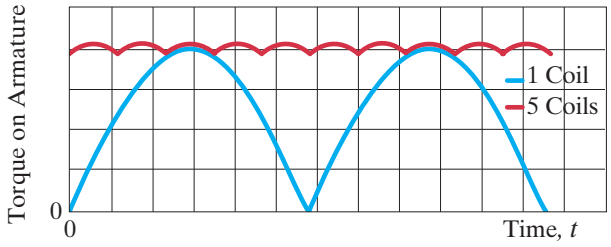
FIGURE 27–27 (right) Commutator-brush arrangement in a dc motor ensures alternation of the current in the armature to keep rotation continuous in one direction. The commutators are attached to the motor shaft (Fig. 27–26) and turn with it, whereas the brushes remain stationary.

Most motors contain several coils, called *windings*, each connected to a different portion of the armature, Fig. 27–28a. Current flows through each coil only during a small part of a revolution, at the time when its orientation results in the maximum torque. In this way, a motor produces a much steadier torque than can be obtained from a single coil (Fig. 27–28b).

FIGURE 27–28 (a) Motor with many windings. (b) Graph comparing the torque on a motor armature with one coil as in Fig. 27–26 (blue) vs. five coils (red).



(a)

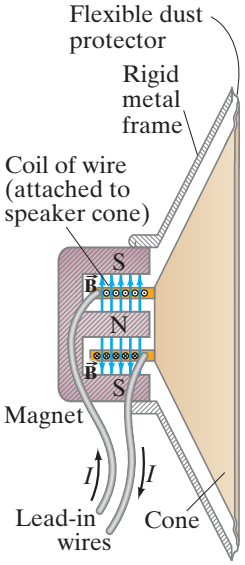


(b)

PHYSICS APPLIED
AC motor

An **ac motor**, with ac current as input, can work without commutators since the current itself alternates. Many motors use wire coils to produce the magnetic field (electromagnets) instead of a permanent magnet. Indeed the design of most motors is more complex than described here, but the general principles remain the same.

FIGURE 27–29 Loudspeaker.



Loudspeakers and Headsets

Loudspeakers, audio headsets and earbuds also work on the principle that a magnet exerts a force on a current-carrying wire. The electrical output of a stereo or TV set is connected to the wire leads of the speaker or earbuds. The speaker leads are connected internally to a coil of wire, which is itself attached to the speaker cone, Fig. 27–29. The speaker cone is usually made of stiffened cardboard and is mounted so that it can move back and forth freely. A permanent magnet is mounted directly in line with the coil of wire. When the alternating current of an audio signal flows through the wire coil, which is free to move within the magnet, the coil experiences a force due to the magnetic field of the magnet. (The force is to the left at the instant shown in Fig. 27–29, RHR-2, page 791) As the current alternates at the frequency of the audio signal, the coil and attached speaker cone move back and forth at the same frequency, causing alternate compressions and rarefactions of the adjacent air, and sound waves are produced that flow outward to be heard by the listener. A speaker thus changes electrical energy into sound energy, and the frequencies and intensities of the emitted sound waves can be an accurate reproduction of the electrical input.

27–7 Discovery and Properties of the Electron

The electron plays a basic role in our understanding of electricity and magnetism today. But its existence was not suggested until the 1890s. We discuss it here because magnetic fields were crucial for measuring its properties.

Toward the end of the nineteenth century, studies were being done on the discharge of electricity through rarefied gases. One apparatus, diagrammed in Fig. 27–30, was a glass tube fitted with electrodes and evacuated so only a small amount of gas remained inside. When a very high voltage was applied to the electrodes, a dark space seemed to extend outward from the cathode (negative electrode) toward the opposite end of the tube; and that far end of the tube would glow. If one or more screens containing a small hole was inserted as shown, the glow was restricted to a tiny spot on the end of the tube. It seemed as though something being emitted by the cathode traveled across to the opposite end of the tube. These "somethings" were named **cathode rays**.

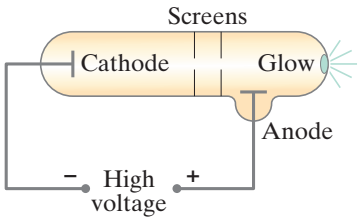


FIGURE 27–30 Discharge tube. In some models, one of the screens is the anode (positive plate).

There was much discussion at the time about what these rays might be. Some scientists thought they might resemble light. But the observation that the bright spot at the end of the tube could be deflected to one side by an electric or magnetic field suggested that cathode rays were charged particles; and the direction of the deflection was consistent with a negative charge. Furthermore, if the tube contained certain types of rarefied gas, the path of the cathode rays was made visible by a slight glow.

Estimates of the charge e of the cathode-ray particles, as well as of their charge-to-mass ratio e/m, had been made by 1897. But in that year, J. J. Thomson (1856–1940) was able to measure e/m directly, using the apparatus shown in Fig. 27–31. Cathode rays are accelerated by a high voltage and then pass between a pair of parallel plates built into the tube. Another voltage applied to the parallel plates produces an electric field $\vec{\mathbf{E}}$, and a pair of coils produces a magnetic field $\vec{\mathbf{B}}$. If $E = B = 0$, the cathode rays follow path b in Fig. 27–31. When only the electric field is present, say with the upper plate positive, the cathode rays are deflected upward as in path a in Fig. 27–31. If only a magnetic field exists, say inward, the rays are deflected downward along path c. These observations are just what is expected for a negatively charged particle. The force on the rays due to the magnetic field is $F = evB$, where e is the charge and v is the velocity of the cathode rays. In the absence of an electric field, the rays are bent into a curved path, so from Newton's second law, $F = ma$ with $a =$ centripetal acceleration, we have

$$evB = m\frac{v^2}{r},$$

and thus

$$\frac{e}{m} = \frac{v}{Br}.$$

The radius of curvature r can be measured and so can B. The velocity v can be found by applying an electric field in addition to the magnetic field. The electric field E is adjusted so that the cathode rays are undeflected and follow path b in Fig. 27–31.

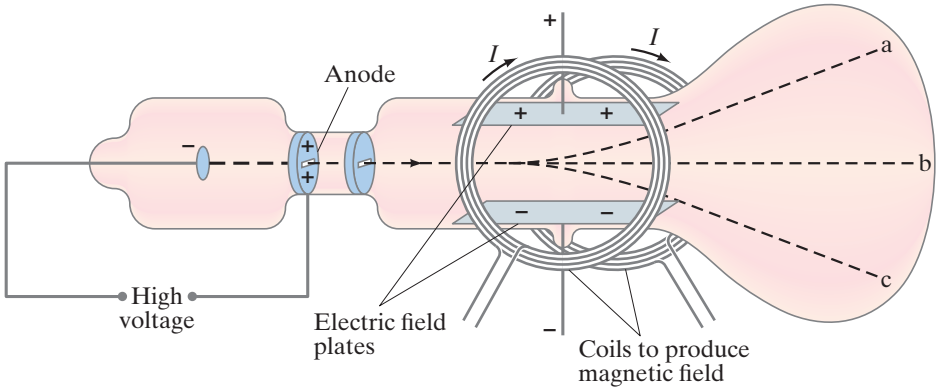


FIGURE 27–31 Cathode rays deflected by electric and magnetic fields.

This is just like the velocity selector of Example 27–10 where the force due to the electric field, $F = eE$, is balanced by the force due to the magnetic field, $F = evB$. Thus $eE = evB$ and $v = E/B$. Substituting $v = E/B$ into the equation at the bottom of the previous page, $e/m = v/Br$, we have

$$\frac{e}{m} = \frac{v}{Br} = \frac{E}{B^2 r}. \qquad \textbf{(27–13)}$$

The quantities on the right side can all be measured, and although e and m could not be determined separately, the ratio e/m could be determined. The accepted value today is $e/m = 1.76 \times 10^{11}$ C/kg. Cathode rays soon came to be called **electrons** ("carriers of electricity").

It is worth noting that the "discovery" of the electron, like many others in science, is not quite so obvious as discovering gold or oil. Should the discovery of the electron be credited to the person who first saw a glow in the tube? Or to the person who first called them cathode rays? Perhaps neither one, for they had no conception of the electron as we know it today. In fact, the credit for the discovery is generally given to Thomson, but not because he was the first to see the glow in the tube. Rather it is because he believed that this phenomenon was due to tiny negatively charged particles and made careful measurements on them. Furthermore he argued that these particles were constituents of atoms, and not ions or atoms themselves as many thought, and he developed an electron theory of matter. His view is close to what we accept today, and this is why Thomson is credited with the "discovery." Note, however, that neither he nor anyone else ever actually saw an electron itself. We discuss this briefly, for it illustrates the fact that discovery in science is not always a clear-cut matter. In fact some philosophers of science think the word "discovery" is often not appropriate, such as in this case.

Electron Charge Measurement

Thomson believed that an electron was not an atom, but rather a constituent, or part, of an atom. Convincing evidence for this came soon with the determination of the charge and the mass of the cathode rays. Thomson's student J. S. Townsend made the first direct (but rough) measurements of e in 1897. But it was the more refined **oil-drop experiment** of Robert A. Millikan (1868–1953) that yielded a precise value for the charge on the electron and showed that charge comes in discrete amounts. In this experiment, tiny droplets of mineral oil carrying an electric charge were allowed to fall under gravity between two parallel plates, Fig. 27–32. The electric field E between the plates was adjusted until the drop was suspended in midair. The downward pull of gravity, mg, was then just balanced by the upward force due to the electric field. Thus $qE = mg$, so the charge $q = mg/E$. The mass of the droplet was determined by measuring its terminal velocity in the absence of the electric field. Often the droplet was charged negatively, but sometimes it was positive, suggesting that the droplet had acquired or lost electrons (by friction, leaving the atomizer). Millikan's painstaking observations and analysis presented convincing evidence that any charge was an integral multiple

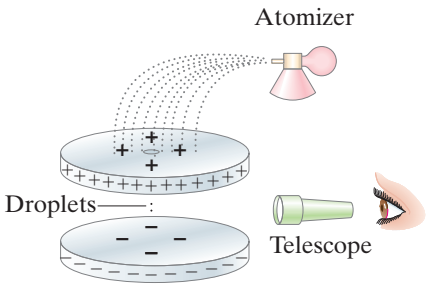


FIGURE 27–32 Millikan's oil-drop experiment.

of a smallest charge, e, that was ascribed to the electron, and that the value of e was 1.6×10^{-19} C. This value of e, combined with the measurement of e/m, gives the mass of the electron to be $(1.6 \times 10^{-19}\,\text{C})/(1.76 \times 10^{11}\,\text{C/kg}) = 9.1 \times 10^{-31}$ kg. This mass is less than a thousandth the mass of the smallest atom, and thus confirmed the idea that the electron is only a part of an atom. The accepted value today for the mass of the electron is

$$m_e = 9.11 \times 10^{-31}\,\text{kg}.$$

The experimental result that any charge is an integral multiple of e means that electric charge is *quantized* (exists only in discrete amounts). We start discussing "quantum theory" in Chapter 37.

*CRT, Revisited

The cathode ray tube (CRT), which once served as the picture tube of TV sets, oscilloscopes, and computer monitors, was discussed in Chapter 23. There, in Fig. 23–28, was a design using electric deflection plates to maneuver the electron beam. Many CRTs used a magnetic field produced by coils to maneuver the electron beam, and operated much like the coils shown in Fig. 27–31.

27–8 The Hall Effect

When a current-carrying conductor is held fixed in a magnetic field, the field exerts a sideways force on the charges moving in the conductor. For example, if electrons move to the right in the rectangular conductor shown in Fig. 27–33a, the inward magnetic field will exert a downward force (Eq. 27–5a) on the electrons $\vec{\mathbf{F}}_B = -e\vec{\mathbf{v}}_d \times \vec{\mathbf{B}}$, where $\vec{\mathbf{v}}_d$ is the drift velocity of the electrons (Section 25–8). Thus the electrons will tend to move nearer to side D than side C, causing a potential difference between sides C and D of the conductor. This potential difference provides an electric field $\vec{\mathbf{E}}_H$ that exerts a force, $e\vec{\mathbf{E}}_H$, on the moving charges that is equal and opposite to the magnetic force $(ev_d B)$. This effect is called the **Hall effect** after E. H. Hall, who discovered it in 1879. The difference of potential produced is called the **Hall emf**.

The electric field due to the separation of charge is called the *Hall field*, $\vec{\mathbf{E}}_H$, and points downward in Fig. 27–33a, as shown. In equilibrium, the force due to this electric field is balanced by the magnetic force $ev_d B$, so

$$eE_H = ev_d B.$$

Hence $E_H = v_d B$. The Hall emf is then (Eq. 23–4b, assuming the conductor is long and thin so E_H is uniform)

$$\mathscr{E}_H = E_H d = v_d B d, \qquad \textbf{(27–14)}$$

where d is the width of the conductor.

A current of negative charges moving to the right is equivalent to positive charges moving to the left, at least for most purposes. But the Hall effect can distinguish these two. As can be seen in Fig. 27–33b, positive particles moving to the left are deflected downward, so that the bottom surface is positive relative to the top surface. The Hall emf is the reverse of part (a). Indeed, the direction of the emf in the Hall effect first revealed that it is negative particles (not positive ones) that move in metal conductors, and much later that positive "holes" move in p-type semiconductors. (Section 40–8).

Because the Hall emf is proportional to B, the Hall effect can be used to measure magnetic fields. A device to do so, called a *Hall probe*, is first calibrated with known magnetic fields. Then, for the same current, its emf output will be a measure of B. When B is known, the Hall effect can be used to measure the drift speed of charge carriers. Such measurements also allow us to determine the density of charge carriers in a material.

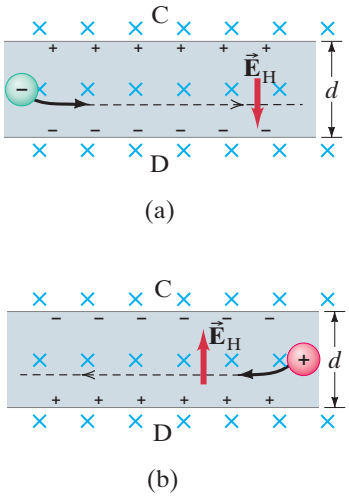
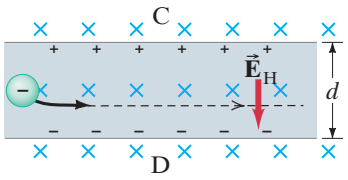


FIGURE 27–33 The Hall effect. (a) Negative charges moving to the right as the current. (b) Positive charges moving to the left as the current.



(a)

FIGURE 27–33a (Repeated here for Example 27–13.)

EXAMPLE 27–13 **Drift speed using the Hall effect.** A long copper strip 1.8 cm wide ($= d$) and thickness 1.0 mm is placed in a 1.2-T magnetic field as in Fig. 27–33a. When a steady current of 15 A passes through it, the Hall emf is measured to be 1.02 μV. Determine the drift speed of the electrons and the density of free (conducting) electrons (number per unit volume) in the copper.

APPROACH We use Eq. 27–14 to obtain the drift speed, and Eq. 25–13 of Chapter 25 to find the density of conducting electrons.

SOLUTION The drift speed (Eq. 27–14) is

$$v_d = \frac{\mathscr{E}_H}{Bd} = \frac{1.02 \times 10^{-6}\,\text{V}}{(1.2\,\text{T})(1.8 \times 10^{-2}\,\text{m})} = 4.7 \times 10^{-5}\,\text{m/s}.$$

The density of charge carriers n is obtained from Eq. 25–13, $I = nev_d A$, where A is the cross-sectional area through which the current I flows. Then

$$n = \frac{I}{ev_d A} = \frac{15\,\text{A}}{(1.6 \times 10^{-19}\,\text{C})(4.7 \times 10^{-5}\,\text{m/s})(1.8 \times 10^{-2}\,\text{m})(1.0 \times 10^{-3}\,\text{m})}$$
$$= 11 \times 10^{28}\,\text{m}^{-3}.$$

This value for the density of free electrons in copper, $n = 11 \times 10^{28}$ per m^3, is the experimentally measured value. It represents *more* than one free electron per atom, which as we saw in Example 25–14 is $8.4 \times 10^{28}\,\text{m}^{-3}$.

27–9 Mass Spectrometer

PHYSICS APPLIED
The mass spectrometer

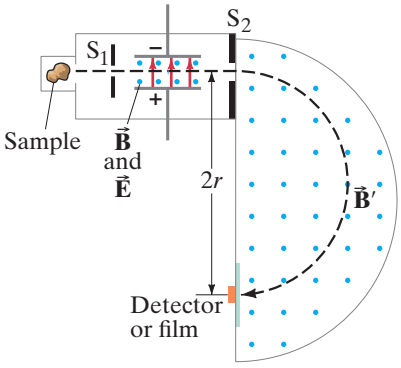


FIGURE 27–34 Bainbridge-type mass spectrometer. The magnetic fields B and B' point out of the paper (indicated by the dots).

A **mass spectrometer** is a device to measure masses of atoms. It is used today not only in physics but also in chemistry, geology, and medicine, often to identify atoms (and their concentration) in given samples. Ions can be produced by heating the sample, or by using an electric current. As shown in Fig. 27–34, the ions (mass m, charge q) pass through slit S_1 and enter a region (before S_2) where there are crossed ($\perp$) electric and magnetic fields. Ions follow a straight-line path in this region if the electric force qE (upward on a positive ion) is just balanced by the magnetic force qvB (downward on a positive ion): that is, if $qE = qvB$, or

$$v = \frac{E}{B}.$$

Only those ions whose speed is $v = E/B$ will pass through undeflected and emerge through slit S_2 (just as in the "velocity selector" of Example 27–10). In the semicircular region, after S_2, there is only a magnetic field, B', so the ions follow a circular path. The radius of the circular path is found from their mark on film, or by detectors, if B' is fixed. If instead r is fixed by the position of a detector, then B' is varied until detection occurs. Newton's second law, $\Sigma F = ma$, applied to an ion moving in a circle under the influence only of the magnetic field B' gives $qvB' = mv^2/r$. Since $v = E/B$, we have

$$m = \frac{qB'r}{v} = \frac{qBB'r}{E}. \qquad \textbf{(27–15)}$$

All the quantities on the right side are known or can be measured, and thus m can be determined.

Historically, the masses of many atoms were measured this way. When a pure substance was used, it was sometimes found that two or more closely spaced marks would appear on the film. For example, neon produced two marks whose radii corresponded to atoms of 20 and 22 atomic mass units (u). Impurities were ruled out and it was concluded that there must be two types of neon with different masses. These different forms were called **isotopes**. It was soon found that most elements are mixtures of isotopes, and the difference in mass is due to different numbers of neutrons (discussed in Chapter 41).

EXAMPLE 27–14 **Mass spectrometry.** Carbon atoms of atomic mass 12.0 u are found to be mixed with an unknown element. In a mass spectrometer with fixed B', the carbon traverses a path of radius 22.4 cm and the unknown's path has a 26.2-cm radius. What is the unknown element? Assume the electric charge on both types of ion is the same (maybe $+1e$ or $-1e$).

APPROACH We assume the carbon and unknown atoms pass through the same electric and magnetic fields, so the velocity selector selects the same speed. Hence their masses are proportional to the radius of their respective paths (see Eq. 27–15).

SOLUTION We write a ratio for the masses, using Eq. 27–15:

$$\frac{m_x}{m_C} = \frac{qBB'r_x/E}{qBB'r_C/E} = \frac{r_x}{r_C}$$
$$= \frac{26.2\text{ cm}}{22.4\text{ cm}} = 1.17.$$

Thus $m_x = 1.17 \times 12.0\text{ u} = 14.0\text{ u}$. The other element is probably nitrogen (see the Periodic Table, inside the back cover).

NOTE The unknown could also be an isotope such as carbon-14 $\left({}^{14}_{6}\text{C}\right)$. See Appendix F. Further physical or chemical analysis would be needed.

Summary

A magnet has two **poles**, north and south. The north pole is that end which points toward geographic north when the magnet is freely suspended. Like poles of two magnets repel each other, whereas unlike poles attract.

We can picture that a **magnetic field** surrounds every magnet. The SI unit for magnetic field is the **tesla** (T).

Electric currents produce magnetic fields. For example, the lines of magnetic field due to a current in a straight wire form circles around the wire, and the field exerts a force on magnets (or currents) near it.

A magnetic field exerts a force on an electric current. The force on an infinitesimal length of wire $d\vec{\boldsymbol{\ell}}$ carrying a current I in a magnetic field $\vec{\mathbf{B}}$ is

$$d\vec{\mathbf{F}} = I\, d\vec{\boldsymbol{\ell}} \times \vec{\mathbf{B}}. \qquad \textbf{(27–4)}$$

If the field $\vec{\mathbf{B}}$ is uniform over a straight length $\vec{\boldsymbol{\ell}}$ of wire, then the force is

$$\vec{\mathbf{F}} = I\vec{\boldsymbol{\ell}} \times \vec{\mathbf{B}} \qquad \textbf{(27–3)}$$

which has magnitude

$$F = I\ell B \sin\theta \qquad \textbf{(27–1)}$$

where θ is the angle between magnetic field $\vec{\mathbf{B}}$ and the current direction. The direction of the force is perpendicular to the wire and to the magnetic field, and is given by the right-hand rule. This relation serves as the definition of magnetic field $\vec{\mathbf{B}}$.

Similarly, a magnetic field $\vec{\mathbf{B}}$ exerts a force on a charge q moving with velocity $\vec{\mathbf{v}}$ given by

$$\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}. \qquad \textbf{(27–5a)}$$

The magnitude of the force is

$$F = qvB \sin\theta, \qquad \textbf{(27–5b)}$$

where θ is the angle between $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$.

The path of a charged particle moving perpendicular to a uniform magnetic field is a circle.

If both electric and magnetic fields $\left(\vec{\mathbf{E}} \text{ and } \vec{\mathbf{B}}\right)$ are present, the force on a charge q moving with velocity $\vec{\mathbf{v}}$ is

$$\vec{\mathbf{F}} = q\vec{\mathbf{E}} + q\vec{\mathbf{v}} \times \vec{\mathbf{B}}. \qquad \textbf{(27–7)}$$

The torque on a current loop in a magnetic field $\vec{\mathbf{B}}$ is

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{\mu}} \times \vec{\mathbf{B}}, \qquad \textbf{(27–11)}$$

where $\vec{\boldsymbol{\mu}}$ is the **magnetic dipole moment** of the loop or loops:

$$\vec{\boldsymbol{\mu}} = NI\vec{\mathbf{A}}. \qquad \textbf{(27–10)}$$

Here N is the number of loops carrying current I in the coil and $\vec{\mathbf{A}}$ is a vector perpendicular to the plane of the loop (use right-hand rule, fingers along current in loop) and has magnitude equal to the area of the loop.

The force or torque exerted on a current-carrying wire by a magnetic field is the basis for operation of many devices, such as **motors** and **loudspeakers**, and in **galvanometers** (used in analog electric meters).

The measurement of the charge-to-mass ratio (e/m) of the electron was done using magnetic and electric fields. The charge e on the electron was first measured in the Millikan oil-drop experiment and then its mass was obtained from the measured value of the e/m ratio.

In the **Hall effect**, moving charges in a conductor placed in a magnetic field are forced to one side, producing an emf between the two sides of the conductor.

A **mass spectrometer** uses magnetic and electric fields to measure the mass of ions.

Questions

1. A compass needle is not always balanced parallel to the Earth's surface, but one end may dip downward. Explain.
2. Explain why the Earth's "north pole" is really a magnetic south pole. Indicate how north and south magnetic poles were defined and how we can tell experimentally that the north pole is really a south magnetic pole.
3. Draw the magnetic field lines around a straight section of wire carrying a current horizontally to the left.
4. A horseshoe magnet is held vertically with the north pole on the left and south pole on the right. A wire passing between the poles, equidistant from them, carries a current directly away from you. In what direction is the force on the wire? Explain.
5. In the relation $\vec{\mathbf{F}} = I\vec{\ell} \times \vec{\mathbf{B}}$, which pairs of the vectors $(\vec{\mathbf{F}}, \vec{\ell}, \vec{\mathbf{B}})$ are always at 90°? Which can be at other angles?
6. The magnetic field due to current in wires in your home can affect a compass. Discuss the effect in terms of currents, including if they are ac or dc.
7. In Fig. 27–35, charged particles move in the vicinity of a current-carrying wire. For each charged particle, the arrow indicates the initial direction of motion of the particle, and the + or − indicates the sign of the charge. For each of the particles, indicate the direction of the initial magnetic force due to the magnetic field produced by the wire. Explain.

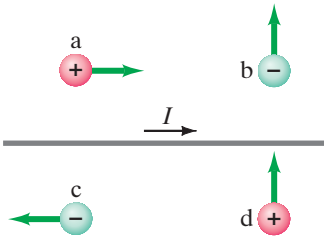


FIGURE 27–35
Question 7.

8. If a negatively charged particle enters a region of uniform magnetic field which is perpendicular to the particle's velocity, will the kinetic energy of the particle increase, decrease, or stay the same? Explain your answer. (Neglect gravity and assume there is no electric field.)
9. Note that the pattern of magnetic field lines surrounding a bar magnet is similar to that of the electric field around an electric dipole. From this fact, predict how the magnetic field will change with distance (*a*) when near one pole of a very long bar magnet, and (*b*) when far from a magnet as a whole.
10. Three particles, a, b, and c, enter a magnetic field and follow paths as shown in Fig. 27–36. What can you say about the charge on each particle? Explain.

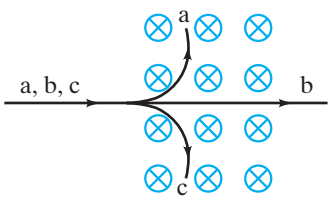


FIGURE 27–36
Question 10.

11. Describe the trajectory of a negatively charged particle in the velocity selector of Fig. 27–22 if its speed exceeds E/B. What is its trajectory if $v < E/B$? Would it make any difference if the particle were positively charged?
12. Can you set a resting electron into motion with a steady magnetic field? With an electric field? Explain.
13. A positively charged particle in a nonuniform magnetic field follows the trajectory shown in Fig. 27–37. For each of the labeled points (a, b, c, d, e), indicate the direction of the magnetic field near the path, assuming the path is always in the plane of the page, and indicate the relative magnitudes of the field near each point. Explain your answers.

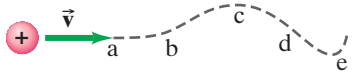


FIGURE 27–37
Question 13.

14. A charged particle is moving in a circle under the influence of a uniform magnetic field. If an electric field that points in the same direction as the magnetic field is turned on, describe the path the charged particle will take.
15. A beam of electrons is directed toward a horizontal wire carrying a current from left to right (Fig. 27–38). In what direction is the beam deflected?

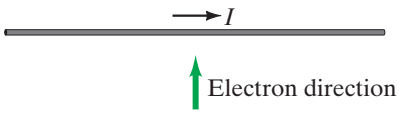


FIGURE 27–38
Question 15.

16. A charged particle moves in a straight line through a particular region of space. Could there be a nonzero magnetic field in this region? If so, give two possible situations.
17. If a moving charged particle is deflected sideways in some region of space, can we conclude, for certain, that $\vec{\mathbf{B}} \neq 0$ in that region? Explain.
18. How could you tell whether moving electrons in a certain region of space are being deflected by an electric field or by a magnetic field (or by both)?
19. How could you make a compass without using iron or other ferromagnetic material?
20. Describe how you could determine the dipole moment of a bar magnet or compass needle.
21. In what positions (if any) will a current loop placed in a uniform magnetic field be in (*a*) stable equilibrium, and (*b*) unstable equilibrium?
22. The force on a moving particle in a magnetic field is the idea behind **electromagnetic pumping**. It can be used to pump metallic fluids (such as sodium) and to pump blood in artificial heart machines. A basic design is shown in Fig. 27–39. For blood, an electric field is applied perpendicular to a blood vessel and to the magnetic field. Explain in detail how ions in the blood are caused to move. Do positive and negative ions feel a force in the same direction? Explain.

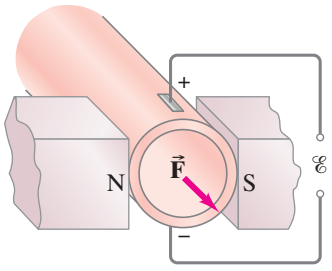


FIGURE 27–39
Electromagnetic pumping in a blood vessel.
Question 22.

23. A rectangular piece of semiconductor is inserted in a magnetic field and a battery is connected to its ends as shown in Fig. 27–40. When a sensitive voltmeter is connected between points a and b, it is found that point a is at a higher potential than b. What is the sign of the charge carriers in this semiconductor material?

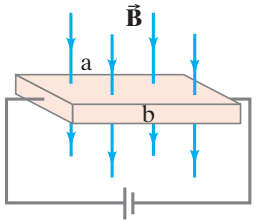
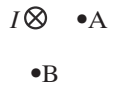


FIGURE 27–40
Question 23.

24. Two ions have the same mass, but one is singly ionized and the other is doubly ionized. How will their positions on the film of a mass spectrometer (Fig. 27–34) differ? Explain.

MisConceptual Questions

1. A current in a wire points into the page as shown at the right. In which direction is the magnetic field at point A (choose below)?





(*a*) (*b*) (*c*) (*d*) (*e*) None of these.

2. In which direction (see above) is the magnetic field at point B?

3. How should you cut a bar magnet to create two magnetic poles for which the north pole is twice as strong as the south pole?

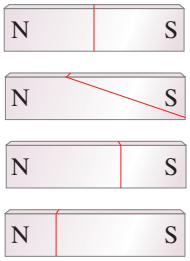
(*a*) Cut the magnet in half at the midpoint between the north and south ends.

(*b*) Cut the magnet on a diagonal as shown.

(*c*) Cut the magnet so 2/3 of the magnet ends up with the N side.

(*d*) Cut the magnet so 2/3 of the magnet ends up with the S side.

(*e*) No cut can meet this request.



4. A wire carries a current directly away from you. Which way do the magnetic field lines produced by this wire point?

(*a*) They point parallel to the wire in the direction of the current.

(*b*) They point parallel to the wire opposite the direction of the current.

(*c*) They point toward the wire.

(*d*) They point away from the wire.

(*e*) They make clockwise circles around the wire.

(*f*) They make counterclockwise circles around the wire.

5. As a proton moves through space, it creates

(*a*) an electric field only.

(*b*) a magnetic field only.

(*c*) both an electric field and a magnetic field.

(*d*) nothing; the electric field and magnetic fields cancel each other out.

6. Which statements about the force on a charged particle placed in a magnetic field are true?

(*a*) A magnetic force can be exerted only if the particle is moving.

(*b*) The force is a maximum if the particle is moving in the direction of the field.

(*c*) The force causes the particle to gain kinetic energy.

(*d*) The direction of the force is along the magnetic field.

(*e*) A magnetic field always exerts a force on a charged particle.

7. When a charged particle moves parallel to the direction of a magnetic field, the particle travels in a

(*a*) straight line.

(*b*) circular path.

(*c*) helical path.

(*d*) hysteresis loop.

8. Compare the magnetic force on a proton and on an electron, each traveling with the same speed in the same direction and perpendicular to the magnetic field.

(*a*) The magnitude and direction of the force is the same on the proton and the electron.

(*b*) The direction of the force is the same on the proton and the electron, but the force on the proton is larger.

(*c*) The direction of the force is the same on the proton and the electron, but the force on the electron is larger.

(*d*) The magnitude of the force is the same on the proton and the electron, but the direction of the force is opposite.

(*e*) There is no force on either the proton or the electron.

9. A proton enters a uniform magnetic field that is perpendicular to the proton's velocity (Fig. 27–41). What happens to the kinetic energy of the proton?

(*a*) It increases.

(*b*) It decreases.

(*c*) It stays the same.

(*d*) It depends on the velocity direction.

(*e*) It depends on the *B* field direction.

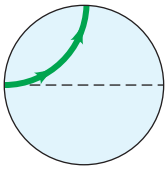


FIGURE 27–41
MisConceptual Question 9.

10. For a charged particle, a constant magnetic field can be used to change
(*a*) only the direction of the particle's velocity.
(*b*) only the magnitude of the particle's velocity.
(*c*) both the magnitude and direction of the particle's velocity.
(*d*) None of the above.

11. Which of the following statements about the force on a charged particle due to a magnetic field are not valid?
(*a*) It depends on the particle's charge.
(*b*) It depends on the particle's velocity.
(*c*) It depends on the strength of the external magnetic field.
(*d*) It acts at right angles to the direction of the particle's motion.
(*e*) None of the above; all of these statements are valid.

12. A magnetic field exists in the space surrounding a metallic wire that carries a steady electric current. But an electric field is not present in that space because
(*a*) moving charges create magnetic fields but not electric fields.
(*b*) the wire contains equal amounts of positive and negative charges.
(*c*) electrons and atoms move in different directions in the wire.
(*d*) the magnetic field cancels out the electric field.

13. Two particles follow the paths shown in Fig. 27–42. They move in a uniform magnetic field $\vec{\mathbf{B}}$ directed into the page. What is the charge on particle 1, and on particle 2?
(*a*) +, + (*b*) −, + (*c*) +, − (*d*) −, −
(*e*) None of the above.



FIGURE 27–42
MisConceptual Question 13.

Problems

27–1 Magnets and Magnetic Fields

1. (II) You go on a hike in a straight line, 35° east of north, using a compass and heading toward a point 3.0 km away according to a map. If you forget to correct for the difference between true north on the map and magnetic north, how far from your destination would you be after 3.0 km? Assume a magnetic declination of +13° for California. A positive declination means magnetic north is east of true geographic north.

2. (II) If you approximate the Earth's magnetic field as the field of a magnetic dipole, the field will decrease roughly as $1/r^3$, just as for an electric dipole field (Section 21–11, and see also Section 28–6 for confirmation). If the magnetic field is 0.50×10^{-4} T at the Earth's surface, estimate the strength of the magnetic field at an altitude of (*a*) 0.20 Earth radii and (*b*) 2.0 Earth radii.

27–3 Force on Electric Current in Magnetic Field

3. (I) (*a*) What is the force per meter of length on a straight wire carrying a 7.40-A current when perpendicular to a 0.90-T uniform magnetic field? (*b*) What if the angle between the wire and field is 35.0°?

4. (I) How much current is flowing in a wire 4.20 m long if the maximum force on it is 0.625 N when placed in a uniform 0.0800-T field?

5. (I) A 240-m length of a nearly straight wire stretches between two towers and carries a 150-A current. Determine the magnitude of the force on the wire due to the Earth's magnetic field of 5.0×10^{-5} T which makes an angle of 68° with the wire.

6. (I) A 2.6-m length of horizontal wire carries a 4.5-A current toward the south. The dip angle of the Earth's magnetic field makes an angle of 35° to the wire. Estimate the magnitude of the magnetic force on the wire due to the Earth's magnetic field of 5.5×10^{-5} T.

7. (I) The magnetic force per meter on a wire is measured to be only 55% of its maximum possible value. What is the angle between the wire and the magnetic field?

8. (II) The force on a wire is a maximum of 7.50×10^{-2} N when placed between the pole faces of a magnet. The current flows horizontally to the right and the magnetic field is vertical. The wire is observed to "jump" toward the observer when the current is turned on. (*a*) What type of magnetic pole is the horizontal upper pole face? (*b*) If the pole faces have a diameter of 10.0 cm, estimate the current in the wire if the field is 0.180 T. (*c*) If the wire is tipped so that it makes an angle of 10.0° with the horizontal, what force will it now feel? [*Hint*: What length of wire will now be in the field?]

9. (II) Suppose a straight 1.00-mm-diameter copper wire could just "float" horizontally in air because of the force due to the Earth's magnetic field $\vec{\mathbf{B}}$, which is horizontal, perpendicular to the wire, and of magnitude 5.0×10^{-5} T. What current would the wire carry? Does the answer seem feasible? Explain briefly.

10. (II) The force on a wire carrying 6.45 A is a maximum of 1.64 N when placed between the pole faces of a magnet. If the pole faces are 55.5 cm in diameter, what is the approximate strength of the magnetic field?

11. (II) A stiff wire 50.0 cm long is bent at a right angle in the middle. One section lies along the z axis and the other is along the line $y = 2x$ in the xy plane. A current of 20.0 A flows in the wire—down the z axis and out the wire in the xy plane. The wire passes through a uniform magnetic field given by $\vec{\mathbf{B}} = (0.285\,\hat{\mathbf{i}})$T. Determine the magnitude and direction of the total force on the wire.

12. (II) A long wire stretches along the x axis and carries a 3.0-A current to the right $(+x)$. The wire is in a uniform magnetic field $\vec{\mathbf{B}} = (0.20\,\hat{\mathbf{i}} - 0.42\,\hat{\mathbf{j}} + 0.25\,\hat{\mathbf{k}})$T. Determine the components of the force on the wire per cm of length.

13. (II) A 2.0-m-long wire carries a current of 7.4 A and is immersed within a uniform magnetic field $\vec{\mathbf{B}}$. When this wire lies along the $+x$ axis, a magnetic force $\vec{\mathbf{F}} = (-2.5\,\hat{\mathbf{j}})$N acts on the wire, and when it lies on the $+y$ axis, the force is $\vec{\mathbf{F}} = (2.5\,\hat{\mathbf{i}} - 5.0\,\hat{\mathbf{k}})$N. Find $\vec{\mathbf{B}}$.

14. (II) A current-carrying circular loop of wire (radius r, current I) is partially immersed in a magnetic field of constant magnitude B_0 directed out of the page as shown in Fig. 27–43. Determine the net force on the loop due to the field in terms of θ_0. (Note that θ_0 points to the dashed line, above which $B = 0$.)

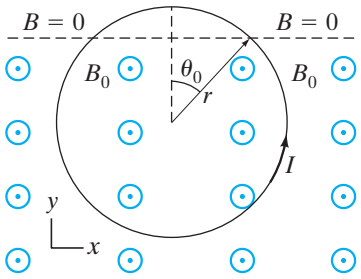


FIGURE 27–43 Problem 14.

15. (III) A curved wire, connecting two points a and b, lies in a plane perpendicular to a uniform magnetic field $\vec{\mathbf{B}}$ and carries a current I. Show that the resultant magnetic force on the wire, no matter what its shape, is the same as that on a straight wire connecting the two points carrying the same current I. See Fig. 27–44.

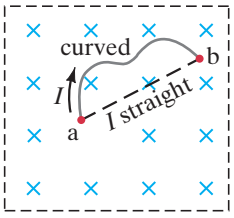


FIGURE 27–44 Problem 15.

16. (III) A circular loop of wire, of radius r, carries current I. It is placed in a magnetic field whose straight lines seem to diverge from a point a distance d below the loop on its axis. (That is, the field makes an angle θ with the loop at all points, Fig. 27–45, where $\tan\theta = r/d$.) Determine the force on the loop.

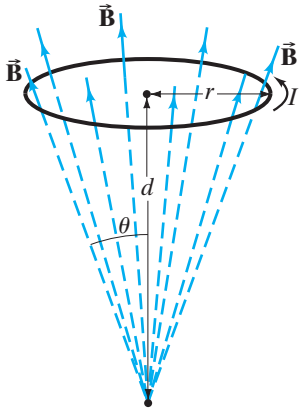


FIGURE 27–45 Problem 16.

27–4 Force on Charge Moving in Magnetic Field

17. (I) Determine the magnetic magnitude and direction of the force on an electron traveling 7.75×10^5 m/s horizontally to the east in a vertically upward magnetic field of strength 0.65 T.

18. (I) An electron is projected vertically upward with a speed of 1.70×10^6 m/s into a uniform magnetic field of 0.550 T that is directed horizontally away from the observer. Describe the electron's path in this field.

19. (I) A small bird (mass = 20.0 grams) has acquired a charge of 15.0 μC. How fast would the bird have to fly for the magnetic force due to the Earth's 0.5-G field to equal 1/10 the bird's weight? Is this a realistic speed? Assume a 90° angle between the bird's velocity and the magnetic field.

20. (I) Alpha particles (charge $q = +2e$, mass $m = 6.6 \times 10^{-27}$ kg) move at 2.2×10^6 m/s. What magnetic field strength would be required to bend them into a circular path of radius $r = 0.14$ m?

21. (I) Find the direction of the force on a negative charge for each diagram shown in Fig. 27–46, where $\vec{\mathbf{v}}$ (green) is the velocity of the charge and $\vec{\mathbf{B}}$ (blue) is the direction of the magnetic field. (⊗ means the vector points inward. ⊙ means it points outward, toward you.)

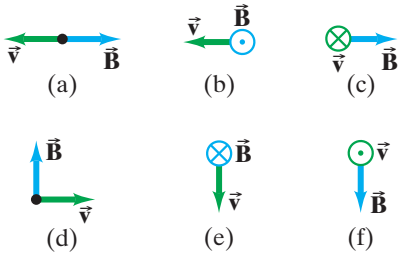


FIGURE 27–46 Problem 21.

22. (II) Determine the direction of $\vec{\mathbf{B}}$ for each case in Fig. 27–47, where $\vec{\mathbf{F}}$ represents the maximum magnetic force on a positively charged particle moving with velocity $\vec{\mathbf{v}}$.

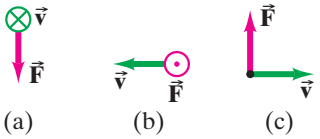


FIGURE 27–47 Problem 22.

23. (II) A 720-KeV (kinetic energy) proton enters a 0.20-T field, in a plane perpendicular to the field. What is the radius of its path? See Section 23–8.

24. (II) An electron experiences the greatest force as it travels 2.8×10^6 m/s in a magnetic field when it is moving northward. The force is vertically upward and of magnitude 7.4×10^{-13} N. What is the magnitude and direction of the magnetic field?

25. (II) A proton moves through a region of space where there is a magnetic field $\vec{\mathbf{B}} = (0.56\,\hat{\mathbf{i}} + 0.42\,\hat{\mathbf{j}})$ T and an electric field $\vec{\mathbf{E}} = (3.0\,\hat{\mathbf{i}} - 4.2\,\hat{\mathbf{j}}) \times 10^3$ V/m. At a given instant, the proton's velocity is $\vec{\mathbf{v}} = (6.0\,\hat{\mathbf{i}} + 3.0\,\hat{\mathbf{j}} - 5.0\,\hat{\mathbf{k}}) \times 10^3$ m/s. Determine the components of the total force on the proton.

26. (II) An electron experiences a force $\vec{\mathbf{F}} = (3.8\,\hat{\mathbf{i}} - 2.7\,\hat{\mathbf{j}}) \times 10^{-13}$ N when passing through a magnetic field $\vec{\mathbf{B}} = (0.74\,\text{T})\hat{\mathbf{k}}$. Determine the components of the electron's velocity.

27. (II) A particle of charge q moves in a circular path of radius r in a uniform magnetic field $\vec{\mathbf{B}}$. If the magnitude of the magnetic field is doubled, and the kinetic energy of the particle remains constant, what happens to the angular momentum of the particle?

28. (II) An electron moves with velocity $\vec{\mathbf{v}} = (6.0\,\hat{\mathbf{i}} - 7.0\,\hat{\mathbf{j}}) \times 10^4$ m/s in a magnetic field $\vec{\mathbf{B}} = (-0.80\,\hat{\mathbf{i}} + 0.60\,\hat{\mathbf{j}})$ T. Determine the magnitude and direction of the force on the electron.

29. (II) A helium ion ($Q = +2e$) whose mass is 6.6×10^{-27} kg is accelerated by a voltage of 5200 V. (*a*) What is its speed? (*b*) What will be its radius of curvature if it moves in a plane perpendicular to a uniform 0.340-T field? (*c*) What is its period of revolution?

30. (II) For a particle of mass m and charge q moving in a circular path in a magnetic field B, (*a*) show that its kinetic energy is proportional to r^2, the square of the radius of curvature of its path. (*b*) Show that its angular momentum is $L = qBr^2$, around the center of the circle.

31. (II) A proton (mass m_p), a deuteron $(m = 2m_p,\ q = e)$, and an alpha particle $(m = 4m_p,\ q = 2e)$ are accelerated by the same potential difference V and then enter a uniform magnetic field $\vec{\mathbf{B}}$, where they move in circular paths perpendicular to $\vec{\mathbf{B}}$. Determine the radius of the paths for the deuteron and alpha particle in terms of that for the proton.

32. (II) An electron enters a uniform magnetic field $B = 0.28\ \text{T}$ at a 45° angle to $\vec{\mathbf{B}}$. Determine the radius r and pitch p (distance between loops) of the electron's helical path assuming its speed is $2.2 \times 10^6\ \text{m/s}$. See Fig. 27–48.

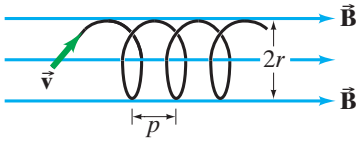


FIGURE 27–48
Problem 32.

33. (II) A particle with charge q and momentum p, initially moving along the x axis, enters a region where a uniform magnetic field $\vec{\mathbf{B}} = B_0\hat{\mathbf{k}}$ extends over a width $x = \ell$ as shown in Fig. 27–49. The particle is deflected a distance d in the $+y$ direction as it traverses the field. Determine (*a*) whether q is positive or negative, and (*b*) the magnitude of its momentum p in terms of q, B_0, ℓ, and d.

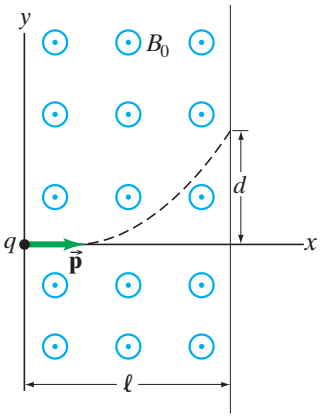


FIGURE 27–49
Problem 33.

34. (II) The path of protons emerging from an accelerator must be bent by 90° by a "bending magnet" so as not to strike a barrier in their path a distance d from their exit hole in the accelerator. Show that the field $\vec{\mathbf{B}}$ in the bending magnet, which we assume is uniform and can extend over an area $d \times d$, must have magnitude $B \geq (2mK/e^2d^2)^{\frac{1}{2}}$, where m is the mass of a proton and K is its kinetic energy.

35. (III) A 3.40-g bullet moves with a speed of 155 m/s perpendicular to the Earth's magnetic field of $5.00 \times 10^{-5}\ \text{T}$. If the bullet possesses a net charge of $18.5 \times 10^{-9}\ \text{C}$, by what distance will it be deflected from its path due to the Earth's magnetic field after it has traveled 1.25 km?

36. (III) A proton moving with speed $v = 1.8 \times 10^5\ \text{m/s}$ in a field-free region abruptly enters an essentially uniform magnetic field $B = 0.850\ \text{T}\,(\vec{\mathbf{B}} \perp \vec{\mathbf{v}})$. If the proton enters the magnetic field region at a 45° angle as shown in Fig. 27–50, (*a*) at what angle does it leave, and (*b*) at what distance x does it exit the field?

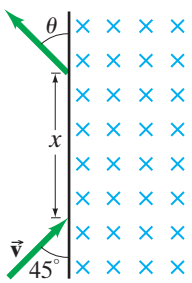


FIGURE 27–50
Problem 36.

37. (III) A particle with charge $+q$ and mass m travels in a uniform magnetic field $\vec{\mathbf{B}} = B_0\hat{\mathbf{k}}$. At time $t = 0$, the particle's speed is v_0 and its velocity vector lies in the xy plane directed at an angle of 30° with respect to the y axis as shown in Fig. 27–51. At a later time $t = t_\alpha$, the particle will cross the x axis at $x = \alpha$. In terms of q, m, v_0, and B_0, determine (*a*) α, and (*b*) t_α.

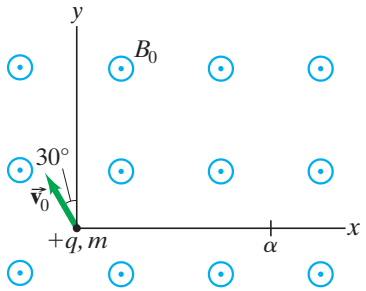


FIGURE 27–51
Problem 37.

27–5 Torque on a Current Loop; Magnetic Moment

38. (I) A 13.0-cm-diameter circular loop of wire is placed with the plane of the loop parallel to the uniform magnetic field between the pole pieces of a large magnet. When 3.80 A flows in the coil, the torque on it is $0.185\ \text{m} \cdot \text{N}$. What is the magnetic field strength?

39. (I) A single square loop of wire 22.0 cm on a side is placed with its face parallel to a 1.30-T magnetic field as in Fig. 27–23b. When 5.70 A flows in the coil, what is the torque on it?

40. (I) How much work is required to rotate the current loop (Fig. 27–23) in a uniform magnetic field $\vec{\mathbf{B}}$ from (*a*) $\theta = 0°\,(\vec{\boldsymbol{\mu}} \| \vec{\mathbf{B}})$ to $\theta = 180°$, (*b*) $\theta = 90°$ to $\theta = -90°$?

41. (II) A circular coil 18.0 cm in diameter and containing twelve loops lies flat on the ground. The Earth's magnetic field at this location has magnitude $5.50 \times 10^{-5}\ \text{T}$ and points into the Earth at an angle of 54.0° below a line pointing due north. If a 7.10-A clockwise current passes through the coil, (*a*) determine the torque on the coil; (*b*) which edge of the coil rises up: north, east, south, or west?

42. (II) Show that the magnetic dipole moment μ of an electron orbiting the proton nucleus of a hydrogen atom is related to the orbital momentum L of the electron by

$$\mu = \frac{e}{2m}L.$$

43. (II) A 24-loop circular coil 22 cm in diameter lies in the xy plane. The current in each loop of the coil is 7.6 A clockwise, and an external magnetic field $\vec{\mathbf{B}} = (0.55\hat{\mathbf{i}} + 0.60\hat{\mathbf{j}} - 0.65\hat{\mathbf{k}})\ \text{T}$ passes through the coil. Determine: (*a*) the magnetic moment of the coil, $\vec{\boldsymbol{\mu}}$; (*b*) the torque on the coil due to the external magnetic field; (*c*) the potential energy U of the coil in the field (take the same zero for U as we did in our discussion of Fig. 27–23).

44. (III) Suppose a nonconducting rod of length d carries a uniformly distributed charge Q. It is rotated with angular velocity ω about an axis perpendicular to the rod at one end, Fig. 27–52. Show that the magnetic dipole moment of this rod is $\frac{1}{6}Q\omega d^2$. [*Hint*: Consider the motion of each infinitesimal length of the rod.]

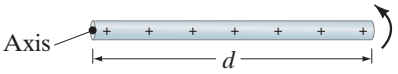


FIGURE 27–52
Problem 44.

27–6 Motors, Galvanometers

45. (I) If the current to an electric motor drops by 18%, by what factor does the output torque change?

46. (I) A galvanometer needle deflects full scale for a 63.0-μA current. What current will give full-scale deflection if the magnetic field weakens to 0.700 of its original value?

47. (I) If the restoring spring of a galvanometer weakens by 15% over the years, what current will give full-scale deflection if it originally required 25 μA?

27–7 Discovery of Electron

48. (I) What is the value of q/m for a particle that moves in a circle of radius 8.0 mm in a 0.46-T magnetic field if a crossed 320-V/m electric field will make the path straight?

49. (II) (*a*) What is the velocity of a beam of electrons that go undeflected when passing through crossed (perpendicular) electric and magnetic fields of magnitude 1.52×10^4 V/m and 2.60×10^{-3} T, respectively? (*b*) What is the radius of the electron orbit if the electric field is turned off?

50. (II) An oil drop whose mass is determined to be 3.8×10^{-15} kg is held at rest between two large plates separated by 1.0 cm as in Fig. 27–32. If the potential difference between the plates is 340 V, how many excess electrons does this oil drop have?

27–8 Hall Effect

51. (II) A **Hall probe**, consisting of a rectangular slab of current-carrying material, is calibrated by placing it in a known magnetic field of magnitude 0.10 T. When the field is oriented normal to the slab's rectangular face, a Hall emf of 12 mV is measured across the slab's width. The probe is then placed in a magnetic field of unknown magnitude B, and a Hall emf of 63 mV is measured. Determine B assuming that the angle θ between the unknown field and the plane of the slab's rectangular face is (*a*) $\theta = 90°$, and (*b*) $\theta = 60°$.

52. (II) A Hall probe used to measure magnetic field strengths consists of a rectangular slab of material (free-electron density n) with width d and thickness t, carrying a current I along its length ℓ. The slab is immersed in a magnetic field of magnitude B oriented perpendicular to its rectangular face (of area ℓd), so that a Hall emf $\mathcal{E}_H$ is produced across its width d. The probe's *magnetic sensitivity*, defined as $K_H = \mathcal{E}_H/IB$, indicates the magnitude of the Hall emf achieved for a given applied magnetic field and current. A slab with a large K_H is a good candidate for use as a Hall probe. (*a*) Show that $K_H = 1/ent$. Thus, a good Hall probe has small values for both n and t. (*b*) As possible candidates for the material used in a Hall probe, consider (i) a typical metal ($n \approx 1 \times 10^{29}/\text{m}^3$) and (ii) a (doped) semiconductor ($n \approx 3 \times 10^{22}/\text{m}^3$). Given that a semiconductor slab can be manufactured with a thickness of 0.15 mm, how thin (nm) should a metal slab be to yield a K_H value equal to that of the semiconductor slab? Compare this metal slab thickness with the 0.3-nm size of a typical metal atom. (*c*) For the typical semiconductor slab described in part (*b*), what is the expected value for $\mathcal{E}_H$ when $I = 100$ mA and $B = 0.1$ T?

53. (II) A long copper strip is 3.0 cm wide and 1.0 mm thick. When it carries a steady 42-A current in a 0.80-T magnetic field it produces a 6.5-μV Hall emf. Determine: (*a*) the Hall field in the conductor; (*b*) the drift speed of the conduction electrons; (*c*) the density of free electrons in the metal.

54. (II) In a probe that uses the Hall effect to measure magnetic fields, a 12.0-A current passes through a 1.50-cm-wide strip ($= d$ in Fig. 27–33) of sodium metal that is 1.20 mm thick. If the Hall emf is 1.86 μV, what is the magnitude of the magnetic field (take it perpendicular to the flat face of the strip)? Assume one free electron per atom of Na, and take its specific gravity to be 0.971.

55. (III) The Hall effect can be used to measure **blood flow speed** because the blood contains ions that constitute an electric current. (*a*) Does the sign of the ions influence the emf? Explain. (*b*) Determine the flow speed in an artery 3.3 mm in diameter if the measured emf across the width of the artery is 0.13 mV and B is 0.070 T. (In actual practice, an alternating magnetic field is used.)

27–9 Mass Spectrometer

56. (I) In a mass spectrometer, germanium atoms have radii of curvature equal to 21.0, 21.6, 21.9, 22.2, and 22.8 cm. The largest radius corresponds to an atomic mass of 76 u. What are the atomic masses of the other isotopes?

57. (II) One form of mass spectrometer accelerates ions by a voltage V before they enter a magnetic field B. The ions are assumed to start from rest. Show that the mass of an ion is $m = qB^2R^2/2V$, where R is the radius of the ions' path in the magnetic field and q is their charge.

58. (II) Suppose the electric field between the electric plates in the mass spectrometer of Fig. 27–34 is 2.84×10^4 V/m and the magnetic fields are $B = B' = 0.58$ T. The source contains carbon isotopes of mass numbers 12, 13, and 14 from a long-dead piece of a tree. (To estimate atomic masses, multiply by 1.67×10^{-27} kg.) (*a*) How far apart are the marks formed by the singly charged ions of each type on a detector or photographic film? (*b*) What if the ions were doubly charged? (*c*) Does it matter if the ion charge is positive (lost electrons) or negative (gained electrons)? Explain.

59. (II) An unknown particle moves in a straight line through crossed electric and magnetic fields with $E = 1.5$ kV/m and $B = 0.036$ T. If the electric field is turned off, the particle moves in a circular path of radius $r = 2.4$ cm. What might the two lightest possible particles be?

60. (II) Protons would move in a circle of radius 5.10 cm perpendicular to a 0.585-T magnetic field. What value of electric field could make their paths straight? In what direction must the electric field point?

61. (II) A mass spectrometer is monitoring air pollutants. It is difficult, however, to separate molecules of nearly equal mass such as CO (28.0106 u) and N_2 (28.0134 u). How large a radius of curvature must a spectrometer have (Fig. 27–34) if these two molecules are to be separated on the detector by 0.50 mm?

General Problems

62. A proton and an electron have the same kinetic energy upon entering a region of constant magnetic field. What is the ratio of the radii of their circular paths?

63. A motor run by a 9.0-V battery has a 20-turn square coil with sides of length 5.0 cm and total resistance 24 Ω. When spinning, the magnetic field felt by the wire in the coil is 0.020 T. What is the maximum torque on the motor?

64. Two stiff parallel wires a distance d apart in a horizontal plane act as rails to support a light metal rod of mass m (perpendicular to each rail), Fig. 27–53. A magnetic field $\vec{\mathbf{B}}$, directed vertically upward (outward in diagram), acts throughout. At $t = 0$, a constant current I begins to flow through the system. Determine the speed of the rod, which starts from rest at $t = 0$, as a function of time (*a*) assuming no friction between the rod and the rails, and (*b*) if the coefficient of friction is μ_k. (*c*) In which direction does the rod move, east or west, if the current through it heads north?

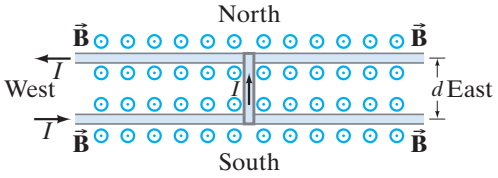


FIGURE 27–53 Looking down on a rod sliding on rails. Problems 64 and 65.

65. Suppose the rod in Fig. 27–53 (Problem 64) has mass $m = 0.40$ kg and length 28 cm and the current through it is $I = 36$ A. If the coefficient of static friction is $\mu_s = 0.50$, determine the minimum magnetic field $\vec{\mathbf{B}}$ (not necessarily vertical) that will just cause the rod to slide. Give the magnitude of $\vec{\mathbf{B}}$ and its direction relative to the vertical (= out of the page).

66. Near the equator, the Earth's magnetic field points almost horizontally to the north and has magnitude $B = 0.50 \times 10^{-4}$ T. What should be the magnitude and direction for the velocity of an electron if its weight is to be exactly balanced by the magnetic force?

67. A uniform conducting rod of length ℓ and mass m sits atop a fulcrum, which is placed a distance $\ell/4$ from the rod's left-hand end and is immersed in a uniform magnetic field of magnitude B directed into the page (Fig. 27–54). An object whose mass M is 7.0 times greater than the rod's mass is hung from the rod's left-hand end. What current (direction and magnitude) should flow through the rod in order for it to be "balanced" (i.e., be at rest horizontally) on the fulcrum? (Flexible connecting wires which exert negligible force on the rod are not shown.)

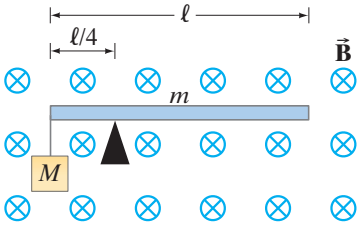


FIGURE 27–54 Problem 67.

68. Protons with momentum 4.8×10^{-21} kg · m/s are magnetically steered clockwise in a circular path 2.7 m in diameter. Determine the magnitude and direction of the field in the magnets surrounding the beam pipe.

69. The **cyclotron** (Fig. 27–55) is a device used to accelerate elementary particles such as protons to high speeds. Particles starting at point A with some initial velocity travel in semicircular orbits in the magnetic field B. The particles are accelerated to higher speeds each time they pass through the gap between the metal "dees," where there is an electric field E. (There is no electric field inside the hollow metal dees where the protons move in circular paths.) The electric field changes direction each half-cycle, owing to an ac voltage $V = V_0 \sin 2\pi ft$, so that the particles are increased in speed at each passage through the gap. (*a*) Show that the frequency f of the voltage must be $f = Bq/2\pi m$, where q is the charge on the particles and m their mass. (*b*) Show that the kinetic energy of the particles increases by $2qV_0$ each revolution, assuming that the gap is small. (*c*) If the radius of the cyclotron is 2.0 m and the magnetic field strength is 0.50 T, what will be the maximum kinetic energy of accelerated protons in MeV?

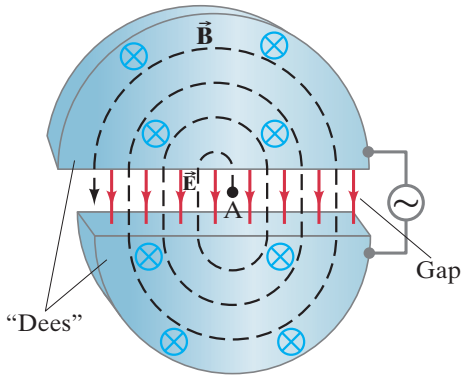


FIGURE 27–55 A cyclotron. Problem 69.

70. Magnetic fields are very useful in particle accelerators for "beam steering"; that is, the magnetic fields can be used to change the direction of the beam of charged particles without altering their speed (Fig. 27–56). Show how this could work with a beam of protons. What happens to protons that are not moving with the speed for which the magnetic field was designed? If the field extends over a region 3.9 cm wide and has a magnitude of 0.32 T, by approximately what angle θ will a beam of protons traveling at 1.8×10^6 m/s be bent?

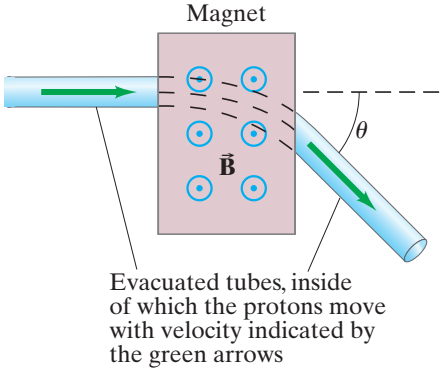


FIGURE 27–56 Problem 70.

71. The magnetic field B at the center of a circular coil of wire carrying a current I (as in Fig. 27–9) is

$$B = \frac{\mu_0 NI}{2r},$$

where N is the number of loops in the coil and r is its radius. Imagine a simple model in which the Earth's magnetic field of about $1\,\text{G}\,(= 1 \times 10^{-4}\,\text{T})$ near the poles is produced by a single current loop around the equator. Roughly estimate the current this loop would carry.

72. A square loop of aluminum wire is 20.0 cm on a side. It is to carry 15.0 A and rotate in a uniform 1.35-T magnetic field as shown in Fig. 27–57. (*a*) Determine the minimum diameter of the wire so that it will not fracture from tension or shear. Assume a safety factor of 8. (See Table 12–2.) (*b*) What is the resistance of a single loop of this wire?

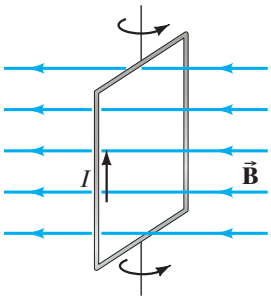


FIGURE 27–57 Problem 72.

73. A sort of "projectile launcher" is shown in Fig. 27–58. A large current moves in a closed loop composed of fixed rails, a power supply, and a very light, almost frictionless plate (blue-green) touching the rails. A magnetic field is perpendicular to the plane of the circuit. If the bar has a length $\ell = 36$ cm, a mass of 45 g, and is placed in a field of 1.5 T, what constant current flow is needed to accelerate the bar from rest to 28 m/s in a distance of 1.0 m? In what direction must the magnetic field point?

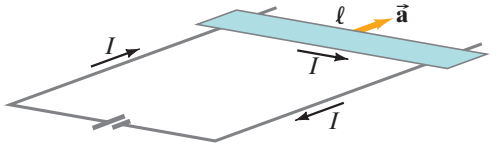


FIGURE 27–58 Problem 73.

74. (*a*) What value of magnetic field would make a beam of electrons, traveling to the west at a speed of 4.2×10^6 m/s, go undeflected through a region where there is a uniform electric field of 12,000 V/m pointing south? (*b*) What is the direction of the magnetic field if it is perpendicular to the electric field? (*c*) What is the frequency of the circular orbit of the electrons if the electric field is turned off?

75. A proton follows a spiral path through a gas in a uniform magnetic field of 0.010 T, perpendicular to the plane of the spiral, as shown in Fig. 27–59. In two successive loops, at points P and Q, the radii are 10.0 mm and 8.5 mm, respectively. Calculate the change in the kinetic energy of the proton as it travels from P to Q, in eV.

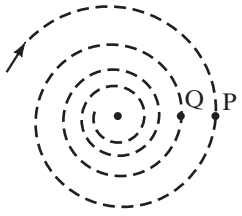


FIGURE 27–59 Problem 75.

76. Suppose the Earth's magnetic field at the equator has magnitude $0.50 \times 10^{-4}\,\text{T}$ and a northerly direction at all points. Estimate the speed a singly ionized uranium ion $(m = 238\,\text{u},\ q = e)$ would need to circle the Earth 5.0 km above the equator. Can you ignore gravity? [Ignore relativity.]

77. The power cable for an electric trolley (Fig. 27–60) carries a horizontal current of 330 A toward the east. The Earth's magnetic field has a strength $5.0 \times 10^{-5}\,\text{T}$ and makes an angle of dip of 22° at this location. Calculate the magnitude and direction of the magnetic force on a 15-m length of this cable.

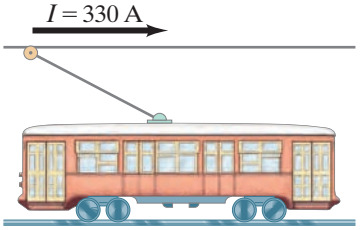


FIGURE 27–60 Problem 77.

78. **Zeeman effect.** In the Bohr model of the hydrogen atom, the electron is held in its circular orbit of radius r about its proton nucleus by electrostatic attraction. If the atoms are placed in a weak magnetic field $\vec{\mathbf{B}}$, the frequency of electrons revolving (in a circle) in a plane perpendicular to $\vec{\mathbf{B}}$ is changed by an amount

$$\Delta f = \pm \frac{eB}{4\pi m}$$

where e and m are the charge and mass of an electron. (*a*) Derive this result, assuming the force due to $\vec{\mathbf{B}}$ is much less than that due to electrostatic attraction of the nucleus. (*b*) What does the $\pm$ sign indicate?

79. The net force on a current loop whose face is perpendicular to a uniform magnetic field is zero (Fig. 27–61) because contributions to the net force from opposite sides of the loop cancel. However, if the field varies in magnitude from one side of the loop to the other, then there can be a net force on the loop. Consider a square loop with sides whose length is a, located with one side at $x = b$ in the xy plane (Fig. 27–61). A magnetic field is directed along z, with a magnitude in the region shown that varies with x according to

$$B = B_0\left(1 + \frac{x}{b}\right).$$

If the current in the loop circulates counterclockwise (that is, the magnetic dipole moment of the loop is along the z axis), find an expression for the net force on the loop.

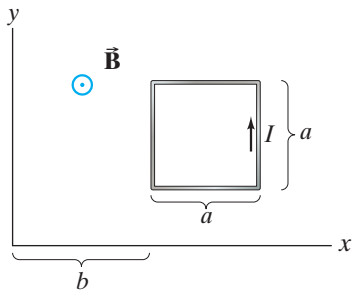


FIGURE 27–61 Problem 79.

80. An electron is moving north at a constant speed of 3.0×10^4 m/s. (*a*) In what direction should an electric field point if the electron is to be accelerated to the east? (*b*) In what direction should a magnetic field point if the electron is to be accelerated to the west? (*c*) If the electric field of part *a* has a strength of 380 V/m, what magnetic field (magnitude and direction) will produce zero net force on the electron? (*d*) If the electron in part *c* is moving faster than 3.0×10^4 m/s, in which direction will it be accelerated? What if it is moving slower than 3.0×10^4 m/s? (*e*) Now consider electrons that move perpendicular to both a magnetic field and to an electric field, which are perpendicular to each other. If only electrons with speeds of 5.5×10^4 m/s go straight through undeflected, what is the ratio of the magnitudes of electric field to magnetic field? Without knowing the value of the electric field, can you know the value of the magnetic field?

81. The rectangular loop of wire shown in Fig. 27–23 has mass m and carries current I. Show that if the loop is oriented at an angle $\theta \ll 1$ (in radians), then when it is released it will execute simple harmonic motion about $\theta = 0$. Calculate the period of the motion.

82. A particle of charge q moves in a circular path of radius r perpendicular to a uniform magnetic field B. Determine its linear momentum in terms of the quantities given.

83. An airplane has acquired a net charge of 1160 μC. If the Earth's magnetic field of 5.0×10^{-5} T is perpendicular to the airplane's velocity of magnitude 120 m/s, determine the force on the airplane.

84. Muons (mass = 207 times the electron mass, same charge $-e$) are accelerated horizontally by 65 kV. They then pass through a uniform magnetic field B for a distance of 3.8 cm, which deflects them upward so they reach a detector 22 cm away, 11 cm above their original direction. Estimate the value of B.

85. In a simple device for measuring the magnitude B of a magnetic field, a conducting rod (length $d = 1.0$ m, mass $m = 150$ g) hangs from a friction-free pivot and is oriented so that its axis of rotation is aligned with the direction of the magnetic field to be measured. Thin flexible wires (which exert negligible force on the rod) carry a current $I = 12$ A, which causes the rod to deflect an angle θ with respect to the vertical, where it remains at rest (Fig. 27–62). (*a*) Is the current flowing upward (toward the pivot) or downward in Fig. 27–62? (*b*) If $\theta = 13°$, determine B. (*c*) What is the largest magnetic field magnitude that can be measured using this device?

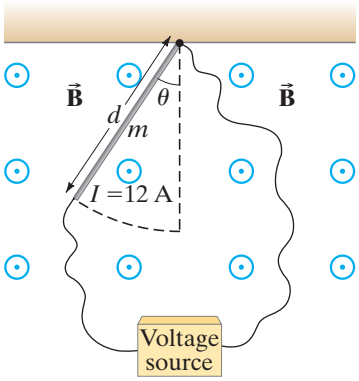
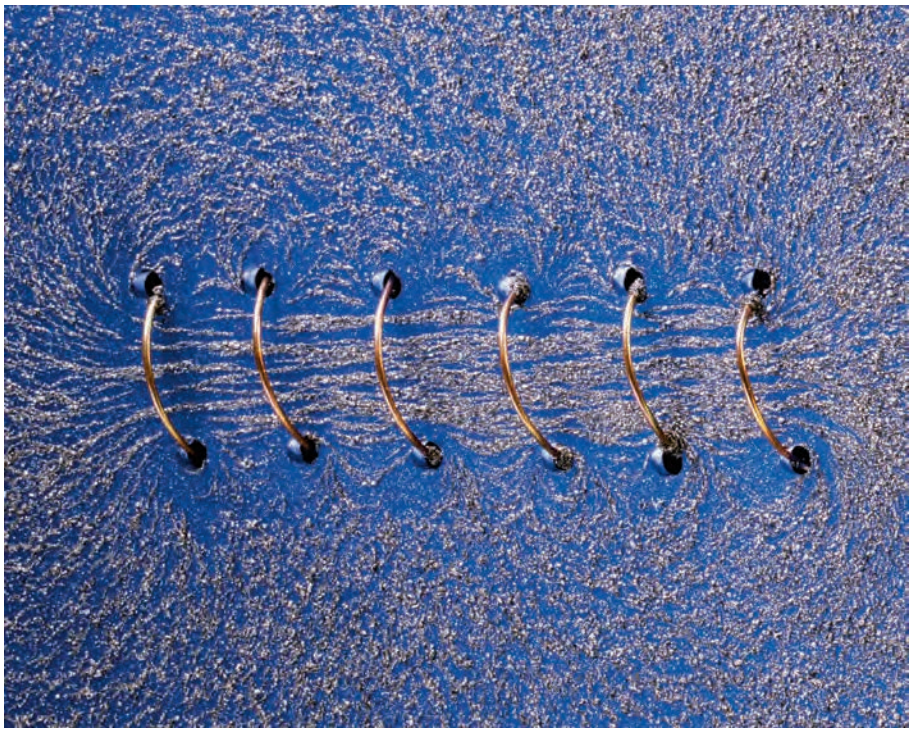


FIGURE 27–62 Problem 85.

ANSWERS TO EXERCISES

A: Near the poles, where the field lines are closer together.
B: Circles, pointing counterclockwise.
C: (*b*), (*c*), (*d*).
D: 0.15 N.
E: (*b*), (*c*), (*d*).
F: Zero.
G: Smaller.
H: Negative; the helical path would rotate in the opposite direction (still going to the right).
I: (*d*).



A long coil of wire with many closely spaced loops is called a solenoid. When a long solenoid carries an electric current, a nearly uniform magnetic field is produced within the loops as suggested by the alignment of the iron filings in this photo. The magnitude of the field inside a solenoid is readily found using Ampère's law, one of the great general laws of electromagnetism, relating magnetic fields and electric currents. We examine these connections in detail in this Chapter, as well as other means for producing magnetic fields.

CHAPTER 28

Sources of Magnetic Field

CHAPTER-OPENING QUESTION—Guess now!

Which of the following will produce a magnetic field?

(a) An electric charge at rest.
(b) A moving electric charge.
(c) An electric current.
(d) The voltage of a battery not connected to anything.
(e) Any piece of iron.
(f) A piece of any metal.

In the previous Chapter, we discussed the effects (forces and torques) that a magnetic field has on electric currents and on moving electric charges. We also saw that magnetic fields are produced not only by magnets but also by electric currents (Oersted's great discovery). It is this aspect of magnetism, the production of magnetic fields, that we discuss in this Chapter. We will now see how to determine the magnetic field for some simple situations, and discuss some general relations between magnetic fields and their sources, electric current. Most elegant is Ampère's law. We also study the Biot-Savart law, which can be very helpful for solving practical problems.

CONTENTS

28 – 1 Magnetic Field Due to a Straight Wire
28 – 2 Force between Two Parallel Wires
28 – 3 Definitions of the Ampere and the Coulomb
28 – 4 Ampère's Law
28 – 5 Magnetic Field of a Solenoid and a Toroid
28 – 6 Biot-Savart Law
28 – 7 Magnetic Field Due to a Single Moving Charge
28 – 8 Magnetic Materials—Ferromagnetism
28 – 9 Electromagnets and Solenoids—Applications
28–10 Magnetic Fields in Magnetic Materials; Hysteresis
*28–11 Paramagnetism and Diamagnetism

28–1 Magnetic Field Due to a Straight Wire

We saw in Section 27–2 that the magnetic field due to the electric current in a long straight wire forms circles with the wire at the center (Fig. 28–1). You might expect that the field strength at a given point would be greater if the current flowing in the wire were greater; and that the field would be less at points farther from the wire. This is indeed the case. Careful experiments show that the magnetic field B due to the current in a long straight wire is directly proportional to the current I in the wire and inversely proportional to the distance r from the wire:

$$B \propto \frac{I}{r}.$$

This relation is valid as long as r, the perpendicular distance to the wire, is much less than the distance to the ends of the wire (i.e., the wire is long).

The proportionality constant is written as $\mu_0/2\pi$, so

$$B = \frac{\mu_0}{2\pi}\frac{I}{r}. \qquad \text{[near a long straight wire]} \quad \textbf{(28–1)}$$

The value of the constant μ_0, which is called the **permeability of free space**,[†] is $\mu_0 = 4\pi \times 10^{-7}\,\text{T}\cdot\text{m/A}$.

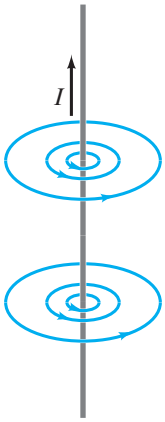
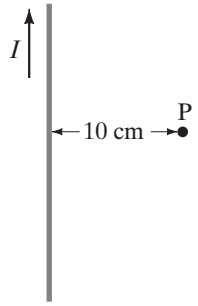


FIGURE 28–1 Same as Fig. 27–8c. Magnetic field lines around a long straight wire carrying an electric current I.

FIGURE 28–2 Example 28–1.



EXAMPLE 28–1 Calculation of $\vec{B}$ near a wire. An electric wire carries a dc current of 25 A vertically upward. What is the magnetic field due to this current at a point P, 10 cm due north of the wire (Fig. 28–2)?

APPROACH We assume the wire is much longer than the 10-cm distance to the point P so we can apply Eq. 28–1.

SOLUTION According to Eq. 28–1:

$$B = \frac{\mu_0 I}{2\pi r} = \frac{(4\pi \times 10^{-7}\,\text{T}\cdot\text{m/A})(25\,\text{A})}{(2\pi)(0.10\,\text{m})} = 5.0 \times 10^{-5}\,\text{T},$$

or 0.50 G. By the right-hand rule (Table 27–1, page 791), the magnetic field due to the current points to the west (into the page in Fig. 28–2) at point P.

NOTE The magnetic field at P produced by this current has about the same magnitude as Earth's, so a compass at P would not point north but to the northwest.

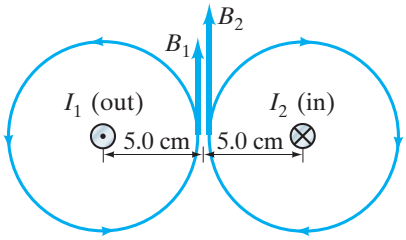
NOTE Most electrical wiring is ac, which changes too rapidly to affect a compass. Besides, even for dc, wiring in buildings consists of cables with two wires in each cable. Since the two wires carry current in opposite directions, their magnetic fields cancel to a large extent, but may still affect sensitive electronic devices.

CAUTION

A compass, near a dc current, may not point north

EXERCISE A In Chapter 25 we saw that a typical lightning bolt produces a 100-A current for 0.2 s (Example 25–10). Estimate the magnetic field 10 m from a lightning bolt. Would it have a significant effect on a compass?

FIGURE 28–3 Example 28–2. Wire 1 carrying current I_1 out towards us, and wire 2 carrying current I_2 into the page, produce magnetic fields whose lines are circles around their respective wires.



EXAMPLE 28–2 Magnetic field midway between two currents. Two parallel straight wires 10.0 cm apart carry dc currents in opposite directions (Fig. 28–3). Current $I_1 = 5.0\,\text{A}$ is out of the page, and $I_2 = 7.0\,\text{A}$ is into the page. Determine the magnitude and direction of the magnetic field halfway between the two wires.

APPROACH The magnitude of the field produced by each wire is calculated from Eq. 28–1. The direction of *each* wire's field is determined with the right-hand rule. The total field is the vector sum of the two fields at the midway point.

[†]The constant μ_0 is chosen in this way so that Ampère's law (Section 28–4), which is considered a fundamental law of physics, has a simple and elegant form. Also, in the new SI (System of Units), described in Section 1–4, the value of μ_0 is no longer exact, but does correspond to the 4π here to 10 decimal places, or 1 part in 10^{10}. For convenience, we usually will use the traditional 4π in calculations so cancellation of terms is easy.

SOLUTION The magnetic field lines due to current I_1 form circles around the wire of I_1, and right-hand-rule-1 (Fig. 27–8d) tells us they point counterclockwise around the wire. The field lines due to I_2 form circles around the wire of I_2 and point clockwise, Fig. 28–3. At the midpoint, both fields point upward in Fig. 28–3, and so add together. The midpoint is 0.050 m from each wire. From Eq. 28–1 the magnitudes of B_1 and B_2 are

$$B_1 = \frac{\mu_0 I_1}{2\pi r} = \frac{(4\pi \times 10^{-7}\,\mathrm{T\cdot m/A})(5.0\,\mathrm{A})}{2\pi(0.050\,\mathrm{m})} = 2.0 \times 10^{-5}\,\mathrm{T};$$

$$B_2 = \frac{\mu_0 I_2}{2\pi r} = \frac{(4\pi \times 10^{-7}\,\mathrm{T\cdot m/A})(7.0\,\mathrm{A})}{2\pi(0.050\,\mathrm{m})} = 2.8 \times 10^{-5}\,\mathrm{T}.$$

The total field is *up* with a magnitude of

$$B = B_1 + B_2 = 4.8 \times 10^{-5}\,\mathrm{T}.$$

EXERCISE B Suppose both I_1 and I_2 point into the page in Fig. 28–3. What then is the magnetic field midway between the two wires?

CONCEPTUAL EXAMPLE 28–3 **Magnetic field due to four wires.** Figure 28–4 shows four long parallel wires which carry equal dc currents into or out of the page as shown. In which configuration, (a) or (b), is the magnetic field greater at the center of the square?

RESPONSE It is greater in (a). The arrows illustrate the directions of the field produced by each wire; check it out, using right-hand-rule-1 to confirm these results. The net field at the center is the superposition of the four fields (which are of equal magnitude), which will point to the left in (a) and is zero in (b).

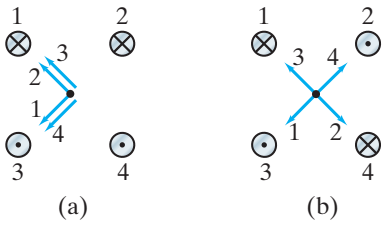


FIGURE 28–4 Example 28–3.

28–2 Force between Two Parallel Wires

We have seen that a wire carrying a current produces a magnetic field (magnitude given by Eq. 28–1 for a long straight wire). Also, a current-carrying wire feels a force when placed in a magnetic field (Section 27–3, Eq. 27–1). Thus, we expect that two current-carrying wires will exert a force on each other.

Consider two long parallel wires separated by a distance d, as in Fig. 28–5a. They carry dc currents I_1 and I_2, respectively. Each current produces a magnetic field that is "felt" by the other, so each must exert a force on the other. For example, the magnetic field B_1 produced by I_1 in Fig. 28–5 is given by Eq. 28–1, which at the location of wire 2 points into the page and has magnitude

$$B_1 = \frac{\mu_0}{2\pi}\frac{I_1}{d}.$$

See Fig. 28–5b, where the field due *only* to I_1 is shown. According to Eq. 27–2, the force F_2 exerted by B_1 on a length ℓ_2 of wire 2, carrying current I_2, has magnitude

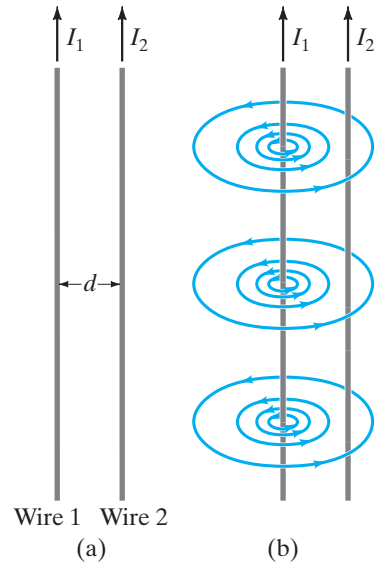
$$F_2 = I_2 \ell_2 B_1.$$

Note that the force on I_2 is due only to the field produced by I_1. Of course, I_2 also produces a field; but it does not exert a force on itself. We substitute B_1 into the formula for F_2 and find that the force on a length ℓ_2 of wire 2 is

$$F_2 = \frac{\mu_0}{2\pi}\frac{I_1 I_2}{d}\ell_2. \qquad \text{[parallel wires]} \quad \textbf{(28–2)}$$

If we use right-hand-rule-1 of Fig. 27–8d, we see that the lines of B_1 are as shown in Fig. 28–5b. Then using right-hand-rule-2 of Fig. 27–11c, we see that the force exerted on I_2 will be to the left in Fig. 28–5b. (See page 791 for right-hand-rules.)

FIGURE 28–5 (a) Two parallel conductors carrying dc currents I_1 and I_2. (b) Magnetic field $\vec{\mathbf{B}}_1$ produced by I_1. (Field produced by I_2 is not shown.) $\vec{\mathbf{B}}_1$ points into page at position of I_2.



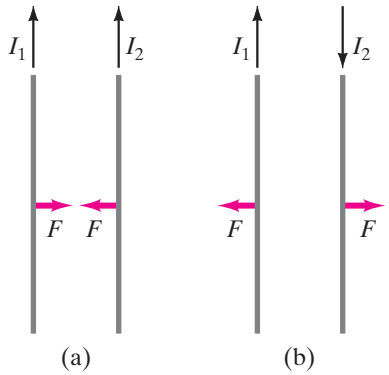


FIGURE 28–6 (a) Parallel currents in the same direction exert an attractive force on each other. (b) Antiparallel currents (in opposite directions) exert a repulsive force on each other.

That is, I_1 exerts an attractive force on I_2 (Fig. 28–6a). This is true as long as the currents are in the same direction. If I_2 is in the opposite direction from I_1, right-hand-rule-2 indicates that the force is in the opposite direction. That is, I_1 exerts a repulsive force on I_2 (Fig. 28–6b).

Reasoning similar to that above shows that the magnetic field produced by I_2 exerts an equal but opposite force on I_1. We expect this to be true also from Newton's third law. Thus, as shown in Fig. 28–6, parallel currents in the same direction attract each other, whereas parallel currents in opposite directions repel.

EXAMPLE 28–4 **Force between two current-carrying wires.** The two wires of a 2.0-m-long appliance cord are 3.0 mm apart and carry a current of 8.0 A dc. Calculate the force one wire exerts on the other.

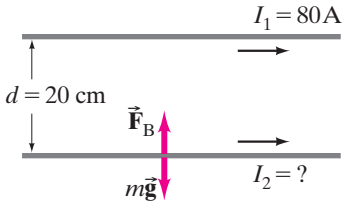
APPROACH Each wire is in the magnetic field of the other when the current is on, so we can apply Eq. 28–2.

SOLUTION Equation 28–2 gives

$$F = \frac{\mu_0}{2\pi}\frac{I_1 I_2}{d}\ell_2 = \frac{(4\pi \times 10^{-7}\,\mathrm{T\cdot m/A})(8.0\,\mathrm{A})^2(2.0\,\mathrm{m})}{(2\pi)(3.0\times 10^{-3}\,\mathrm{m})} = 8.5\times 10^{-3}\,\mathrm{N}.$$

The currents are in opposite directions (one toward the appliance, the other away from it), so the force would be repulsive and tend to spread the wires apart.

FIGURE 28–7 Example 28–5.



EXAMPLE 28–5 **Suspending a wire with a current.** A horizontal wire carries a current $I_1 = 80$ A dc. A second parallel wire 20 cm below it (Fig. 28–7) must carry how much current I_2 so that it doesn't fall due to gravity? The lower wire has a mass of 0.12 g per meter of length.

APPROACH If wire 2 is not to fall under gravity, which acts downward, the magnetic force on it must be upward. This means that the current in the two wires must be in the same direction (Fig. 28–6). We can find the current I_2 by equating the magnitudes of the magnetic force and the gravitational force on the wire.

SOLUTION The force of gravity on wire 2 is downward. For each 1.0 m of wire length, the gravitational force has magnitude

$$F = mg = (0.12\times 10^{-3}\,\mathrm{kg/m})(1.0\,\mathrm{m})(9.8\,\mathrm{m/s^2}) = 1.18\times 10^{-3}\,\mathrm{N}.$$

The magnetic force on wire 2 must be upward, and Eq. 28–2 gives

$$F = \frac{\mu_0}{2\pi}\frac{I_1 I_2}{d}\ell$$

where $d = 0.20$ m and $I_1 = 80$ A. We solve this for I_2 and set the two force magnitudes equal (letting $\ell = 1.0$ m):

$$I_2 = \frac{2\pi d}{\mu_0 I_1}\left(\frac{F}{\ell}\right) = \frac{2\pi(0.20\,\mathrm{m})}{(4\pi\times 10^{-7}\,\mathrm{T\cdot m/A})(80\,\mathrm{A})}\frac{(1.18\times 10^{-3}\,\mathrm{N/m})}{(1.0\,\mathrm{m})} = 15\,\mathrm{A}.$$

28–3 Definitions of the Ampere and the Coulomb

You may have wondered how the constant μ_0 in Eq. 28–1 could be exactly $4\pi\times 10^{-7}\,\mathrm{T\cdot m/A}$. Here is how it happened. With an older definition of the ampere, the unit of electric current, μ_0 was measured experimentally to be very close to this value. Then, until recently, μ_0 was *defined* to be exactly $4\pi\times 10^{-7}\,\mathrm{T\cdot m/A}$. This could not have been done if the ampere had been defined independently, so it was then defined in terms of the magnetic field B it produced using the defined value of μ_0.

In particular, in the traditional SI we use the force between two parallel current-carrying wires, Eq. 28–2, to define the ampere precisely. If $I_1 = I_2 = 1\,\text{A}$ exactly, and the two wires are exactly 1 m apart, then

$$\frac{F}{\ell} = \frac{\mu_0}{2\pi}\frac{I_1 I_2}{d} = \frac{(4\pi \times 10^{-7}\,\text{T}\cdot\text{m/A})}{(2\pi)}\frac{(1\,\text{A})(1\,\text{A})}{(1\,\text{m})} = 2 \times 10^{-7}\,\text{N/m},$$

exactly. Thus, *one* **ampere** *can be defined as that current flowing in each of two long parallel wires exactly 1 m apart, which results in a force of exactly* 2×10^{-7} *N per meter of length of each wire.*

This has long been the precise definition of the ampere,[†] and because it is readily reproducible, is called an "operational definition." The **coulomb** is defined in terms of the ampere as being *exactly* one ampere-second: $1\,\text{C} = 1\,\text{A}\cdot\text{s}$.

Operational Definitions

This may seem a roundabout way of defining quantities. The reason is the desire for **operational definitions**, definitions of quantities that can be readily reproducible given a definite set of operations to carry out. For example, the unit of charge, the coulomb, could be defined in terms of the force between two equal charges after defining a value for ϵ_0 or k in Eqs. 21–1 or 21–2. But to carry out an actual experiment to measure the force between two charges is difficult: any desired amount of charge is not easily obtained precisely, and charge tends to leak from objects into the air. The amount of current in a wire, on the other hand, can be varied accurately and continuously (by putting a variable resistor in a circuit). Thus the force between two current-carrying conductors is easier to measure precisely. This is why we first define the ampere, and then define the coulomb in terms of the ampere.

28–4 Ampère's Law

The relation between the current in a long straight wire and the magnetic field it produces is given by Eq. 28–1, Section 28–1. This equation is valid *only* for a long straight wire. Is there a general relation between a current in a wire of any shape and the magnetic field around it? The answer is yes: the French scientist André Marie Ampère (1775–1836) proposed such a relation shortly after Oersted's discovery. Consider any (arbitrary) closed path around a current, as shown in Fig. 28–8, and imagine this path as being made up of short segments each of length $\Delta\ell$. First, we take the product of the length of each segment times the component of magnetic field $\vec{\mathbf{B}}$ parallel to that segment (call this component $B_\parallel$). If we now sum all these terms, according to Ampère, the result will be equal to μ_0 times the net current I_{encl} that passes through the surface enclosed by the path:

$$\sum B_\parallel \,\Delta\ell = \mu_0 I_{\text{encl}}.$$

The lengths $\Delta\ell$ are chosen small enough so that $B_\parallel$ is essentially constant along each length. The sum $(=\Sigma)$ must be made over a *closed path*; and I_{encl} is the net current passing through the surface bounded by this closed path (orange in Fig. 28–8). In the limit $\Delta\ell \to 0$, this relation becomes

$$\oint \vec{\mathbf{B}}\cdot d\vec{\ell} = \mu_0 I_{\text{encl}}, \qquad \textbf{(28–3)}$$

where $d\vec{\ell}$ is an infinitesimal length vector and the vector dot product assures that the parallel component of $\vec{\mathbf{B}}$ is taken. Equation 28–3 is known as **Ampère's law**. The integrand in Eq. 28–3 is taken around a closed path (a line integral) and I_{encl} is the current passing through the space enclosed by the chosen path or loop.

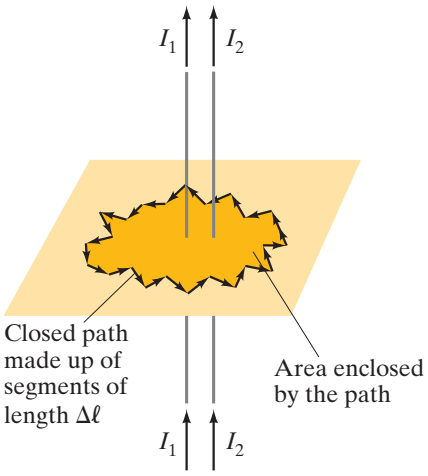


FIGURE 28–8 Arbitrary path enclosing electric currents, for Ampère's law. The path is broken down into segments of length $\Delta\ell$. The total current enclosed by the path shown is $I_{\text{encl}} = I_1 + I_2$.

AMPÈRE'S LAW

[†]In the new SI (Section 1–4), 1 ampere is still 1 coulomb/s, and the Coulomb is defined by assigning the electron the exact value of $e = 1.602176634 \times 10^{-19}$ C.

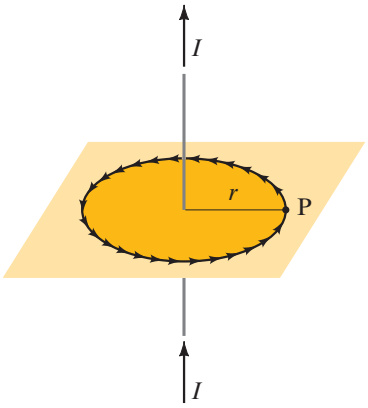


FIGURE 28–9 Circular path of radius r.

We can check Ampère's law by applying it to the simple case of a long straight wire carrying a current I, which we've already examined, and which served as an inspiration for Ampère himself. Let us find the magnitude of $\vec{\mathbf{B}}$ at point P, a distance r from the wire in Fig. 28–9. The magnetic field lines are circles with the wire at their center. As the path to be used in Eq. 28–3, we choose a convenient one: a circle of radius r, because at any point on this path, $\vec{\mathbf{B}}$ will be tangent to this circle. All points on the path are the same distance from the wire, so by *symmetry* we expect B to have the same magnitude at each point. For any short segment of the circle (Fig. 28–9), $\vec{\mathbf{B}}$ will be parallel to that segment, so $B_{\|} = B$. Then (setting $I_{\text{encl}} = I$) Ampère's law gives

$$\mu_0 I = \oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$$

$$= \oint B\, d\ell = B \oint d\ell = B(2\pi r),$$

where $\oint d\ell = 2\pi r$, the circumference of the circle. We solve for B and obtain

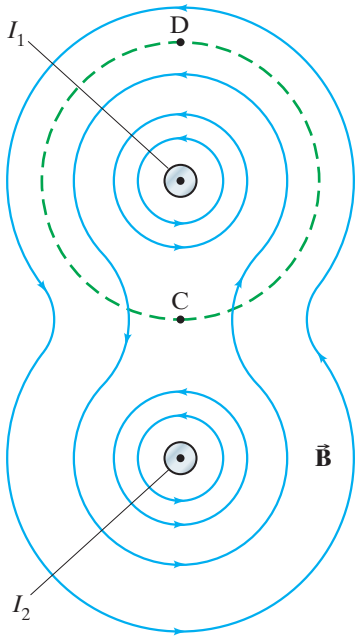
$$B = \frac{\mu_0 I}{2\pi r}.$$

This is just Eq. 28–1 for the magnetic field near a long straight wire, so Ampère's law agrees with experiment in this case.

A great many experiments indicate that Ampère's law is valid in general. As with Gauss's law for the electric field, its practical value to calculate the magnetic field is limited, however, mainly to simple or symmetric situations. Its importance is that it relates the magnetic field to the current in a direct and mathematically elegant way. Ampère's law is considered one of the basic laws of electricity and magnetism. It is valid for any situation where the currents and fields are steady and not changing in time, and no magnetic materials are present.

We now can see why the constant in Eq. 28–1 is written $\mu_0/2\pi$: because then only μ_0 appears in Eq. 28–3, rather than, say, $2\pi C$ if we had used C in Eq. 28–1. In this way, the more fundamental equation, Ampère's law, has the simpler form.

FIGURE 28–10 Magnetic field lines around two long parallel wires whose equal currents, I_1 and I_2, are coming out of the paper toward the viewer.



It should be noted that the $\vec{\mathbf{B}}$ in Ampère's law is not necessarily due only to the current I_{encl}. Ampère's law, like Gauss's law for the electric field, is valid in general. $\vec{\mathbf{B}}$ is the field at each point in space along the chosen path due to all sources—including the current I enclosed by the path, but also due to any other sources. For example, the field surrounding two parallel current-carrying wires is the vector sum of the fields produced by each, and the field lines are shown in Fig. 28–10. If the path chosen for the integral (Eq. 28–3) is a circle centered on one of the wires with radius less than the distance between the wires (the dashed line in Fig. 28–10), only the current (I_1) in the encircled wire is included on the right side of Eq. 28–3. $\vec{\mathbf{B}}$ on the left side of the equation must be the total $\vec{\mathbf{B}}$ at each point due to both wires. Note also that $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$ for the path shown in Fig. 28–10 is the same whether the second wire is present or not (in both cases, it equals $\mu_0 I_1$ according to Ampère's law). How can this be? It can be so because the fields due to the two wires partially cancel one another at some points between them, such as point C in the diagram $(\vec{\mathbf{B}} = 0$ at a point midway between the wires if $I_1 = I_2)$; at other points, such as D in Fig. 28–10, the fields add together to produce a larger field. In the *sum*, $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$, these effects just balance so that $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_1$, whether the second wire is there or not. The integral $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$ will be the same in each case, even though $\vec{\mathbf{B}}$ will not be the same at every point for each of the two cases.

EXAMPLE 28–6 **Field inside and outside a wire.** A long straight cylindrical wire conductor of radius R carries a dc current I of uniform current density in the conductor. Determine the magnetic field due to this current at (a) points outside the conductor ($r > R$), and (b) points inside the conductor ($r < R$). See Fig. 28–11. Assume that r, the radial distance from the axis, is much less than the length of the wire. (c) If $R = 2.0\,\text{mm}$ and $I = 60\,\text{A}$, what is B at $r = 1.0\,\text{mm}$, $r = 2.0\,\text{mm}$, and $r = 3.0\,\text{mm}$?

APPROACH We can use *symmetry*: Because the wire is long, straight, and cylindrical, we expect from symmetry that the magnetic field must be the same at all points that are the same distance from the center of the conductor. There is no reason why any such point should have preference over others at the same distance from the wire (they are physically equivalent), so B must have the same value at all points the same distance from the center. This is what we mean by **symmetry**. We also expect $\vec{\mathbf{B}}$ to be tangent to circles around the wire (Fig. 28–1), so we choose a circular path of integration as we did in Fig. 28–9.

SOLUTION (a) We apply Ampère's law, integrating around a circle ($r > R$) centered on the wire (Fig. 28–11a), and then $I_{\text{encl}} = I$:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = B(2\pi r) = \mu_0 I_{\text{encl}}$$

or

$$B = \frac{\mu_0 I}{2\pi r}, \qquad [r > R]$$

which is the same result as for a thin wire, Eq. 28–1.

(b) Inside the wire ($r < R$), we again choose a circular path concentric with the cylinder; we expect $\vec{\mathbf{B}}$ to be tangential to this path, and again, because of the symmetry, it will have the same magnitude at all points on the circle. The current enclosed in this case is less than I by a factor of the ratio of the areas:

$$I_{\text{encl}} = I\frac{\pi r^2}{\pi R^2}.$$

So Ampère's law gives

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_{\text{encl}}$$

$$B(2\pi r) = \mu_0 I\left(\frac{\pi r^2}{\pi R^2}\right)$$

so

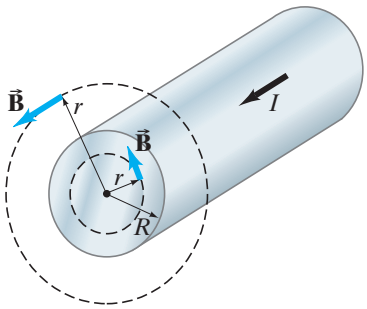
$$B = \left(\frac{\mu_0 I}{2\pi R^2}\right) r. \qquad [r < R]$$

The field is zero at the center of the conductor and increases linearly with r until $r = R$; beyond $r = R$, B decreases as $1/r$. This is shown in Fig. 28–11b. Note that these results are valid only at points close to the center of the conductor as compared to its length. For a current to flow, there must be connecting wires (to a battery, say), and the field due to these conducting wires, if not very far away, will destroy the assumed symmetry.

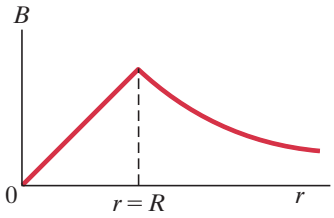
(c) At $r = 2.0\,\text{mm}$, the surface of the wire, $r = R$, so

$$B = \frac{\mu_0 I}{2\pi R} = \frac{(4\pi \times 10^{-7}\,\text{T}\cdot\text{m/A})(60\,\text{A})}{(2\pi)(2.0 \times 10^{-3}\,\text{m})} = 6.0 \times 10^{-3}\,\text{T}.$$

We saw in (b) that inside the wire B is linear in r. So at $r = 1.0\,\text{mm}$, B will be half what it is at $r = 2.0\,\text{mm}$ or $3.0 \times 10^{-3}\,\text{T}$. Outside the wire, B falls off as $1/r$, so at $r = 3.0\,\text{mm}$ it will be two-thirds as great as at $r = 2.0\,\text{mm}$, or $B = 4.0 \times 10^{-3}\,\text{T}$. To check, we use our result in (a), $B = \mu_0 I/2\pi r$, which gives the same result.



(a)



(b)

FIGURE 28–11 Magnetic field inside and outside a cylindrical conductor (Example 28–6).

CAUTION
Connecting wires can destroy assumed symmetry

PHYSICS APPLIED
Coaxial cable (shielding)

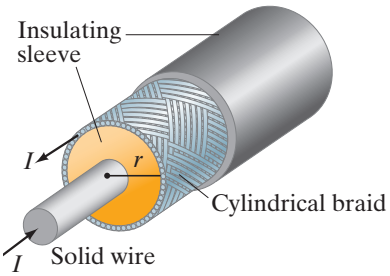


FIGURE 28–12 Coaxial cable. Example 28–7.

CONCEPTUAL EXAMPLE 28–7 **Coaxial cable.** A *coaxial cable* is a single wire surrounded by a cylindrical metallic braid, as shown in Fig. 28–12. The two conductors are separated by an insulator. The central wire carries current to the other end of the cable, and the outer braid carries the return current and is usually considered ground. Describe the magnetic field (*a*) in the space between the conductors, and (*b*) outside the cable.

RESPONSE (*a*) In the space between the conductors, we can apply Ampère's law for a circular path around the center wire, just as we did for the case shown in Figs. 28–9 and 28–11. The magnetic field lines will be concentric circles centered on the center of the wire, and the magnitude is given by Eq. 28–3 or 28–1,

$$B = \mu_0 I/2\pi r$$

where r is the distance from the center. The current in the outer conductor has no bearing on this result (it is not enclosed by the integral's loop).
(*b*) Outside the cable, we can draw a similar circular path, for we expect the field to have the same cylindrical symmetry. Now, however, there are two currents enclosed by the path, and they add up to zero. The field outside the cable is zero:

$$B = 0.$$

A key feature of coaxial cables is that they are self-shielding: no stray magnetic fields exist outside the cable. The outer cylindrical conductor also shields external electric fields from coming in (see also Example 21–15). This makes them ideal for carrying signals near sensitive equipment. Audiophiles use coaxial cables between stereo equipment components and to the loudspeakers.

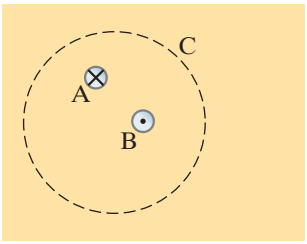
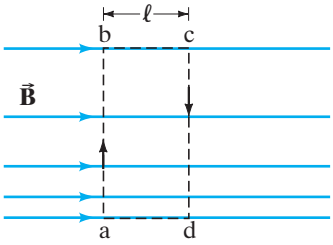


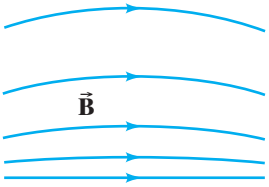
FIGURE 28–13 Exercise C.

EXERCISE C In Fig. 28–13, A and B are wires each carrying a 3.0-A current but in opposite directions. On the circle C, which statement is true? (*a*) $B = 0$; (*b*) $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = 0$; (*c*) $B = 3\mu_0$; (*d*) $B = -3\mu_0$; (*e*) $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = 6\mu_0$.

FIGURE 28–14 Example 28–8.



(a) (impossible)



(b) (possible)

EXAMPLE 28–8 **A nice use for Ampère's law.** Use Ampère's law to show that in any region of space where there are no currents the magnetic field cannot be both unidirectional and nonuniform as shown in Fig. 28–14a.

APPROACH The wider spacing of lines near the top of Fig. 28–14a indicates the field $\vec{\mathbf{B}}$ has a smaller magnitude at the top than it does lower down. We apply Ampère's law to the rectangular path abcd shown dashed in Fig. 28–14a.

SOLUTION Because no current is enclosed by the chosen path, Ampère's law gives $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$ is *zero*. The integral along sections ab and cd is zero, since $\vec{\mathbf{B}} \perp d\vec{\boldsymbol{\ell}}$. But along bc and da

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = B_{\text{bc}}\ell - B_{\text{da}}\ell = (B_{\text{bc}} - B_{\text{da}})\ell,$$

which is *not* zero since the field B_{bc} along the path bc is less than the field B_{da} along path da. Hence we have a contradiction: $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$ cannot be both zero (since $I = 0$) and nonzero. Thus we have shown that a nonuniform unidirectional field is not consistent with Ampère's law. A nonuniform field whose direction also changes, as in Fig. 28–14b, is consistent with Ampère's law (convince yourself this is so), and possible. The fringing of a permanent magnet's field (Fig. 27–7) has this shape.

PROBLEM SOLVING

Ampère's Law

1. Ampère's law, like Gauss's law, is always a valid statement. But as a calculation tool it is limited mainly to systems with a high degree of symmetry. The first step in applying Ampère's law is to identify useful **symmetry**.
2. Choose an integration path that reflects the symmetry (see the Examples). Search for paths where B has constant magnitude along the entire path or along segments of the path. Make sure your integration path passes through the point where you wish to evaluate the magnetic field.
3. Use symmetry to determine the direction of $\vec{\mathbf{B}}$ along the integration path. With a smart choice of path, $\vec{\mathbf{B}}$ will be either parallel or perpendicular to the path.
4. Determine the enclosed current, I_{encl}. Be careful with signs. Let the fingers of your right hand curl along the direction of $\vec{\mathbf{B}}$ so that your thumb shows the direction of positive current. If you have a solid conductor and your integration path does not enclose the full current, you can use the current density (current per unit area) multiplied by the enclosed area (as in Example 28–6).

28–5 Magnetic Field of a Solenoid and a Toroid

A long coil of wire consisting of many loops (or turns) of wire carrying a dc current is called a **solenoid**. The electric current in each loop produces a magnetic field as was shown in Fig. 27–9, and the total magnetic field is the sum of the fields due to each current loop, Fig. 28–15a. The direction of the magnetic field produced by the solenoid at any point is found using right-hand-rule-1, Fig. 27–8d. In Fig. 28–15a, we see the field due to a solenoid when the loops are far apart. Near each loop of wire, the field lines are very nearly circles as for a straight wire. Between any two loops, the fields due to each loop tend to cancel. Toward the center of the solenoid, the fields add up to give a field that can be large and fairly uniform. For a long solenoid with closely packed loops, the field is nearly uniform and parallel to the solenoid axis within the entire cross section, as shown in Fig. 28–15b. The field outside the solenoid is very small compared to the field inside, except near the ends. Note that the same number of field lines that are concentrated inside the solenoid, spread out into the vast open space outside, so the field is much weaker outside.

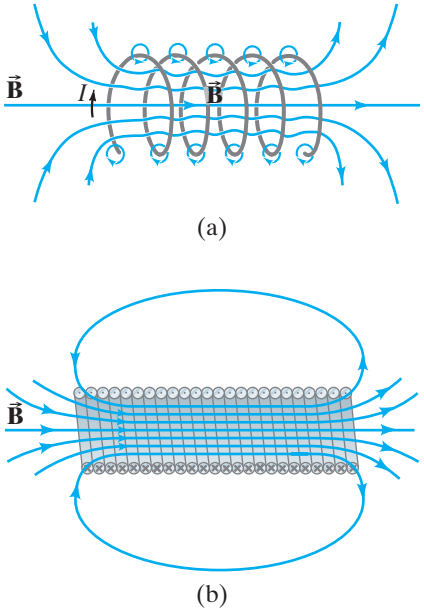


FIGURE 28–15 Magnetic field due to a solenoid: (a) a few loosely spaced loops; (b) for many closely spaced loops, the field is nearly uniform.

We now use Ampère's law to determine the magnetic field inside a very long (ideally, infinitely long) closely packed solenoid. We choose the path abcd shown in red in Fig. 28–16, far from either end, for applying Ampère's law. We will consider this path as made up of four segments, the sides of the rectangle: ab, bc, cd, da. Then the left side of Eq. 28–3, Ampère's law, becomes

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \int_a^b \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} + \int_b^c \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} + \int_c^d \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} + \int_d^a \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}.$$

The field outside the solenoid is so small as to be negligible compared to the field inside. So we set the first term in this sum equal to zero. Furthermore, $\vec{\mathbf{B}}$ is perpendicular to the segments bc and da inside the solenoid, and is nearly zero between and outside the wire loops, so these terms too are zero.

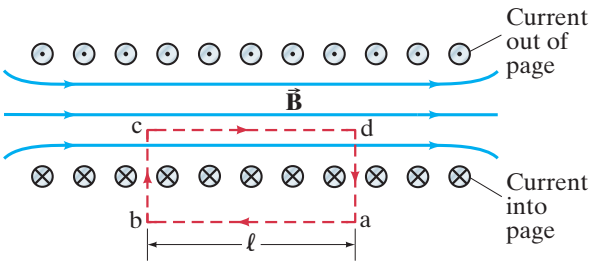


FIGURE 28–16 Cross-sectional view into a solenoid. The magnetic field inside is straight except at the ends. Red dashed lines indicate the path chosen for use in Ampère's law. ⊙ and ⊗ are electric current direction (in the wire loops) out of the page and into the page.

Therefore we have reduced the integral to the segment cd where $\vec{\mathbf{B}}$ is the nearly uniform field inside the solenoid, and is parallel to $d\vec{\boldsymbol{\ell}}$, so

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \int_c^d \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = B\ell,$$

where ℓ is the length cd. Now, for the right side of Ampère's law, Eq. 28–3, we determine the current enclosed by our path abcd. If a current I flows in the wire of the solenoid, the total current enclosed by our path abcd is NI where N is the number of loops our path encircles (five in Fig. 28–16). Thus Ampère's law gives us

$$B\ell = \mu_0 NI.$$

If we let $n = N/\ell$ be the *number of loops per unit length*, then

$$B = \mu_0 nI. \qquad \text{[solenoid]} \quad \textbf{(28–4)}$$

This is the magnitude of the magnetic field within a solenoid. Note that B depends only on the number of loops per unit length, n, and the current I. The field does not depend on position within the solenoid, so B is uniform. This is strictly true only for an infinite solenoid, but it is a good approximation for tightly wound real ones for points not close to the ends.

EXERCISE D A thin 10-cm-long solenoid used for fast electromechanical switching has 400 loops of wire and carries a current of 2.0 A. Estimate the field inside near the center.

The *direction* of the magnetic field inside a solenoid is found using right-hand-rule-1, Fig. 27–8d (see also Figs. 27–9 and 27–10), and is as shown for a tightly wound solenoid in Fig. 28–16.

A close look at Fig. 28–15 shows that the field outside of a solenoid is much like that of a bar magnet, Fig. 27–4. Indeed, a solenoid acts like a magnet, with one end acting as a north pole and the other as a south pole, depending on the direction of the current in the loops. Since magnetic field lines leave the north pole of a magnet, the north pole of the solenoid in Fig. 28–15b or 28–16 is on the right.

Solenoids have many practical applications, and we discuss some of them later in this Chapter, Section 28–9.

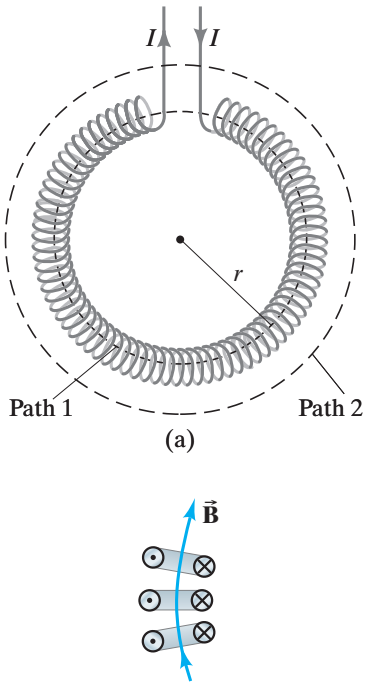


FIGURE 28–17 (a) A toroid. (b) A section of the toroid showing direction of the current for three loops: ⊙ means current toward you, ⊗ means current away from you.

EXAMPLE 28–9 **Toroid.** A toroid is like a solenoid bent into the shape of a circle as shown in Fig. 28–17a. Use Ampère's law to determine the magnetic field (*a*) inside the toroid's coil, and (*b*) outside the toroid.

APPROACH The magnetic field lines inside the toroid will be circles concentric with the toroid. (If you think of the toroid as a solenoid bent into a circle, the field lines bend along with the solenoid.) The direction of $\vec{\mathbf{B}}$ is clockwise. As our path of integration, we choose one of the field lines at radius r inside the toroid as shown by the dashed line labeled "path 1" in Fig. 28–17a. We make this choice to use the *symmetry* of the situation, so B will be tangent to the path and will have the same magnitude at all points along the path. (B is not necessarily the same across the whole cross section of the toroid.) This chosen path encloses *all* the current loops; if there are N loops, each carrying current I, then $I_{\text{encl}} = NI$.

SOLUTION (*a*) Ampère's law applied along this path 1 gives

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_{\text{encl}}$$

$$B(2\pi r) = \mu_0 NI,$$

where N is the total number of loops and I is the current in each of the loops. Thus

$$B = \frac{\mu_0 NI}{2\pi r}.$$

The magnetic field B is not uniform within the toroid: it is largest along the inner edge (where r is smallest) and smallest at the outer edge. However, if the toroid is large, but thin (so that the difference between the inner and outer radii is small compared to the average radius), the field will be essentially uniform within the toroid. In this case, the formula for B reduces to that for a straight solenoid $B = \mu_0 nI$ where $n = N/(2\pi r)$ is the number of loops per unit length.

(*b*) Outside the toroid, we choose our path of integration to be a circle concentric with the toroid, "path 2" in Fig. 28–17a. This path encloses N loops carrying current I in one direction and N loops carrying the same current in the opposite direction. (Figure 28–17b shows the directions of the current for the parts of the loop on the inside and outside of the toroid.) Thus the net current enclosed by path 2 is zero. For a very tightly packed toroid, all points on path 2 are equidistant from the toroid and equivalent, so we expect B to be the same at all points along the path. Hence, Ampère's law gives

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_{\text{encl}}$$

$$B(2\pi r) = 0$$

or

$$B = 0.$$

For a path taken at a radius *smaller* than the inner edge of the toroid (in the "hole of the donut"), $I_{\text{encl}} = 0$ and $B = 0$. So there is no field exterior to a very tightly wound toroid. The magnetic field is all inside the loops.

28–6 Biot-Savart Law

The usefulness of Ampère's law for determining the magnetic field $\vec{\mathbf{B}}$ due to particular electric currents is restricted to situations where the symmetry of the given currents allows us to evaluate $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$ readily. This does not, of course, invalidate Ampère's law nor does it reduce its fundamental importance. Recall the electric case, where Gauss's law is considered fundamental but is limited in its use for actually calculating $\vec{\mathbf{E}}$. We must often determine the electric field $\vec{\mathbf{E}}$ by another method: summing over contributions due to infinitesimal charge elements dq via Coulomb's law: $dE = (1/4\pi\epsilon_0)(dq/r^2)$. A magnetic equivalent to this infinitesimal form of Coulomb's law would be helpful for currents that do not have great symmetry. Such a law was developed by Jean Baptiste Biot (1774–1862) and Felix Savart (1791–1841) shortly after Oersted's discovery in 1820 that a current produces a magnetic field.

According to Biot and Savart, a current I flowing in any path can be considered as many tiny (infinitesimal) current elements, such as in the wire of Fig. 28–18. If $d\vec{\boldsymbol{\ell}}$ represents any infinitesimal length along which the current is flowing, then the magnetic field, $d\vec{\mathbf{B}}$, at any point P in space, due to this element of current, is given by

$$d\vec{\mathbf{B}} = \frac{\mu_0 I}{4\pi} \frac{d\vec{\boldsymbol{\ell}} \times \hat{\mathbf{r}}}{r^2}, \qquad \textbf{(28–5)} \quad \textit{Biot-Savart law}$$

where $\vec{\mathbf{r}}$ is the displacement vector from the element $d\vec{\boldsymbol{\ell}}$ to the point P, and $\hat{\mathbf{r}} = \vec{\mathbf{r}}/r$ is the unit vector (magnitude = 1) in the direction of $\vec{\mathbf{r}}$ (see Fig. 28–18).

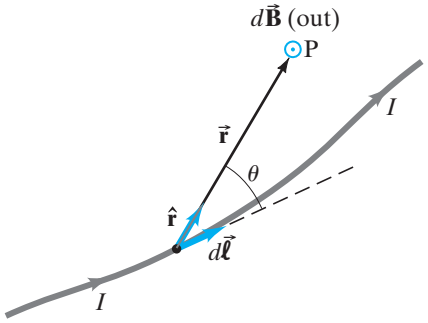


FIGURE 28–18 Biot-Savart law: the field at P due to current element $Id\vec{\boldsymbol{\ell}}$ is $d\vec{\mathbf{B}} = (\mu_0 I/4\pi)(d\vec{\boldsymbol{\ell}} \times \hat{\mathbf{r}}/r^2)$.

Equation 28–5 is known as the **Biot-Savart law**. The magnitude of $d\vec{\mathbf{B}}$ is

$$dB = \frac{\mu_0 I d\ell \sin\theta}{4\pi r^2}, \qquad \textbf{(28–6)}$$

where θ is the angle between $d\vec{\boldsymbol{\ell}}$ and $\vec{\mathbf{r}}$ (Fig. 28–18). The total magnetic field at point P is then found by summing (integrating) over all current elements:

$$\vec{\mathbf{B}} = \int d\vec{\mathbf{B}} = \frac{\mu_0 I}{4\pi} \int \frac{d\vec{\boldsymbol{\ell}} \times \hat{\mathbf{r}}}{r^2}.$$

Note that this is a *vector* sum. The Biot-Savart law is the magnetic equivalent of Coulomb's law in its infinitesimal form. It is even an inverse square law, like Coulomb's law.

An important difference between the Biot-Savart law and Ampère's law (Eq. 28–3) is that in Ampère's law $\left[\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_{\text{encl}}\right]$, $\vec{\mathbf{B}}$ is not necessarily due only to the current enclosed by the path of integration. But in the Biot-Savart law the field $d\vec{\mathbf{B}}$ in Eq. 28–5 is due only, and entirely, to the current element $Id\vec{\boldsymbol{\ell}}$. To find the total $\vec{\mathbf{B}}$ at any point in space, it is necessary to include *all* currents.

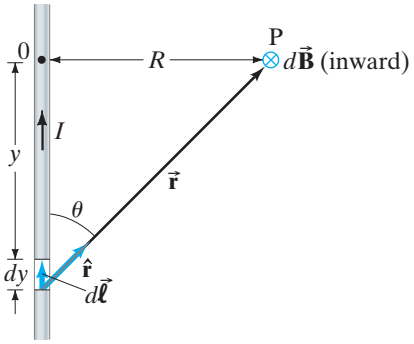


FIGURE 28–19 Determining $\vec{\mathbf{B}}$ due to a long straight wire using the Biot-Savart law. Example 28–10.

EXAMPLE 28–10 **$\vec{\mathbf{B}}$ due to current I in a straight wire.** For the field near a long straight wire carrying a current I, show that the Biot-Savart law gives the same result as Eq. 28–1, $B = \mu_0 I/2\pi r$.

APPROACH We calculate the magnetic field in Fig. 28–19 at point P, which is a perpendicular distance R from an infinitely long wire. The current is moving upwards, and both $d\vec{\boldsymbol{\ell}}$ and $\hat{\mathbf{r}}$, which appear in the cross product of Eq. 28–5, are in the plane of the page. Hence the direction of the field $d\vec{\mathbf{B}}$ due to each element of current must be directed into the plane of the page as shown (right-hand rule for the cross product $d\vec{\boldsymbol{\ell}} \times \hat{\mathbf{r}}$). Thus all the $d\vec{\mathbf{B}}$ have the same direction at point P, and add up to give $\vec{\mathbf{B}}$ the same direction consistent with our previous results (Figs. 28–1 and 28–11).

SOLUTION The magnitude of $\vec{\mathbf{B}}$ will be

$$B = \frac{\mu_0 I}{4\pi}\int_{y=-\infty}^{+\infty}\frac{dy\sin\theta}{r^2},$$

where $dy = d\ell$ and $r^2 = R^2 + y^2$. Note that we are integrating over y (the position of each point of the wire) so R is considered constant. Note also that we measure y as positive upward from point 0, so for the current element shown in Fig. 28–19 $y < 0$. Both y and θ are variables, but they are not independent. In fact, $y = -R/\tan\theta$, so dy is given by (see Appendices B–1 and B–2)

$$dy = -R\frac{d}{d\theta}\left(\frac{1}{\tan\theta}\right)d\theta = +R\frac{1}{\tan^2\theta}\frac{d}{d\theta}(\tan\theta)\,d\theta$$
$$= \frac{R\,d\theta}{\tan^2\theta\cos^2\theta} = \frac{R\,d\theta}{\sin^2\theta} = \frac{R\,d\theta}{(R/r)^2} = \frac{r^2\,d\theta}{R}.$$

From Fig. 28–19 we can see that $y = -\infty$ corresponds to $\theta = 0$ and that $y = +\infty$ corresponds to $\theta = \pi$ radians. So our integral becomes

$$B = \frac{\mu_0 I}{4\pi}\frac{1}{R}\int_{\theta=0}^{\pi}\sin\theta\,d\theta = -\frac{\mu_0 I}{4\pi R}\cos\theta\Big|_0^{\pi} = \frac{\mu_0 I}{2\pi R}.$$

This is just Eq. 28–1 for the field near a long wire, where R has been used instead of r.

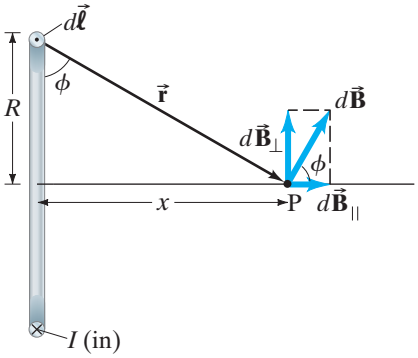


FIGURE 28–20 Determining $\vec{\mathbf{B}}$ due to a current loop. Example 28–11.

EXAMPLE 28–11 **Current loop.** Determine $\vec{\mathbf{B}}$ for points on the axis of a circular loop of wire of radius R carrying a current I, Fig. 28–20.

APPROACH For an element of current at the top of the loop, the magnetic field $d\vec{\mathbf{B}}$ at point P on the axis is perpendicular to $\vec{\mathbf{r}}$ as shown, and has magnitude (Eq. 28–5)

$$dB = \frac{\mu_0 I\,d\ell}{4\pi r^2}$$

since $d\vec{\boldsymbol{\ell}}$ is perpendicular to $\vec{\mathbf{r}}$ so $|d\vec{\boldsymbol{\ell}} \times \hat{\mathbf{r}}| = d\ell$. We can break $d\vec{\mathbf{B}}$ down into components $dB_\parallel$ and $dB_\perp$, which are parallel and perpendicular to the axis as shown.

SOLUTION When we sum over all the elements of the loop, *symmetry* tells us that the perpendicular components will cancel on opposite sides, so $B_\perp = 0$. Hence, the total $\vec{\mathbf{B}}$ will point along the axis, and will have magnitude

$$B = B_\parallel = \int dB\cos\phi = \int dB\frac{R}{r} = \int dB\frac{R}{(R^2 + x^2)^{\frac{1}{2}}},$$

where x is the distance of P from the center of the ring, and $r^2 = R^2 + x^2$. Now we put in dB from the equation above (see APPROACH) and integrate around the current loop, noting that all segments $d\vec{\boldsymbol{\ell}}$ of current are the same distance, $(R^2 + x^2)^{\frac{1}{2}}$, from point P:

$$B = \frac{\mu_0 I}{4\pi}\frac{R}{(R^2 + x^2)^{\frac{3}{2}}}\int d\ell = \frac{\mu_0 I R^2}{2(R^2 + x^2)^{\frac{3}{2}}}$$

since $\int d\ell = 2\pi R$, the circumference of the loop.

NOTE At the very center of the loop (where $x = 0$) the field has its maximum value

$$B = \frac{\mu_0 I}{2R}. \qquad \text{[at center of current loop]}$$

EXAMPLE 28–12 **$\vec{\mathbf{B}}$ due to a wire segment.** One quarter of a circular loop of wire carries a current I as shown in Fig. 28–21. The current I enters and leaves on straight segments of wire, as shown; the straight wires are along the radial direction from the center C of the circular portion. Find the magnetic field at point C.

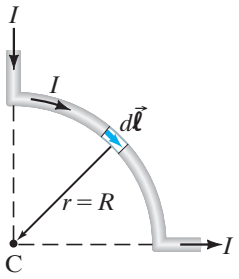


FIGURE 28–21 Example 28–12.

APPROACH The current in the straight sections produces no magnetic field at point C because $d\vec{\ell}$ and $\hat{\mathbf{r}}$ in the Biot-Savart law (Eq. 28–5) are parallel in those straight sections and therefore $d\vec{\ell} \times \hat{\mathbf{r}} = 0$. Each piece $d\vec{\ell}$ of the curved section of wire produces a field $d\vec{\mathbf{B}}$ that points into the page at C (right-hand rule).

SOLUTION The magnitude of each $d\vec{\mathbf{B}}$ due to each $d\ell$ of the circular portion of wire is (we can use Eq. 28–6)

$$dB = \frac{\mu_0 I\, d\ell}{4\pi R^2}$$

where $r = R$ is the radius of the curved section, and $\sin\theta$ in Eq. 28–6 is $\sin 90° = 1$. With $r = R$ for all pieces $d\vec{\ell}$, we integrate over a quarter of a circle.

$$B = \int dB = \frac{\mu_0 I}{4\pi R^2}\int d\ell = \frac{\mu_0 I}{4\pi R^2}\left(\frac{1}{4}2\pi R\right) = \frac{\mu_0 I}{8R}.$$

Magnetic Dipole Field

Recall from Section 27–5 that a current loop, such as that discussed in Example 28–11 (Fig. 28–20), is considered a **magnetic dipole**. We saw there that a current-carrying coil has a magnetic dipole moment

$$\mu = NIA,$$

where A is the area of the coil and N is the number of loops in the coil, each carrying current I. We also saw in Chapter 27 that a magnetic dipole placed in an external magnetic field experiences a torque and possesses potential energy, just like an electric dipole. In Example 28–11, we looked at another aspect of a magnetic dipole: the magnetic field *produced by* a magnetic dipole has magnitude, along the dipole axis, of

$$B = \frac{\mu_0 I R^2}{2\left(R^2 + x^2\right)^{\frac{3}{2}}}.$$

We can write this in terms of the magnetic dipole moment $\mu = IA = I\pi R^2$ (for a single loop $N = 1$):

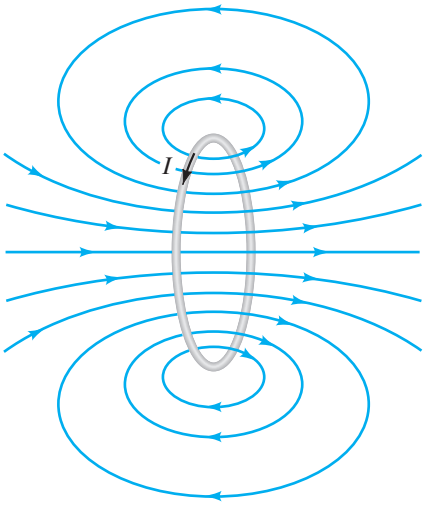
$$B = \frac{\mu_0}{2\pi}\frac{\mu}{\left(R^2 + x^2\right)^{\frac{3}{2}}}. \qquad \text{[magnetic dipole]} \quad \textbf{(28–7a)}$$

(Be careful to distinguish μ for dipole moment from μ_0, the magnetic permeability constant.) For distances far from the loop, $x \gg R$, this becomes

$$B \approx \frac{\mu_0}{2\pi}\frac{\mu}{x^3}. \qquad \left[\begin{array}{c}\text{on axis,}\\ \text{magnetic dipole, } x \gg R\end{array}\right] \quad \textbf{(28–7b)}$$

The magnetic field on the axis of a magnetic dipole decreases with the cube of the distance, just as the electric field does for an electric dipole. B decreases as the cube of the distance also for points not on the axis, although the multiplying factor is not the same. The magnetic field due to a current loop can be determined at various points using the Biot-Savart law and the results are in accord with experiment. The field lines around a current loop are shown in Fig. 28–22.

FIGURE 28–22 Magnetic field due to a circular loop of wire. (Same as Fig. 27–9.)



28–7 Magnetic Field Due to a Single Moving Charge

It might be tempting to use the Biot-Savart law (Eq. 28–5) to apply to a single point charge q moving with a constant velocity $\vec{\mathbf{v}}$. A moving electric charge should indeed produce a magnetic field. We might try to determine such a field by relating the current I (in the Biot-Savart law) to moving charges q in a wire as we did in Chapter 27. In Section 27–4, we wrote a current $I = Nq/t$ where N is the number of charges q passing a given point in time t. Let us let dt be the very brief time for a tiny charge dq to travel a very tiny distance $d\ell$ at velocity $\vec{\mathbf{v}}$, then $\vec{\mathbf{v}} = d\vec{\boldsymbol{\ell}}/dt$ and $I\,d\vec{\boldsymbol{\ell}}$ in Eq. 28–5 is $I\,d\vec{\boldsymbol{\ell}} = Ndq\,d\vec{\boldsymbol{\ell}}/dt = N\vec{\mathbf{v}}dq$. If we substitute this expression into the Biot-Savart law, Eq. 28–5, setting $N = 1$ for the one single charge and integrating over that one moving charge q, we get $\vec{\mathbf{B}} = \int(\mu_0 I\,d\vec{\boldsymbol{\ell}} \times \hat{\mathbf{r}})/4\pi r^2 = \int \mu_0 Ndq\,\vec{\mathbf{v}} \times \hat{\mathbf{r}}/4\pi r^2 = \mu_0 q\vec{\mathbf{v}} \times \hat{\mathbf{r}}/4\pi r^2$.

But there is a problem with our derivation of this equation, and therefore with the equation itself. Our "derivation" started with the Biot-Savart law which works for steady electric currents that do not change in time moving in a significant length of wire, whereas one moving charge is at one spot at a given time and continually changing—it is changing position in time and its magnetic field is changing at any given location. So the equation we "derived" for $\vec{\mathbf{B}}$ above can not be trusted. Here are two more problems. (1) That "derived" equation would be difficult to verify experimentally because a single particle would shoot out of your lab before your instruments could measure B accurately, unless the speed v was very small—in which case B would probably be too small to measure. (2) We expect the magnetic field lines to be circles around the direction of motion of the charged particle, just as for a current in a wire, Figs. 28–1 and 28–19. See Fig. 28–23. At any point in space, B is largest when the moving charge is nearest that point; as the charge speeds away, B decreases at that point. So B is a function of time at any point. Here now is the fun: if you jumped on the particle, you might expect its magnetic field lines to be circles around it, something like in Fig. 28–23 (with $v = 0$). But no. If you are in the reference frame of the charge, the charge is at rest. A charge at rest produces no magnetic field! How could a particle have a magnetic field in one reference frame, and zero magnetic field in another? This contradiction, or paradox, and others like it, helped Einstein to develop his theory of Special Relativity, which resolved this inconsistency. Part of the answer, complicated though it seems, is that a magnetic field in one reference frame can be an electric field in a different reference frame. Observers in two different reference frames, moving relative to each other, will not observe the same $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields.

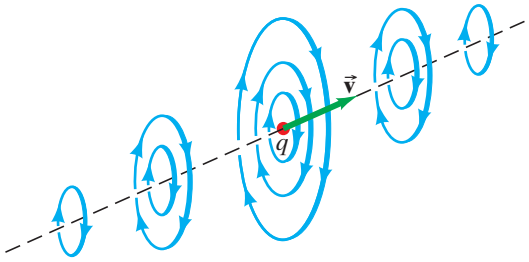


FIGURE 28–23 The magnetic field at one single instant due to a single positive charge q, moving at velocity $\vec{\mathbf{v}}$.

28–8 Magnetic Materials—Ferromagnetism

Magnetic fields can be produced (1) by magnetic materials (magnets) and (2) by electric currents. Common magnetic materials include ordinary magnets, iron cores in motors and electromagnets, recording tape, computer hard drives, and magnetic stripes on credit cards. We saw in Section 27–1 that iron (and a few other materials) can be made into strong magnets. These materials are said to be **ferromagnetic**. We now look into the sources of ferromagnetism.

A bar magnet, with its two opposite poles near either end, resembles an electric dipole (equal-magnitude positive and negative charges separated by a distance). Indeed, a bar magnet is sometimes referred to as a "magnetic dipole."

There are opposite "poles" separated by a distance. And the magnetic field lines of a bar magnet form a pattern much like that for the electric field of an electric dipole: compare Fig. 21–35a with Fig. 27–4.

Microscopic examination reveals that a piece of iron is made up of tiny regions known as **domains**, less than 1 mm in length or width. Each domain behaves like a tiny magnet with a north and a south pole. In an unmagnetized piece of iron, these domains are arranged randomly, Fig. 28–24a. The magnetic effects of the domains cancel each other out, so this piece of iron does not behave as a magnet. In a magnet, the domains are preferentially aligned in one direction as shown in Fig. 28–24b (downward in this case). A magnet can be made from an unmagnetized piece of iron by placing it in a strong magnetic field. (You can make a needle magnetic, for example, by stroking it with one pole of a strong magnet.) The magnetization direction of domains may actually rotate slightly to be more nearly parallel to the external field, and the borders of domains may move so domains with magnetic orientation parallel to the external field grow larger (compare Figs. 28–24a and b).

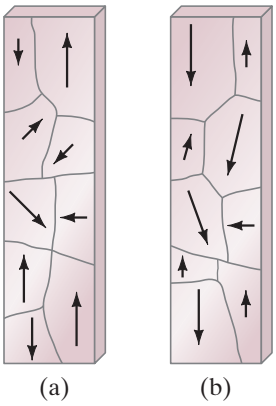


FIGURE 28–24 (a) An unmagnetized piece of iron is made up of domains that are randomly arranged. Each domain is like a tiny magnet; the arrows represent the magnetization direction, with the arrowhead being the N pole. (b) In a magnet, the domains are preferentially aligned in one direction (down in this case), and may be altered in size by the magnetization process.

We can now explain how a magnet can pick up unmagnetized pieces of iron like paper clips. Consider the north pole of a permanent magnet. The magnet's field causes a slight realignment of the domains in the unmagnetized paper clip, so that it becomes a temporary magnet with its south pole facing the north pole of the permanent magnet; thus, attraction results. Similarly, elongated iron filings in a magnetic field acquire aligned domains and align themselves to reveal the shape of the magnetic field, Fig. 28–25.

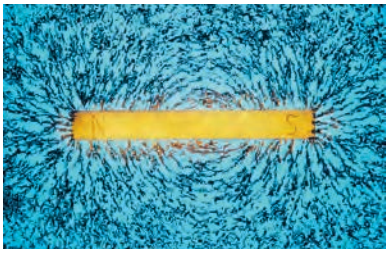


FIGURE 28–25 Iron filings line up along magnetic field lines due to a permanent magnet.

An iron magnet can remain magnetized for a long time, and is referred to as a "permanent magnet." But if you drop a magnet on the floor or strike it with a hammer, you can jar the domains into randomness and the magnet loses some or all of its magnetism. Heating a permanent magnet can also cause loss of magnetism, because raising the temperature increases the random thermal motion of atoms, which tends to randomize the domains. Above a certain temperature known as the **Curie temperature** (1043 K for iron), a magnet cannot be made at all. Iron, nickel, cobalt, gadolinium, and certain alloys are ferromagnetic at room temperature; several other elements and alloys are ferromagnetic only at low temperatures. Most other metals, such as aluminum and copper, do not show any noticeable magnetic effect (but see Section 28–11).

The striking similarity between the fields produced by a bar magnet and by a loop of electric current (Figs. 27–4b and 28–22) offers a clue that perhaps magnetic fields produced by electric currents may have something to do with ferromagnetism. According to an early atomic model, atoms can be roughly visualized as having electrons that orbit around a central nucleus. The electrons are charged, and so constitute an electric current and therefore produce a magnetic field. But the fields due to orbiting electrons end up adding to zero. Electrons themselves produce an additional magnetic field, almost as if they and their electric charge were spinning about their own axes. It is the magnetic field due to electron **spin**[†] that is believed to produce ferromagnetism in most ferromagnetic materials.

It is believed today that probably *all* magnetic fields are caused by electric currents. This means that magnetic field lines always form closed loops, unlike electric field lines which begin on positive charges and end on negative charges.

CAUTION
$\vec{\mathbf{B}}$ *lines form closed loops,*
$\vec{\mathbf{E}}$ *start on* ⊕ *and end on* ⊖

EXERCISE E Return to the Chapter-Opening Question, page 811, and answer it again now. Try to explain why you may have answered differently the first time.

[†]The name "spin" comes from an early suggestion that this intrinsic magnetic moment arises from the electron "spinning" on its axis (as well as "orbiting" the nucleus) to produce the extra field. However this view of a spinning electron is oversimplified and not valid (Chapters 38 and 39).

28–9 Electromagnets and Solenoids—Applications

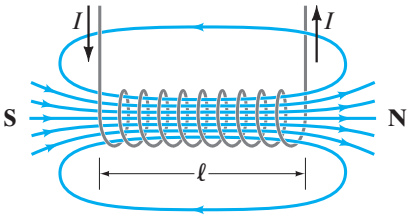
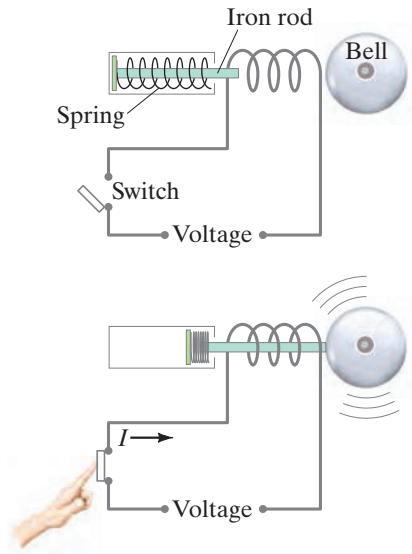


FIGURE 28–26 Magnetic field of a solenoid. The north pole of this solenoid, thought of as a magnet, is on the right, and the south pole is on the left.

PHYSICS APPLIED
Electromagnets and solenoids

FIGURE 28–27 Solenoid used as a doorbell.



PHYSICS APPLIED
Doorbell, car starter

PHYSICS APPLIED
Magnetic circuit breakers

A long coil of wire consisting of many loops of wire is called a **solenoid**, as we saw in Section 28–5. The magnetic field within a solenoid can be fairly large because it is the sum of the fields due to the current in each loop (Fig. 28–26). A solenoid acts like a magnet; one end can be considered the north pole and the other the south pole, depending on the direction of the current in the loops (use the right-hand rule). Since the magnetic field lines leave the north pole of a magnet, the north pole of the solenoid in Fig. 28–26 is on the right.

If a piece of iron is placed inside a solenoid, the magnetic field is increased greatly because the domains of the iron are aligned by the magnetic field produced by the current; that is, the iron becomes a magnet. The resulting magnetic field is the sum of the field due to the solenoid's current and the field due to the iron, and can be hundreds or thousands of times larger than the field due to the current alone (see Section 28–10). Such an iron-core solenoid is an **electromagnet**. The alloys of iron used in electromagnets acquire and lose their magnetism quite readily when the current is turned on or off, and so are referred to as "soft iron." (It is "soft" only in a magnetic sense.) Iron that holds its magnetism even when there is no externally applied field is called "hard iron." Hard iron is used in permanent magnets. Soft iron is usually used in electromagnets so that the field can be turned on and off readily. Whether iron is hard or soft depends on heat treatment, type of alloy, and other factors.

Electromagnets have many practical applications, from use in motors and generators to producing large magnetic fields for research. Sometimes an iron core is not present—the magnetic field then comes only from the current in the wire loops. A large field B in this case requires a large current I, which produces a large amount of waste heat $(P = I^2R)$. But if the current-carrying wires are made of superconducting material kept below the transition temperature (Section 25–9), then very high magnetic fields can be produced, and no electric power is needed to maintain the large current in the superconducting coils. Energy is required, however, to refrigerate the coils at the low temperatures where they superconduct.

Another useful device consists of a solenoid into which a moveable rod of iron is partially inserted, and can be used as a switch. This combination is also referred to as a **solenoid**. One simple use is as a doorbell (Fig. 28–27). When the circuit is closed by pushing the button, the coil effectively becomes a magnet and exerts a force on the iron rod. The rod is pulled into the coil and strikes the bell. A large solenoid is used in the starter of a car: when you engage the starter, you are closing a circuit that not only turns the starter motor, but first activates a solenoid that moves the starter into direct contact with the gears on the engine's flywheel. Solenoids are often used as switches in cars and many other devices. They have the advantage of moving mechanical parts quickly and accurately.

Magnetic Circuit Breakers

Modern circuit breakers that protect houses and buildings from overload and fire contain not only a "thermal" part (bimetallic strip as described in Section 25–6, Fig. 25–19) but also a magnetic sensor. If the current is above a certain level, the magnetic field the current produces pulls an iron plate that breaks the same contact points as in Figs. 25–19b and c. Such magnetic circuit breakers act quickly (< 10 ms), and for buildings are designed to react to the high currents of short circuits, but not shut off for the start-up surges of motors.

In more sophisticated circuit breakers, including ground fault circuit interrupters (GFCIs—discussed in Section 29–9), a solenoid is used. The iron rod of Fig. 28–27, instead of striking a bell, strikes one side of a pair of electric contact points, opening them and opening the circuit. They react very quickly $(\approx 1$ ms$)$ and to very small currents $(\approx 5$ mA$)$ and thus protect humans (not just property) and save lives. They should be installed in every kitchen and bathroom especially.

28–10 Magnetic Fields in Magnetic Materials; Hysteresis

The magnetic field inside a long solenoid is directly proportional to the current, as we saw in Eq. 28–4:

$$B_0 = \mu_0 nI. \tag{28–4}$$

We write B_0 to indicate there is only air inside the solenoid, and n is the number of loops per unit length. If a piece of ferromagnetic material like iron is placed inside a solenoid to form an electromagnet (Section 28–9), the magnetic field increases greatly over that produced by the current in the solenoid loops alone, often by hundreds or thousands of times. This happens because the domains in the iron become aligned by the external field produced by the current in the solenoid coil. The total magnetic field $\vec{\mathbf{B}}$ is then the sum of two terms,

$$\vec{\mathbf{B}} = \vec{\mathbf{B}}_0 + \vec{\mathbf{B}}_\mathrm{M}. \tag{28–8}$$

$\vec{\mathbf{B}}_0$ is the field due to the current in the solenoid coil and $\vec{\mathbf{B}}_\mathrm{M}$ is the additional field due to the magnetic material (typically iron). Often $B_\mathrm{M} \gg B_0$.

The total field can also be written by replacing the constant μ_0 in Eq. 28–4 by another parameter called the **magnetic permeability** μ, which is characteristic of the magnetic material inside the coil. Then

$$B = \mu nI. \tag{28–9}$$

(Do not confuse this μ with $\vec{\boldsymbol{\mu}}$ for magnetic dipole moment.) For ferromagnetic materials, μ is much greater than μ_0. For all other materials, its value is very close to μ_0 (Section 28–11). The value of μ, however, is not constant for ferromagnetic materials; it depends on the value of the "external field" B_0, as the following experiment shows.

Hysteresis

Measurements on magnetic materials often use a **torus** or **toroid**, which is like a long solenoid bent into the shape of a donut (Fig. 28–28), so practically all the lines of $\vec{\mathbf{B}}$ remain within the toroid. Consider a toroid with an iron core that is initially unmagnetized and there is no current in the wire loops. Then the current I is slowly increased, and B_0 (which is due only to I) increases linearly with I. The total field B also increases, but follows the curved line shown in Fig. 28–29 which is a graph of total B vs. B_0. Initially, point a on the graph, the domains are randomly oriented. As B_0 increases, the domains become more and more aligned until at point b, nearly all are aligned. The iron is said to be approaching **saturation**. Next, suppose current in the coil is reduced, so the field B_0 decreases. If the current (and B_0) is reduced to zero, point c in Fig. 28–30, the domains do *not* become completely random. Instead, some permanent magnetism remains in the iron core. If the current is increased in the opposite direction, enough domains can be turned around so the total B becomes zero at point d. As the reverse current is increased further, the iron approaches saturation in the opposite direction, point e in Fig. 28–30. Finally, if the current is again reduced to zero (point f) and then increased in the original direction, the total field follows the path efgb, again approaching saturation at point b.

Notice that the field did not pass through the origin (point a) in this cycle. The fact that the curve does not retrace itself on the same path is called **hysteresis**. The curve bcdefgb is called a **hysteresis loop**. In such a cycle, much energy is transformed to thermal energy (friction) due to realigning of the domains. [It can be shown that the energy dissipated in this way is proportional to the area of the hysteresis loop.]

At points c and f, the iron core is magnetized even though there is no current in the coils. These points correspond to a permanent magnet. For a permanent magnet, it is desired that ac and af be as large as possible. Materials for which this is true are said to have high **retentivity**.

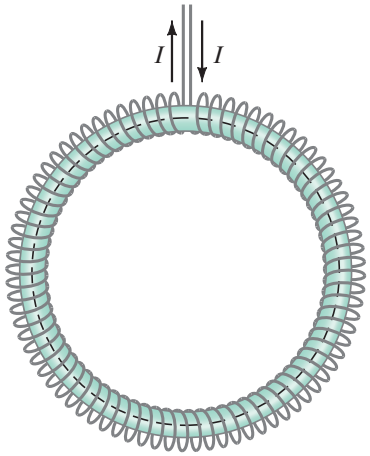


FIGURE 28–28 Iron-core toroid.

FIGURE 28–29 Total magnetic field B in an iron-core toroid as a function of the external field B_0 (B_0 is caused by the current I in the coil). We use gauss ($1\,\mathrm{G} = 10^{-4}\,\mathrm{T}$) so that labels are clear.

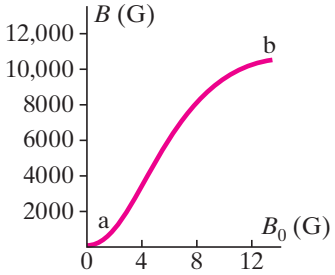
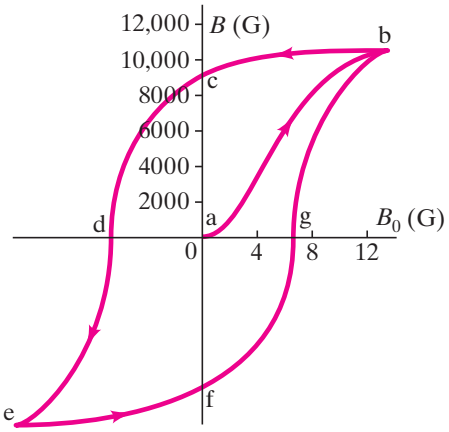


FIGURE 28–30 Hysteresis curve.



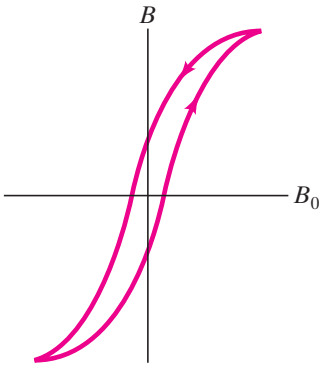
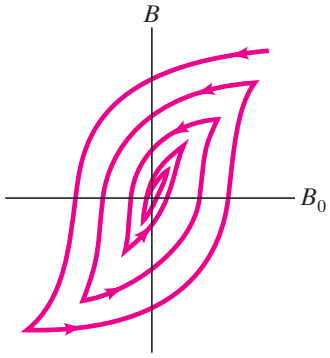


FIGURE 28–31 Hysteresis curve for soft iron.

FIGURE 28–32 Successive hysteresis loops during demagnetization.



Iron alloys with a broad hysteresis curve as in Fig. 28–30 hold their magnetism even when the current is off; they are referred to as "hard iron," and make good permanent magnets. On the other hand, a hysteresis curve such as that in Fig. 28–31 occurs for "soft iron," which is preferred for electromagnets and transformers (Section 29–6) since the field can be more readily switched off, and the field can be reversed with less loss of energy.

A magnetized ferromagnetic material can be demagnetized—that is, made unmagnetized. This can be done by reversing the magnetizing current repeatedly while decreasing its magnitude. This results in the curve of Fig. 28–32. Watches, the heads of a tape recorder, and other objects made of iron are demagnetized in this way. The alternating magnetic field due to a handheld demagnetizer is strong when the demagnetizer is placed near the object and decreases as it is moved slowly away. Computer hard disks, magnetic tape, and other magnetic storage devices can be erased and ruined by a magnetic field, as can the magnetic strips on credit cards.

*28–11 Paramagnetism and Diamagnetism

All materials are magnetic to at least a tiny extent. Nonferromagnetic materials fall into two principal classes: *paramagnetic*, in which the magnetic permeability μ is very slightly greater than μ_0; and *diamagnetic*, in which μ is very slightly less than μ_0. The ratio of μ to μ_0 for any material is called the **relative permeability** K_m:

$$K_m = \frac{\mu}{\mu_0}.$$

Another useful parameter is the **magnetic susceptibility** χ_m defined as

$$\chi_m = K_m - 1.$$

Paramagnetic substances have $K_m > 1$ and $\chi_m > 0$, whereas diamagnetic substances have $K_m < 1$ and $\chi_m < 0$. See Table 28–1, and note how small the effect is.

TABLE 28–1 Paramagnetism and Diamagnetism: Magnetic Susceptibilities

Paramagnetic substance	χ_m	Diamagnetic substance	χ_m
Aluminum	2.3×10^{-5}	Copper	-9.8×10^{-6}
Calcium	1.9×10^{-5}	Diamond	-2.2×10^{-5}
Magnesium	1.2×10^{-5}	Gold	-3.6×10^{-5}
Oxygen (STP)	2.1×10^{-6}	Lead	-1.7×10^{-5}
Platinum	2.9×10^{-4}	Nitrogen (STP)	-5.0×10^{-9}
Tungsten	6.8×10^{-5}	Silicon	-4.2×10^{-6}

The difference between paramagnetic and diamagnetic materials can be understood theoretically at the molecular level on the basis of whether or not the molecules have a permanent magnetic dipole moment. One type of **paramagnetism** occurs in materials whose molecules (or ions) have a permanent magnetic dipole moment.[†] In the absence of an external field, the molecules are randomly oriented and no magnetic effects are observed. However, when an external magnetic field is applied, say, by putting the material in a solenoid, the applied field exerts a torque on the magnetic dipoles (Section 27–5), tending to align them parallel to the field. The total magnetic field (external plus that due to aligned magnetic dipoles) will be slightly greater than B_0. The thermal motion of the molecules

[†]Other types of paramagnetism also occur whose origin is different from that described here, such as in metals where free electrons can contribute.

reduces the alignment, however. A useful quantity is the **magnetization vector**, $\vec{\mathbf{M}}$, which is useful for any magnetic material (even ferromagnetism), defined as the magnetic dipole moment per unit volume,

$$\vec{\mathbf{M}} = \frac{\vec{\boldsymbol{\mu}}}{V},$$

where $\vec{\boldsymbol{\mu}}$ is the *magnetic dipole moment* of the sample and V its volume. (This $\vec{\boldsymbol{\mu}}$, also used on page 823, is magnetic moment, and *not* the μ and μ_0 that stands for the constant *magnetic permeability* as on pages 827 and 828.) It is found experimentally that M is directly proportional to the external magnetic field B (tending to align the dipoles) and inversely proportional to the kelvin temperature T (tending to randomize dipole directions). This is called *Curie's law*, after Pierre Curie (1859–1906), who first noted it:

$$M = C\frac{B}{T},$$

where C is a constant. If the ratio B/T is very large (B very large or T very small) Curie's law is no longer accurate; as B is increased (or T decreased), the magnetization approaches some maximum value, M_{max}. This makes sense, of course, since M_{max} corresponds to complete alignment of all the permanent magnetic dipoles. However, even for very large magnetic fields, $\approx$2.0 T, deviations from Curie's law are normally noted only at very low temperatures, on the order of a few kelvins.

Ferromagnetic materials, as mentioned in Section 28–7, are no longer ferromagnetic above a characteristic temperature called the Curie temperature (1043 K for iron). Above this Curie temperature, they generally are paramagnetic.

Diamagnetic materials (for which the magnetic permeability μ is slightly less than μ_0) are made up of molecules that have no permanent magnetic dipole moment. When an external magnetic field is applied, magnetic dipoles are induced, but the induced magnetic dipole moment is in the direction *opposite* to that of the field: the effect of the external field—in the crude model of electrons orbiting nuclei—is to increase the "orbital" speed of electrons revolving in one direction, and to decrease the speed of electrons revolving in the other direction. The net result is a net dipole moment opposing the external field. Hence the total field will be slightly less than the external field. Diamagnetism is present in all materials, but is weaker even than paramagnetism and so is overwhelmed by paramagnetic and ferromagnetic effects in materials that display these other forms of magnetism.

Summary

The magnetic field B at a distance r from a long straight wire is directly proportional to the current I in the wire and inversely proportional to r:

$$B = \frac{\mu_0}{2\pi}\frac{I}{r}. \qquad \textbf{(28–1)}$$

The magnetic field lines are circles centered on the straight wire.

The force that one long current-carrying wire exerts on a second parallel current-carrying wire 1 m away serves as the definition of the ampere unit, and ultimately of the coulomb as well.

Ampère's law states that the line integral of the magnetic field $\vec{\mathbf{B}}$ around any closed path is equal to μ_0 times the total net current I_{encl} enclosed by that path:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_{\text{encl}}. \qquad \textbf{(28–3)}$$

The magnetic field inside a long tightly wound solenoid is

$$B = \mu_0 nI \qquad \textbf{(28–4)}$$

where n is the number of loops per unit length and I is the current in each loop.

The **Biot-Savart law** is useful for determining the magnetic field due to a known arrangement of currents. It states that

$$d\vec{\mathbf{B}} = \frac{\mu_0 I}{4\pi}\frac{d\vec{\boldsymbol{\ell}} \times \hat{\mathbf{r}}}{r^2}, \qquad \textbf{(28–5)}$$

where $d\vec{\mathbf{B}}$ is the contribution to the total field at some point P due to a current I along an infinitesimal length $d\vec{\boldsymbol{\ell}}$ of its path, and $\hat{\mathbf{r}}$ is the unit vector along the direction of the displacement vector $\vec{\mathbf{r}}$ from $d\vec{\boldsymbol{\ell}}$ to P. The total field $\vec{\mathbf{B}}$ will be the integral over all $d\vec{\mathbf{B}}$.

Iron and a few other materials that are **ferromagnetic** can be made into strong permanent magnets. Ferromagnetic materials are made up of tiny **domains**—each a tiny magnet—which are preferentially aligned in a permanent magnet, but randomly aligned in a nonmagnetized sample.

When iron or other ferromagnetic material is placed in a magnetic field B_0 due to a current (as inside a solenoid or toroid) the iron becomes magnetized. When the current is turned off, the material remains magnetized; when the current is increased in the opposite direction, a graph of the total field B versus B_0 is a **hysteresis loop**, and the fact that the curve does not retrace itself is called **hysteresis**.

[*All materials exhibit some magnetic effects. Nonferromagnetic materials have much smaller **paramagnetic** or **diamagnetic** properties.]

Questions

1. The magnetic field due to current in wires in your home may affect a compass. Discuss the problem in terms of currents, depending on whether they are ac or dc, and their distance away.
2. Compare and contrast the magnetic field due to a long straight current and the electric field due to a long straight line of electric charge at rest (Section 21–7).
3. Two insulated long wires carrying equal currents I cross at right angles to each other. Describe the magnetic force one exerts on the other.
4. A horizontal wire carries a large current. A second wire carrying a current in the same direction is suspended below it. Can the current in the upper wire hold the lower wire in suspension against gravity? Under what conditions will the lower wire be in equilibrium?
5. (*a*) Write Ampère's law for a path that surrounds both conductors in Fig. 28–10. (*b*) Repeat, assuming the lower current, I_2, is in the opposite direction $(I_2 = -I_1)$.
6. Explain why a field such as that shown in Fig. 28–14b is consistent with Ampère's law. Could the lines curve upward instead of downward?
7. What would be the effect on B inside a long solenoid if the solenoid's length was halved along with a doubling in the total number of loops?
8. Use the Biot-Savart law to show that the field of the current loop in Fig. 28–21 is correct as shown for points off the axis.
9. Do you think $\vec{\mathbf{B}}$ will be the same for all points in the plane of the current loop of Fig. 28–22? Explain.
10. Why does twisting the lead-in wires to electrical devices reduce the magnetic effects of the leads?
11. Compare the Biot-Savart law with Coulomb's law. What are the similarities and differences?
12. How might you measure the magnetic dipole moment of the Earth?
13. A type of magnetic switch similar to a solenoid is a **relay** (Fig. 28–33). A relay is an electromagnet (the iron rod inside the coil does not move) which, when activated, attracts a strip of iron on a pivot. Design a relay to close an electrical switch. A relay is used when you need to switch on a circuit carrying a very large current but you do not want that large current flowing through the main switch. For example, a car's starter switch is connected to a relay so that the large current needed for the starter doesn't pass to the dashboard switch.

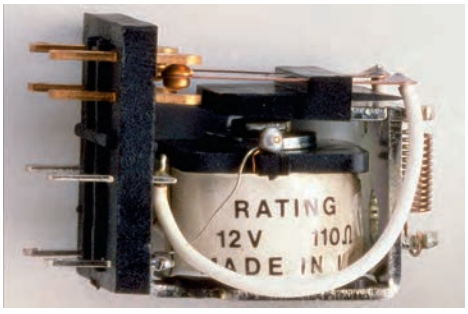


FIGURE 28–33 Question 13.

14. A heavy block of iron, at rest, is attracted by a heavy magnet. Before striking the magnet the block has acquired considerable kinetic energy. (*a*) What is the source of this kinetic energy? (*b*) When the block strikes the magnet, some of the latter's domains may be jarred into randomness; describe the energy transformations.
15. A bar magnet exerts a force on a paper clip (page 825) or another bar magnet only because its magnetic field is non-uniform. To see why, show that a uniform magnetic field exerts no force on a bar magnet (see Fig. 27–7 on page 784). Also note there is no net force on the magnetic dipole in Fig. 27–23.
16. Will a magnet attract any metallic object, such as those made of aluminum or copper? (Try it and see.) Why is this so?
17. An unmagnetized nail will not attract an unmagnetized paper clip. However, if one end of the nail is in contact with a magnet, the other end *will* attract a paper clip. Explain.
18. Can an iron rod attract a magnet? Can a magnet attract an iron rod? What must you consider to answer these questions?
19. Why will either pole of a magnet attract an unmagnetized piece of iron?
20. Suppose you have three iron rods, two of which are magnetized but the third is not. How would you determine which two are the magnets without using any additional objects?
21. Two iron bars attract each other no matter which ends are placed close together. Are both magnets? Explain.
22. Sketch a magnet with domains as in Fig. 28–24b whose magnetic field would be in the opposite direction and half as strong as the magnet in Fig. 28–24b.

*23. Can all materials be considered (*a*) diamagnetic, (*b*) paramagnetic, (*c*) ferromagnetic? Explain.

*24. Describe the magnetization curve for (*a*) a paramagnetic substance and (*b*) a diamagnetic substance, and compare to that for a ferromagnetic substance (Fig. 28–30).

*25. Fill in each blank with the correct choice of ferromagnetic, paramagnetic, or diamagnetic, or state the sentence is not true for any of these terms.

(*a*) For _____ materials placed in a strong magnetic field, the induced dipole moment points in the opposite direction of the external field.

(*b*) _____ materials have stronger magnetic fields than either _____ or _____ materials.

(*c*) _____ materials are no longer _____ above the Curie temperature.

(*d*) All materials are _____.

MisConceptual Questions

1. Which of the following statements is false? The magnetic field of a current-carrying wire
(*a*) is directed circularly around the wire.
(*b*) decreases inversely with the distance from the wire.
(*c*) exists only if the current in the wire is changing.
(*d*) depends on the magnitude of the current.

2. A horizontal current-carrying wire, free to move in Earth's gravitational field, is suspended directly above a parallel, current-carrying wire. In what direction is the current in the lower wire?
(*a*) Same direction as the upper one.
(*b*) Opposite to the direction in the upper one.
(*c*) Zero.
(*d*) Not enough information given.

3. Two parallel wires are vertical. The one on the left carries a 10-A current upward. The other carries 5-A current downward. Compare the magnitude of the force that each wire exerts on the other.
(*a*) The wire on the left carries twice as much current, so it exerts twice the force on the right wire as the right one exerts on the left one.
(*b*) The wire on the left exerts a smaller force. It creates a magnetic field twice that due to the wire on the right; and therefore has less energy to cause a force on the wire on the right.
(*c*) The two wires exert the same force on each other.
(*d*) Not enough information; we need the length of the wire.

4. What would be the effect on B inside a long solenoid if (i) the diameter of all the loops was doubled, (ii) the spacing between loops instead was doubled?
(*a*) Nothing; nothing.
(*b*) Halved; nothing.
(*c*) Doubled; nothing.
(*d*) Nothing; doubled.
(*e*) Nothing; halved.

5. What would be the effect on B inside a long solenoid if the solenoid's length was doubled along with a doubling in the total number of loops?
(*a*) Doubled.
(*b*) Four times larger.
(*c*) Halved.
(*d*) $\frac{1}{4}$ as large.
(*e*) No change.

6. Two long straight current-carrying wires are perpendicular, and pass near each other but do not touch. What can you say about the magnetic force between the two wires?
(*a*) There is no force between the two wires.
(*b*) Near where the wires cross, the force is about the same as for parallel wires.
(*c*) Each wire exerts a circularly directed force on the other.

7. Suppose the cylindrical conductor of Fig. 28–11a has a concentric cylindrical hollow cavity inside it (so it looks like a pipe). What can you say about $\vec{\mathbf{B}}$ in the cavity?
(*a*) $B \propto r$. (*b*) $B \propto 1/r$. (*c*) $B \propto 1/R$. (*d*) $B = 0$.

8. A mosquito carries a positive charge and is flying parallel to a wire carrying current in the same direction. The wire has no net charge. Which statement(s) below is/are true?
(*a*) There is no electric or magnetic force on the mosquito.
(*b*) The mosquito is attracted to the wire.
(*c*) The mosquito is repelled by the wire.
(*d*) There is a force on the mosquito, but it is not toward or away from the wire.

9. A thick straight copper wire carries a dc current. What is true about the magnetic field inside the wire?
(*a*) The field is strongest just inside the surface and drops to zero at the center.
(*b*) The field is zero inside the wire.
(*c*) The field is greatest at the center and drops off toward the surface.
(*d*) The field is uniform but not zero inside the wire.

10. Four wires are positioned at the corners of a square, Fig. 28–34. Each wire carries a current (note directions), all four of the same magnitude. What is the direction of the magnetic field at point P at the center of the square?

(a) (d)
(b) (e) $B = 0$.
(c)

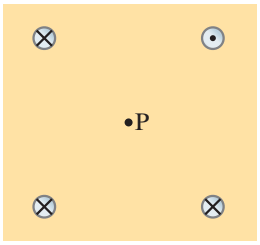


FIGURE 28–34 MisConceptual Question 10.

11. The magnetic field at distance x from the current in a long, straight conductor has magnitude B. What is the magnitude of the field at twice the distance from the conductor?
(*a*) $2B$.
(*b*) B.
(*c*) $B/4$.
(*d*) $4B$.
(*e*) $B/2$.

12. As an electron moves through space, it creates
(*a*) an electric field only.
(*b*) a magnetic field only.
(*c*) both an electric field and a magnetic field.
(*d*) nothing; the electric and magnetic fields cancel each other out.

13. Two long concentric pipes are cylindrical conductors as shown in Fig. 28–35. The inner pipe carries 5 A into the page and the outer pipe carries 5 A out of the page. In which region or regions does the magnetic field have the lowest magnitude?

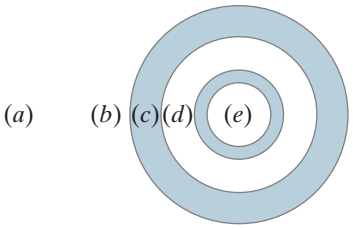


FIGURE 28–35 MisConceptual Question 13.

14. Which of the following will reduce the magnetic field in a moderately strong magnet? Select as many answers as are true.
(*a*) Heat the magnet above the Curie temperature.
(*b*) Heat the magnet, but keep below the Curie temperature.
(*c*) Put the magnet in an alternating magnetic field that gradually decreases.
(*d*) Hold the magnet in a very strong magnetic field for an hour and then remove the magnet.
(*e*) Strike the magnet sharply.

Problems

28–1 and 28–2 Straight Wires, Magnetic Field, and Force

1. (I) Jumper cables used to start a stalled vehicle often carry a 65-A current. How strong is the magnetic field 5.0 cm from one cable? Compare to the Earth's magnetic field $(5.0 \times 10^{-5}\,\text{T})$.
2. (I) If an electric wire is allowed to produce a magnetic field no larger than that of the Earth $(0.50 \times 10^{-4}\,\text{T})$ at a distance of 12 cm from the wire, what is the maximum dc current the wire can carry?
3. (I) Determine the magnitude and direction of the force between two long parallel wires 35 m long and 4.0 cm apart, each carrying 35 A in the same direction.
4. (I) A long straight vertical wire carrying an upward 22-A current exerts an attractive force per unit length of $7.8 \times 10^{-4}\,\text{N/m}$ on a second parallel wire 9.0 cm away. What current (magnitude and direction) flows in the second wire?
5. (II) In Fig. 28–36, a long straight wire carries current I out of the page toward you. Indicate, with appropriate arrows, the direction and (relative) magnitude of $\vec{\mathbf{B}}$ at each of the points C, D, and E in the plane of the page.

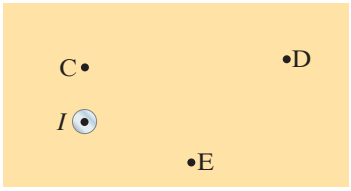


FIGURE 28–36 Problem 5.

6. (II) An experiment on the Earth's magnetic field is being carried out 1.00 m from an electric cable. What is the maximum allowable dc current in the cable if the experiment is to be accurate to $\pm 5.0\%$?
7. (II) Two long thin parallel wires 13.0 cm apart carry 25-A currents in the same direction. Determine the magnetic field vector at a point 10.0 cm from one wire and 6.0 cm from the other (Fig. 28–37). [*Hint*: You could try using the law of cosines, Appendix A.]

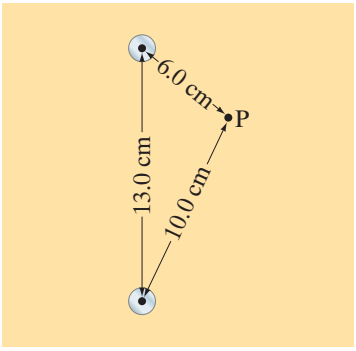


FIGURE 28–37 Problem 7.

8. (II) A horizontal compass is placed 24 cm due south from a straight vertical wire carrying a 48-A dc current downward. In what direction does the compass needle point at this location? Assume the horizontal component of the Earth's field at this point is $0.45 \times 10^{-4}\,\text{T}$ and the magnetic declination is 0°.
9. (II) A long horizontal wire carries 24.0 A of dc current due north. What is the net magnetic field 20.0 cm due west of the wire if the Earth's field there points downward, 38° below the horizontal, and has magnitude $5.0 \times 10^{-5}\,\text{T}$?
10. (II) Determine the magnetic field midway between two long straight wires 2.0 cm apart in terms of the current I in one when the other carries 25 A dc. Assume these currents are (*a*) in the same direction, and (*b*) in opposite directions.
11. (II) A continuous straight stream of protons passes a given point in space at a rate of 2.5×10^{9} protons/s. What magnetic field do they produce 1.8 m from the beam?
12. (II) Two straight parallel wires are separated by 6.0 cm. There is a 2.0-A dc current flowing in the first wire. If the magnetic field strength is found to be zero between the two wires at a distance of 2.2 cm from the first wire, what is the magnitude and direction of the current in the second wire?
13. (II) Two long straight wires each carry a dc current I out of the page toward the viewer, Fig. 28–38. Indicate, with appropriate arrows, the direction of $\vec{\mathbf{B}}$ at each of the points 1 to 6 in the plane of the page. State if the field is zero at any of the points.

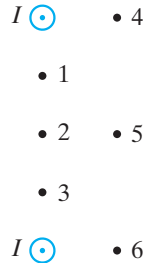


FIGURE 28–38 Problem 13.

14. (II) A long pair of insulated wires serves to conduct 35.0 A of dc current to and from an instrument. If the wires are of negligible diameter but are 2.8 mm apart, what is the magnetic field 10.0 cm from their midpoint, in their plane (Fig. 28–39)? Compare to the magnetic field of the Earth.

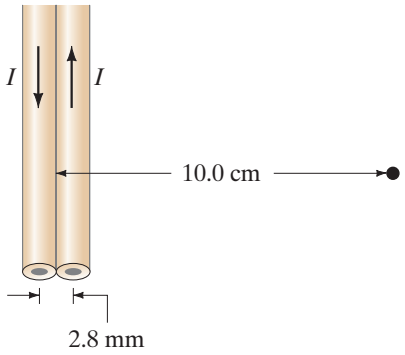


FIGURE 28–39 Problems 14 and 15.

15. (II) A third wire is placed in the plane of the two wires shown in Fig. 28–39 parallel and just to the right. If it carries 25.0 A upward, what force per meter of length does it exert on each of the other two wires? Assume it is 2.8 mm from the nearest wire, center to center.
16. (II) A dc power line carries a current of 85 A west along the tops of 8.5-m-high poles. (*a*) What is the magnitude and direction of the magnetic field produced by this wire at the ground directly below? How does this compare with the Earth's field of about $\frac{1}{2}$ G? (*b*) Where would the current's magnetic field cancel the Earth's field?
17. (II) A compass needle points 13° E of N outdoors. However, when it is placed 16.0 cm to the east of a vertical wire inside a building, it points 55° E of N. What is the magnitude and direction of the current in the wire? The Earth's field there is $0.50 \times 10^{-4}\,\text{T}$ and is horizontal.

18. (II) A rectangular loop of wire is placed next to a straight wire, as shown in Fig. 28–40. There is a dc current of 3.5 A in both wires. Determine the magnitude and direction of the net force on the loop.

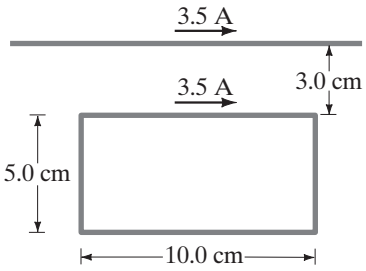


FIGURE 28–40
Problem 18.

19. (II) Let two long parallel wires, a distance d apart, carry equal dc currents I in the same direction. One wire is at $x = 0$, the other at $x = d$, Fig. 28–41. Determine $\vec{\mathbf{B}}$ along the x axis between the wires as a function of x.

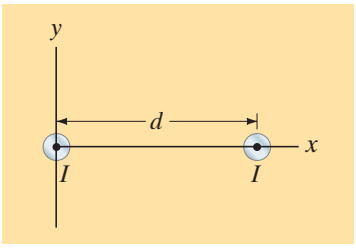


FIGURE 28–41
Problems 19 and 20.

20. (II) Repeat Problem 19 if the wire at $x = 0$ carries twice the current ($2I$) as the other wire, and in the opposite direction.

21. (II) Two long wires are oriented so that they are perpendicular to each other. At their closest, they are 20.0 cm apart (Fig. 28–42). What is the magnitude of the magnetic field at a point midway between them if the top one carries a current of 18.0 A and the bottom one carries 12.0 A?

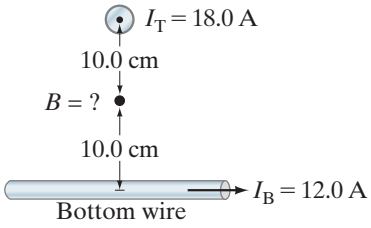


FIGURE 28–42
Problem 21.

22. (II) Two long parallel wires are 15 cm apart. The dc current in wire A is 20.0 A and that in B is 30.0 A. (*a*) In what relative direction (same or opposite) must the currents flow for there to be a point between the wires where $\vec{\mathbf{B}}$ is zero? (*b*) How far from wire A is that point?

23. (II) Two long parallel wires 8.20 cm apart carry 19.5-A dc currents in the same direction. Determine the magnetic field vector at a point P, 12.0 cm from one wire and 13.0 cm from the other. See Fig. 28–43. [*Hint*: Use the law of cosines. See Appendix A or inside rear cover.]

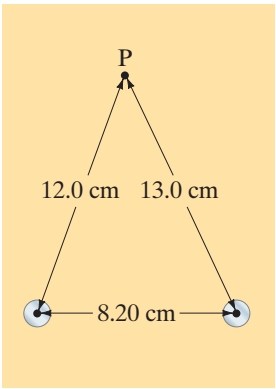


FIGURE 28–43
Problem 23.

24. (III) A very long flat conducting strip of width ℓ and negligible thickness lies in a horizontal plane and carries a uniform current I across its cross section. (*a*) Show that at points a distance y directly above its center, the field is given by

$$B = \frac{\mu_0 I}{\pi \ell} \tan^{-1} \frac{\ell}{2y},$$

assuming the strip is infinitely long. [*Hint*: Divide the strip into many thin "wires," and sum (integrate) over these.] (*b*) What value does B approach for $y \gg \ell$? Does this make sense? Explain.

25. (III) A triangular loop of side length a carries a current I (Fig. 28–44). If this loop is placed a distance d away from a very long straight wire carrying a current I', determine the force on the loop due to I'.

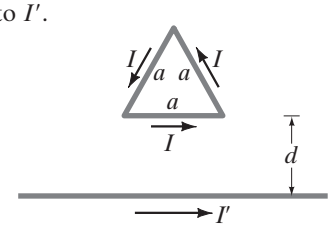


FIGURE 28–44
Problem 25.

28–3 Definitions of Ampere and Coulomb

26. (I) Consider the fundamental units to be kilograms, meters, seconds, and amperes. Write the following units in terms of these fundamental units: (*a*) coulombs, (*b*) volts, (*c*) teslas, (*d*) ohms.

28–4 and 28–5 Ampère's Law; Solenoids and Toroids

27. (I) A 40.0-cm-long solenoid 1.35 cm in diameter is to produce a field of 0.385 mT at its center. How much current should the solenoid carry if it has 675 turns of wire?

28. (I) A 32-cm-long solenoid, 1.8 cm in diameter, is to produce a 0.30-T magnetic field at its center. If the maximum current is 5.4 A, how many turns must the solenoid have?

29. (I) A 2.5-mm-diameter copper wire carries a 28-A dc current (uniform across its cross section). Determine the magnetic field: (*a*) at the surface of the wire; (*b*) inside the wire, 0.50 mm below the surface; (*c*) outside the wire 2.5 mm from the surface.

30. (II) A toroid (Fig. 28–17) has a 50.0-cm inner diameter and a 54.0-cm outer diameter. It carries a 25.0-A dc current in its 785 loops. Determine the range of values for B inside the toroid.

31. (II) A 20.0-m-long copper wire, 2.00 mm in diameter including insulation, is tightly wrapped in a single layer with adjacent loops touching, to form a solenoid of diameter 2.50 cm (outer edge). What is (*a*) the length of the solenoid and (*b*) the field at the center when the current in the wire is 15.0 A?

32. (II) (*a*) Use Eq. 28–1, and the vector nature of $\vec{\mathbf{B}}$, to show that the magnetic field lines around two long parallel wires carrying equal currents $I_1 = I_2$ are as shown in Fig. 28–10. (*b*) Draw the equipotential lines around two stationary positive electric charges. (*c*) Are these two diagrams similar? Identical? Why or why not?

33. (II) A 650-turn horizontal solenoid is 15 cm long. The current in its loops is 38 A. A straight wire cuts through the center of the solenoid, along a 3.0-cm diameter. This wire carries a 22-A current downward (and is connected by other wires that don't concern us). What is the force on this wire assuming the solenoid's field points due east?

34. (III) A coaxial cable consists of a solid inner conductor of radius R_1, surrounded by a concentric cylindrical tube of inner radius R_2 and outer radius R_3 (Fig. 28–45). The conductors carry equal and opposite currents I_0 distributed uniformly across their cross sections. Determine the magnetic field at a distance R from the axis for: (*a*) $R < R_1$; (*b*) $R_1 < R < R_2$; (*c*) $R_2 < R < R_3$; (*d*) $R > R_3$. (*e*) Let $I_0 = 1.50$ A, $R_1 = 1.00$ cm, $R_2 = 2.00$ cm, and $R_3 = 2.50$ cm. Graph B from $R = 0$ to $R = 3.00$ cm.

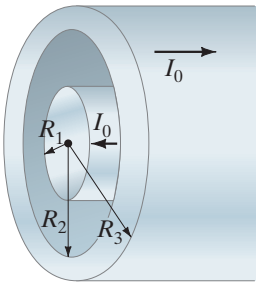


FIGURE 28–45
Problems 34 and 35.

35. (III) Suppose the current in the coaxial cable of Problem 34, Fig. 28–45, is not uniformly distributed, but instead the current density j varies linearly with distance from the center: $j_1 = C_1 R$ for the inner conductor and $j_2 = C_2 R$ for the outer conductor. Each conductor still carries the same total current I_0, in opposite directions. Determine the magnetic field in terms of I_0 in the same four regions of space as in Problem 34.

28–6 Biot-Savart Law

36. (I) The Earth's magnetic field is essentially that of a magnetic dipole. If the field near the North Pole is about 1.0×10^{-4} T, what will it be (approximately) 13,000 km above the surface at the North Pole?

37. (II) A wire, in a plane, has the shape shown in Fig. 28–46, two arcs of a circle connected by radial lengths of wire. Determine $\vec{\mathbf{B}}$ at point C in terms of R_1, R_2, θ, and the current I.

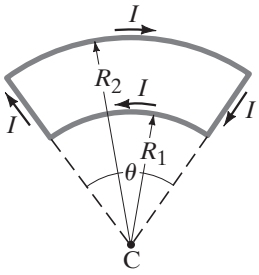


FIGURE 28–46
Problem 37.

38. (II) A circular conducting ring of radius R is connected to two exterior straight wires at two ends of a diameter (Fig. 28–47). The current I splits into unequal portions as shown (unequal resistance) while passing through the ring. What is $\vec{\mathbf{B}}$ at the center of the ring?

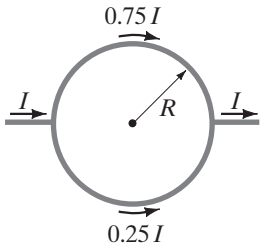


FIGURE 28–47
Problem 38.

39. (II) A small loop of wire of radius 1.8 cm is placed at the center of a wire loop with radius 25.0 cm. The planes of the loops are perpendicular to each other, and a 6.0-A current flows in each. Estimate the torque the large loop exerts on the smaller one. What simplifying assumption did you make?

40. (II) A wire is formed into the shape of two half circles connected by equal-length straight sections as shown in Fig. 28–48. A current I flows in the circuit clockwise as shown. Determine (*a*) the magnitude and direction of the magnetic field at the center, C, and (*b*) the magnetic dipole moment of the circuit.

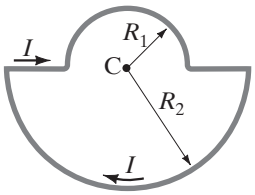


FIGURE 28–48
Problem 40.

41. (II) A single point charge q is moving with velocity $\vec{\mathbf{v}}$. Show that the Biot-Savart law predicts the magnetic field $\vec{\mathbf{B}}$ produced at a point P, whose position vector relative to the charge is $\vec{\mathbf{r}}$ (Fig. 28–49), is given by

$$\vec{\mathbf{B}} = \frac{\mu_0}{4\pi}\frac{q\vec{\mathbf{v}} \times \vec{\mathbf{r}}}{r^3}.$$

This result is only an estimate, and a reasonable one only for low speeds v. (See Section 28–7.)

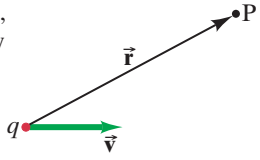


FIGURE 28–49
Problem 41.

42. (II) A nonconducting circular disk, of radius R, carries a uniformly distributed electric charge Q. The plate is set spinning with angular velocity ω about an axis perpendicular to the plate through its center (Fig. 28–50). Determine (*a*) its magnetic dipole moment and (*b*) the magnetic field at points on its axis a distance x from its center; (*c*) does Eq. 28–7b apply in this case for $x \gg R$?

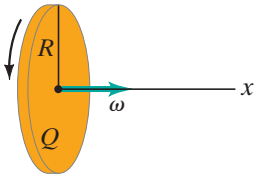


FIGURE 28–50
Problem 42.

43. (II) Consider a straight section of wire of length d, as in Fig. 28–51, which carries a current I. (*a*) Show that the magnetic field at a point P a distance R from the wire along its perpendicular bisector is

$$B = \frac{\mu_0 I}{2\pi R}\frac{d}{(d^2 + 4R^2)^{\frac{1}{2}}}.$$

(*b*) Show that this is consistent with Example 28–10 for an infinite wire.

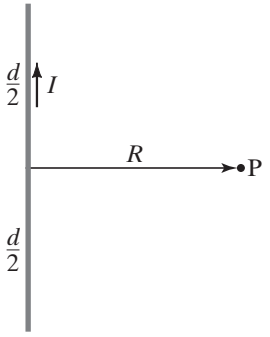


FIGURE 28–51
Problem 43.

44. (II) A segment of wire of length ℓ carries a dc current I as shown in Fig. 28–52. (*a*) Show that for points along the positive x axis (the axis of the wire), such as point Q, the magnetic field $\vec{\mathbf{B}}$ is zero. (*b*) Determine a formula for the field at points along the y axis, such as point P.

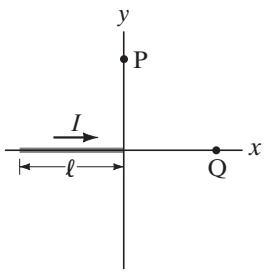


FIGURE 28–52
Problem 44.

45. (III) Use the result of Problem 44 to find the magnetic field at point P in Fig. 28–53 due to the current in the square loop.

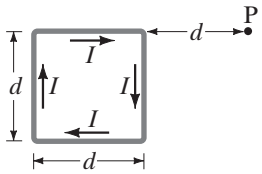


FIGURE 28–53 Problem 45.

46. (III) A wire is bent into the shape of a regular polygon with n sides whose vertices are a distance R from the center. (See Fig. 28–54, which shows the special case of $n = 6$.) If the wire carries a current I_0, (*a*) determine the magnetic field at the center; (*b*) if n is allowed to become very large ($n \to \infty$), show that the formula in part (*a*) reduces to that for a circular loop (Example 28–11).

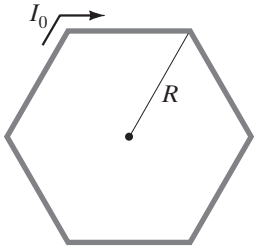


FIGURE 28–54 Problem 46.

47. (III) Start with the result of Example 28–11 for the magnetic field along the axis of a single circular loop of wire carrying a dc current I to obtain the field along the central axis inside a very long solenoid with n turns per meter (Eq. 28–4) that stretches from $+\infty$ to $-\infty$.

48. (III) A single rectangular loop of wire, with sides a and b, carries a dc current I. An xy coordinate system has its origin at the lower left corner of the rectangle with the x axis parallel to side b (Fig. 28–55) and the y axis parallel to side a. Determine the magnetic field B at all points (x, y) within the loop.

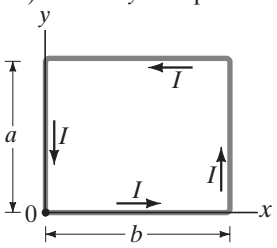


FIGURE 28–55
Problem 48.

49. (III) A square loop of wire, of side d, carries a current I. (*a*) Determine the magnetic field B at points on a line (call it the x axis) perpendicular to the plane of the square which passes through the center of the square (Fig. 28–56). Express B as a function of x, the distance from the center of the square. (*b*) For $x \gg d$, does the square appear to be a magnetic dipole? If so, what is its dipole moment?

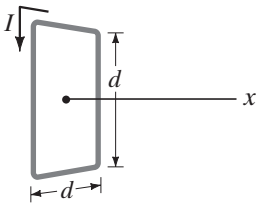


FIGURE 28–56
Problem 49.

28–8 Magnetic Materials—Ferromagnetism

50. (II) An iron atom has a magnetic dipole moment of about $1.8 \times 10^{-23}\ \text{A} \cdot \text{m}^2$. (*a*) Determine the dipole moment of an iron bar 15 cm long, 1.2 cm wide, and 1.0 cm thick, if it is 100 percent saturated. (*b*) What torque would be exerted on this bar when placed in a 0.80-T field acting at right angles to the bar?

28–10 Magnetic Materials; Hysteresis

51. (I) A long thin iron-core solenoid has 350 loops of wire per meter, and a 350-mA current flows through the wire. If the permeability of the iron is $3000\mu_0$, what is the total field B inside the solenoid?

52. (I) A large thin toroid has 315 loops of wire per meter wrapped around an iron core. A 3.0-A dc current flows through the wire. If the relative permeability of the iron is $\mu/\mu_0 = 2200$, what is the total field B inside the toroid?

53. (II) The following are some values of B and B_0 for a piece of annealed iron as it is being magnetized (note different units):

$B_0(10^{-4}\,\text{T})$	0.0	0.13	0.25	0.50	0.63	0.78	1.0	1.3
B(T)	0.0	0.0042	0.010	0.028	0.043	0.095	0.45	0.67
$B_0(10^{-4}\,\text{T})$	1.9	2.5	6.3	13.0	130	1300	10,000	
B(T)	1.01	1.18	1.44	1.58	1.72	2.26	3.15	

Determine the magnetic permeability μ for each value and plot a graph of μ versus B_0.

54. (II) An iron-core solenoid is 38 cm long and 1.8 cm in diameter, and has 640 turns of wire. The magnetic field inside the solenoid is 2.2 T when 42 A flows in the wire. What is the permeability μ at this high field strength?

General Problems

55. Three long parallel wires are 3.5 cm from one another. (Looking along them, they are at three corners of an equilateral triangle.) The current in each wire is 9.50 A, but its direction in wire M is opposite to that in wires N and P (Fig. 28–57). Determine the magnetic force per unit length on each wire due to the other two.

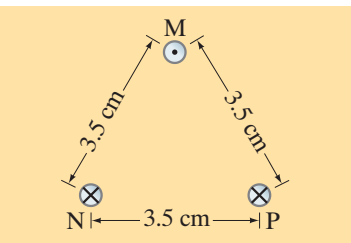


FIGURE 28–57
Problems 55, 56, and 57.

56. In Fig. 28–57, determine the magnitude and direction of the magnetic field midway between points M and N.

57. In Fig. 28–57 the top wire is 1.00-mm-diameter copper wire and is suspended in air due to the two magnetic forces from the bottom two wires. The current is 35.0 A in each of the two bottom wires. Calculate the required current in the suspended wire (M).

58. Two long straight parallel wires are 18 cm apart. Wire A carries 3.0-A current. Wire B's current is 6.0 A in the same direction. (*a*) Determine the magnetic field due to wire A at the position of wire B. (*b*) Determine the magnetic field due to wire B at the position of wire A. (*c*) Are these two magnetic fields equal and opposite? Why or why not? (*d*) Determine the force on wire A due to wire B, and the force on wire B due to wire A. Are these two forces equal and opposite? Why or why not?

59. A rectangular loop of wire carries a 2.0-A dc current and lies in a plane which also contains a very long straight wire carrying a 10.0-A current as shown in Fig. 28–58. Determine (*a*) the net force and (*b*) the net torque on the loop due to the straight wire.

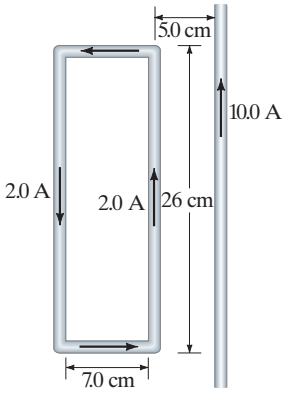


FIGURE 28–58
Problem 59.

60. A very large flat conducting sheet of thickness t carries a uniform current density $\vec{\mathbf{j}}$ throughout (Fig. 28–59). Determine the magnetic field (magnitude and direction) at a distance y above the plane. (Assume the plane is infinitely long and wide.)

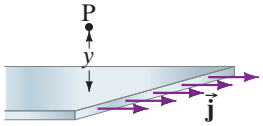


FIGURE 28–59
Problem 60.

61. A long horizontal wire carries a current of 42 A. A second wire, made of 1.00-mm-diameter copper wire and parallel to the first, is kept in suspension magnetically 5.0 cm below (Fig. 28–60). (*a*) Determine the magnitude and direction of the current in the lower wire. (*b*) Is the lower wire in stable equilibrium? (*c*) Repeat parts (*a*) and (*b*) if the second wire is suspended 5.0 cm *above* the first due to the first's magnetic field.

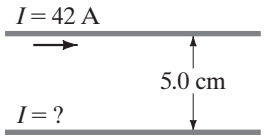


FIGURE 28–60
Problem 61.

62. **Helmholtz coils** are two identical circular coils having the same radius R and the same number of turns N, separated by a distance equal to the radius R and carrying the same dc current I in the same direction. (See Fig. 28–61.) They are used in scientific instruments to generate nearly uniform magnetic fields. (They can be seen in the photo, Fig. 27–19.) (*a*) Determine the magnetic field B at points x along the line joining their centers. Let $x = 0$ at the center of one coil, and $x = R$ at the center of the other. (*b*) Show that the field midway between the coils is particularly uniform by showing that $\frac{dB}{dx} = 0$ and $\frac{d^2B}{dx^2} = 0$ at the midpoint between the coils. (*c*) If $R = 10.0\,\text{cm}$, $N = 85$ turns and $I = 3.0\,\text{A}$, what is the field at the midpoint between the coils, $x = R/2$?

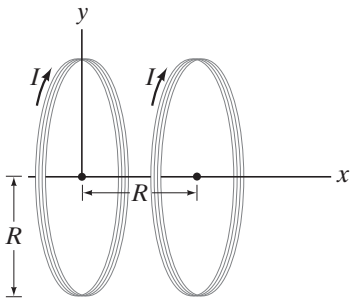


FIGURE 28–61
Problems 62 and 63.

63. A set of Helmholtz coils (see Problem 62, Fig. 28–61) have a radius $R = 10.0\,\text{cm}$ and are separated by a distance $R = 10.0\,\text{cm}$. Each coil has 85 loops carrying a current $I = 2.0\,\text{A}$. (*a*) Determine the total magnetic field B along the x axis (the center line for the two coils) in steps of 0.2 cm from the center of one coil $(x = 0)$ to the center of the other $(x = R)$. (*b*) Graph B as a function of x. (*c*) By what % does B vary from $x = 5.0\,\text{cm}$ to $x = 6.0\,\text{cm}$?

64. Near the Earth's poles the magnetic field is about 1 G $(1 \times 10^{-4}\,\text{T})$. Imagine a simple model in which the Earth's field is produced by a single current loop around the equator. Estimate roughly the current this loop would carry.

65. A 175-g model airplane charged to 18.0 mC and traveling at 2.8 m/s passes within 7.8 cm of a wire, nearly parallel to its path, carrying a 25-A dc current. What acceleration (in g's) does this interaction give the airplane?

66. Four long straight parallel wires located at the corners of a square of side d carry equal currents I_0 perpendicular to the page as shown in Fig. 28–62. Determine the magnitude and direction of $\vec{\mathbf{B}}$ at the center C of the square.

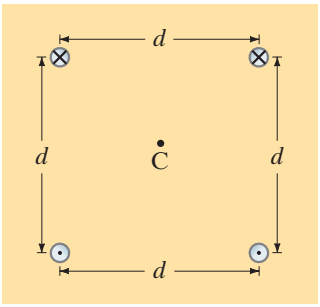


FIGURE 28–62
Problem 66.

67. Determine the magnetic field at a point P due to a very long wire with a square bend as shown in Fig. 28–63. The point P is halfway between the two corners.

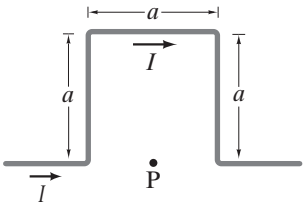


FIGURE 28–63
Problem 67.

68. A thin 12-cm-long solenoid has a total of 370 turns of wire and carries a current of 2.0 A. Calculate the field inside the solenoid near the center.

69. You have 1.0 kg of copper and want to make a practical solenoid that produces the greatest possible magnetic field for a given voltage. Should you make your copper wire long and thin, short and fat, or something else? Consider other variables, such as solenoid diameter, length, and so on. Explain your reasoning.

70. A small solenoid (radius r_a) is inside a larger solenoid (radius $r_b > r_a$). They are coaxial with n_a and n_b turns per unit length, respectively. The solenoids carry the same current, but in opposite directions. Let r be the radial distance from the common axis of the solenoids. If the magnetic field inside the inner solenoid $(r < r_a)$ is to be in the opposite direction as the field between the solenoids $(r_a < r < r_b)$, but have half the magnitude, determine the required ratio n_b/n_a.

71. Find B at the center of the 5.0-cm-radius semicircle (point P) in Fig. 28–64. The straight wires extend a great distance outward to the left and carry a current $I = 9.0\,\text{A}$.



FIGURE 28–64
Problem 71.

72. The design of a magneto-optical atom trap requires a magnetic field B that is directly proportional to position x along an axis. Such a field perturbs the absorption of laser light by atoms in the manner needed to spatially confine atoms in the trap. Let us demonstrate that "anti-Helmholtz" coils will provide the required field $B = Cx$, where C is a constant. Anti-Helmholtz coils consist of two identical circular wire coils, each with radius R and N turns, carrying current I in opposite directions (Fig. 28–65). The coils share a common axis (defined as the x axis with $x = 0$ at the midpoint (0) between the coils). Assume that the centers of the coils are separated by a distance equal to the radius R of the coils. (*a*) Show that the magnetic field at position x along the x axis is given by

$$B(x) = \frac{4\mu_0 NI}{R}\left\{\left[4+\left(1-\frac{2x}{R}\right)^2\right]^{-\frac{3}{2}} - \left[4+\left(1+\frac{2x}{R}\right)^2\right]^{-\frac{3}{2}}\right\}.$$

(*b*) For small excursions from the origin where $|x| \ll R$, show that the magnetic field is given by $B \approx Cx$, where the constant $C = 48\,\mu_0 NI/25\sqrt{5}\,R^2$. (*c*) For optimal atom trapping, dB/dx should be about 0.15 T/m. Assume an atom trap uses anti-Helmholtz coils with $R = 4.0$ cm and $N = 150$. What current should flow through the coils? [Coil separation equal to coil radius, as assumed in this Problem, is not a strict requirement for anti-Helmholtz coils.]

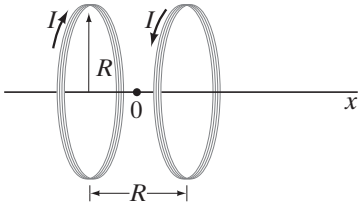


FIGURE 28–65
Problem 72.

73. A bolt of lightning strikes a metal flag pole, one end of which is anchored in the ground. Estimate the force the Earth's magnetic field can exert on the flag pole while the lightning-induced current flows. See Example 25–10.

74. You want to get an idea of the magnitude of magnetic fields produced by overhead power lines. You estimate that a transmission wire is about 12 m above the ground. The local power company tells you that the lines operate at 145 kV and provide a maximum of 45 MW to the local area. Estimate the maximum magnetic field you might experience walking under one such power line, and compare to the Earth's field. [For an ac current, values are rms, and the magnetic field will be changing.]

75. A circular current loop of radius 15 cm containing 220 turns carries a current of 2.0 A. Its center is at the origin and its axis lies along the x axis. Calculate the magnetic field B at a point x on the x axis for $x = -40$ cm to $+40$ cm in steps of 2 cm and make a graph of B as a function of x.

76. Suppose that an electromagnet uses a coil 2.0 m in diameter made from square copper wire 2.0 mm on a side. The power supply produces 35 V at a maximum power output of 1.5 kW. (*a*) How many turns are needed to run the power supply at maximum power? (*b*) What is the magnetic field strength at the center of the coil? (*c*) If you use a greater number of turns and this same power supply, will a greater magnetic field result? Explain.

77. Two long straight aluminum wires, each of diameter 0.42 mm, carry the same current but in opposite directions. They are suspended by 0.50-m-long strings as shown in Fig. 28–66. If the suspension strings make an angle of 3.0° with the vertical and are hanging freely, what is the current in the wires?

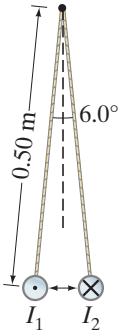


FIGURE 28–66
Problem 77.

78. A small but rigid U-shaped wire carrying a 5.0-A current (Fig. 28–67) is placed inside a solenoid. The solenoid is 15.0 cm long and has 600 loops of wire, and the current in each loop is 7.0 A. What is the net force on the U-shaped wire?

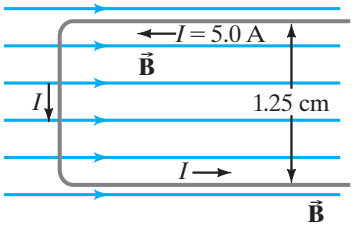


FIGURE 28–67
Problem 78.

79. Part of a long, thin insulated straight wire is formed into a single circular loop of radius R (Fig. 28–68) and carries a current I. (*a*) What is the magnitude and direction of the magnetic field at the center of the loop? (*b*) If the plane of the loop is twisted 90 degrees so that the plane is perpendicular to the straight part of the wire (i.e., in the yz plane) what is the magnitude and direction of the field now at the center of the loop?

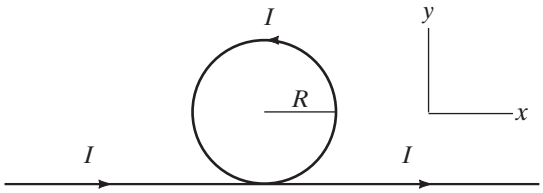


FIGURE 28–68 Problem 79.

80. A toroid is fabricated with a circular shape and loops with a square cross section as shown in Fig. 28–69. The cross section of a loop is a square of side 6.0 cm. The inner radius of the whole circular toroid is 3.0 m. There are 320 loops of wire which carry a 45-A dc current using a nearby power supply at 20.0 V. The arrows show the current flow in and out of the toroid. The current flows up at the inner diameter and down at the outer diameter. (*a*) Calculate the strength of the magnetic field at the center of the square's cross section at point P. (*b*) Is the magnetic field pointing clockwise or counterclockwise? (*c*) The square cross sectional area of the *wire* is uniformly 0.10 cm^2. What is the resistivity of the wire?

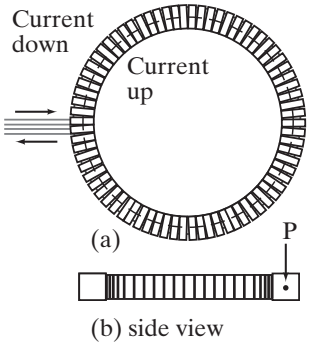


FIGURE 28–69
Problem 80.

ANSWERS TO EXERCISES

A: 2×10^{-6} T; not at this distance, and then only briefly.
B: 0.8×10^{-5} T, up.
C: (*b*).
D: 1.0×10^{-2} T.
E: (*b*), (*c*).

One of the great laws of physics is Faraday's law of induction, which says that a changing magnetic flux produces an induced emf. This photo shows a bar magnet moving into (or out of) a coil of wire, and the galvanometer registers an induced current. This phenomenon of electromagnetic induction is the basis for many practical devices, including generators, alternators, transformers, magnetic recording on tape or disk (hard drive), and computer memory.



CHAPTER 29

Electromagnetic Induction and Faraday's Law

CONTENTS

29–1 Induced EMF
29–2 Faraday's Law of Induction; Lenz's Law
29–3 EMF Induced in a Moving Conductor
29–4 Electric Generators
29–5 Back EMF and Counter Torque; Eddy Currents
29–6 Transformers and Transmission of Power
29–7 A Changing Magnetic Flux Produces an Electric Field
*29–8 Information Storage: Magnetic and Semiconductor
*29–9 Applications of Induction: Microphone, Seismograph, GFCI

CHAPTER-OPENING QUESTION—Guess now!

In the photograph above, the bar magnet is inserted down into the coil of wire, and is left there for 1 minute; then it is pulled up and out from the coil. What would an observer watching the galvanometer see?

(a) No change (pointer stays on zero): without a battery there is no current to detect.
(b) A small current flows while the magnet is inside the coil of wire.
(c) A current spike as the magnet enters the coil, and then nothing.
(d) A current spike as the magnet enters the coil, and then a steady small current.
(e) A current spike as the magnet enters the coil, then nothing (pointer at zero), then a current spike in the opposite direction as the magnet exits the coil.

In Chapter 27, we discussed two ways in which electricity and magnetism are related: (1) an electric current produces a magnetic field; and (2) a magnetic field exerts a force on an electric current or on a moving electric charge. These discoveries were made in 1820–1821. Scientists then began to wonder: if electric currents produce a magnetic field, is it possible that a magnetic field can produce an electric current? Ten years later the American Joseph Henry (1797–1878) and the Englishman Michael Faraday (1791–1867) independently found that it was possible. Henry actually made the discovery first. But Faraday published his results earlier and investigated the subject in more detail. We now discuss this phenomenon and some of its world-changing applications including the electric generator.

29–1 Induced EMF

In his attempt to produce an electric current from a magnetic field, Faraday used an apparatus like that shown in Fig. 29–1. A coil of wire, X, was connected to a battery. The current that flowed through X produced a magnetic field that was intensified by the ring-shaped iron core around which the wire was wrapped. Faraday hoped that a strong steady current in X would produce a great enough magnetic field to produce a current in a second coil Y wrapped on the same iron ring.

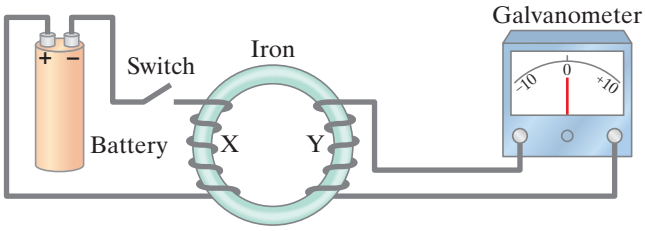


FIGURE 29–1 Faraday's experiment to induce an emf. Note that the galvanometer zero is at the center of the scale and the needle deflects left or right, depending on the direction of the current.

This second circuit, Y, contained a galvanometer to detect any current but contained no battery. He met no success with constant currents. But the long-sought effect was finally observed when Faraday noticed the galvanometer in circuit Y deflect strongly at the moment he closed the switch in circuit X. And the galvanometer deflected strongly in the opposite direction when he opened the switch in X. A constant current in X produced a constant magnetic field which produced *no* current in Y. Only when the current in X was starting or stopping was a current produced in Y.

Faraday concluded that although a constant magnetic field produces no current in a conductor, a *changing* magnetic field can produce an electric current. Such a current is called an **induced current**. When the magnetic field through coil Y changes, a current occurs in Y as if there were a source of emf in circuit Y. We therefore say that

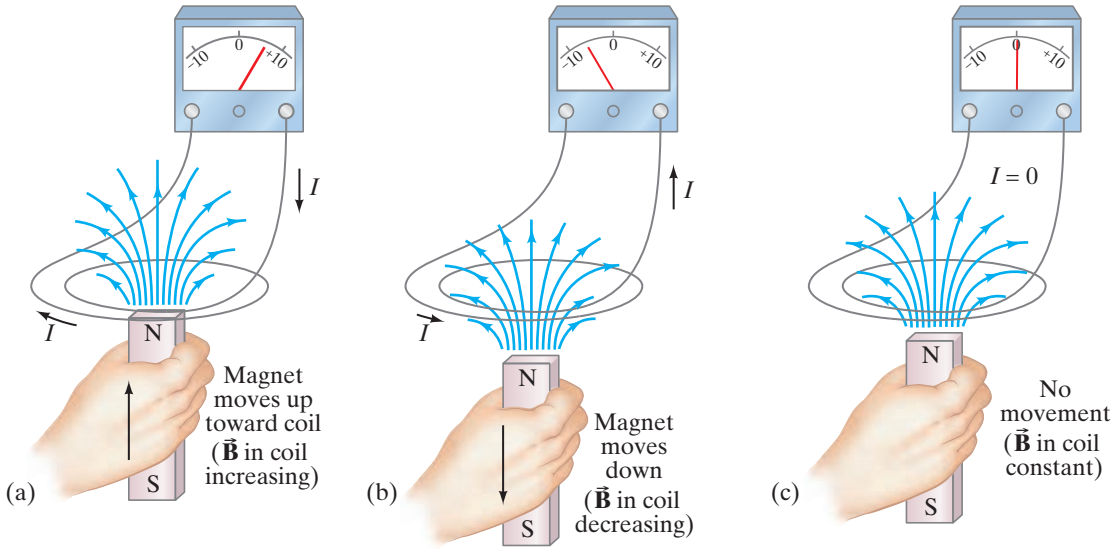
a changing magnetic field induces an emf.

CAUTION
Changing $\vec{\mathbf{B}}$, not $\vec{\mathbf{B}}$ itself, induces current

Faraday did further experiments on **electromagnetic induction**, as this phenomenon is called. For example, Fig. 29–2 shows that if a magnet is moved quickly into a coil of wire, a current is induced in the wire. If the magnet is quickly removed, a current is induced in the opposite direction ($\vec{\mathbf{B}}$ through the coil decreases). Furthermore, if the magnet is held steady and the coil of wire is moved toward or away from the magnet, again an emf is induced and a current flows. Motion or change is required to induce an emf. It doesn't matter whether the magnet or the coil moves. It is their *relative motion* that counts.

CAUTION
Relative motion—magnet or coil moving induces current

FIGURE 29–2 (a) A current is induced when a magnet is moved toward a coil, momentarily increasing the magnetic field through the coil. (b) The induced current is opposite when the magnet is moved away from the coil ($\vec{\mathbf{B}}$ decreases). In (c), no current is induced if the magnet does not move relative to the coil. It is the relative motion that counts here: the magnet can be held steady and the coil moved, which also induces an emf.



EXERCISE A Return to the Chapter-Opening Question, page 838, and answer it again now. Try to explain why you may have answered differently the first time.

29–2 Faraday's Law of Induction; Lenz's Law

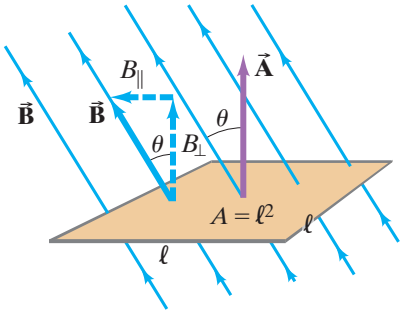


FIGURE 29–3 Determining the flux through a flat loop of wire. This loop is square, of side ℓ and area $A = \ell^2$.

FIGURE 29–4 Magnetic flux Φ_B is proportional to the number of lines of $\vec{\mathbf{B}}$ that pass through the loops of a coil (here with 3 loops).

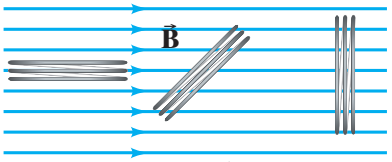
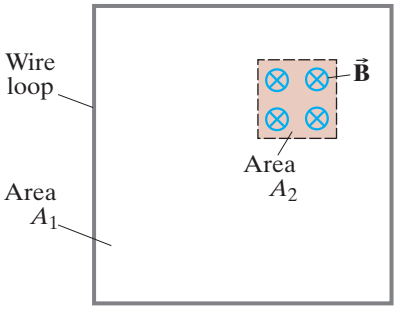


FIGURE 29–5 Example 29–1.



Faraday investigated quantitatively what factors influence the magnitude of the emf induced. He found first of all that the more rapidly the magnetic field changes, the greater the induced emf. He also found that the induced emf depends on the area of the circuit loop (and also the angle it makes with $\vec{\mathbf{B}}$). In fact, it is found that the emf is proportional to the rate of change of the **magnetic flux**, Φ_B, passing through the circuit or loop of area A. Magnetic flux for a uniform magnetic field through a loop of area A is defined in the same way we did for electric flux in Chapter 22, as

$$\Phi_B = B_\perp A = BA\cos\theta = \vec{\mathbf{B}}\cdot\vec{\mathbf{A}}. \qquad [\vec{\mathbf{B}} \text{ uniform}] \quad \textbf{(29–1a)}$$

Here $B_\perp$ is the component of the magnetic field $\vec{\mathbf{B}}$ perpendicular to the face of the loop, and θ is the angle between $\vec{\mathbf{B}}$ and the vector $\vec{\mathbf{A}}$ (representing the area) whose direction is perpendicular to the face of the loop. These quantities are shown in Fig. 29–3 for a square loop of side ℓ whose area is $A = \ell^2$. If the area is of some other shape, or $\vec{\mathbf{B}}$ is not uniform, the magnetic flux can be written

$$\Phi_B = \int \vec{\mathbf{B}}\cdot d\vec{\mathbf{A}}, \qquad \textbf{(29–1b)}$$

where the integral is taken over the area A.†

We saw in Chapter 27 that the lines of $\vec{\mathbf{B}}$ (like lines of $\vec{\mathbf{E}}$) can be drawn so that the number of lines per unit area is proportional to the field strength. Then the flux Φ_B can be thought of as being proportional to the *total number of lines passing through the area enclosed by the loop*. This is illustrated in Fig. 29–4, where three wire loops of a coil are viewed from the side (on edge). For $\theta = 90°$, no magnetic field lines pass through the loop and $\Phi_B = 0$, whereas Φ_B is a maximum when $\theta = 0°$. The unit of magnetic flux is the tesla-meter^2, and is called a **weber**: $1\ \text{Wb} = 1\ \text{T}\cdot\text{m}^2$.

CONCEPTUAL EXAMPLE 29–1 **Determining flux.** A square loop of wire encloses area A_1 as shown in Fig. 29–5. A uniform magnetic field $\vec{\mathbf{B}}$ perpendicular to the loop extends over the area A_2. What is the magnetic flux through the loop A_1?

RESPONSE We assume that the magnetic field is zero outside the area A_2. The total magnetic flux through area A_1 is the flux through area A_2, which by Eq. 29–1a for a uniform field is BA_2, plus the flux through the remaining area $(= A_1 - A_2)$, which is zero because $B = 0$. So the total flux is $\Phi_B = BA_2 + 0(A_1 - A_2) = BA_2$. It is *not* equal to BA_1 because $\vec{\mathbf{B}}$ is not uniform over A_1.

With our definition of flux, Eqs. 29–1, we can write down the results of Faraday's investigations: The emf $\mathscr{E}$ induced in a circuit is equal to the rate of change of magnetic flux through the circuit:

FARADAY'S LAW OF INDUCTION

$$\mathscr{E} = -\frac{d\Phi_B}{dt}. \qquad [1\ \text{loop}] \quad \textbf{(29–2a)}$$

This fundamental result is known as **Faraday's law of induction**, and is one of the basic laws of electromagnetism.

†The integral is taken through an ordinary area that is bounded by a closed curve, such as a circle or square. The area is not an entirely enclosed surface (like a football) as we used in Gauss's law, Chapter 22.

If the circuit contains N loops that are closely wrapped so the same flux passes through each, the emfs induced in each loop add together, so the total emf is

$$\mathscr{E} = -N\frac{d\Phi_B}{dt}. \qquad [N \text{ loops}] \quad \textbf{(29–2b)}$$

FARADAY'S LAW OF INDUCTION

EXAMPLE 29–2 **A loop of wire in a magnetic field.** A square loop of wire of side $\ell = 5.0\,\text{cm}$ is in a uniform magnetic field $B = 0.16\,\text{T}$. What is the magnetic flux in the loop (*a*) when $\vec{\mathbf{B}}$ is perpendicular to the face of the loop and (*b*) when $\vec{\mathbf{B}}$ is at an angle of 30° to the area $\vec{\mathbf{A}}$ of the loop? (*c*) What is the magnitude of the average current in the loop, if it has a resistance of 0.012 Ω, while it is rotated from position (*b*) to position (*a*) in 0.14 s?

APPROACH We use the definition $\Phi_B = \vec{\mathbf{B}} \cdot \vec{\mathbf{A}}$ to calculate the magnetic flux. Then we use Faraday's law of induction to find the induced emf in the coil, and from that the induced current $(I = \mathscr{E}/R)$.

SOLUTION The area of the coil is $A = \ell^2 = (5.0 \times 10^{-2}\,\text{m})^2 = 2.5 \times 10^{-3}\,\text{m}^2$, and the direction of $\vec{\mathbf{A}}$ is perpendicular to the face of the loop (Fig. 29–3).
(*a*) $\vec{\mathbf{B}}$ is perpendicular to the coil's face, and thus parallel to $\vec{\mathbf{A}}$ (Fig. 29–3), so

$$\Phi_B = \vec{\mathbf{B}} \cdot \vec{\mathbf{A}}$$
$$= BA\cos 0° = (0.16\,\text{T})(2.5 \times 10^{-3}\,\text{m}^2)(1) = 4.0 \times 10^{-4}\,\text{T}\cdot\text{m}^2$$

or $4.0 \times 10^{-4}\,\text{Wb}$.
(*b*) The angle θ between $\vec{\mathbf{B}}$ and $\vec{\mathbf{A}}$ is 30°, and $\cos 30° = 0.866$, so

$$\Phi_B = \vec{\mathbf{B}} \cdot \vec{\mathbf{A}}$$
$$= BA\cos\theta = (0.16\,\text{T})(2.5 \times 10^{-3}\,\text{m}^2)\cos 30° = 3.5 \times 10^{-4}\,\text{T}\cdot\text{m}^2$$

or $3.5 \times 10^{-4}\,\text{Wb}$, a bit less than in part (*a*).
(*c*) The magnitude of the induced emf (Eq. 29–2a) during the 0.14-s time interval is

$$\mathscr{E} = \frac{\Delta\Phi_B}{\Delta t} = \frac{(4.0 \times 10^{-4}\,\text{T}\cdot\text{m}^2) - (3.5 \times 10^{-4}\,\text{T}\cdot\text{m}^2)}{0.14\,\text{s}} = 3.6 \times 10^{-4}\,\text{V}.$$

Before and after the loop rotates, when it is at rest, the emf is zero. The current in the wire loop (Ohm's law) while it is rotating averages

$$I = \frac{\mathscr{E}}{R} = \frac{3.6 \times 10^{-4}\,\text{V}}{0.012\,\Omega} = 0.030\,\text{A} = 30\,\text{mA}.$$

The minus signs in Eqs. 29–2a and b are there to remind us in which direction the induced emf acts. Experiments show that

a current produced by an induced emf creates a magnetic field that opposes the original change in magnetic flux.

This is known as **Lenz's law**. Be aware that we are now discussing two distinct magnetic fields: (1) the changing magnetic field or flux that induces the current, and (2) the magnetic field produced by the induced current (all currents produce a magnetic field). The second (induced) field opposes the *change* in the first.

⚠ **CAUTION**
Distinguish two different magnetic fields

Lenz's law can be said another way:

an induced emf is always induced in a direction that opposes the original change in flux that caused it,

and this is valid even if no current can flow (as when a circuit is not complete).

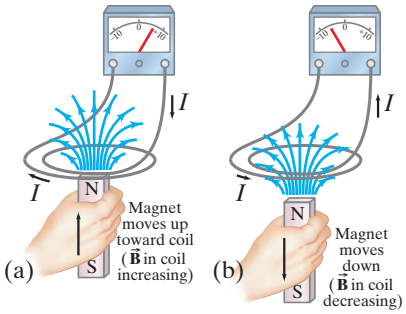


FIGURE 29–2 (Repeated.)

Let us apply Lenz's law to the relative motion between a magnet and a coil, Fig. 29–2 on page 839 (repeated here, but smaller, in the margin). The changing flux through the coil when the magnet moves in or out induces an emf in the coil, producing a current. This induced current produces its own magnetic field. In Fig. 29–2a the distance between the coil and the magnet decreases. The magnet's magnetic field (and number of field lines) through the coil increases, and therefore the flux increases. The magnetic field of the magnet points upward. To oppose the upward increase of the magnet's field through the coil, the new induced magnetic field produced by the induced current needs to point *downward* inside the coil. Thus, Lenz's law tells us the current moves as shown in Fig. 29–2a (use the right-hand rule). In Fig. 29–2b, the flux *decreases* (because the magnet is moved away and B through the coil decreases), so the induced current in the coil produces an *upward* magnetic field through the coil that is "trying" to maintain the status quo. Thus the current in Fig. 29–2b is in the opposite direction from Fig. 29–2a.

It is important to note that an emf is induced whenever there is a change in *flux* through the coil, and we now consider some more possibilities.

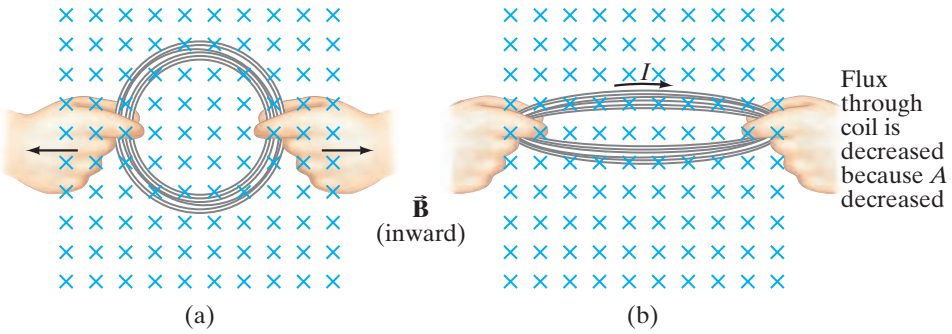


FIGURE 29–6 A current can be induced by changing the coil's area, even though B doesn't change. Here the area A is reduced by pulling on the sides of the coil: the *flux* through the coil is reduced as we go from (a) to (b). The brief induced current acts in the direction shown so as to try to maintain the original flux $(\Phi = BA)$ by producing its own magnetic field into the page. That is, as area A decreases, the current acts to increase B in the original (inward) direction.

Magnetic flux $\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = \int B \cos\theta \, dA$, so an emf can be induced in three ways: (1) by a changing magnetic field B; (2) by changing the area A of the loop in the field; or (3) by changing the loop's orientation θ with respect to the field. Figures 29–1 and 29–2 showed case 1. Examples of cases 2 and 3 are shown in Figs. 29–6 and 29–7.

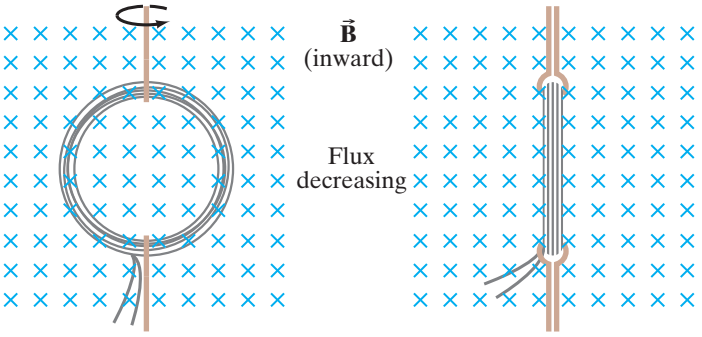


FIGURE 29–7 A current can be induced by rotating a coil in a magnetic field. The flux through the coil changes from (a) to (b) because θ (in Eq. 29–1a, $\Phi = BA\cos\theta$) went from $0°$ $(\cos\theta = 1)$ to $90°$ $(\cos\theta = 0)$.

FIGURE 29–8 Example 29–3: An induction stove, heating soup but not burning the hand.



CONCEPTUAL EXAMPLE 29–3 **Induction stove.** In an induction stove (Fig. 29–8), an ac current exists in a coil that is the "burner" (a burner that never gets hot). Why will it heat a metal pan, usually iron, but not a glass container?

RESPONSE The current is ac, so it induces a changing magnetic field that passes through the pan bottom. This changing magnetic field induces a current in the pan, and since the metal pan offers electric resistance, electric energy is transformed to thermal energy which heats the pan and its contents. If the pan is iron, magnetic hysteresis due to the changing current produces additional heating. A glass container offers such high resistance that little current is induced and little energy is transferred $(P = V^2/R)$.

PROBLEM SOLVING

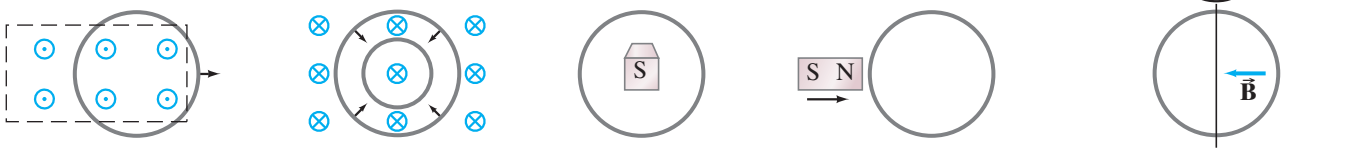
Lenz's Law

Lenz's law is used to determine the direction of the (conventional) electric current *induced* in a loop due to a change in magnetic flux inside the loop. (The loop may already be carrying its own ordinary current.) To produce an *induced* current you need

(a) a closed conducting loop, and

(b) an external magnetic flux through the loop that is changing in time.

1. Determine whether the magnetic flux ($\Phi_B = BA\cos\theta$) inside the loop is decreasing, increasing, or unchanged.
2. The magnetic field due to the *induced* current: (a) points in the same direction as the external field if the flux is decreasing; (b) points in the opposite direction from the external field if the flux is increasing; or (c) is zero if the flux is not changing.
3. Once you know the direction of the induced magnetic field, use the right-hand-rule-1 (page 791, Chapter 27) to find the direction of the induced current that would give this induced $\vec{\mathbf{B}}$.
4. Always keep in mind that there are two magnetic fields: (1) an external field whose flux must be changing if it is to induce an electric current, and (2) a magnetic field produced by the induced current.
5. If a wire is already carrying an electric current, the total current while the magnetic field is changing will be the algebraic sum of the original current and the induced current.



(a) Pulling a round loop to the right out of a magnetic field which points out of the page

(b) Shrinking a loop in a magnetic field pointing into the page

(c) S magnetic pole moving from below, up toward the loop

(d) N magnetic pole moving toward loop in the plane of the loop

(e) Rotating the loop by pulling the left side toward us and pushing the right side in; the magnetic field points from right to left

FIGURE 29–9 Example 29–4.

CONCEPTUAL EXAMPLE 29–4 **Practice with Lenz's law.** In which direction is the current induced in the circular loop for each situation in Fig. 29–9?

RESPONSE In (*a*), the magnetic field pointing out of the page initially passes through the entire loop. If you pull the loop out of the field, magnetic flux through the loop decreases (to zero). So the induced current will be in a direction to maintain the decreasing flux through the loop: the current will be counterclockwise to produce a magnetic field outward (toward the reader).

(*b*) The external field is into the page. The coil area gets smaller, so the flux will decrease; hence the induced current will be clockwise, producing its own field into the page to oppose the flux decrease.

(*c*) Magnetic field lines point into the S pole of a magnet, so as the magnet moves toward us and the loop, the magnet's field points into the page and is getting stronger. The current in the loop will be induced in the counterclockwise direction in order to produce a field $\vec{\mathbf{B}}$ *out* of the page.

(*d*) The magnetic field is in the plane of the loop, so no magnetic field lines pass through the loop and the flux through the loop is zero throughout the process; hence there is no change in flux and no induced emf or current in the loop.

(*e*) Initially there is no flux through the loop: the $\vec{\mathbf{B}}$ field is in the plane of the loop. When you start to rotate the loop, the external field through the loop begins increasing to the left. To counteract this change in flux, the loop will have current induced in a counterclockwise direction so as to produce its own field to the right.

CAUTION

Magnetic field created by induced current opposes change in external flux, not necessarily opposing the external field

EXERCISE B What is the direction of the induced current in the circular loop due to the current shown in each part of Fig. 29–10?



FIGURE 29–10 Exercise B.

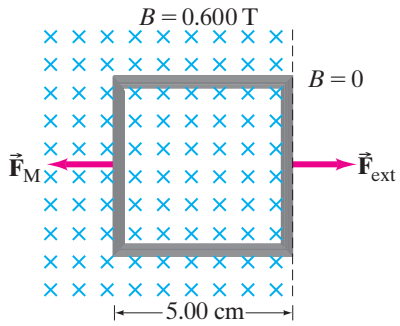


FIGURE 29–11 Example 29–5. The square coil in a magnetic field $B = 0.600\,\text{T}$ is pulled abruptly to the right to a region where $B = 0$. (The forces shown are discussed in the alternate solution at the end of Example 29–5.)

EXAMPLE 29–5 **Pulling a coil from a magnetic field.** A 100-loop square coil of wire, with side $\ell = 5.00\,\text{cm}$ and total resistance $R = 100\,\Omega$, is positioned perpendicular to a uniform magnetic field $B = 0.600\,\text{T}$, as shown in Fig. 29–11. It is quickly pulled from the field at constant speed (moving perpendicular to $\vec{\mathbf{B}}$) to a region where B drops abruptly to zero. At $t = 0$, the right edge of the coil is at the edge of the field. It takes 0.100 s for the whole coil to reach the field-free region. Determine (*a*) the rate of change in flux through one loop of the coil, and (*b*) the total emf and current induced in the 100-loop coil. (*c*) How much energy is dissipated in the coil? (*d*) What was the average force required (F_{ext})?

APPROACH We start by finding how the magnetic flux, $\Phi_B = BA\cos 0° = BA$, changes during the time interval $\Delta t = 0.100\,\text{s}$. Faraday's law then gives the induced emf and Ohm's law gives the current.

SOLUTION (*a*) The coil's area is $A = \ell^2 = (5.00 \times 10^{-2}\,\text{m})^2 = 2.50 \times 10^{-3}\,\text{m}^2$. The flux through one loop is initially $\Phi_B = BA = (0.600\,\text{T})(2.50 \times 10^{-3}\,\text{m}^2) = 1.50 \times 10^{-3}\,\text{T}\cdot\text{m}^2$. After 0.100 s, the flux is zero. The rate of change in flux is constant (because the coil is square), and for one loop is equal to

$$\frac{\Delta\Phi_B}{\Delta t} = \frac{0 - (1.50 \times 10^{-3}\,\text{T}\cdot\text{m}^2)}{0.100\,\text{s}} = -1.50 \times 10^{-2}\,\text{T}\cdot\text{m}^2/\text{s}.$$

(*b*) The emf induced (Eq. 29–2) in the 100-loop coil during this 0.100-s interval is

$$\mathscr{E} = -N\frac{\Delta\Phi_B}{\Delta t} = -(100)(-1.50 \times 10^{-2}\,\text{T}\cdot\text{m}^2/\text{s}) = 1.50\,\text{V}.$$

The current is found by applying Ohm's law to the 100-Ω coil:

$$I = \frac{\mathscr{E}}{R} = \frac{1.50\,\text{V}}{100\,\Omega} = 1.50 \times 10^{-2}\,\text{A} = 15.0\,\text{mA}.$$

By Lenz's law, the current must be clockwise to produce more $\vec{\mathbf{B}}$ into the page and thus oppose the decreasing flux into the page.

(*c*) The total energy dissipated in the coil is the product of the power ($= I^2R$) and the time:

$$E = Pt = I^2Rt = (1.50 \times 10^{-2}\,\text{A})^2(100\,\Omega)(0.100\,\text{s}) = 2.25 \times 10^{-3}\,\text{J}.$$

(*d*) We can use the result of part (*c*) and apply the work-energy principle: the energy dissipated E is equal to the work W needed to pull the coil out of the field (Chapters 7 and 8). Because $W = \bar{F}_{\text{ext}}\,d$ where $d = 5.00\,\text{cm}$, then

$$\bar{F}_{\text{ext}} = \frac{W}{d} = \frac{2.25 \times 10^{-3}\,\text{J}}{5.00 \times 10^{-2}\,\text{m}} = 0.0450\,\text{N}.$$

Alternate Solution (*d*) We can also calculate the force directly using $\vec{\mathbf{F}} = I\vec{\boldsymbol{\ell}} \times \vec{\mathbf{B}}$, Eq. 27–3, which here for constant $\vec{\mathbf{B}}$ is $F = I\ell B$. The force the magnetic field exerts on the top and bottom sections of the square coil of Fig. 29–10 are in opposite directions and cancel each other. The magnetic force $\vec{\mathbf{F}}_{\text{M}}$ exerted on the left vertical section of the square coil acts to the left as shown because the current is up (clockwise, as in part *b*). The right side of the loop is in the region where $\vec{\mathbf{B}} = 0$. Hence the external force, to the right, needed to just overcome the magnetic force to the left (on $N = 100$ loops) is

$$F_{\text{ext}} = NI\ell B = (100)(0.0150\,\text{A})(0.0500\,\text{m})(0.600\,\text{T}) = 0.0450\,\text{N},$$

which is the same answer, confirming our use of energy conservation above.

EXAMPLE 29–6 **Initial dc current in the coil.** Suppose the 100 loops of wire in Fig. 29–11, Example 29–5, are connected to an emf that produces a counterclockwise 60.0-mA dc current in the wire. When the square coil is pulled from the magnetic field, what then is the current in the wire?

SOLUTION As the square loop is pulled to the right, out of the magnetic field, a 15.0-mA clockwise current is induced in the wire as we saw in Example 29–5. So the current in the wire is $(60.0 - 15.0)\,\text{mA} = 45.0\,\text{mA}$. Once the whole coil is outside the magnetic field, the current returns to 60.0 mA.

29–3 EMF Induced in a Moving Conductor

Another way to induce an emf is shown in Fig. 29–12a, and this situation helps illuminate the nature of the induced emf. Assume that a uniform magnetic field $\vec{\mathbf{B}}$ is perpendicular to the area bounded by the U-shaped conductor and the movable rod resting on it. If the rod is made to move at a speed v to the right, it travels a distance $dx = v\,dt$ in a time dt. Therefore, the area of the loop increases by an amount $dA = \ell\,dx = \ell v\,dt$ in a time dt. By Faraday's law there is an induced emf $\mathcal{E}$ whose magnitude is given by

$$\mathcal{E} = \frac{d\Phi_B}{dt} = \frac{B\,dA}{dt} = \frac{B\ell v\,dt}{dt} = B\ell v. \qquad \textbf{(29–3)}$$

The induced current is clockwise (to counter the increasing flux).

Equation 29–3 is valid as long as B, ℓ, and v are mutually perpendicular. (If they are not, we use only the components of each that are mutually perpendicular.) An emf induced on a conductor moving in a magnetic field is sometimes called *motional emf*.

We can also obtain Eq. 29–3 without using Faraday's law. We saw in Chapter 27 that a charged particle moving perpendicular to a magnetic field B with speed v experiences a force $\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$ (Eq. 27–5a). When the rod of Fig. 29–12a moves to the right with velocity $\vec{\mathbf{v}}$, the electrons in the rod also move with this velocity. Therefore, since $\vec{\mathbf{v}} \perp \vec{\mathbf{B}}$, each electron feels a force $F = qvB$, which acts up the page as the red arrow in Fig. 29–12b shows. If the rod is not in contact with the U-shaped conductor, electrons would collect at the upper end of the rod, leaving the lower end positive (see signs in Fig. 29–12b). There must thus be an induced emf. If the rod is in contact with the U-shaped conductor (Fig. 29–12a), the electrons will flow into the U. There will then be a clockwise (conventional) current in the loop. To calculate the emf, we determine the work W needed to move a charge q from one end of the rod to the other against this potential difference: $W = \text{force} \times \text{distance} = (qvB)(\ell)$. The emf equals the work done per unit charge, so $\mathcal{E} = W/q = qvB\ell/q = B\ell v$, the same result† as from Faraday's law above, Eq. 29–3.

EXERCISE C In what direction will the electrons flow in Fig. 29–12 if the rod moves to the left, decreasing the area of the current loop?

EXAMPLE 29–7 **ESTIMATE** **Does a moving airplane develop a large emf?** An airplane travels 1000 km/h in a region where the Earth's magnetic field is about 5×10^{-5} T and is nearly vertical (Fig. 29–13). What is the potential difference induced between the wing tips that are 70 m apart?

APPROACH We consider the wings to be a 70-m-long conductor moving through the Earth's magnetic field. We use Eq. 29–3 to get the emf.

SOLUTION Since $v = 1000\,\text{km/h} = 280\,\text{m/s}$, and $\vec{\mathbf{v}} \perp \vec{\mathbf{B}}$, we have

$$\mathcal{E} = B\ell v = (5 \times 10^{-5}\,\text{T})(70\,\text{m})(280\,\text{m/s}) \approx 1\,\text{V}.$$

NOTE Not much to worry about.

EXAMPLE 29–8 **Electromagnetic blood-flow measurement.** The rate of blood flow in our body's vessels can be measured using the apparatus shown in Fig. 29–14, since blood contains charged ions. Suppose that the blood vessel is 2.0 mm in diameter, the magnetic field is 0.080 T, and the measured emf is 0.10 mV. What is the flow velocity v of the blood?

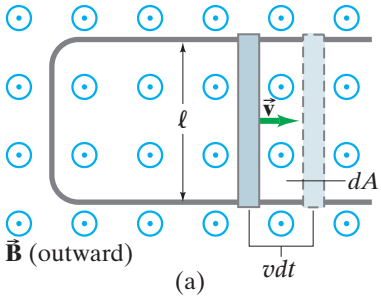
APPROACH The magnetic field $\vec{\mathbf{B}}$ points horizontally from left to right (N pole toward S pole). The induced emf acts over the width $\ell = 2.0\,\text{mm}$ of the blood vessel, perpendicular to $\vec{\mathbf{B}}$ and $\vec{\mathbf{v}}$ (Fig. 29–14), just as in Fig. 29–12. We can then use Eq. 29–3 to get v.

SOLUTION We solve for v in Eq. 29–3:

$$v = \frac{\mathcal{E}}{B\ell} = \frac{(1.0 \times 10^{-4}\,\text{V})}{(0.080\,\text{T})(2.0 \times 10^{-3}\,\text{m})} = 0.63\,\text{m/s}.$$

NOTE In actual practice, an alternating current is used to produce an alternating magnetic field. The induced emf is then alternating (ac).

†This force argument, which is basically the same as for the Hall effect (Section 27–8), explains this one way of inducing an emf. It does not explain the general case of electromagnetic induction.



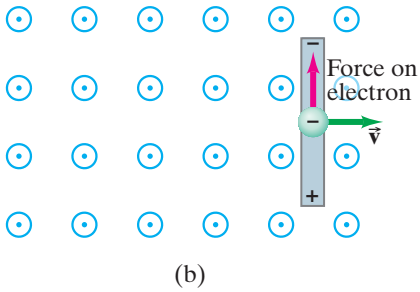


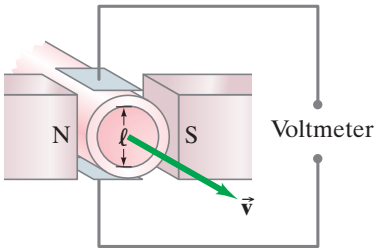
FIGURE 29–12 (a) A conducting rod is moved to the right on the smooth surface of a U-shaped conductor in a uniform magnetic field $\vec{\mathbf{B}}$ that points out of the page. The induced current is clockwise. (b) Force on an electron in the metal rod (moving to the right) is upward due to $\vec{\mathbf{B}}$ pointing out of the page. Hence electrons can collect at the top of the rod, leaving + charge at the bottom.

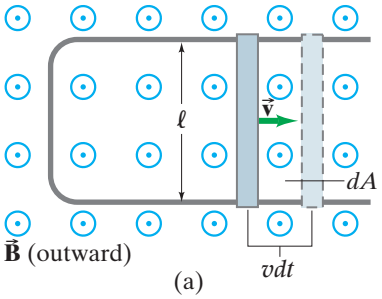
FIGURE 29–13 Example 29–7.



PHYSICS APPLIED
Blood-flow measurement

FIGURE 29–14 Measurement of blood velocity from the induced emf. Example 29–8.





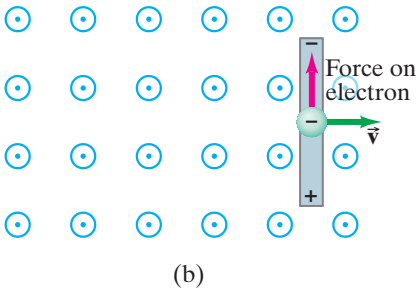


FIGURE 29–12 (Repeated.)

EXAMPLE 29–9 **Force on the rod.** To make the rod of Fig. 29–12a move to the right at constant speed v, you need to apply an external force on the rod to the right. (*a*) Explain and determine the magnitude of the required force. (*b*) What external power is needed to move the rod? (Do not confuse this external force on the rod with the upward force on the electrons shown in Fig. 29–12b.)

APPROACH When the rod moves to the right, electrons flow upward in the rod according to the right-hand rule. So the conventional current is downward in the rod. We can see this also from Lenz's law: the outward magnetic flux through the loop is increasing, so the induced current must oppose the increase. Thus the current is clockwise so as to produce a magnetic field into the page (right-hand rule), opposite to the outward $\vec{\mathbf{B}}$ shown. The magnetic force on the moving rod is $\vec{\mathbf{F}} = I\vec{\ell} \times \vec{\mathbf{B}}$ for a constant $\vec{\mathbf{B}}$ (Eq. 27–3). The right-hand rule tells us this magnetic force is to the left, and is thus a "drag force" opposing our effort to move the rod to the right.

SOLUTION (*a*) The magnitude of the external force, to the right, needs to balance the magnetic force $F = I\ell B$, to the left. The current $I = \mathscr{E}/R = B\ell v/R$ (see Eq. 29–3), and the resistance R is that of the whole circuit: the rod and the U-shaped conductor. The force F required to move the rod is thus

$$F = I\ell B = \left(\frac{B\ell v}{R}\right)\ell B = \frac{B^2\ell^2}{R}v.$$

If B, ℓ, and R are constant, then a constant speed v is produced by a constant external force. (Constant R implies that the parallel rails have negligible resistance.)

(*b*) The external power needed to move the rod for constant R is

$$P_{\text{ext}} = Fv = \frac{B^2\ell^2v^2}{R}.$$

The power dissipated in the resistance is $P = I^2R$. With $I = \mathscr{E}/R = B\ell v/R$,

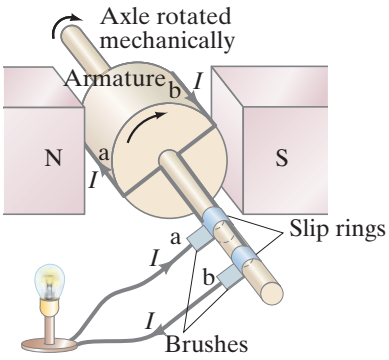
$$P_{\text{R}} = I^2R = \frac{B^2\ell^2v^2}{R},$$

so the power input equals the power dissipated in the resistance at any moment.

29–4 Electric Generators

We discussed alternating currents (ac) briefly in Section 25–7. Now we examine how ac is generated: by an **electric generator** or **dynamo**, an indispensable practical result of Faraday's discovery. A generator is the main way electricity is produced for practical purposes (some by batteries and solar cells), and it does so by transforming mechanical energy into electric energy (this is just the opposite of what a motor does).

FIGURE 29–15 An ac generator.



A simplified diagram of an **ac generator** is shown in Fig. 29–15. A generator consists of many loops of wire (only one is shown) wound on an **armature** that can rotate in a magnetic field. The axle is turned by some mechanical means (falling water at a dam, steam turbine, car motor belt), and an emf is induced in the rotating coil. An ac electric current is thus the *output* of a generator. Suppose in Fig. 29–15 that the armature is rotating clockwise; then $\vec{\mathbf{F}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}}$ applied to charged particles in the wire (or Lenz's law) tells us that the (conventional) current in the wire labeled b on the armature is outward, toward us; therefore the current is outward through brush b. Each brush is fixed and presses against a *continuous* slip ring that rotates with the armature. After one-half revolution, wire b will be where wire a is now in Fig. 29–15, and the current then at brush b will be inward. Thus the current produced is alternating (AC).

Let us assume the loop is being made to rotate in a uniform magnetic field $\vec{\mathbf{B}}$ with constant angular velocity ω. From Faraday's law (Eq. 29–2a), the induced emf is

$$\mathscr{E} = -\frac{d\Phi_B}{dt} = -\frac{d}{dt}\int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = -\frac{d}{dt}[BA\cos\theta]$$

where A is the area of the loop and θ is the angle between $\vec{\mathbf{B}}$ and $\vec{\mathbf{A}}$.

Since $\omega = d\theta/dt$, then $\theta = \theta_0 + \omega t$. We choose $\theta_0 = 0$, so

$$\mathscr{E} = -BA\frac{d}{dt}(\cos\omega t) = BA\omega\sin\omega t.$$

If the rotating coil contains N loops,

$$\mathscr{E} = NBA\omega\sin\omega t = \mathscr{E}_0\sin\omega t. \qquad \textbf{(29–4)}$$

Thus the output emf is sinusoidal (Fig. 29–16) with amplitude $\mathscr{E}_0 = NBA\omega$. Such a rotating coil in a magnetic field is the basic operating principle of an *ac generator*. The rms output (Section 25–7) is

$$\mathscr{E}_{\text{rms}} = \frac{NB\omega A}{\sqrt{2}} = \frac{\mathscr{E}_0}{\sqrt{2}}.$$

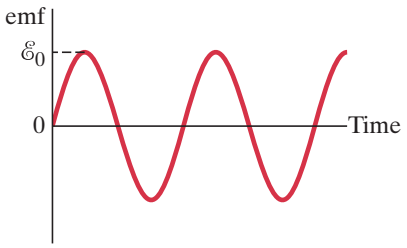
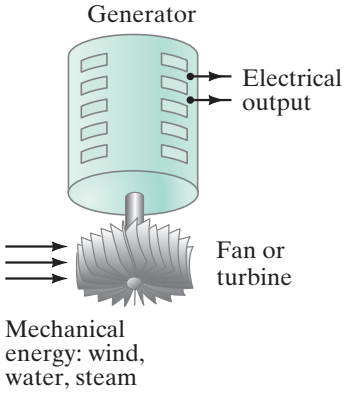


FIGURE 29–16 An ac generator produces an alternating current. The output emf $\mathscr{E} = \mathscr{E}_0\sin\omega t$, where $\mathscr{E}_0 = NAB\omega$ (Eq. 29–4).

The frequency $f\ (= \omega/2\pi)$ is 60 Hz for general use in the United States, Canada, and a few other countries, whereas 50 Hz is used in Europe and most other countries. Most of the power generated in the United States is done at steam plants, where the burning of fossil fuels (coal, oil, natural gas) boils water to produce high-pressure steam that turns a turbine connected to the generator axle, Fig. 29–17. Turbines can also be turned by water gushing from beneath a dam (hydroelectric) and by wind, Section 20–11 and Table 20–3. At nuclear power plants, the nuclear energy released is used to produce steam to turn turbines. Indeed, a heat engine (Chapter 20) connected to a generator is the principal means of generating electric power. The frequency of 60 Hz or 50 Hz is maintained very precisely by power companies, and in doing Problems, we will assume it is as precise as other numbers given.

PHYSICS APPLIED
Power plants

FIGURE 29–17 A generator axle can be turned by various mechanical means to produce electricity: steam, wind, and water gushing out at the base of a dam.



EXAMPLE 29–10 **An ac generator.** The armature of a 60-Hz ac generator rotates in a 0.15-T magnetic field. If the area of the coil is $2.0\times10^{-2}\,\text{m}^2$, how many loops must the coil contain if the peak output is to be $\mathscr{E}_0 = 170\,\text{V}$?

APPROACH From Eq. 29–4 we see that the maximum emf is $\mathscr{E}_0 = NBA\omega$.

SOLUTION We solve Eq. 29–4 for N with $\omega = 2\pi f = (6.28)(60\,\text{s}^{-1}) = 377\,\text{s}^{-1}$ (a useful number to remember):

$$N = \frac{\mathscr{E}_0}{BA\omega} = \frac{170\,\text{V}}{(0.15\,\text{T})(2.0\times10^{-2}\,\text{m}^2)(377\,\text{s}^{-1})} = 150\ \text{turns}.$$

A **dc generator** is much like an ac generator, except the slip rings are replaced by split-ring commutators, Fig. 29–18a, just as in a dc motor (Section 27–6). The output of a dc generator is as shown by the red curves, and can be smoothed out by placing a capacitor in parallel with the output† (Section 26–5). More common is the use of many armature windings, as in Fig. 29–18b, which produces a smoother output.

PHYSICS APPLIED
DC generator

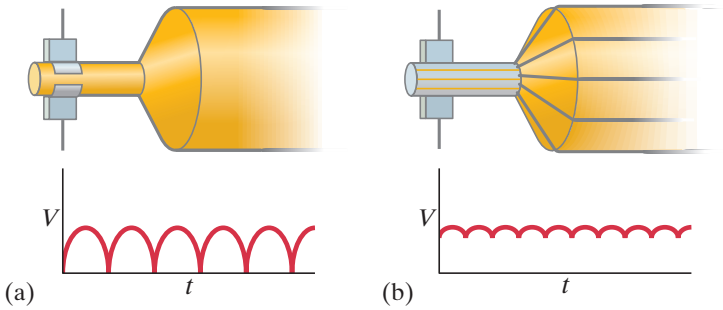
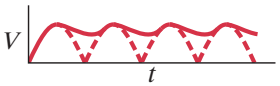


FIGURE 29–18 (a) A dc generator with one set of commutators, and (b) a dc generator with many sets of commutators and windings.

†A capacitor tends to store charge and, if the time constant RC is long enough, helps to smooth out the voltage as shown in the figure to the right.



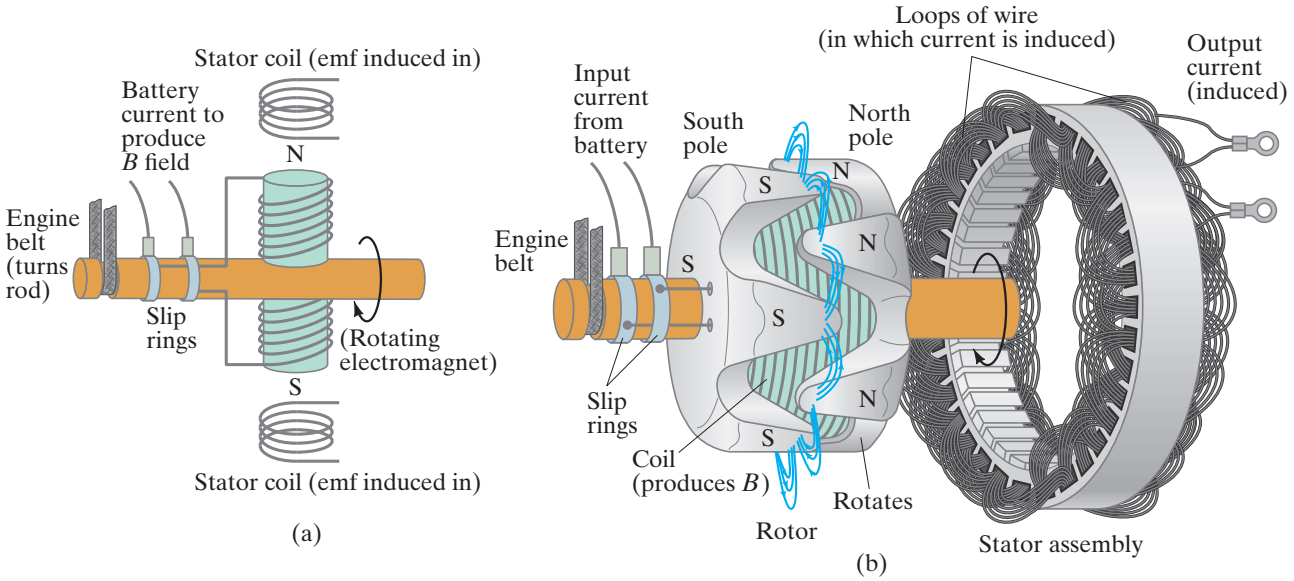


FIGURE 29–19 (a) Simplified schematic diagram of an alternator. The input current to the rotor from the battery is connected through continuous slip rings. Sometimes the rotor electromagnet is replaced by a permanent magnet (no input current). (b) Actual shape of an alternator. The rotor is made to turn by a belt from the engine. The current in the wire coil of the rotor (green) produces a magnetic field inside it along its axis that points horizontally from left to right (not shown), thus making north and south poles of the odd-shaped metal plates attached at either end. These end plates are made with triangular fingers that are bent over the coil—hence there are N and S poles quite close to one another, with magnetic field lines between them as shown by the blue lines. As the rotor turns, these field lines pass through the fixed stator coils (shown on the right for clarity, but in fact the rotor rotates inside the stator), inducing a current in them, which is the output.

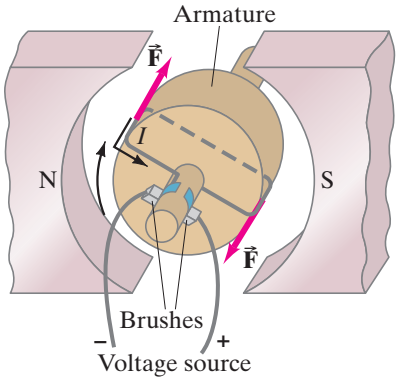
PHYSICS APPLIED
Alternators

Automobiles used to use dc generators. Today they mainly use **alternators**, which avoid the problems of wear and electrical arcing (sparks) across the split-ring commutators of dc generators. Alternators differ from generators in that an electromagnet, called the **rotor**, is fed by current from the battery and is made to rotate by a belt from the engine. The magnetic field of the turning rotor passes through a surrounding set of stationary coils called the **stator** (Fig. 29–19), inducing an alternating current in the stator coils, which is the output. This ac output is changed to dc for charging the battery by the use of semiconductor diodes, which allow current flow in one direction only. Note: the battery sends a current to the rotor, and the mechanical work done by the car engine produces *more* current to send back to the battery.

29–5 Back EMF and Counter Torque; Eddy Currents

Back EMF, in a Motor

FIGURE 27–26 (Repeated.) A simple dc motor.



A motor turns and produces mechanical energy when a current is made to flow in it. From our description in Section 27–6 of a simple dc motor, you might expect that the armature in Fig. 27–26 (shown again here in the margin) would accelerate indefinitely due to the torque on it. However, as the armature of the motor turns, the magnetic flux through the coil changes and an emf is generated. This induced emf acts to oppose the motion (Lenz's law) and is called the **back emf** or **counter emf**. The greater the speed of the motor, the greater the back emf. A motor normally rotates and does work on something ("the load"), but if there were no load to push (or rotate), the motor's speed would increase until the back emf equaled the input voltage. When there is a mechanical load, the speed of the motor is limited also by that load. The back emf will then be less than the external applied voltage. The greater the mechanical load, the slower the motor rotates and the lower is the back emf ($\mathscr{E} \propto \omega$, Eq. 29–4).

EXAMPLE 29–11 **Back emf in a motor.** The armature windings of a dc motor have a resistance of 5.0 Ω. The motor is connected to a 120-V line, and when the motor reaches full speed against its normal load, the back emf is 108 V. Calculate (*a*) the current into the motor when it is just starting up, and (*b*) the current when the motor reaches full speed.

APPROACH As the motor is just starting up, it is turning very slowly, so there is negligible back emf. The only voltage is the 120-V line. The current is given by Ohm's law with $R = 5.0\ \Omega$. At full speed, we must include as emfs both the 120-V applied emf and the opposing back emf.

SOLUTION (*a*) At start up, the current is controlled by the 120 V applied to the coil's 5.0-Ω resistance. By Ohm's law,

$$I = \frac{V}{R} = \frac{120\ \text{V}}{5.0\ \Omega} = 24\ \text{A}.$$

(*b*) When the motor is at full speed, the back emf must be included in the equivalent circuit shown in Fig. 29–20. In this case, Ohm's law (or Kirchhoff's rule, Chapter 26) gives

$$120\ \text{V} - 108\ \text{V} = I(5.0\ \Omega).$$

Therefore

$$I = \frac{12\ \text{V}}{5.0\ \Omega} = 2.4\ \text{A}.$$

NOTE This result shows that the current can be very high when a motor first starts up. This is why the lights in your house may dim when the motor of the refrigerator (or other large motor) starts up. The large initial refrigerator current causes the voltage to the lights and at the outlets to drop because the house wiring has resistance and there is noticeable voltage drop across it when large currents are drawn.

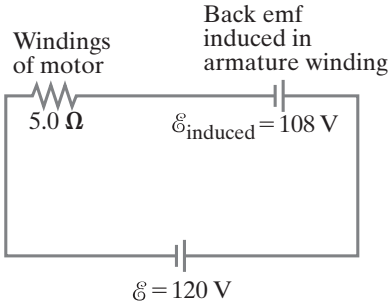


FIGURE 29–20 Circuit of a motor showing induced back emf. Example 29–11.

CONCEPTUAL EXAMPLE 29–12 **Motor overload.** When using an appliance such as a blender, an electric drill, or a sewing machine, if the appliance is overloaded or jammed so that the motor slows appreciably or stops while the power is still connected, the device can burn out and be ruined. Explain why this happens.

PHYSICS APPLIED
Burning out a motor

RESPONSE The motors are designed to run at a certain speed for a given applied voltage, and the designer must take the expected back emf into account. If the rotation speed is reduced, the back emf will not be as high as expected ($\mathcal{E} \propto \omega$, Eq. 29–4). The current will increase, and may become large enough that the windings of the motor heat up and may melt, ruining the motor.

Counter Torque, in a Generator

In a generator, the situation is the reverse of that for a motor. As we saw, the mechanical turning of the armature induces an emf in the loops, which is the output. If the generator rotates but is not connected to an external circuit, the emf exists at the terminals but there is no current. In this case, it takes little effort to turn the armature. But if the generator *is* connected to a device that draws current, then a current flows in the coils of the armature. Because this current-carrying coil is in an external magnetic field, there will be a torque exerted on it (as in a motor), and this torque opposes the motion (use right-hand-rule-2, page 791; for the force on a wire in Fig. 29–15). This is called a **counter torque**. The greater the electrical load—that is, the more current that is drawn—the greater will be the counter torque. Hence the external applied torque will have to be greater to keep the generator turning. This makes sense from the conservation of energy principle. More mechanical energy input is needed to produce more electrical energy output.

EXERCISE D A bicycle headlight is powered by a generator that is turned by the bicycle wheel. (*a*) If you speed up, how does the power to the light change? (*b*) Does the generator resist being turned as the bicycle's speed increases, and if so how?

Eddy Currents

Induced currents are not always confined to well-defined paths such as in wires. Consider, for example, the rotating metal wheel in Fig. 29–21a. An external magnetic field is applied across a limited area of the wheel as shown and points into the page. The section of wheel in the magnetic field has an emf induced in it because the conductor is moving, carrying electrons with it. The flow of induced (conventional) current in the wheel is upward in the region of the magnetic field (Fig. 29–21b), and the current follows a downward return path outside that region. Why? According to Lenz's law, the induced currents oppose the change that causes them. Consider the part of the rotating wheel labeled c in Fig. 29–21b, where the magnetic field is zero but is just about to enter a region where $\vec{\mathbf{B}}$ points into the page. To oppose this inward increase in magnetic field, the induced current is counterclockwise to produce a field pointing out of the page (right-hand-rule-1). Similarly, region d is about to move to e, where $\vec{\mathbf{B}}$ is zero; hence the current is clockwise to produce an inward field opposed to the decreasing flux inward. These currents are referred to as **eddy currents**. They can be present in any conductor that is moving across a magnetic field or through which the magnetic flux is changing.

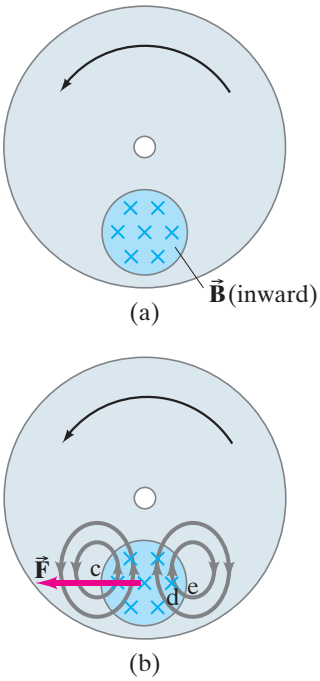


FIGURE 29–21 Production of eddy currents in a rotating wheel. The wide gray lines in (b) indicate induced current.

In Fig. 29–21b, the magnetic field exerts a force $\vec{\mathbf{F}}$ on the induced currents it has created, and that force opposes the rotational motion. Eddy currents can be used in this way as a smooth braking device on, say, a rapid-transit railroad car. In order to stop the car, an electromagnet can be turned on that applies its field either to the wheels or to the moving steel rail below. Eddy currents can also be used to dampen (reduce) the oscillation of a vibrating system, which is referred to as **magnetic damping**.

Eddy currents, however, can be a problem. For example, eddy currents induced in the armature of a motor or generator produce heat ($P = I\mathscr{E}$) and waste energy. To reduce the eddy currents, the armatures are *laminated*; that is, they are made of very thin sheets of iron that are well insulated from one another (used also in transformers, Fig. 29–24). The total path length of the eddy currents is confined to each slab, which increases the total resistance; hence the current is less and there is less wasted energy.

PHYSICS APPLIED
Metal detector

Walk-through metal detectors (Fig. 29–22) use electromagnetic induction and eddy currents to detect metal objects. Several coils are situated in the walls of the walk-through at different heights. In one technique, the coils are given brief pulses of current, hundreds or thousands of times per second. When a person passes through the walk-through, any metal object being carried will have eddy currents induced in it, and the small magnetic field produced by that eddy current can be detected, setting off an alert or alarm.

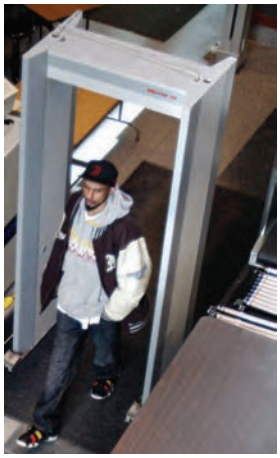


FIGURE 29–22
Metal detector.

29–6 Transformers and Transmission of Power

A transformer is a device for increasing or decreasing an ac voltage. Transformers are found everywhere: on utility poles (Fig. 29–23) to reduce the high voltage from the electric company to a usable voltage in houses (120 V or 240 V); in chargers for cell phones, laptops, and other electrical devices; in your car to give the needed high voltage to the spark plugs; and in many other applications. A **transformer** consists of two coils of wire known as the **primary** and **secondary** coils. The two coils can be interwoven (with insulated wire); or they can be linked by an iron core which is laminated to minimize eddy-current losses (Section 29–5), as shown in Fig. 29–24. Transformers are designed so that (nearly) all the magnetic flux produced by the current in the primary coil also passes through the secondary coil, and we assume this is true in what follows. We also assume that energy losses (in resistance and hysteresis) can be ignored—a good approximation for real transformers, which are often better than 99% efficient.



FIGURE 29–23 Repairing a step-down transformer on a utility pole.

When an ac voltage is applied to the primary coil, the changing magnetic field it produces will induce an ac voltage of the same frequency in the secondary coil. However, the voltage will be different according to the number of "turns" or loops in each coil. From Faraday's law, the voltage or emf induced in the secondary coil is

$$V_S = N_S \frac{d\Phi_B}{dt},$$

where N_S is the number of turns in the secondary coil, and $d\Phi_B/dt$ is the rate at which the magnetic flux changes through each coil.

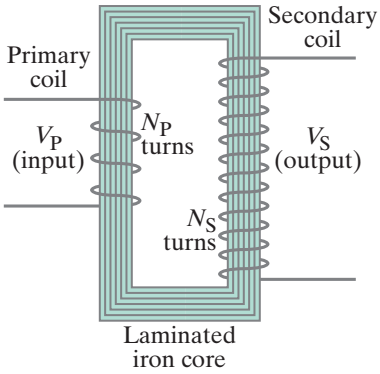
The input primary voltage, V_P, is related to the rate at which the flux changes through it,

$$V_P = N_P \frac{d\Phi_B}{dt},$$

where N_P is the number of turns in the primary coil. This follows because the changing flux produces a back emf, $N_P\, d\Phi_B/dt$, in the primary that exactly balances the applied voltage V_P if the resistance of the primary can be ignored (Kirchhoff's rules). We divide these two equations, assuming little or no flux is lost, to find

$$\frac{V_S}{V_P} = \frac{N_S}{N_P}. \qquad \textbf{(29–5)}$$

FIGURE 29–24 Step-up transformer $(N_P = 4,\ N_S = 12)$.



This **transformer equation** tells how the secondary (output) voltage is related to the primary (input) voltage; V_S and V_P in Eq. 29–5 can be the rms values (Section 25–7) for both, or peak values for both. DC voltages don't work in a transformer because there would be no changing magnetic flux.

If the secondary coil contains more loops than the primary coil $(N_S > N_P)$, we have a **step-up transformer**. The secondary voltage is greater than the primary voltage. For example, if the secondary coil has twice as many turns as the primary coil, then the secondary voltage will be twice that of the primary voltage. If N_S is less than N_P, we have a **step-down transformer**.

Although ac voltage can be increased (or decreased) with a transformer, we don't get something for nothing. Energy conservation tells us that the power output can be no greater than the power input. A well-designed transformer can be greater than 99% efficient, so little energy is lost to heat. The power output thus essentially equals the power input. Since power $P = IV$ (Eq. 25–6), we have

$$I_P V_P = I_S V_S,$$

or

$$\frac{I_S}{I_P} = \frac{N_P}{N_S}. \qquad \textbf{(29–6)}$$

EXAMPLE 29–13 **Cell phone charger.** The charger for a cell phone contains a transformer that reduces 120-V (or 240-V) ac to 5.0-V ac to charge the 3.7-V battery (Section 26–4). (It also contains diodes to change the 5.0-V ac to 5.0-V dc.) Suppose the secondary coil contains 30 turns and the charger supplies 700 mA to the cell phone battery. Calculate (*a*) the number of turns in the primary coil, (*b*) the current in the primary, and (*c*) the power transformed.

APPROACH We assume the transformer is ideal, with no flux loss, so we can use Eq. 29–5 and then Eq. 29–6.

SOLUTION (*a*) This is a step-down transformer, and from Eq. 29–5 we have

$$N_P = N_S \frac{V_P}{V_S} = \frac{(30)(120\,\text{V})}{(5.0\,\text{V})} = 720\,\text{turns}.$$

(*b*) From Eq. 29–6

$$I_P = I_S \frac{N_S}{N_P} = (0.70\,\text{A})\left(\frac{30}{720}\right) = 29\,\text{mA}.$$

(*c*) The power transformed is

$$P = I_S V_S = (0.70\,\text{A})(5.0\,\text{V}) = 3.5\,\text{W}.$$

NOTE The power in the primary coil, $P = (0.029\,\text{A})(120\,\text{V}) = 3.5\,\text{W}$, is the same as the power in the secondary coil. This 100% efficiency in power transfer is valid only for our ideal transformer.

EXERCISE E How many turns would you want in the secondary coil of a transformer having $N_P = 400$ turns if it were to reduce the voltage from 120-V ac to 3.0-V ac?

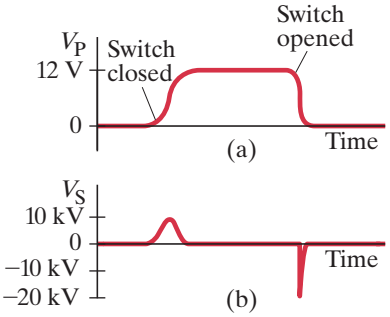


FIGURE 29–25 A dc voltage turned on and off as shown in (a) produces voltage pulses in the secondary (b). Voltage scales in (a) and (b) are not the same.

PHYSICS APPLIED
Transformers in power transmission

A transformer operates only on ac. A dc current in the primary coil does not produce a changing flux in the secondary and therefore induces no emf. However, if a dc voltage is applied to the primary through a switch, at the instant the switch is opened or closed there will be an abrupt change in magnetic flux, and an induced voltage in the secondary. For example, if the dc is turned on and off as shown in Fig. 29–25a, the voltage induced in the secondary is as shown in Fig. 29–25b. Notice that the secondary voltage drops to zero when the dc voltage is steady. This is basically how, in the **ignition system** of an automobile, the high voltage is created to produce the spark across the gap of a spark plug that ignites the gas-air mixture in the engine cylinders. The transformer is referred to simply as an "ignition coil," and transforms the 12-V dc of the battery (when switched down from 12 V to 0 V in the primary) into a spike of as much as 30 kV in the secondary.

Transformers play an important role in the transmission of electricity. Power plants are often situated some distance from metropolitan areas, so electricity must then be transmitted over long distances (Fig. 29–26). There is always some power loss in the transmission lines, and this loss can be minimized if the power is transmitted at high voltage, using transformers, as the following Example shows.

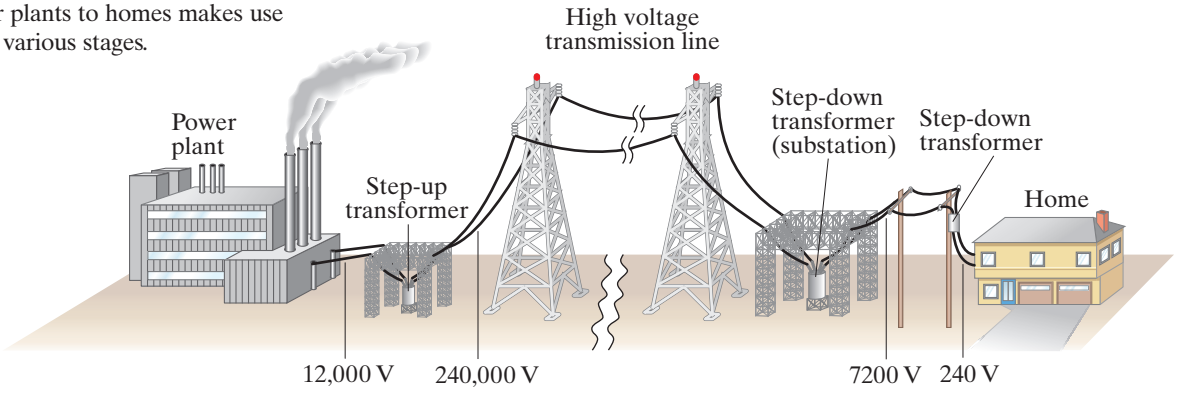


FIGURE 29–26 The transmission of electric power from power plants to homes makes use of transformers at various stages.

EXAMPLE 29–14 **Transmission lines.** An average of 120 kW of electric power is sent to a small town from a power plant 10 km away. The transmission lines have a total resistance of 0.40 Ω. Calculate the power loss if the power is transmitted at (*a*) 240 V and (*b*) 24,000 V.

APPROACH We cannot use $P = V^2/R$ because R is the resistance of the transmission lines and we don't know the voltage drop along them. The given voltages are applied across the lines *plus* the load (the town). But we can determine the current I in the lines $(= P/V)$, and then find the power loss from $P_L = I^2R$, for both cases (*a*) and (*b*).

SOLUTION (*a*) If 120 kW is sent at 240 V, the total current will be

$$I = \frac{P}{V} = \frac{1.2 \times 10^5\,\text{W}}{2.4 \times 10^2\,\text{V}} = 500\,\text{A}.$$

The power loss in the lines, P_L, is then

$$P_L = I^2R = (500\,\text{A})^2(0.40\,\Omega) = 100\,\text{kW}.$$

Thus, over 80% of all the power would be wasted as heat in the power lines!

(*b*) If 120 kW is sent at 24,000 V, the total current will be

$$I = \frac{P}{V} = \frac{1.2 \times 10^5\,\text{W}}{2.4 \times 10^4\,\text{V}} = 5.0\,\text{A}.$$

The power loss in the lines is then

$$P_L = I^2R = (5.0\,\text{A})^2(0.40\,\Omega) = 10\,\text{W},$$

which is less than $\frac{1}{100}$ of 1%. At high voltage we get a far better efficiency!

NOTE We see that the higher voltage results in less current, and thus less power is wasted as heat in the transmission lines. It is for this reason that power is usually transmitted at very high voltages, as high as 700 kV.

The great advantage of ac, and a major reason it is in nearly universal use,[†] is that the voltage can easily be stepped up or down by a transformer. The output voltage of an electric generating plant is stepped up prior to transmission—so less energy is wasted in the transmission line. Upon arrival in a city, that high voltage is stepped down in stages at electric substations prior to distribution. The voltage in lines along city streets is typically 2400 V or 7200 V, and then is stepped down to 240 V or 120 V for home use by transformers (Figs. 29–23 and 29–26).

Another use of a transformer is for fluorescent lights which require a very high voltage initially to ionize the gas inside the bulb. The high voltage is obtained using a step-up transformer, called a **ballast**, and can be replaced independently of the bulb in many fluorescent light fixtures. When the ballast starts to fail, the tube is slow to light. Replacing the bulb will not solve the problem. In newer compact fluorescent bulbs designed to replace incandescent bulbs, the ballast (transformer) is part of the bulb, and is very small.

PHYSICS APPLIED
Fluorescent lightbulb ballast

[†]DC transmission along wires does exist, and has some advantages (if the current is constant, there is no induced current in nearby conductors as there is with ac). But boosting to high voltage and down again at the receiving end requires more complicated electronics.

Wireless Transmission of Power—Inductive Charging

PHYSICS APPLIED
Charging by induction

Many devices with rechargeable batteries, like cell phones, cordless phones, and even electric cars, can be recharged using a direct metal contact between the device and the charger. But devices can also be charged "wirelessly" by induction, without the need for exposed electric contacts. The electric toothbrush shown in Fig. 29–27 sits on a plastic base. Inside the base is a "primary coil" connected to an ac outlet. Inside the toothbrush is a "secondary coil" in which a current is induced by the changing magnetic field produced by the changing current in the primary coil. The current induced in the secondary coil charges the rechargeable batteries. (Not an option for ordinary AA or AAA batteries which are *not* rechargeable.) This is a simple transformer—except there is no iron (to contain the field lines), so you can separate the two parts (toothbrush and charger) and brush your teeth.

Inductive charging is also a possible means for recharging an electric car's batteries. A primary coil may be installed in the floor of the garage, and the secondary coil underneath the car (photo page 869).

Wireless transmission of power must be done over short distances to maintain a reasonable efficiency. Wireless transmission of signals (information) can be done over great distances (Section 31–10) because even fairly low power signals can be detected, and it is the information in the signal voltages that counts, not power.

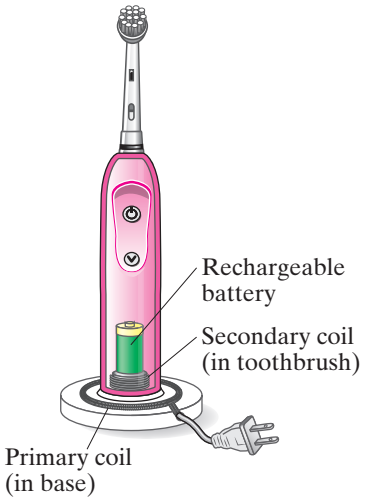


FIGURE 29–27 This electric toothbrush contains rechargeable batteries which are being recharged as it sits on its base. Charging occurs from a primary coil in the base to a secondary coil in the toothbrush. The toothbrush can be lifted from its base when you want to brush your teeth.

29–7 A Changing Magnetic Flux Produces an Electric Field

We saw in Section 25–8 that when an electric current flows in a wire, there is an electric field in the wire that does the work of moving the electrons in the wire. In this Chapter we have seen that a changing magnetic flux induces a current in a wire, and that would imply that there is an *electric field* in the wire induced by the changing magnetic flux. Thus we come to the important conclusion that

a changing magnetic flux produces an electric field.

This result applies not only to wires and other conductors, but is a general result that applies to any region in space. Indeed, an electric field will be produced (= induced) at any point in space where there is a changing magnetic field. Such electric fields are *not static*, as are the electric fields due to electric charges at rest.

Faraday's Law—General Form

We can put these ideas into mathematical form by generalizing our relation between an electric field $\vec{\mathbf{E}}$ and the potential difference V_{ab} between two points a and b: $V_{ab} = \int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$ (Eq. 23–4a) where $d\vec{\boldsymbol{\ell}}$ is an element of displacement along the path of integration. The emf $\mathscr{E}$ induced in a circuit is equal to the work done per unit charge by the electric field, which equals the integral of $\vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$ around a closed path:

$$\mathscr{E} = \oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}. \qquad \textbf{(29–7)}$$

We combine this with Eq. 29–2a, to obtain a more elegant and general form of Faraday's law

FARADAY'S LAW (general form)

$$\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{d\Phi_B}{dt} \qquad \textbf{(29–8)}$$

which relates the changing magnetic flux to the electric field it produces. The integral on the left is taken around a path enclosing the area through which the magnetic flux Φ_B is changing. This more elegant statement of Faraday's law (Eq. 29–8) is valid not only in conductors, but in any region of space. To illustrate this, let us take an Example.

EXAMPLE 29–15 **$\vec{\mathbf{E}}$ produced by changing $\vec{\mathbf{B}}$.** A magnetic field $\vec{\mathbf{B}}$ between the pole faces of an electromagnet is nearly uniform at any instant over a circular area of radius r_0 as shown in Figs. 29–28a and b. The current in the windings of the electromagnet is increasing in time so that $\vec{\mathbf{B}}$ changes in time at a constant rate $d\vec{\mathbf{B}}/dt$ at each point. Beyond the circular region $(r > r_0)$, we assume $\vec{\mathbf{B}} = 0$ at all times. Determine the electric field $\vec{\mathbf{E}}$ at any point P a distance r from the center of the circular area due to the changing $\vec{\mathbf{B}}$.

APPROACH The changing magnetic flux through a circle of radius r, shown dashed in Fig. 29–28b, will produce an emf around this circle. Because all points on the dashed circle are equivalent physically, the electric field too will show this symmetry and will be in the plane perpendicular to $\vec{\mathbf{B}}$. Thus we can expect $\vec{\mathbf{E}}$ to be perpendicular to $\vec{\mathbf{B}}$ and to be tangent to the circle of radius r. The direction of $\vec{\mathbf{E}}$ will be as shown in Fig. 29–28b and c, since by Lenz's law the induced $\vec{\mathbf{E}}$ needs to be capable of producing a current that generates a magnetic field opposing the original change in $\vec{\mathbf{B}}$. By *symmetry*, we also expect $\vec{\mathbf{E}}$ to have the same magnitude at all points on the circle of radius r.

SOLUTION We take the circle shown in Fig. 29–28b as our path of integration in Eq. 29–8. We ignore the minus sign so we can concentrate on magnitude since we already found the direction of $\vec{\mathbf{E}}$ from Lenz's law, and obtain

$$E(2\pi r) = (\pi r^2)\frac{dB}{dt}, \qquad [r < r_0]$$

since $\Phi_B = BA = B(\pi r^2)$ at any instant. We solve for E and obtain

$$E = \frac{r}{2}\frac{dB}{dt}. \qquad [r < r_0]$$

This expression is valid up to the edge of the circle $(r \le r_0)$, beyond which $\vec{\mathbf{B}} = 0$. If we now consider a point where $r > r_0$, the flux through a circle of radius r is $\Phi_B = \pi r_0^2 B$. Then Eq. 29–8 gives

$$E(2\pi r) = \pi r_0^2 \frac{dB}{dt} \qquad [r > r_0]$$

or

$$E = \frac{r_0^2}{2r}\frac{dB}{dt}. \qquad [r > r_0]$$

Thus the magnitude of the induced electric field increases linearly from zero at the center of the magnet to $E = (dB/dt)(r_0/2)$ at the edge, and then decreases inversely with distance in the region beyond the edge of the magnetic field. The electric field lines are circles as shown in Fig. 29–28c. A graph of E vs. r is shown in Fig. 29–28d.

EXERCISE F Consider the magnet shown in Fig. 29–28 with a radius $r_0 = 6.0\,\text{cm}$. If the magnetic field changes uniformly from 0.040 T to 0.090 T in 0.18 s, what is the magnitude of the resulting electric field at (*a*) $r = 3.0\,\text{cm}$ and (*b*) $r = 9.0\,\text{cm}$?

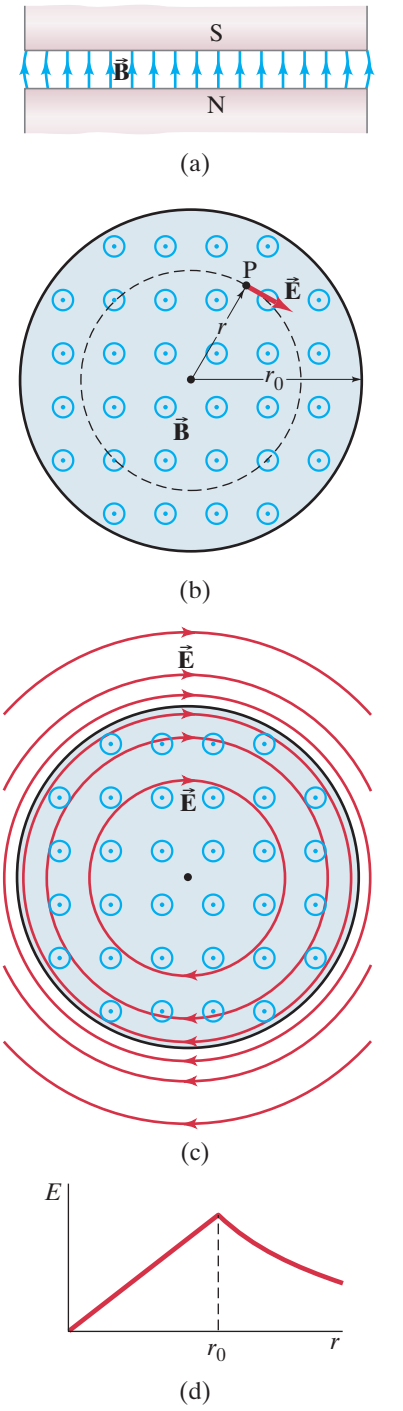


FIGURE 29–28 Example 29–15. (a) Side view of nearly constant $\vec{\mathbf{B}}$. (b) Top view, for determining the electric field $\vec{\mathbf{E}}$ at point P. (c) Lines of $\vec{\mathbf{E}}$ produced by increasing $\vec{\mathbf{B}}$ (pointing outward). (d) Graph of E vs. r.

Forces Due to Changing $\vec{\mathbf{B}}$ are Nonconservative

Example 29–15 illustrates an important difference between electric fields produced by changing magnetic fields from electric fields produced by electric charges at rest (electrostatic fields). Electric field lines produced in the electrostatic case (Chapters 21 to 24) start and stop on electric charges. But the electric field lines produced by a changing magnetic field are continuous; they form closed loops. This distinction goes even further and is an important one. In the electrostatic case, the potential difference between two points is given by (Eq. 23–4a)

$$V_{ba} = V_b - V_a = -\int_a^b \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}.$$

If the integral is around a closed loop, so points a and b are the same, then $V_{ba} = 0$.

Hence the integral of $\vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$ around a closed path is zero:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = 0. \qquad \text{[electrostatic field]}$$

This followed from the fact that the electrostatic force (Coulomb's law) is a conservative force, and so a potential energy function could be defined. Indeed, the relation above, $\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = 0$, tells us that the work done per unit charge around any closed path is zero (or the work done between any two points is independent of path—see Chapter 8), which is a property only of a *conservative force.* But in the nonelectrostatic case, when the electric field is produced by a changing magnetic field, the integral around a closed path is *not* zero, but is given by Eq. 29–8:

$$\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{d\Phi_B}{dt}.$$

We thus come to the conclusion that the forces due to changing magnetic fields are *nonconservative*. We are not able therefore to define a potential energy, or potential function, at a given point in space for the nonelectrostatic case. Although static electric fields are *conservative fields*, the electric field produced by a changing magnetic field is a **nonconservative field**.

*29–8 Information Storage: Magnetic and Semiconductor

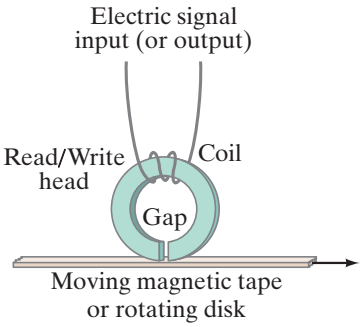
*Magnetic Storage: Read/Write on Tape and Disks

PHYSICS APPLIED
Hard drive

Recording and playback on tape or disk is done by magnetic **heads**.[†] Magnetic tapes contain a thin layer of ferromagnetic oxide on a thin plastic tape. Computer **hard drives** (HD) store digital information (applications and data): they have a thin layer of ferromagnetic material on the surface of each rotating disk or platter, Fig. 29–29a. During recording of an audio or video signal on tape, or "writing" on a hard drive, the voltage is sent to the recording head which acts as a tiny electromagnet (Fig. 29–29b) that magnetizes the tiny section of tape or disk passing the narrow gap in the head at each instant. During playback, or "reading" of an HD, the changing magnetism of the moving tape or disk at the gap causes corresponding changes in the magnetic field within the soft-iron head, which in turn induces an emf in the coil (Faraday's law). This induced emf is the output signal that can be processed by the computer, or for audio can be amplified and sent to a loudspeaker (for video to a monitor or TV).



(a)



(b)

FIGURE 29–29 (a) Photo of a hard drive showing several platters and read/write heads that can quickly move from the edge of the disk to the center. (b) Read/Write (playback/recording) head for disk or tape. In writing or recording, the electric input signal to the head, which acts as an electromagnet, magnetizes the passing tape or disk. In reading or playback, the changing magnetic field of the passing tape or disk induces a changing magnetic field in the head, which in turn induces in the coil an emf that is the output signal.

Audio and video signals may be **analog**, varying continuously in amplitude over time: the variation in degree of magnetization at sequential points reflects the variation in amplitude and frequency of the audio or video signal. In modern equipment, analog signals (say from a microphone) are electronically converted to **digital**—which means a series of **bits**, each of which is a "1" or a "0", that forms a **binary code** as discussed in Section 23–9. (Recall also, 8 bits in a row = 1 **byte**.) Computers process only digital information.

[†]CDs and DVDs are "read" by a laser light beam—an *optical drive* (Section 39–9).

*Semiconductor Memory: DRAM, Flash

Basic to your computer is its **random access memory (RAM)**. This is where the information you are working with at any given time is temporarily stored and manipulated by you. Each data storage location can be accessed and read (or written) directly and quickly, so you don't have to wait. In contrast, hard drives, tape, flash and external devices are more permanent **storage**, and they are much slower to access because the data must be searched for, sequentially, such as along the circular tracks of hard drives (Fig. 29–29a). Programs, applications, and data that you want to use are imported by the computer into the RAM from their more permanent (and more slowly accessed) storage area.[†]

RAM is based on semiconductor technology, storing the binary bits ("0" or "1") as electric charge or voltage. Some computers may use semiconductors also for long-term storage ("flash memory") in place of a hard drive.

A common type of RAM is **dynamic random access memory** or **DRAM**, which uses arrays of transistors known as MOSFETs (metal-oxide semiconductor field-effect transistors). Transistors will be discussed in Section 40–10, but we already encountered them in Section 23–10 about TV screen addressing, Fig. 23–33, which we show again here, Fig. 29–30. A MOSFET transistor in RAM serves basically as an on-off switch: the voltage on the **gate** terminal acts to control the conductivity between the **source** and the **drain** terminals, thus allowing current to flow (or not) between them.

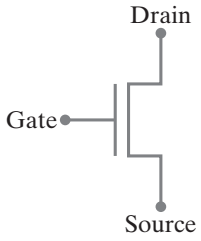
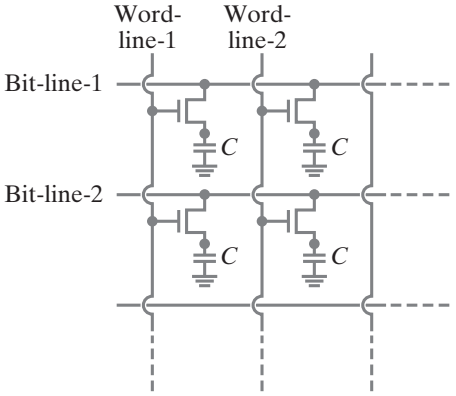


FIGURE 29–30 Symbol for a MOSFET transistor. The gate acts to attract or repel charge, and thus open or close the connection along the semiconductor that connects source and drain.

Each memory "cell," which in DRAM consists of one transistor and a capacitor, stores one bit (= a "0" or a "1"). Each cell is extremely small physically, less than 100 nm across.[‡] Typical DRAM *chips* (integrated circuits) contain billions of these memory cells. To see how they work, we look at a tiny part, the simple four-cell array shown in Fig. 29–31. One side of each cell capacitor is grounded; the other side is connected to the transistor source. The drain of each transistor is connected to a very thin conducting wire or "line," a **bit-line**, that runs across the array of cells. Each gate is connected to a **word-line**. A particular bit is a "1" or a "0" depending on whether the capacitor of the cell is charged to a voltage V (maybe 3 V) or is at zero (uncharged, or at a very small voltage).

FIGURE 29–31 A tiny 2×2 cell, part of a simple DRAM array. The word-lines and bit-lines do not touch each other where they cross.



To **write** data, say on the upper left cell in Fig. 29–31, word-line-1 is given a high enough pulse of voltage to "turn on" the transistor: that is, the gate voltage is high and attracts charge which allows bit-line-1 and the capacitor to be connected. Thus charge can flow from bit-line-1 to the capacitor, charging it either to V or to 0, depending on the bit-line-1 voltage at that moment, thus writing a "1" or a "0".

The lower left cell in Fig. 29–31 can be written at the same time by setting bit-line-2 voltage to V or zero.

Now let us see a simple way to **read** a cell. In order to read the data stored ("1" or "0") on the *upper left cell*, a voltage of about $\frac{1}{2}V$ is given to bit-line-1. Then word-line-1 provides enough voltage to the gate to turn on the transistor (in fact all of them in the first column) and connect bit-line-1 to the capacitor. The capacitor, if uncharged (= "0"), will now drag charge from bit-line-1 and the bit-line voltage will drop *below* $\frac{1}{2}V$. If the capacitor is already charged to V (= "1"), the connection to bit-line-1 will raise bit-line-1's voltage to *above* $\frac{1}{2}V$. A sensor at the end of bit-line-1 will detect either change in voltage (increase means it reads a "1," decrease a "0"). All cells connected to one word-line are read at the same moment. The capacitor voltage has been altered by the small charge flow during the reading process. So that cell or bit which has just been read needs to be written again, or "refreshed."

[†]Computer specifications may use "memory" for the random access (fast) memory, and "storage" for the long-term (and slower access) information on hard drives, flash drives, and related devices.

[‡]At 100 nm, 10^5 bits can fit along a 1-cm line, $(1\text{ cm}/100\text{ nm}) = 10^{-2}\text{ m}/10^{-7}\text{ m}$. So a square, 1 cm on a side, can hold $10^5 \times 10^5 = 10^{10} = 10$ Gbits ≈ 1 GB (gigabyte). Actually, cells are even smaller than 100 nm: a $(30\text{ nm})^2$ cell can hold $\approx$10 GB in a 1 cm^2 area.

The transistors are imperfect switches and allow the charge on the tiny capacitors in each cell to be "leaky" and lose charge fairly quickly, so every cell has to be read and rewritten (refreshed) many times per second. The D in DRAM stands for this "dynamic" refreshing action. If the power is turned off, the capacitors lose their charge and the data are lost. DRAM is thus referred to as being **volatile** memory. A hard drive, on the other hand, keeps its (magnetic) memory even when the electric power is off and is called **nonvolatile** memory (doesn't "evaporate").

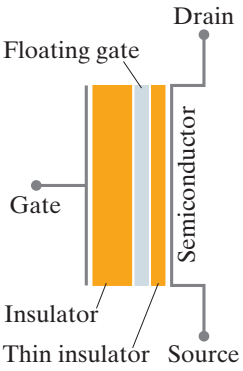


FIGURE 29–32 A floating gate nonvolatile memory cell (NVM).

Flash memory is also made of semiconductor material on tiny "chips." The transistor structures are more complicated, and are able to keep the data even without power so they are nonvolatile. Each MOSFET contains a second gate (the **floating gate**) insulated on both faces, and can hold charge for many years. Charged or not corresponds to a "1" or a "0" bit. Figure 29–32 is a diagram of such an **NVM** (nonvolatile memory) cell. The floating gate is insulated from the standard gate and the semiconductor connecting source and drain. A positive voltage on the gate forces electrons in the semiconductor (at 0 V) to pass through the thin insulator into the floating gate by a process of quantum mechanical **tunneling** (discussed in Section 38–10). This charge is stored on the floating gate as a "1" bit. The erase process is done by applying the opposite voltage to force electrons to tunnel out of the floating gate, returning it to the uncharged state (= a "0" bit). The erase process is slow (milliseconds vs. ns for DRAM), so erasure is done in large blocks of memory. Flash memory[†] is slower to read or write, and is too slow to use as RAM. Instead, flash memory can be used in place of a hard drive as general storage in computers and tablets, and may be called a "solid state device" (**SSD**). Flash is also used for flash drives, memory cards (such as SD cards), thumb drives, cell phone and portable player memory, and external computer memory.

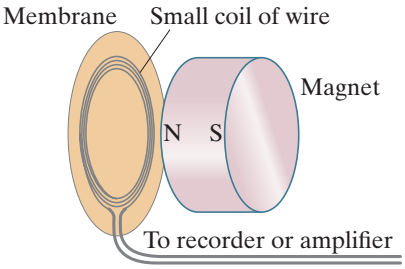
Magnetoresistive RAM (**MRAM**) is a recent development, involving (again) magnetic properties. One cell (storing one bit) consists of two tiny ferromagnetic plates (separated by an insulator), one of which is permanently magnetized. The other plate can be magnetized in one direction or the other, for a "1" or a "0", by current in nearby wires. Cell size is a bit large, but MRAM is fast and nonvolatile (no power and no refresh needed) and therefore has the potential to be used as any type of memory.

*29–9 Applications of Induction: Microphone, Seismograph, GFCI

*Microphone

The condenser microphone was discussed in Section 24–2. Many other types operate on the principle of induction. In one form, a microphone is just the inverse of a loudspeaker (Section 27–6). A small coil of wire attached to a membrane is suspended close to a small permanent magnet, as shown in Fig. 29–33. When sound waves strike the membrane, the coil moves in the magnetic field and this motion induces an emf in the moving wire coil. The frequency of the induced emf will be just that of the impinging sound waves, and this emf is the "signal" that can be amplified and sent to loudspeakers or a recorder.

FIGURE 29–33 Diagram of a microphone that works by induction.



PHYSICS APPLIED
Credit card

*Card Stripe Reader

Many credit cards have a semiconductor chip that is more difficult to fraudulently copy than those with a magnetic stripe on the back of the card that passes over a read head like Fig. 29–29b. The chip or magnetic stripe contains personal information about your account and connects by telephone line for approval from your credit card company.

[†]Why the name "Flash"? It may come from the erase process: large blocks erased "in a flash," and/or because the earliest floating gate memories were erased by a flash of UV light which ejected the stored electrons.

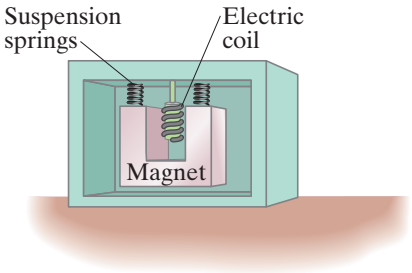


FIGURE 29–34 One type of seismograph, in which the electric coil is fixed to the case and moves with the Earth's surface. The magnet, suspended by springs, has inertia and does not move instantaneously with the coil (and case), so there is relative motion between magnet and coil.

*Seismograph

PHYSICS APPLIED
Seismograph

In geophysics, a **seismograph** measures the intensity of earthquake waves using a magnet and a coil of wire. Either the magnet or the coil is fixed to the case, and the other is inertial (suspended by a spring; Fig. 29–34). The relative motion of magnet and coil when the Earth shakes induces an emf output.

*Ground Fault Circuit Interrupter (GFCI)

Fuses and circuit breakers (Sections 25–6 and 28–9) protect buildings from electricity-induced fire, and apparatus from damage, due to undesired high currents. But they do not turn off the current until it is very much greater than that which can cause permanent damage to humans or death ($\approx$100 mA). If fast enough, they may protect humans in some cases, such as very high currents due to short circuits. A **ground fault circuit interrupter** (GFCI) is meant above all to protect humans.

PHYSICS APPLIED
GFCI

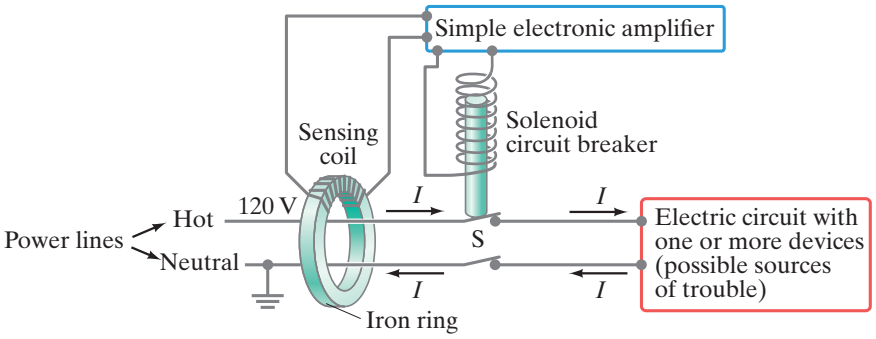


FIGURE 29–35 A ground fault circuit interrupter (GFCI).

Electromagnetic induction is the physical basis of a GFCI. Figure 29–35 shows the two wires carrying ac current to an electrical circuit or device (red box) passing through a small iron ring. Around the ring are many loops of thin wire that serve as a sensing coil. Under normal conditions (no ground fault), the current moving in the hot power wire is exactly balanced by the returning current in the neutral wire. If something goes wrong and the hot wire touches the ungrounded metal case of the device or appliance, some of the entering current can pass through a person who touches the case and then to ground (a **ground fault**). Then the return current in the neutral wire will be less than the entering current in the hot wire, so there is a *net current* passing through the GFCI's iron ring. Because the current is ac, it is changing and that current difference produces a changing magnetic field in the iron ring, thus inducing an emf in the sensing coil wrapped around the iron. For example, if a device draws 8.0 A, and there is a ground fault through a person of 100 mA (= 0.1 A), then 7.9 A will appear in the neutral wire. The emf induced in the sensing coil by this 100-mA difference is amplified by a simple transistor circuit and sent to its own solenoid circuit breaker that opens the circuit at the switch S, thus protecting lives.

If the outer case of the faulty device is grounded, the difference is even higher when there is a fault, and the GFCI trips very quickly.

GFCIs can sense current differences as low as 5 mA and react in 1 ms (10^{-3} s), saving lives. They can be small enough to fit as a wall outlet (Fig. 29–36a), or as a plug-in unit into which you plug a hair dryer or toaster (Fig. 29–36b). It is especially important to have GFCIs installed in kitchens, bathrooms, outdoors, and near swimming pools, where people are most in danger of touching ground. GFCIs always have a "test" button (to be sure the GFCI itself works) and a "reset" button (after it goes off).

FIGURE 29–36 (a) A GFCI wall outlet. GFCIs can be recognized because they have "test" and "reset" buttons. (b) Add-on GFCI that plugs into outlet.

(a)



(b)



Summary

The **magnetic flux** passing through a loop is equal to the product of the area of the loop times the perpendicular component of the (uniform) magnetic field: $\Phi_B = B_\perp A = BA\cos\theta$. If $\vec{\mathbf{B}}$ is not uniform, then

$$\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}. \qquad \textbf{(29–1b)}$$

If the magnetic flux through a coil of wire changes in time, an emf is induced in the coil. The magnitude of the induced emf equals the time rate of change of the magnetic flux through the loop times the number N of loops in the coil:

$$\mathscr{E} = -N\frac{d\Phi_B}{dt}. \qquad \textbf{(29–2b)}$$

This is **Faraday's law of induction**.

The induced emf can produce a current whose magnetic field opposes the original change in flux (**Lenz's law**).

It follows from Faraday's law that a straight wire of length ℓ moving with speed v perpendicular to a magnetic field B has an emf induced between its ends equal to:

$$\mathscr{E} = B\ell v. \qquad \textbf{(29–3)}$$

An electric **generator** changes mechanical energy into electric energy. Its operation is based on Faraday's law: a coil of wire is made to rotate uniformly by mechanical means in a magnetic field, and the changing flux through the coil induces a sinusoidal current, which is the output of the generator.

A motor, which operates in the reverse of a generator, acts a bit like a generator in that a **back emf** is induced in its rotating coil. Because this back emf opposes the input voltage, it can act to limit the current in a motor coil. Similarly, a generator acts somewhat like a motor in that a **counter torque** acts on its rotating coil.

A **transformer**, which is a device to change the magnitude of an ac voltage, consists of a primary coil and a secondary coil. The changing flux due to an ac voltage in the primary coil induces an ac voltage in the secondary coil. In a 100% efficient transformer, the ratio of output to input voltages (V_S/V_P) equals the ratio of the number of turns N_S in the secondary to the number N_P in the primary:

$$\frac{V_S}{V_P} = \frac{N_S}{N_P}. \qquad \textbf{(29–5)}$$

The ratio of secondary to primary current is in the inverse ratio of turns:

$$\frac{I_S}{I_P} = \frac{N_P}{N_S}. \qquad \textbf{(29–6)}$$

Faraday's law also tells us that *a changing magnetic flux produces an electric field*. The mathematical relation is

$$\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{d\Phi_B}{dt} \qquad \textbf{(29–8)}$$

and is the general form of Faraday's law. The integral on the left is taken around the loop through which the magnetic flux Φ_B is changing.

[*Read/write heads for computer drives and tape, as well as microphones, ground fault circuit interrupters, and seismographs, are all applications of electromagnetic induction. Computer memory and information storage can be done using magnetic drives or by semiconductor memory (RAM, flash).]

Questions

1. What would be the advantage, in Faraday's experiments (Fig. 29–1), of using coils with many turns?
2. What is the difference between magnetic flux and magnetic field?
3. Suppose you are holding a circular ring of wire in front of you and (*a*) suddenly thrust a magnet, south pole first, away from you toward the center of the circle. Is a current induced in the wire? (*b*) Is a current induced when the magnet is held steady within the ring? (*c*) Is a current induced when you withdraw the magnet? For each yes answer, specify the direction. Explain your answers.
4. In what direction will the current flow in Fig. 29–12a if the rod moves to the left, which decreases the area of the loop?
5. (*a*) If the north pole of a thin flat magnet moves on a table toward a loop, also flat on the table (Fig. 29–37), in what direction is the induced current in the loop? Assume the magnet is the same thickness as the wire. (*b*) What if the magnet is four times thicker than the wire loop? Explain your answers.

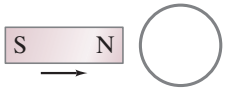


FIGURE 29–37 Question 5.

6. Two loops of wire are moving in the vicinity of a very long straight wire carrying a steady current I as shown in Fig. 29–38. Is there a force between the two loops? If so, in what direction? Explain.

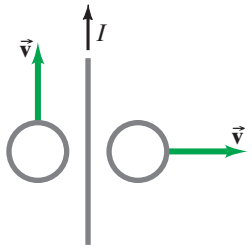


FIGURE 29–38 Questions 6 and 7.

7. Answer Question 6 if the current I in Fig. 29–38 is decreasing.
8. Suppose you are looking along a line through the centers of two circular (but separate) wire loops, one behind the other. A battery is suddenly connected to the front loop, establishing a clockwise current. (*a*) Will a current be induced in the second loop? (*b*) If so, when does this current start? (*c*) When does it stop? (*d*) In what direction is this current? (*e*) Is there a force between the two loops? (*f*) If so, in what direction?

9. The battery mentioned in Question 8 is disconnected. Will a current be induced in the second loop? If so, when does it start and stop? In what direction is this current?

10. In Fig. 29–39, determine the direction of the induced current in resistor R_A (*a*) when coil B is moved toward coil A, (*b*) when coil B is moved away from A, (*c*) when the resistance R_B is increased but the coils remain fixed. Explain your answers.

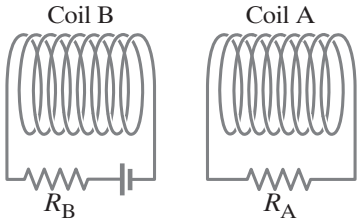


FIGURE 29–39 Question 10.

11. What is the advantage of placing the two insulated electric wires carrying ac close together or even twisted about each other?

12. In situations where a small signal must travel over a distance, a *shielded cable* is used in which the signal wire is surrounded by an insulator and then enclosed by a cylindrical conductor (shield) carrying the return current (Fig. 28–12). Why is a "shield" necessary?

13. Use Fig. 29–15 plus the right-hand rules to show why the counter torque in a generator *opposes* the motion.

14. The use of higher-voltage lines in homes—say, 600 V or 1200 V—would reduce energy waste. Why are they not used?

15. An enclosed transformer has four wire leads coming from it. (*a*) How would you know which wires paired with which? (*b*) How could you determine the ratio of turns on the two coils without taking the transformer apart?

16. A transformer designed for a 120-V ac input will often "burn out" if connected to a 120-V dc source (but not if ac). Explain. [*Hint*: The resistance of the primary coil is usually very low.]

17. Explain why, exactly, the lights may dim briefly when a refrigerator motor starts up. When an electric heater is turned on, the lights may stay dimmed as long as the heater is on. Explain the difference.

18. Consider a large magnet whose north and south poles are close together to form a nearly uniform magnetic field, as in Fig. 27–7. If an aluminum sheet is held between the poles, it requires some force to pull it out of the magnetic field even though the sheet is not ferromagnetic and does not touch the pole faces. Explain.

19. A bar magnet is held above the floor and dropped (Fig. 29–40). In case (a), the magnet falls through a wire loop. In case (b), there is nothing between the magnet and the floor. How will the speeds of the magnets compare? Explain.

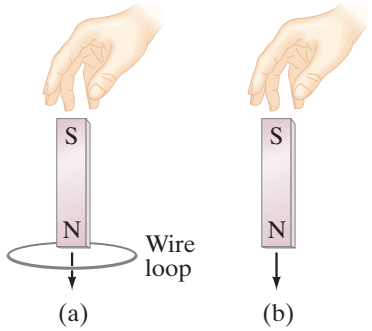


FIGURE 29–40 Question 19 and MisConceptual Question 5.

20. Will an eddy current brake (Fig. 29–21) work on a copper or aluminum wheel, or must the wheel be ferromagnetic? Explain.

21. Eddy currents can be used to sort solid waste for **recycling**. The waste is first ground into tiny pieces and iron removed with a magnet. The waste then is allowed to slide down an incline over permanent magnets. How will this aid in the separation of nonferrous metals (Al, Cu, Pb, brass) from nonmetallic materials?

22. The pivoted metal bar with slots in Fig. 29–41 falls much more quickly through a magnetic field than does a solid bar. Explain.

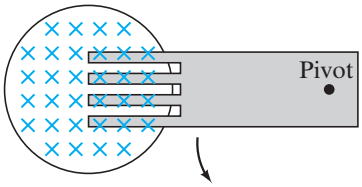


FIGURE 29–41 Question 22.

23. A metal bar, pivoted at one end, oscillates freely in the absence of a magnetic field; but in a magnetic field, its oscillations are quickly damped out. Explain. (This **magnetic damping** is used in a number of practical devices.)

*24. A bar magnet falling inside a vertical metal tube reaches a terminal velocity even if the tube is evacuated so that there is no air resistance. Explain.

*25. Since a magnetic microphone is basically like a loudspeaker, could a loudspeaker (Section 27–6) actually serve as a microphone? That is, could you speak into a loudspeaker and obtain an output signal that could be amplified? Explain. Discuss, in light of your response, how a microphone and loudspeaker differ in construction.

MisConceptual Questions

1. A coil of wire rests in the plane of the page while a magnetic field is directed into the page. A clockwise current is induced
(*a*) when the magnetic field gets stronger.
(*b*) when the size of the coil decreases.
(*c*) when the coil is moved sideways across the page.
(*d*) when the magnetic field is tilted so it is no longer perpendicular to the page.

2. A wire loop moves at constant velocity without rotation through a constant magnetic field. The induced current in the loop will be
(*a*) clockwise.
(*b*) counterclockwise.
(*c*) zero.
(*d*) We need to know the orientation of the loop relative to the magnetic field.

3. A square loop moves to the right from an area where $\vec{\mathbf{B}} = 0$, completely through a region containing a uniform magnetic field directed into the page (Fig. 29–42), and then out to $\vec{\mathbf{B}} = 0$, after point L. A current is induced in the loop
(*a*) only as it passes line J.
(*b*) only as it passes line K.
(*c*) only as it passes line L.
(*d*) as it passes line J or line L.
(*e*) as it passes all three lines.

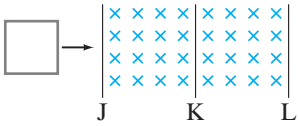


FIGURE 29–42
MisConceptual Question 3.

4. Two loops of wire are moving in the vicinity of a very long straight wire carrying a steady current (Fig. 29–43). Find the direction of the induced current in each loop.

For C:
(*a*) Clockwise.
(*b*) Counterclockwise.
(*c*) Zero.
(*d*) Alternating (ac).

For D:
(*a*) Clockwise.
(*b*) Counterclockwise.
(*c*) Zero.
(*d*) Alternating (ac).

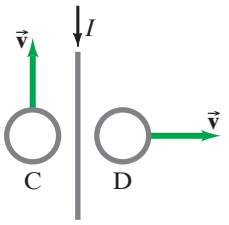


FIGURE 29–43
MisConceptual Question 4.

5. If there is induced current in Question 19 (see Fig. 29–40), wouldn't that cost energy? Where would that energy come from in case (a)?
(*a*) Induced current doesn't need energy.
(*b*) Energy conservation is violated.
(*c*) There is less kinetic energy.
(*d*) There is more gravitational potential energy.

6. A nonconducting plastic hoop is held in a magnetic field that points out of the page (Fig. 29–44). As the strength of the field increases,
(*a*) an induced emf will be produced that causes a clockwise current.
(*b*) an induced emf will be produced that causes a counterclockwise current.
(*c*) an induced emf will be produced but no current.
(*d*) no induced emf will be produced.

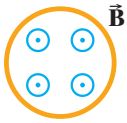


FIGURE 29–44
MisConceptual Question 6

7. A long straight wire carries a current I as shown in Fig. 29–45. A small loop of wire rests in the plane of the page. Which of the following will *not* induce a current in the loop?
(*a*) Increasing the current in the straight wire.
(*b*) Moving the loop in a direction parallel to the wire.
(*c*) Rotating the loop so that it becomes perpendicular to the plane of the page.
(*d*) Moving the loop farther from the wire without rotating it.
(*e*) Moving the loop farther from the wire while rotating it.

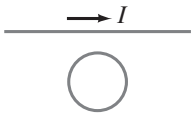


FIGURE 29–45
MisConceptual Question 7.

8. When a generator is used to produce electric current, the resulting electric energy originates from which source?
(*a*) The generator's magnetic field.
(*b*) Whatever rotates the generator's axle.
(*c*) The resistance of the generator's coil.
(*d*) Back emf.
(*e*) Empty space.

9. Which of the following will *not* increase a generator's voltage output?
(*a*) Rotating the generator faster.
(*b*) Increasing the area of the coil.
(*c*) Rotating the magnetic field so that it is more closely parallel to the generator's rotation axis.
(*d*) Increasing the magnetic field through the coil.
(*e*) Increasing the number of turns in the coil.

10. Two separate but nearby coils of wire are mounted along the same axis. A power supply controls the flow of current in the first coil, and thus the magnetic field it produces. The second coil is connected only to an ammeter. The ammeter will indicate that a current is flowing in the second coil
(*a*) whenever a current flows in the first coil.
(*b*) only when a steady current flows in the first coil.
(*c*) only when the current in the first coil changes.
(*d*) only if the second coil is connected to the power supply by rewiring it to be in series with the first coil.

11. Which of the following can a good transformer accomplish?
(*a*) Changing voltage but not current.
(*b*) Changing current but not voltage.
(*c*) Changing power.
(*d*) Changing both current and voltage.

12. The alternating electric current at a wall outlet is most commonly produced by
(*a*) a connection to rechargeable batteries.
(*b*) a coil that is rotating in a magnetic field.
(*c*) accelerating electrons between oppositely charged capacitor plates.
(*d*) using an electric motor.
(*e*) alternately heating and cooling a wire.

13. Which of the following statements about transformers is false?
(*a*) Transformers work using ac current or dc current.
(*b*) If the current in the secondary is higher, the voltage is lower.
(*c*) If the voltage in the secondary is higher, the current is lower.
(*d*) If no flux is lost, the product of the voltage and the current is the same in the primary and secondary coils.

Problems

29–1 and 29–2 Faraday's Law of Induction

1. (I) The magnetic flux through a coil of wire containing two loops changes at a constant rate from −68 Wb to +48 Wb in 0.42 s. What is the emf induced in the coil?

2. (I) The rectangular loop shown in Fig. 29–46 is pushed into the magnetic field which points inward. In what direction is the induced current?



FIGURE 29–46
Problem 2.

3. (I) A 22.0-cm-diameter round loop of wire is initially oriented perpendicular to a 1.5-T magnetic field. The loop is rotated so that its plane is parallel to the field direction in 0.35 s. What is the average induced emf in the loop?

4. (II) A 22.0-cm-diameter coil consists of 36 turns of circular copper wire 2.6 mm in diameter. A uniform magnetic field, perpendicular to the plane of the coil, changes at a rate of 8.65×10^{-3} T/s. Determine (*a*) the current in the loop, and (*b*) the rate at which thermal energy is produced.

5. (II) A 16-cm-diameter circular loop of wire is placed in a 0.65-T magnetic field. (*a*) When the plane of the loop is perpendicular to the field lines, what is the magnetic flux through the loop? (*b*) The plane of the loop is rotated until it makes a 35° angle with the field lines. What is the angle θ in Eq. 29–1a for this situation? (*c*) What is the magnetic flux through the loop at this angle?

6. (II) A fixed 10.8-cm-diameter wire coil is perpendicular to a magnetic field 0.48 T pointing up. In 0.14 s, the field is changed to 0.25 T pointing down. What is the average induced emf in the coil?

7. (II) If the solenoid in Fig. 29–47 is being pulled away from the loop shown, in what direction is the induced current in the loop? Explain.

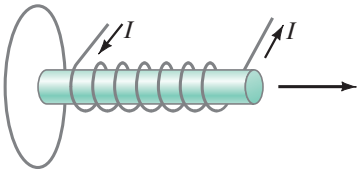


FIGURE 29–47
Problem 7.

8. (II) (*a*) If the resistance of the resistor in Fig. 29–48 is slowly increased, what is the direction of the current induced in the small circular loop inside the larger loop? (*b*) What would it be if the small loop were placed outside the larger one, to the left? Explain your answers.

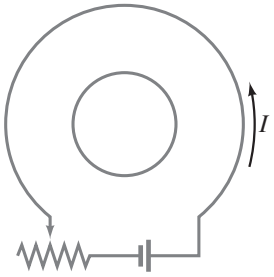


FIGURE 29–48
Problem 8.

9. (II) A circular loop in the plane of the paper lies in a 0.65-T uniform magnetic field pointing into the paper. The loop's diameter changes from 20.0 cm to 8.0 cm in 0.50 s. What is (*a*) the direction of the induced current, (*b*) the magnitude of the average induced emf, and (*c*) the average induced current if the coil resistance is 2.5 Ω?

10. (II) Part of a single rectangular loop of wire with dimensions shown in Fig. 29–49 is situated inside a region of uniform magnetic field of 0.650 T. The total resistance of the loop is 0.250 Ω. Calculate the force required to pull the loop from the field (to the right) at a constant velocity of 3.40 m/s. Neglect gravity.

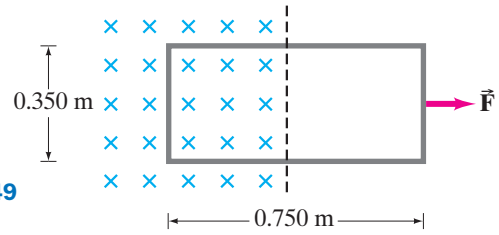


FIGURE 29–49
Problem 10.

11. (II) What is the direction of the induced current in the circuit loop due to the current shown in each part of Fig. 29–50? Explain why.

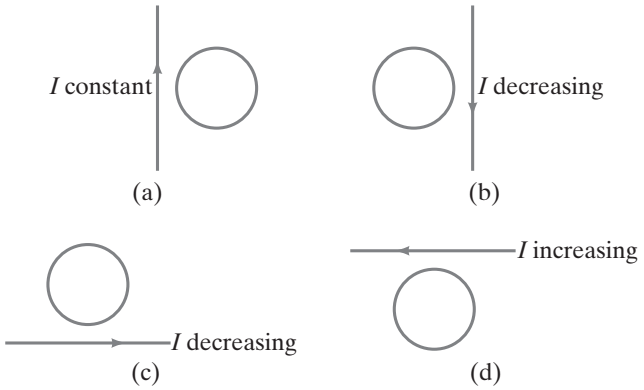


FIGURE 29–50 Problem 11.

12. (II) A circular wire loop of radius $r = 12$ cm is in a uniform magnetic field $B = 0.400$ T with its plane perpendicular to the direction of the field. If the field magnitude begins to decrease at a rate of -0.010 T/s, at what rate should r be increasing at this instant so that the induced emf within the loop is zero?

13. (II) The magnetic flux through each loop of a 75-loop coil is given by $(9.6t - 0.51t^3) \times 10^{-2}\ \text{T} \cdot \text{m}^2$, where the time t is in seconds. (*a*) Determine the emf $\mathcal{E}$ as a function of time. (*b*) What is $\mathcal{E}$ at $t = 1.0$ s and $t = 4.0$ s?

14. (II) A 25-cm-diameter circular loop of wire has a resistance of 135 Ω. It is initially in a 0.40-T magnetic field, with its plane perpendicular to $\vec{\mathbf{B}}$, but is removed from the field in 120 ms. Calculate the electric energy dissipated in this process.

15. (II) The area of an elastic circular loop decreases at a constant rate, $dA/dt = -3.50 \times 10^{-2}\ \text{m}^2/\text{s}$. The loop is in a magnetic field $B = 0.35$ T whose direction is perpendicular to the plane of the loop. At $t = 0$, the loop has area $A = 0.285\ \text{m}^2$. Determine the induced emf at $t = 0$, and at $t = 2.00$ s.

16. (II) Suppose the radius of the elastic loop in Problem 15 increases at a constant rate, $dr/dt = 3.75$ cm/s. Determine the emf induced in the loop at $t = 0$ and at $t = 1.00$ s.

17. (II) The magnetic field perpendicular to a single 13.2-cm-diameter circular loop of copper wire decreases uniformly from 0.760 T to zero. If the wire is 2.25 mm in diameter, how much charge moves past a point in the coil during this operation?

18. (II) A single circular loop of wire is placed inside a long solenoid with its plane perpendicular to the axis of the solenoid. The area of the loop is A_1 and that of the solenoid, which has n turns per unit length, is A_2. A current $I = I_0 \cos \omega t$ flows in the solenoid turns. What is the induced emf in the small loop?

19. (II) We are looking down on an elastic conducting loop with resistance $R = 2.0\ \Omega$, immersed in a magnetic field. The field's magnitude is uniform spatially, but varies with time t according to $B(t) = \alpha t$, where $\alpha = 0.60\ \text{T/s}$. The area A of the loop also increases at a constant rate, according to $A(t) = A_0 + \beta t$, where $A_0 = 0.50\ \text{m}^2$ and $\beta = 0.70\ \text{m}^2/\text{s}$. Find the magnitude and direction (clockwise or counterclockwise, when viewed from above the page) of the induced current within the loop at time $t = 2.0\ \text{s}$ if the magnetic field (*a*) is parallel to the plane of the loop to the right; (*b*) is perpendicular to the plane of the loop, down.

20. (III) A power line carrying a sinusoidally varying current with frequency $f = 60\ \text{Hz}$ and peak value $I_0 = 125\ \text{A}$ runs at a height of 7.0 m across a farmer's land (Fig. 29–51). The farmer, not appreciating this eyesore, decides to strike back and constructs a vertical 2.0-m-high 2000-turn rectangular wire coil below the power line. The farmer hopes to use the induced voltage in this coil to power 120-V electrical equipment, which requires a sinusoidally varying voltage with frequency $f = 60\ \text{Hz}$ and peak value $V_0 = 170\ \text{V}$. Estimate the length ℓ of the coil needed. Would this be stealing?

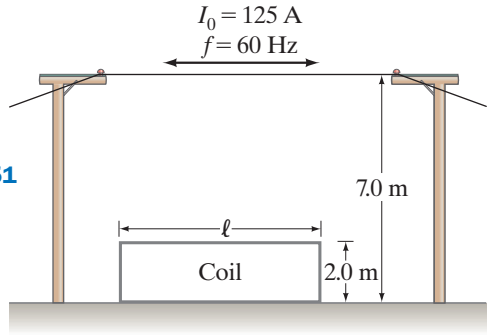


FIGURE 29–51
Problem 20.

21. (III) (*a*) Determine the magnetic flux through a square loop of side a (Fig. 29–52) if one side is parallel to, and a distance b from, a straight wire that carries a current I. (*b*) If the loop is pulled away from the wire at speed v, what emf is induced in it? (*c*) Does the induced current flow clockwise or counterclockwise? (*d*) Determine the force F required to pull the loop away.

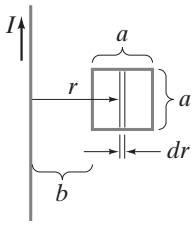


FIGURE 29–52
Problems 21 and 22.

22. (III) Determine the emf induced in the square loop in Fig. 29–52 if the loop stays at rest and the current in the straight wire is given by $I(t) = (15.0\ \text{A}) \sin(2200\, t)$ where t is in seconds. The distance a is 12.0 cm, and b is 15.0 cm.

29–3 Motional EMF

23. (I) The moving rod in Fig. 29–12b is 15.6 cm long and generates an emf of 120 mV while moving in a 0.90-T magnetic field. What is its speed?

24. (I) The moving rod in Fig. 29–12b is 12.0 cm long and is pulled at a speed of 15.0 cm/s. If the magnetic field is 0.650 T, calculate the emf developed.

25. (II) In Fig. 29–12a, the rod moves with a speed of 1.6 m/s on rails 30.0 cm apart. The rod has a resistance of 2.5 Ω. The magnetic field is 0.35 T, and the resistance of the U-shaped conductor is 24.0 Ω at a given instant. Calculate (*a*) the induced emf, (*b*) the current in the U-shaped conductor, and (*c*) the external force needed to keep the rod's velocity constant at that instant.

26. (II) If the U-shaped conductor in Fig. 29–12a has resistivity ρ, whereas that of the moving rod is negligible, derive a formula for the current I as a function of time. Assume the rod starts at the bottom of the U at $t = 0$, and moves with uniform speed v in the magnetic field B. The cross-sectional area of the rod and all parts of the U is A.

27. (II) In Fig. 29–12, the moving rod has a resistance of 0.25 Ω and moves on rails 20.0 cm apart. The stationary U-shaped conductor has negligible resistance. When a force of 0.350 N is applied to the rod, it moves to the right at a constant speed of 1.30 m/s. What is the magnetic field?

28. (II) Suppose that the U-shaped conductor and connecting rod in Fig. 29–12a are oriented vertically (but still in contact) so that the rod is falling due to the gravitational force. Find the terminal speed of the rod if it has mass $m = 3.6$ grams, length $\ell = 18$ cm, and resistance $R = 0.0013\ \Omega$. It is falling in a uniform horizontal field $B = 0.080\ \text{T}$. Neglect the resistance of the U-shaped conductor, and friction.

29. (II) A conducting rod rests on two long frictionless parallel rails in a magnetic field $\vec{\mathbf{B}}$ ($\perp$ to the rails and rod) as in Fig. 29–53. (*a*) If the rails are horizontal and the rod is given an initial push, will the rod travel at constant speed even though a magnetic field is present? (*b*) Suppose at $t = 0$, when the rod has speed $v = v_0$, the two rails are connected electrically by a wire from point a to point b. Assuming the rod has resistance R and the rails have negligible resistance, determine the speed of the rod as a function of time. Discuss your answer.

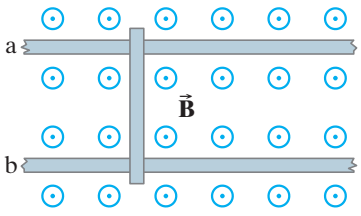


FIGURE 29–53
Problems 29 and 30.

30. (III) Suppose a conducting rod (mass m, resistance R) rests on two frictionless and resistanceless parallel rails a distance ℓ apart in a uniform magnetic field $\vec{\mathbf{B}}$ ($\perp$ to the rails and to the rod) as in Fig. 29–53. At $t = 0$, the rod is at rest and a source of emf is connected to the points a and b. Determine the speed of the rod as a function of time if (*a*) the source puts out a constant current I, (*b*) the source puts out a constant emf $\mathscr{E}_0$. (*c*) Does the rod reach a terminal speed in either case? If so, what is it?

31. (III) A short section of wire, of length a, is moving with velocity $\vec{\mathbf{v}}$, parallel to a very long wire carrying a current I as shown in Fig. 29–54. The near end of the wire section is a distance b from the long wire. Assuming the vertical wire is very long compared to $a + b$, determine the emf between the ends of this short section. Assume $\vec{\mathbf{v}}$ is (a) in the same direction as I, (b) in the opposite direction to I.

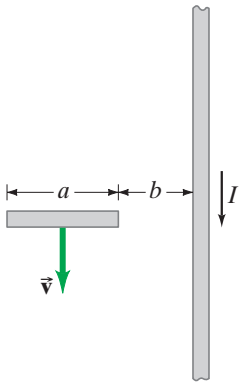


FIGURE 29–54
Problem 31.

29–4 Generators

32. (I) A simple generator is designed to generate a peak output voltage of 24.0 V. The square armature consists of windings that are 5.15 cm on a side and rotates in a field of 0.380 T at a rate of 60.0 rev/s. How many loops of wire should be wound on the square armature?
33. (II) The generator of a car idling at 875 rpm produces 12.4 V. What will the output be at a rotation speed of 1850 rpm assuming nothing else changes?
34. (II) A simple generator has a 440-loop square coil 22.0 cm on a side. How fast must it turn in a 0.550-T field to produce a 120-V peak output?
35. (II) A 280-loop circular armature coil with a diameter of 10.0 cm rotates at 120 rev/s in a uniform magnetic field of strength 0.45 T. (a) What is the rms voltage output of the generator? (b) What would you do to the rotation frequency in order to double the rms voltage output?
36. (II) A generator rotates at 85 Hz in a magnetic field of 0.030 T. It has 950 turns and produces an rms voltage of 150 V and an rms current of 60.0 A. (a) What is the peak current produced? (b) What is the area of each turn of the coil?
37. (II) Calculate the peak output voltage of a simple generator whose square armature windings are 5.50 cm on a side; the armature contains 125 loops and rotates in a field of 0.200 T at a rate of 120 rev/s.
38. (III) A simple generator has a square armature 6.0 cm on a side. The armature has 85 turns of 0.65-mm-diameter copper wire and rotates in a 0.65-T magnetic field. The generator is used to power a lightbulb rated at 12.0 V and 25.0 W. At what rate should the generator rotate to provide 12.0 V to the bulb? Consider the resistance of the wire on the armature.

29–5 Back EMF, Counter Torque; Eddy Current

39. (I) The back emf in a dc motor is 72 V when operating at 1400 rpm. What would be the back emf at 2500 rpm if the magnetic field is unchanged?
40. (I) A dc motor has an armature resistance of 2.65 Ω. If it draws 7.20 A when running at full speed and connected to a 120-V line, how large is the back emf?
41. (II) What will be the current in the motor of Example 29–11 if the load causes it to run at half speed?
42. (II) The back emf in a dc motor is 85 V when the motor is operating at 1400 rpm. How would you change the motor's magnetic field if you wanted to reduce the back emf to 75 V when the motor was running at 2300 rpm?
43. (II) A dc generator is rated at 16 kW, 250 V, and 56 A when it rotates at 1200 rpm. The resistance of the armature windings is 0.40 Ω. (a) Calculate the "no-load" voltage at 1200 rpm (when there is no circuit hooked up to the generator). (b) Calculate the full-load voltage (i.e., at 56 A) when the generator is run at 750 rpm. Assume that the magnitude of the magnetic field remains constant.

29–6 Transformers

[Assume 100% efficiency, unless stated otherwise.]

44. (I) A transformer is designed to change 117 V into 13,500 V, and there are 172 turns in the primary coil. How many turns are in the secondary coil?
45. (I) A transformer has 330 primary turns and 1120 secondary turns. The input voltage is 120 V and the output current is 15.0 A. What are the output voltage and input current?
46. (I) A transformer has 680 turns in the primary coil and 85 in the secondary coil. What kind of transformer is this, and by what factor does it change the voltage? By what factor does it change the current?
47. (I) Neon signs require 12 kV for their operation. To operate from a 240-V line, what must be the ratio of secondary to primary turns of the transformer? What would the voltage output be if the transformer were connected in reverse?
48. (II) A model-train transformer plugs into 120-V ac and draws 0.35 A while supplying 6.5 A to the train. (a) What voltage is present across the tracks? (b) Is the transformer step-up or step-down?
49. (II) If 75 MW of power at 45 kV (rms) arrives at a town from a generator via transmission lines of total resistance 3.0 Ω, calculate (a) the emf at the generator end of the lines, and (b) the fraction of the power generated that is wasted in the lines.
50. (II) Assume a voltage source supplies an ac voltage of amplitude V_0 between its output terminals. If the output terminals are connected to an external circuit, and an ac current of amplitude I_0 flows out of the terminals, then the equivalent resistance of the external circuit is $R_{eq} = V_0/I_0$. (a) If a resistor R is connected directly to the output terminals, what is R_{eq}? (b) If a transformer with N_P and N_S turns in its primary and secondary, respectively, is placed between the source and the resistor as shown in Fig. 29–55, what is R_{eq}? [Transformers can be used in ac circuits to alter the apparent resistance of circuit elements, such as loudspeakers, in order to maximize transfer of power.]

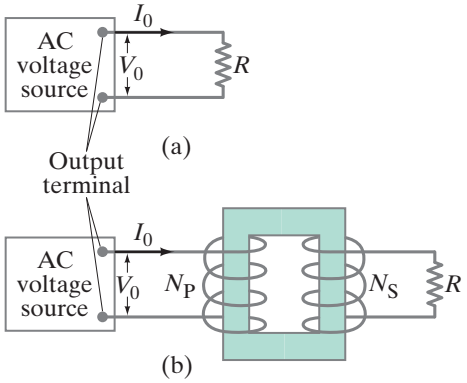


FIGURE 29–55
Problem 50.

51. (II) For the transmission of electric power from power plant to home, as depicted in Fig. 29–26, where the electric power sent by the plant is 100 kW, about how far away could the house be from the power plant before power loss is 50%? Assume the wires have a resistance per unit length of $4 \times 10^{-5}\ \Omega/\text{m}$.

52. (II) For the electric power transmission system shown in Fig. 29–26, what is the ratio N_S/N_P for (*a*) the step-up transformer, (*b*) the step-down transformer next to the home?

53. (III) Design a dc transmission line that can generate 625 MW of electricity and transmit it 485 km with only a 2.0% loss. The wires are to be made of aluminum and the voltage is 660 kV.

54. (III) Suppose 785 kW is to arrive at a shopping mall over two 0.100-Ω lines. Estimate how much power is saved if the voltage is stepped up from 120 V to 1200 V and then down again, rather than simply transmitting at 120 V. Assume the transformers are each 99% efficient.

29–7 Changing Φ_B Produces $\vec{E}$

55. (III) In a circular region, there is a uniform magnetic field $\vec{\mathbf{B}}$ pointing into the page (Fig. 29–56). An *xy* coordinate system has its origin at the circular region's center. A free positive point charge $+Q = 1.0\,\mu C$ is initially at rest at a position $x = +10\,cm$ on the *x* axis. If the magnitude of the magnetic field is now decreased at a rate of $-0.10\,T/s$, what force (magnitude and direction) will act on $+Q$?

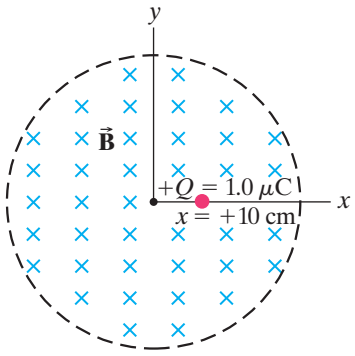


FIGURE 29–56 Problem 55.

56. (III) The **betatron**, a device used to accelerate electrons to high energy, consists of a circular vacuum tube placed in a magnetic field (Fig. 29–57), into which electrons are injected. The electromagnet produces a field that (1) keeps the electrons in their circular orbit inside the tube, and (2) increases the speed of the electrons when *B* changes. (*a*) Explain how the electrons are accelerated. (See Fig. 29–57.) (*b*) In what direction are the electrons moving in Fig. 29–57 (give directions as if looking down from above)? (*c*) Should *B* increase or decrease to accelerate the electrons? (*d*) The magnetic field is actually 60-Hz ac; show that the electrons can be accelerated only during $\frac{1}{4}$ of a cycle $\left(\frac{1}{240}\,s\right)$. (During this time they make hundreds of thousands of revolutions and acquire very high energy.)

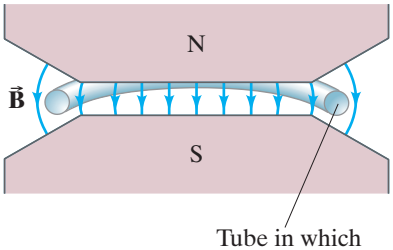


FIGURE 29–57
Problems 56 and 57.

57. (III) Show that the electrons in a betatron, Problem 56 and Fig. 29–57, are accelerated at constant radius if the magnetic field B_0 at the position of the electron orbit in the tube is equal to half the average value of the magnetic field (B_{avg}) over the area of the circular orbit at each moment: $B_0 = \frac{1}{2}B_{avg}$. (This is the reason the pole faces have a rather odd shape, as indicated in Fig. 29–57.)

58. (III) Find a formula for the net electric field in the moving rod of Problem 30 as a function of time for each case, (*a*) and (*b*).

General Problems

59. A square loop 27.0 cm on a side has a resistance of 8.80 Ω. It is initially in a 0.755-T magnetic field, with its plane perpendicular to $\vec{\mathbf{B}}$, but is removed from the field in 40.0 ms. Calculate the electric energy dissipated in this process.

60. Power is generated at 24 kV at a generating plant located 75 km from a town that requires 65 MW of power at 12 kV. Two transmission lines from the plant to the town each have a resistance of 0.10 Ω/km. What should the output voltage of the transformer at the generating plant be for an overall transmission efficiency of 98.5%, assuming a perfect transformer?

61. A circular loop of area $12\,m^2$ encloses a magnetic field perpendicular to the plane of the loop; its magnitude is $B(t) = (8.0\,T/s)t$. The loop is connected to a 7.5-Ω resistor and a 6.5-pF capacitor in series. When fully charged, how much charge is stored on the capacitor?

62. The primary windings of a transformer which has an 85% efficiency are connected to 110-V ac. The secondary windings are connected across a 2.4-Ω, 75-W lightbulb. (*a*) Calculate the current through the primary windings of the transformer. (*b*) Calculate the ratio of the number of primary windings of the transformer to the number of secondary windings of the transformer.

63. A pair of power transmission lines each have a 0.80-Ω resistance and carry 680 A over 9.0 km. If the rms input voltage is 42 kV, calculate (*a*) the voltage at the other end, (*b*) the power input, (*c*) power loss in the lines, and (*d*) the power output.

64. Show that the power loss in transmission lines, P_L, is given by $P_L = (P_T)^2 R_L/V^2$, where P_T is the power transmitted to the user, *V* is the delivered voltage, and R_L is the resistance of the power lines.

65. A high-intensity desk lamp is rated at 35 W but requires only 12 V. It contains a transformer that converts 120-V household voltage. (*a*) Is the transformer step-up or step-down? (*b*) What is the current in the secondary coil when the lamp is on? (*c*) What is the current in the primary coil? (*d*) What is the resistance of the bulb when on?

66. Two parallel resistanceless rails are 32 cm apart on a 6.0° ramp. They are joined at the bottom by a 0.60-Ω resistor. At the top a copper bar of mass 0.040 kg (ignore its resistance) is laid across the rails. Assuming a vertical 0.65-T magnetic field, what is the terminal (steady) velocity of the bar as it slides frictionlessly down the rails?

67. A coil with 160 turns, a radius of 5.0 cm, and a resistance of 12 Ω surrounds a solenoid with 230 turns/cm and a radius of 4.5 cm (Fig. 29–58). The current in the solenoid changes at a constant rate from 0 to 2.0 A in 0.10 s. Calculate the magnitude and direction of the induced current through the resistor in the outer coil.

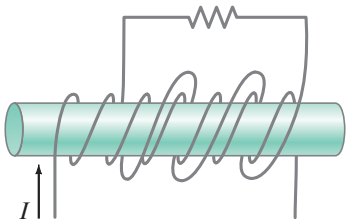


FIGURE 29–58 Problem 67.

68. A **search coil** for measuring B (also called a **flip coil**) is a small coil with N turns, each of cross-sectional area A. It is connected to a so-called **ballistic galvanometer**, which is a device to measure the total charge Q that passes through it in a short time. The flip coil is placed in the magnetic field to be measured with its face perpendicular to the field. It is then quickly rotated 180° about a diameter. Show that the total charge Q that flows in the induced current during this short "flip" time is proportional to the magnetic field B. In particular, show that B is given by

$$B = \frac{QR}{2NA}$$

where R is the total resistance of the circuit, including that of the coil and that of the ballistic galvanometer which measures the charge Q.

69. A ring with a radius of 3.0 cm and a resistance of 0.035 Ω is rotated about an axis through its diameter by 90° in a magnetic field of 0.23 T perpendicular to that axis. What is the largest number of electrons that would flow past a fixed point in the ring as this process is accomplished?

70. A flashlight can be made that is powered by the induced current from a magnet moving through a coil of wire. The coil and magnet are inside a plastic tube that can be shaken by your hand causing the magnet to move back and forth through the coil. Assume the magnet has a maximum field strength of 0.05 T. Make reasonable assumptions and specify the size of the coil and the number of turns necessary to light a standard 1-watt, 3-V flashlight bulb.

71. Conceptual Example 29–12 states that an overloaded motor may burn out due to high currents. Suppose you have a blender with an internal resistance of 3.0 Ω. (*a*) At 120 V, what is the initial current through the blender? (*b*) The blender is rated at 2.0 A for continuous use. What is the back emf of the blender? (*c*) At what rate is heat dissipated in the blender during normal use? (*d*) If the blender jams and stops turning, at what rate is heat dissipated in the motor coils?

72. While demonstrating Faraday's law to her class, a physics professor inadvertently moves the gold ring on her finger from a location where a 0.60-T magnetic field points along her finger to a zero-field location in 45 ms. The 1.5-cm-diameter ring has a resistance and mass of 55 μΩ and 15 g, respectively. (*a*) Estimate the thermal energy produced in the ring due to the flow of induced current. (*b*) Find the temperature rise of the ring, assuming all of the thermal energy produced goes into increasing the ring's temperature. The specific heat of gold is 129 J/kg · C°.

73. Suppose you are looking at two wire loops in the plane of the page as shown in Fig. 29–59. When the switch S is closed in the left-hand coil, (*a*) what is the direction of the induced current in the other loop? (*b*) What is the situation after a "long" time? (*c*) What is the direction of the induced current in the right-hand loop if that loop is quickly pulled horizontally to the right? (*d*) Suppose the right-hand loop also has a switch like the left-hand loop. The switch in the left-hand loop has been closed a long time when the switch in the right-hand loop is closed. What happens in this case? Explain each answer.

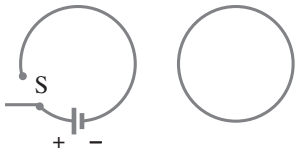


FIGURE 29–59
Problem 73.

74. A small electric car overcomes a 250-N friction force when traveling 35 km/h. The electric motor is powered by ten 12-V batteries connected in series and is coupled directly to the wheels whose diameters are 58 cm. The 320 armature coils are rectangular, 12 cm by 15 cm, and rotate in a 0.54-T magnetic field. (*a*) How much current does the motor draw to produce the required torque? (*b*) What is the back emf? (*c*) How much power is dissipated in the coils? (*d*) What percent of the input power is used to drive the car?

75. What is the energy dissipated as a function of time in a circular loop of 18 turns of wire having a radius of 10.0 cm and a resistance of 2.0 Ω if the plane of the loop is perpendicular to a magnetic field given by

$$B(t) = B_0 e^{-t/\tau}$$

with $B_0 = 0.50$ T and $\tau = 0.10$ s?

76. A thin metal rod of length ℓ rotates with angular velocity ω about an axis through one end (Fig. 29–60). The rotation axis is perpendicular to the rod and is parallel to a uniform magnetic field $\vec{\mathbf{B}}$. Determine the emf developed between the ends of the rod.

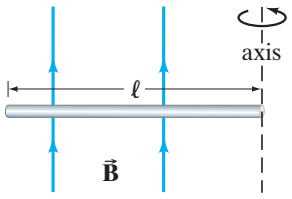


FIGURE 29–60
Problem 76.

77. The magnetic field of a "shunt-wound" dc motor is produced by field coils placed in parallel with the armature coils. Suppose that the field coils have a resistance of 32.0 Ω and the armature coils 3.00 Ω. The back emf at full speed is 105 V when the motor is connected to 115 V dc. (*a*) Draw the equivalent circuit for the situations when the motor is just starting and when it is running full speed. (*b*) What is the total current drawn by the motor at start up? (*c*) What is the total current drawn when the motor runs at full speed?

78. Apply Faraday's law, in the form of Eq. 29–8, to show that the static electric field between the plates of a parallel-plate capacitor cannot drop abruptly to zero at the edges, but must, in fact, fringe. Use the path shown dashed in Fig. 29–61. [*Hint*: Assume the contrary: that there is no fringing. Show that this assumption leads to a contradiction.]

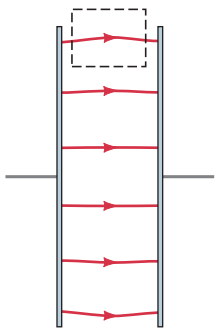


FIGURE 29–61
Problem 78.

79. A circular metal disk of radius R rotates with angular velocity ω about an axis through its center perpendicular to its face. The disk rotates in a uniform magnetic field B whose direction is parallel to the rotation axis. Determine the emf induced between the center and the edges of the disk.

80. What is the magnitude and direction of the electric field at each point in the rotating disk of Problem 79?

81. A circular-shaped circuit of radius r, containing a resistance R and capacitance C, is situated with its plane perpendicular to a spatially uniform magnetic field $\vec{\mathbf{B}}$ directed into the page (Fig. 29–62). Starting at time $t = 0$, the voltage difference $V_{ba} = V_b - V_a$ across the capacitor plates is observed to increase with time t according to $V_{ba} = V_0(1 - e^{-t/\tau})$, where V_0 and τ are positive constants. Determine dB/dt, the rate at which the magnetic field magnitude changes with time. Is B becoming larger or smaller as time increases?

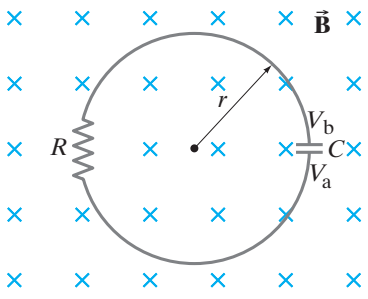


FIGURE 29–62
Problem 81.

82. In a certain region of space near Earth's surface, a uniform horizontal magnetic field of magnitude B exists above a level defined to be $y = 0$. Below $y = 0$, the field abruptly becomes zero (Fig. 29–63). A vertical square wire loop has resistivity ρ, mass density ρ_m, diameter d, and side length ℓ. It is initially at rest with its lower horizontal side at $y = 0$ and is then allowed to fall under gravity, with its plane perpendicular to the direction of the magnetic field. (*a*) While the loop is still partially immersed in the magnetic field (as it falls into the zero-field region), determine the magnetic "drag" force that acts on it at the moment when its speed is v. (*b*) Assume that the loop achieves a constant terminal velocity v_T before its upper horizontal side exits the field. Determine a formula for v_T. (*c*) If the loop is made of copper and $B = 0.80$ T, find v_T.

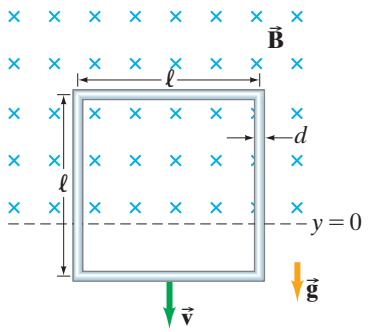


FIGURE 29–63
Problem 82.

83. A single, square loop of wire just fits inside a long, conventional solenoid of radius R with n turns per unit length. The plane of the square loop is perpendicular to the axis of the solenoid and far from the ends. The current through the solenoid is given by the expression $I(t) = I_0 \cos(\omega t)$ where I_0 is the maximum current and ω is the applied angular frequency. What is the maximum emf developed by the square loop?

84. Magnetic flux passes through a stationary loop of wire with resistance R and this flux varies from $t = 0$, to $t = T/2$ such that $\Phi(t) = b \sin(2\pi t/T)$ where b is a given constant. Estimate the energy dissipated in the loop during that time.

85. In an experiment, a coil was mounted on a low-friction cart that moved through the magnetic field B of a permanent magnet. The speed of the cart v and the induced voltage V were simultaneously measured, as the cart moved through the magnetic field, using a computer-interfaced motion sensor and a voltmeter. The Table below shows the collected data:

Speed, v (m/s)	0.367	0.379	0.465	0.623	0.630
Induced voltage, V (V)	0.128	0.135	0.164	0.221	0.222

(*a*) Make a graph of the induced voltage, V, vs. the speed, v. Determine a best-fit linear equation for the data. Theoretically, the relationship between V and v is given by $V = BN\ell v$ where N is the number of turns of the coil, B is the magnetic field, and ℓ is the average of the inside and outside widths of the coil. In the experiment, $B = 0.126$ T, $N = 50$, and $\ell = 0.0561$ m. (*b*) Find the % error between the slope of the experimental graph and the theoretical value for the slope. (*c*) For each of the measured speeds v, determine the theoretical value of V and find the % error of each.

ANSWERS TO EXERCISES

A: (*e*).
B: (*a*) Counterclockwise; (*b*) clockwise; (*c*) zero; (*d*) counterclockwise.
C: Electrons flow clockwise (conventional current counterclockwise).
D: (*a*) Increase (brighter light); (*b*) yes; resists more (counter torque).
E: 10 turns.
F: (*a*) 4.2×10^{-3} V/m; (*b*) 5.6×10^{-3} V/m.



This electric car is having its electric batteries charged by induction. The flat device, connected to a cord and plug, contains a coil of wire and so does that part of the car above it. The changing current (ac) in the flat device induces an emf in the car's coil. That induced emf charges the car's batteries. This is an example of mutual inductance (like a transformer). Electric cars can also be charged by a direct connection.

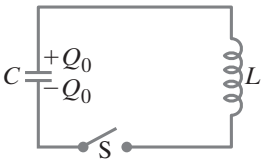
Any coil of wire has a self-inductance, and a changing current in it causes an emf to be induced in its circuit. Such inductors are useful in many circuits.

CHAPTER 30

Inductance, Electromagnetic Oscillations, and AC Circuits

CHAPTER-OPENING QUESTION—Guess now!

Consider a circuit with only a capacitor C and a coil of many loops of wire (called an inductor, L) as shown. If the capacitor is initially charged $(Q = Q_0)$, what will happen when the switch S is closed?



(a) Nothing will happen—the capacitor will remain charged with charge $Q = Q_0$.

(b) The capacitor will quickly discharge and remain discharged $(Q = 0)$.

(c) Current will flow until the positive charge is on the opposite plate of the capacitor, and then will reverse—back and forth.

(d) The energy initially in the capacitor $\left(U_E = \frac{1}{2}Q_0^2/C\right)$ will all transfer to the coil and then remain that way.

(e) The system will quickly transfer half of the capacitor energy to the coil and then remain that way.

We discussed in the last Chapter how a changing magnetic flux through a circuit induces an emf in that circuit. Before that we saw that an electric current produces a magnetic field. Combining these two ideas, we could predict that a changing current in one circuit ought to induce an emf and a current in a second nearby circuit, and even induce an emf in itself. We already saw an example in the previous Chapter (transformers), but now we will treat this effect in a more general way in terms of what we will call mutual inductance and self-inductance. Almost every circuit contains some inductance (sometimes very little), which we have ignored up to now. But in the real world of electronic circuits, inductance is often of great importance. In fact inductance, along with resistance and capacitance, are the three main elements in simple circuits as we discuss later in this Chapter. The concept of inductance also gives us a springboard to treat energy storage in a magnetic field.

CONTENTS

30 – 1 Mutual Inductance
30 – 2 Self-Inductance; Inductors
30 – 3 Energy Stored in a Magnetic Field
30 – 4 *LR* Circuits
30 – 5 *LC* Circuits and Electromagnetic Oscillations
30 – 6 *LC* Oscillations with Resistance (*LRC* Circuit)
30 – 7 AC Circuits and Reactance
30 – 8 *LRC* Series AC Circuit; Phasor Diagrams
30 – 9 Resonance in AC Circuits
30–10 Impedance Matching
*30–11 Three-Phase AC

30–1 Mutual Inductance

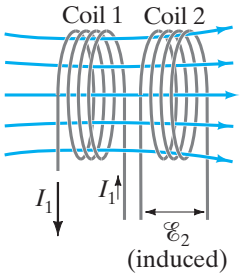


FIGURE 30–1 A changing current in one coil will induce a current in the second coil.

If two coils of wire are near each other, as in Fig. 30–1, a changing current in one will induce an emf in the other. (A transformer—Chapter 29—is but one example.) Let us apply Faraday's law to coil 2: the emf $\mathscr{E}_2$ induced in coil 2 is proportional to the rate of change of magnetic flux passing through it. This changing flux in coil 2 is produced by a changing current I_1 in coil 1, and it is often convenient to express the emf in coil 2 in terms of the current in coil 1.

We let Φ_{21} be the magnetic flux in each loop of coil 2 created by the current in coil 1. If coil 2 contains N_2 closely wrapped loops, then $N_2\Phi_{21}$ is the total flux passing through coil 2. If the two coils are fixed in space, $N_2\Phi_{21}$ is proportional to the current I_1 in coil 1; the proportionality constant is called the **mutual inductance**, M_{21}, defined by

$$M_{21} = \frac{N_2\Phi_{21}}{I_1}. \qquad \textbf{(30–1)}$$

The emf $\mathscr{E}_2$ induced in coil 2 due to a changing current in coil 1 is, by Faraday's law,

$$\mathscr{E}_2 = -N_2\frac{d\Phi_{21}}{dt}.$$

We combine this with Eq. 30–1 rewritten as $\Phi_{21} = M_{21}I_1/N_2$ (and take its derivative) and obtain

$$\mathscr{E}_2 = -M_{21}\frac{dI_1}{dt}. \qquad \textbf{(30–2)}$$

Equation 30–2 relates the change in current in coil 1 to the emf it induces in coil 2. If the two coils are fixed in space, the mutual inductance M_{21} of coil 2 with respect to coil 1 is a "constant" in that it does not depend on I_1; M_{21} depends on "geometric" factors such as the size, shape, number of turns, and relative positions of the two coils, and also on whether iron (or some other ferromagnetic material) is present. For example, if the two coils in Fig. 30–1 are farther apart, fewer lines of flux can pass through coil 2, so M_{21} will be less. For some arrangements, the mutual inductance can be calculated (see Example 30–1). More often it is determined experimentally.

Now consider the reverse situation: when a changing current in coil 2 induces an emf in coil 1. In this case,

$$\mathscr{E}_1 = -M_{12}\frac{dI_2}{dt}$$

where M_{12} is the mutual inductance of coil 1 with respect to coil 2. It is possible to show, although we will not do it here, that $M_{12} = M_{21}$. Hence, for a given arrangement we do not need the subscripts and we can let

$$M = M_{12} = M_{21},$$

so that

$$\mathscr{E}_1 = -M\frac{dI_2}{dt} \qquad \textbf{(30–3a)}$$

and

$$\mathscr{E}_2 = -M\frac{dI_1}{dt}. \qquad \textbf{(30–3b)}$$

The SI unit for mutual inductance is the **henry** (H), where $1\ \text{H} = 1\ \text{V}\cdot\text{s/A} = 1\ \Omega\cdot\text{s}$.

EXAMPLE 30–1 **Solenoid and coil.** A long thin solenoid of length ℓ and cross-sectional area A_1 contains N_1 closely packed turns of wire. Wrapped around it is an insulated coil of N_2 turns, Fig. 30–2. Assume all the flux from coil 1 (the solenoid) passes through coil 2, and calculate the mutual inductance.

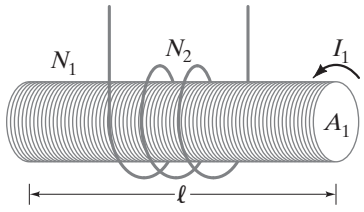


FIGURE 30–2 Example 30–1.

APPROACH We first determine the flux produced by the solenoid, all of which passes uniformly through coil N_2 (as well as N_1), using Eq. 28–4 for the magnetic field inside a solenoid:

$$B = \mu_0 \frac{N_1}{\ell} I_1 = \mu_0 n I_1,$$

where $n = N_1/\ell$ is the number of loops in the solenoid per unit length, and I_1 is the current in the solenoid.

SOLUTION The solenoid is closely packed, so we assume that all the flux in the solenoid stays inside the secondary coil. Then the flux Φ_{21} through coil 2 is

$$\Phi_{21} = BA_1 = \mu_0 \frac{N_1}{\ell} I_1 A_1.$$

So the mutual inductance is

$$M = \frac{N_2 \Phi_{21}}{I_1} = \frac{\mu_0 N_1 N_2 A_1}{\ell}.$$

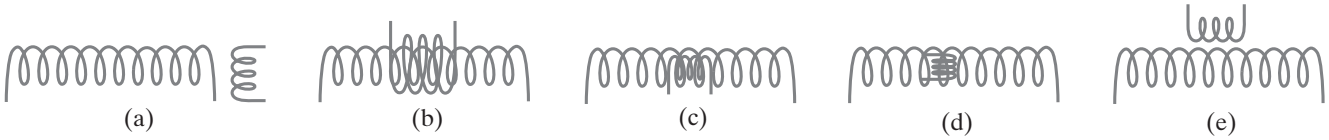
NOTE We calculated M_{21}; if we had tried to calculate M_{12}, it would have been difficult. Given $M_{12} = M_{21} = M$, we did the simpler calculation to obtain M. Note again that M depends only on geometric factors, and not on the currents.

CONCEPTUAL EXAMPLE 30–2 **Reversing the coils.** How would Example 30–1 change if the coil with N_2 turns was inside the solenoid rather than outside the solenoid?

RESPONSE The magnetic field inside the solenoid would be unchanged. The flux through the coil would be BA_2 where A_2 is the area of the coil N_2, not of the solenoid as in Example 30–1. Solving for M would give the same formula except that the area A_1 would be replaced by the area A_2 of coil N_2, and M would be smaller because we have changed the configuration.

EXERCISE A Which solenoid and coil combination shown in Fig. 30–3 has the largest mutual inductance? Assume each solenoid is the same.

FIGURE 30–3 Exercise A.



A transformer is an example of mutual inductance in which the coupling is maximized so that nearly all flux lines pass through both coils. Mutual inductance has other uses as well, including inductive charging of cell phones, electric cars, and other devices with rechargeable batteries, as we discussed in Section 29–6.

Mutual inductance can sometimes be a problem. Any changing current in a circuit can induce an emf in another part of the same circuit or in a different circuit even though the conductors are not in the shape of a coil. The mutual inductance M is usually small unless coils with many loops and/or iron cores are involved. However, in situations where small voltages are being used, problems due to mutual inductance often arise. Shielded cable (or coaxial), in which an inner conductor is surrounded by a cylindrical grounded conductor (Fig. 28–12), is often used to reduce the problem.

30–2 Self-Inductance; Inductors

The concept of inductance applies also to a single isolated coil of N turns. When a changing current passes through a coil (or solenoid), a changing magnetic flux is produced inside the coil, and this in turn induces an emf in that same coil. This induced emf opposes the change in flux (Lenz's law); it is much like the back emf generated in a motor. For example, if the current through the coil is increasing, the increasing magnetic flux induces an emf that opposes the original current and tends to retard its increase. If the current is decreasing in the coil, the decreasing flux induces an emf in the same direction as the current, thus tending to maintain the original current. If this inductance (coil) is in a circuit, it thus can provide a source of emf, in addition to any battery present or other sources of emf.

The magnetic flux Φ_B passing through the N turns of a coil is proportional to the current I in the coil, so we define the **self-inductance** L (in analogy to mutual inductance, Eq. 30–1) as

$$L = \frac{N\Phi_B}{I}. \tag{30–4}$$

Then the emf $\mathscr{E}$ induced in a coil of self-inductance L is, from Faraday's law,

$$\mathscr{E} = -N\frac{d\Phi_B}{dt} = -L\frac{dI}{dt}. \tag{30–5}$$

Like mutual inductance, self-inductance is measured in henrys. The magnitude of L depends on the size and shape of the coil and on the presence of an iron core. Self-inductance (inductance, for short) can be defined, as above, for any circuit or part of a circuit.

An ac circuit always contains some inductance, but often it is quite small unless the circuit contains a coil of many loops or turns. A coil that has significant self-inductance L is called an **inductor**. Inductance is shown on circuit diagrams by the symbol


[inductor symbol]

An inductor may also have significant (ordinary) resistance, which should also be shown in a circuit diagram separately. Inductance can serve a useful purpose in certain circuits. Often, however, inductance is a nuisance in a circuit. Precision resistors are often wire wound and thus would have inductance as well as resistance. The inductance can be minimized by winding the insulated wire back on itself in the opposite sense so that the current going in opposite directions produces little net magnetic flux; this is called a **noninductive winding**.

If an inductor has negligible resistance, it is the inductance (or induced emf) that controls a changing current. If a source of changing or alternating voltage is applied to the coil, this applied voltage will just be balanced by the induced emf of the coil (Eq. 30–5). Thus we can see from Eq. 30–5 that, for a given $\mathscr{E}$, if the inductance L is large, the change in the current will be small, and therefore the current itself if it is ac will be small. The greater the inductance, the less the ac current. An inductance thus acts something like a "resistance" to impede the flow of alternating current. We use the term *reactance* or *impedance* for this quality of an inductor. We will discuss reactance and impedance in ac circuits more fully in Sections 30–7 and 30–8. We shall see that reactance depends not only on the inductance L, but also on the frequency of the ac source. Here we mention one example of its importance. The resistance of the primary in a transformer is usually quite small, perhaps less than 1 Ω. If resistance alone limited the current in a transformer, tremendous currents would flow when a high voltage was applied. Indeed, a dc voltage applied to a transformer can burn it out. It is the induced emf (or reactance) of the coil that limits the current to a reasonable value.

Common inductors have inductances in the range from about 1 μH to about 1 H (where 1 H = 1 henry = 1 $\Omega \cdot$ s).

EXAMPLE 30–3 **Solenoid inductance.** (*a*) Determine a formula for the self-inductance L of a long tightly wrapped solenoid containing N turns of wire in its length ℓ and whose cross-sectional area is A. (*b*) Calculate the value of L if $N = 100$, $\ell = 5.0\,\text{cm}$, $A = 0.30\,\text{cm}^2$ and the solenoid is air filled.

APPROACH To determine the inductance L, it is usually simplest to start with Eq. 30–4, so we need to first determine the flux.

SOLUTION (*a*) According to Eq. 28–4, the magnetic field inside a solenoid (ignoring end effects) is constant: $B = \mu_0 nI$ where $n = N/\ell$. The flux is $\Phi_B = BA = \mu_0 NIA/\ell$, so

$$L = \frac{N\Phi_B}{I} = \frac{\mu_0 N^2 A}{\ell}.$$

(*b*) Since $\mu_0 = 4\pi \times 10^{-7}\,\text{T} \cdot \text{m/A}$, then

$$L = \frac{(4\pi \times 10^{-7}\,\text{T} \cdot \text{m/A})(100)^2(3.0 \times 10^{-5}\,\text{m}^2)}{(5.0 \times 10^{-2}\,\text{m})} = 7.5\,\mu\text{H}.$$

NOTE Magnetic field lines "stray" out of the solenoid (see Fig. 28–15), especially near the ends, so our formula is only an approximation.

CONCEPTUAL EXAMPLE 30–4 **Direction of emf in inductor.** Current passes through the coil in Fig. 30–4 from left to right as shown. (*a*) If the current is increasing with time, in which direction is the induced emf? (*b*) If the current is decreasing in time, what then is the direction of the induced emf?

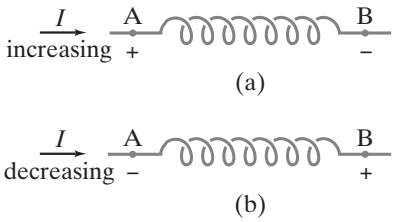
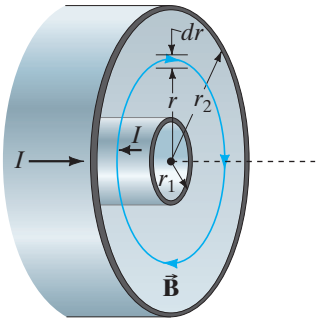


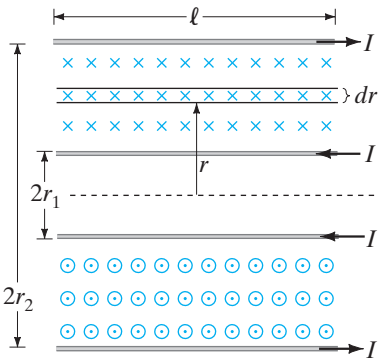
FIGURE 30–4 Example 30–4. The + and − signs refer to the induced emf due to the changing current, as if points A and B were the terminals of a battery (and the coiled loops were the inside of the battery).

RESPONSE (*a*) From Lenz's law we know that the induced emf must oppose the change in magnetic flux. If the current is increasing, so is the magnetic flux. The induced emf acts to oppose the increasing flux, which means it acts like a source of emf that opposes the original source of emf driving the current. So the induced emf in the coil acts to oppose I in Fig. 30–4a. In other words, the inductor might be thought of as a battery with a positive terminal at point A (tending to block the current entering at A), and negative at point B.

(*b*) If the current is decreasing, then by Lenz's law the induced emf acts to increase the flux—like a source of emf reinforcing the original emf. The induced emf acts to increase I in Fig. 30–4b, so in this situation you can think of the induced emf as a battery with its negative terminal at point A to attract more current (conventional, +) to move to the right.



(a)



(b)

FIGURE 30–5 Example 30–5. Coaxial cable: (a) end view, (b) side view (cross section).

EXAMPLE 30–5 **Inductance of coaxial cylinders.** Determine the inductance per unit length of a coaxial cable whose inner conductor has a radius r_1 and the outer conductor has a radius r_2, Fig. 30–5. Assume the conductors are thin hollow tubes so there is no magnetic field within the inner conductor, and the magnetic field inside both thin conductors can be ignored. The conductors carry equal currents I in opposite directions.

APPROACH We need to find the magnetic flux, $\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}$, between the conductors. The lines of $\vec{\mathbf{B}}$ are circles surrounding the inner conductor (only one is shown in Fig. 30–5a). From Ampère's law, $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I$, the magnitude of the field along a circle at a distance r from the center, when the inner conductor carries a current I, is (Examples 28–6 and 28–7):

$$B = \frac{\mu_0 I}{2\pi r}.$$

The magnetic flux through a rectangle of width dr and length ℓ (along the cable, Fig. 30–5b), a distance r from the center, is

$$d\Phi_B = B\,dA = B(\ell\,dr) = \frac{\mu_0 I}{2\pi r}\ell\,dr.$$

SOLUTION The total flux in a length ℓ of cable is

$$\Phi_B = \int d\Phi_B = \frac{\mu_0 I \ell}{2\pi}\int_{r_1}^{r_2}\frac{dr}{r} = \frac{\mu_0 I \ell}{2\pi}\ln\frac{r_2}{r_1}.$$

Since the current I all flows in one direction in the inner conductor, and the same current I all flows in the opposite direction in the outer conductor, we have only one turn, so $N = 1$ in Eq. 30–4. Hence the self-inductance for a length ℓ is

$$L = \frac{\Phi_B}{I} = \frac{\mu_0 \ell}{2\pi}\ln\frac{r_2}{r_1}.$$

The inductance per unit length is

$$\frac{L}{\ell} = \frac{\mu_0}{2\pi}\ln\frac{r_2}{r_1}.$$

Note that the inductance L depends only on geometric factors and not on the current I.

30–3 Energy Stored in a Magnetic Field

When an inductor of inductance L is carrying a current I which is changing at a rate dI/dt, energy is being supplied to the inductor at a rate

$$P = I\mathscr{E} = LI\frac{dI}{dt}$$

where P stands for power and we used[†] Eq. 30–5. Let us calculate the work needed to increase the current in an inductor from zero to some value I. Using this last equation, the work dW done in a time dt is

$$dW = P\,dt = LI\,dI.$$

Then the total work done to increase the current from zero to I is

$$W = \int dW = \int_0^I LI\,dI = \tfrac{1}{2}LI^2.$$

This work done is equal to the energy U stored in the inductor when it is carrying a current I (and we take $U = 0$ when $I = 0$):

$$U = \tfrac{1}{2}LI^2. \qquad \textbf{(30–6)}$$

This can be compared to the energy stored in a capacitor, C, when the potential difference across it is V (see Section 24–4):

$$U = \tfrac{1}{2}CV^2.$$

[†]No minus sign here because we are supplying power to oppose the emf of the inductor.

EXERCISE B What is the inductance of an inductor if it has a stored energy of 1.5 J when there is a current of 2.5 A in it? (*a*) 0.48 H, (*b*) 1.2 H, (*c*) 2.1 H, (*d*) 4.7 H, (*e*) 19 H.

Just as the energy stored in a capacitor can be considered to reside in the electric field between its plates, so the energy in an inductor can be considered to be stored in its magnetic field. To write the energy in terms of the magnetic field, we use the result of Example 30–3, that the inductance of an ideal solenoid (end effects ignored) is $L = \mu_0 N^2 A/\ell$. The magnetic field B inside a solenoid is related to the current I by $B = \mu_0 NI/\ell$ (Eq. 28–4), so

$$U = \tfrac{1}{2}LI^2 = \frac{1}{2}\left(\frac{\mu_0 N^2 A}{\ell}\right)\left(\frac{B\ell}{\mu_0 N}\right)^2$$
$$= \frac{1}{2}\frac{B^2}{\mu_0}A\ell.$$

We can think of this energy as residing in the volume enclosed by the windings, which is $A\ell$. Then the energy per unit volume or **energy density** is

$$u = \text{energy density} = \frac{1}{2}\frac{B^2}{\mu_0}. \qquad \textbf{(30–7)}$$

This formula for the energy density in a magnetic field, which was derived for the special case of a solenoid, can be shown to be valid for any region of space where a magnetic field is present. If a ferromagnetic material is present, μ_0 is replaced by μ. This equation is analogous to that for an electric field, $\frac{1}{2}\epsilon_0 E^2$, Eq. 24–6.

30–4 *LR* Circuits

Any inductor will have some resistance. We represent an inductor by drawing the inductance L and its resistance R separately, as in Fig. 30–6a. The resistance R could also include any other resistance in the circuit. Now we ask, what happens when a battery of voltage V_0 is connected in series to such an LR circuit? At the instant the switch connecting the battery is closed, the current starts to flow. It is opposed by the induced emf in the inductor because of the changing current. As soon as current starts to flow, there is also a voltage drop across the resistance ($V = IR$). Hence, the voltage drop across the inductance is reduced and the current increases less rapidly. The current thus rises gradually, as shown in Fig. 30–6b, and approaches the steady value $I_0 = V_0/R$, where there is no more emf in the inductor (I is no longer changing) so all the voltage drop is across the resistance.

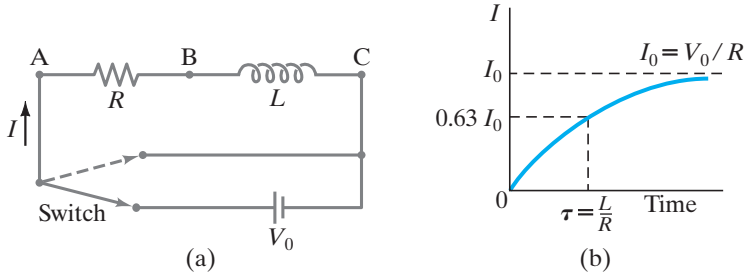


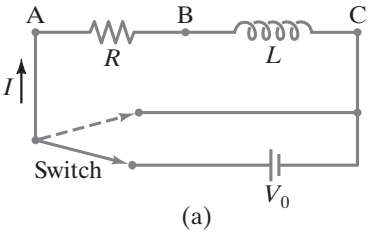
FIGURE 30–6 (a) *LR* circuit; (b) growth of current when connected to battery.

We can show this analytically by applying Kirchhoff's loop rule to the circuit of Fig. 30–6a with the switch closed. The emfs in the circuit are the battery voltage V_0 and the emf $\mathcal{E} = -L(dI/dt)$ in the inductor (Eq. 30–5) opposing the increasing current. Hence the sum of the potential changes around the loop is

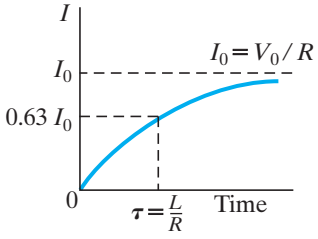
$$V_0 - IR - L\frac{dI}{dt} = 0,$$

where I is the current in the circuit at any instant. We rearrange this to obtain

$$L\frac{dI}{dt} + RI = V_0. \qquad \textbf{(30–8)}$$



(a)



(b)

FIGURE 30–6 (Repeated.) (a) *LR* circuit; (b) growth of current when connected to battery.

This Eq. 30-8 we just obtained

$$L\frac{dI}{dt} + RI = V_0 \qquad \textbf{(30–8)}$$

is a linear differential equation and can be integrated in the same way we did in Section 26–5 for an *RC* circuit. We rewrite Eq. 30–8 and then integrate from $t = 0$, when $I = 0$, to a later time t when the current is I:

$$\int_{I=0}^{I} \frac{dI}{V_0 - IR} = \int_0^t \frac{dt}{L}.$$

Then

$$-\frac{1}{R}\ln\left(\frac{V_0 - IR}{V_0}\right) = \frac{t}{L} \qquad \text{and} \qquad \frac{V_0 - IR}{V_0} = e^{-\frac{Rt}{L}}$$

so

$$I = \frac{V_0}{R}\left(1 - e^{-t/\tau}\right) \qquad \textbf{(30–9)}$$

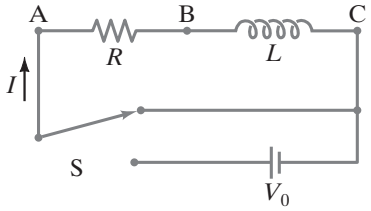
where

$$\tau = \frac{L}{R} \qquad \textbf{(30–10)}$$

is the **time constant** of the *LR* circuit. The symbol τ represents the time required for the current I to reach $(1 - 1/e) = 0.63$ or 63% of its maximum value (V_0/R). Equation 30–9 is plotted in Fig. 30–6b. (Compare to the *RC* circuit, Section 26–5.)

EXERCISE C Show that L/R does have dimensions of time. (See Section 1–7.)

FIGURE 30–7 The switch is flipped quickly so the battery is removed but we still have a circuit. The current at this moment (call it $t = 0$) is I_0.



Now let us flip the switch in Fig. 30–6a so that the battery is taken out of the circuit, and points A and C are connected together as shown in Fig. 30–7 at the moment when the switching occurs (call it $t = 0$) and the current is I_0. Then the differential equation (Eq. 30–8) becomes (since $V_0 = 0$):

$$L\frac{dI}{dt} + RI = 0.$$

We rearrange this equation and integrate:

$$\int_{I_0}^{I} \frac{dI}{I} = -\int_0^t \frac{R}{L}dt$$

where the current is I_0 at $t = 0$, and is I at time t.

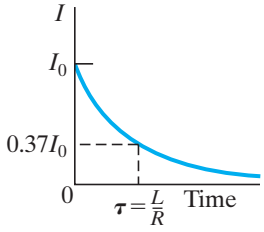
We integrate this last equation to obtain

$$\ln\frac{I}{I_0} = -\frac{R}{L}t$$

or

$$I = I_0 e^{-t/\tau}. \qquad \textbf{(30–11)}$$

FIGURE 30–8 Decay of the current in Fig. 30–7 in time after the battery is switched out of the circuit.



The current thus decays exponentially to zero as shown in Fig. 30–8. Again the time constant is $\tau = L/R$, and is the time for the current to decrease to $0.37\,I_0$.

This analysis shows that there is always some "lag time" or "reaction time" when an electromagnet, for example, is turned on or off. We also see that an *LR* circuit has properties similar to an *RC* circuit (Section 26–5). Unlike the capacitor case, however, the time constant here is *inversely* proportional to *R*.

EXAMPLE 30–6 **An *LR* circuit.** At $t = 0$, a 12.0-V battery is connected in series with a 220-mH inductor and a total resistance of 30 Ω, as shown in Fig. 30–9. (*a*) What is the current at $t = 0$? (*b*) What is the time constant? (*c*) What is the maximum current? (*d*) How long will it take the current to reach half its maximum possible value? (*e*) At this instant, at what rate is energy being delivered by the battery, and (*f*) at what rate is energy being stored in the inductor's magnetic field?

APPROACH We have the situation shown in Figs. 30–6a and b, and we can apply the equations we just developed.

SOLUTION (*a*) The current cannot instantaneously jump from zero to some other value when the switch is closed because the inductor opposes the change $(\mathcal{E}_L = -L(dI/dt))$. Hence just after the switch is closed, I is still zero at $t = 0$ and then begins to increase.

(*b*) The time constant is, from Eq. 30–10, $\tau = L/R = (0.22\,\text{H})/(30\,\Omega) = 7.3\,\text{ms}$.

(*c*) The current reaches its maximum steady value I_0 after a long time (see Fig. 30–6), when $dI/dt = 0$ so $I_0 = V_0/R = 12.0\,\text{V}/30\,\Omega = 0.40\,\text{A}$.

(*d*) We set $I = \frac{1}{2}I_0 = V_0/2R$ in Eq. 30–9, which gives us

$$\tfrac{1}{2} = 1 - e^{-t/\tau}$$

or

$$e^{-t/\tau} = \tfrac{1}{2}.$$

We solve for t:

$$t = \tau \ln 2 = (7.3 \times 10^{-3}\,\text{s})(0.693) = 5.1\,\text{ms}.$$

(*e*) At this instant, $I = \frac{1}{2}I_0 = 200\,\text{mA}$, so the power being delivered by the battery is

$$P = IV = (0.20\,\text{A})(12\,\text{V}) = 2.4\,\text{W}.$$

(*f*) From Eq. 30–6, the energy stored in an inductor L at any instant is

$$U = \tfrac{1}{2}LI^2$$

where I is the current in the inductor $(= 0.20\,\text{A})$ at that instant. The *rate* at which the energy changes is

$$\frac{dU}{dt} = LI\frac{dI}{dt}.$$

We can differentiate Eq. 30–9 to obtain dI/dt, or use the differential equation, Eq. 30–8, directly:

$$\frac{dU}{dt} = I\left(L\frac{dI}{dt}\right) = I(V_0 - RI)$$
$$= (0.20\,\text{A})[12\,\text{V} - (30\,\Omega)(0.20\,\text{A})] = 1.2\,\text{W}.$$

Since only part of the battery's power is feeding the inductor at this instant, where is the rest going?

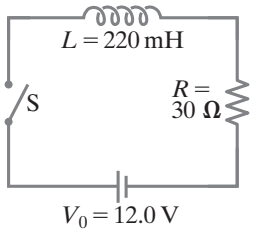


FIGURE 30–9 Example 30–6.

EXERCISE D A resistor in series with an inductor has a time constant of 10 ms. When the same resistor is placed in series with a 5-μF capacitor, the time constant is 5×10^{-6} s. What is the value of the inductor? (*a*) 5 μH; (*b*) 10 μH; (*c*) 5 mH; (*d*) 10 mH; (*e*) not enough information to determine it.

PHYSICS APPLIED
Surge protection

An inductor can act as a **surge protector** for sensitive electronic equipment that can be damaged by high currents. If equipment is plugged into a standard wall plug, a sudden "surge," or increase, in voltage will normally cause a corresponding large change in current and damage the electronics. However, if there is an inductor in series with the voltage to the device, the sudden change in current produces an opposing emf preventing the current from reaching dangerous levels.

30–5 *LC* Circuits and Electromagnetic Oscillations

Any electric circuit can contain the three basic components: resistance, capacitance, and inductance, in addition to a source of emf. (There can also be more complex components, such as diodes or transistors.) We have previously discussed both *RC* and *LR* circuits. Now we look at an *LC* circuit, one that contains only a capacitance C and an inductance, L, Fig. 30–10. This is an idealized circuit in which we assume there is no resistance; in the next Section we will include resistance.

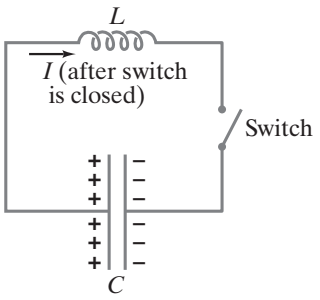


FIGURE 30–10 An *LC* circuit.

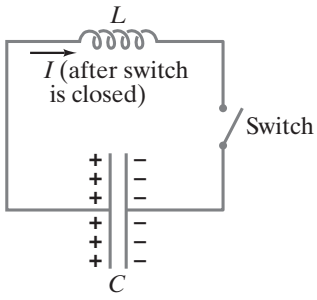


FIGURE 30–10 (Repeated.) An LC circuit.

Let us suppose the capacitor in Fig. 30–10 is initially charged so that one plate has charge Q_0 and the other plate has charge $-Q_0$, and the potential difference across it is $V = Q/C$ (Eq. 24–1). Suppose that at $t = 0$, the switch is closed. The capacitor immediately begins to discharge. As it does so, the current I through the inductor increases. We now apply Kirchhoff's loop rule (sum of potential changes around a loop is zero):

$$-L\frac{dI}{dt} + \frac{Q}{C} = 0.$$

Because charge leaves the positive plate on the capacitor to produce the current I as shown in Fig. 30–10, the charge Q on the (positive) plate of the capacitor is decreasing, so $I = -dQ/dt$. We can then rewrite the above equation as

$$\frac{d^2Q}{dt^2} + \frac{Q}{LC} = 0. \qquad \textbf{(30–12)}$$

This is a familiar differential equation, which has the same form as the equation for simple harmonic motion (Chapter 14, Eq. 14–3). Its solution is *sinusoidal* which we can write like Eq. 14–4 as

$$Q = Q_0 \cos(\omega t + \phi) \qquad \textbf{(30–13)}$$

where Q_0 and ϕ are constants that depend on the initial conditions and $(\omega t + \phi)$ is in radians. [Note: the amplitude Q_0 represents Q at $t = 0$ *only* if the phase angle ϕ is 0.] We insert Eq. 30–13 into Eq. 30–12, noting that $d^2Q/dt^2 = -\omega^2 Q_0 \cos(\omega t + \phi)$; thus

$$-\omega^2 Q_0 \cos(\omega t + \phi) + \frac{1}{LC} Q_0 \cos(\omega t + \phi) = 0$$

or

$$\left(-\omega^2 + \frac{1}{LC}\right)\cos(\omega t + \phi) = 0.$$

This relation can be true for all times t only if $(-\omega^2 + 1/LC) = 0$, which tells us that

$$\omega = 2\pi f = \sqrt{\frac{1}{LC}}. \qquad \textbf{(30–14)}$$

Equation 30–13 shows that the charge on the capacitor in an LC circuit oscillates sinusoidally. The current in the inductor is

$$I = -\frac{dQ}{dt} = \omega Q_0 \sin(\omega t + \phi)$$
$$= I_0 \sin(\omega t + \phi); \qquad \textbf{(30–15)}$$

so the current too is sinusoidal. The maximum value of I is $I_0 = \omega Q_0 = Q_0/\sqrt{LC}$. Equations 30–13 and 30–15 for Q and I when $\phi = 0$ are plotted in Fig. 30–11.

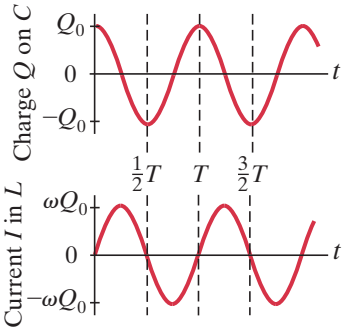


FIGURE 30–11 Charge Q and current I in an LC circuit. The period $T = \frac{1}{f} = \frac{2\pi}{\omega} = 2\pi\sqrt{LC}$.

Now let us look at LC oscillations from the point of view of energy. The energy stored in the electric field of the capacitor at any time t is (see Eq. 24–5):

$$U_E = \frac{1}{2}\frac{Q^2}{C} = \frac{Q_0^2}{2C}\cos^2(\omega t + \phi).$$

The energy stored in the magnetic field of the inductor at the same instant is (Eq. 30–6)

$$U_B = \frac{1}{2}LI^2 = \frac{L\omega^2 Q_0^2}{2}\sin^2(\omega t + \phi) = \frac{Q_0^2}{2C}\sin^2(\omega t + \phi)$$

where we used Eq. 30–14. If we let $\phi = 0$, then at times $t = 0$, $t = \frac{1}{2}T$, $t = T$, and so on (where T is the period $= 1/f = 2\pi/\omega$), we have $U_E = Q_0^2/2C$ and $U_B = 0$. That is, all the energy is stored in the electric field of the capacitor. But at $t = \frac{1}{4}T$, $\frac{3}{4}T$, and so on, $U_E = 0$ and $U_B = Q_0^2/2C$, and so all the energy is stored in the magnetic field of the inductor. At any time t, the total energy is

$$U = U_E + U_B = \frac{1}{2}\frac{Q^2}{C} + \frac{1}{2}LI^2$$
$$= \frac{Q_0^2}{2C}\left[\cos^2(\omega t + \phi) + \sin^2(\omega t + \phi)\right] = \frac{Q_0^2}{2C}. \quad \textbf{(30–16)}$$

Hence the total energy is constant, and energy is conserved.

This LC circuit is an ***LC* oscillator** or **electromagnetic oscillation**. The charge Q oscillates back and forth, from one plate of the capacitor to the other, and repeats this continuously. The current also oscillates back and forth. There are also energy oscillations: when Q is a maximum, the energy is all stored in the electric field of the capacitor; when Q reaches zero, the current I is a maximum and all the energy is stored in the magnetic field of the inductor. Thus the energy in an LC circuit oscillates between the capacitor and the inductor, but the total energy is constant. See Fig. 30–12.

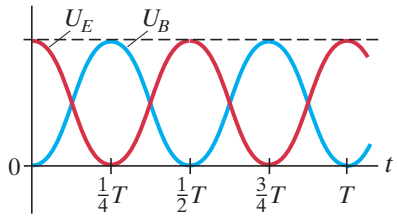


FIGURE 30–12 Energy U_E (red line) and U_B (blue line) stored in the capacitor and the inductor as a function of time. Note how the energy oscillates between electric and magnetic. The dashed line at the top is the (constant) total energy $U = U_E + U_B$.

EXERCISE E Return to the Chapter-Opening Question, page 869, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 30–7 ***LC* circuit.** A 1200-pF capacitor is fully charged by a 500-V dc power supply. It is disconnected from the power supply and is connected, at $t = 0$, to a 75-mH inductor. Determine: (*a*) the initial charge on the capacitor; (*b*) the maximum current; (*c*) the frequency f and period T of oscillation; and (*d*) the total energy oscillating in the system.

APPROACH We use the analysis above, and the definition of capacitance $Q = CV$ (Chapter 24).

SOLUTION (*a*) The 500-V power supply, before being disconnected, charged the capacitor to a charge of

$$Q_0 = CV = (1.2 \times 10^{-9}\,\text{F})(500\,\text{V}) = 6.0 \times 10^{-7}\,\text{C}.$$

(*b*) The maximum current, I_0, is (see Eqs. 30–14 and 30–15)

$$I_0 = \omega Q_0 = \frac{Q_0}{\sqrt{LC}} = \frac{(6.0 \times 10^{-7}\,\text{C})}{\sqrt{(0.075\,\text{H})(1.2 \times 10^{-9}\,\text{F})}} = 63\,\text{mA}.$$

(*c*) Equation 30–14 gives us the frequency:

$$f = \frac{\omega}{2\pi} = \frac{1}{(2\pi\sqrt{LC})} = 17\,\text{kHz},$$

and the period T is

$$T = \frac{1}{f} = 6.0 \times 10^{-5}\,\text{s}.$$

(*d*) Finally the total energy (Eq. 30–16) is

$$U = \frac{Q_0^2}{2C} = \frac{(6.0 \times 10^{-7}\,\text{C})^2}{2(1.2 \times 10^{-9}\,\text{F})} = 1.5 \times 10^{-4}\,\text{J}.$$

30–6 *LC* Oscillations with Resistance (*LRC* Circuit)

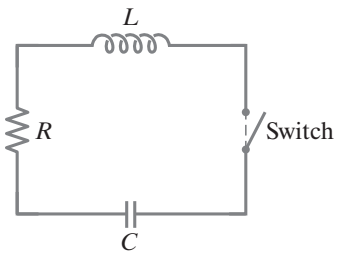


FIGURE 30–13 An *LRC* circuit.

The *LC* circuit discussed in the previous Section is an idealization. There is always some resistance *R* in any circuit, and so we now discuss such a simple *LRC* circuit, Fig. 30–13.

Suppose again that the capacitor is initially given a charge Q_0 and the battery or other source is then removed from the circuit. The switch is closed at $t = 0$. Since there is now a resistance in the circuit, we expect some of the energy to be converted to thermal energy, and so we don't expect undamped oscillations as in a pure *LC* circuit. Indeed, if we use Kirchhoff's loop rule around this circuit, we obtain

$$-L\frac{dI}{dt} - IR + \frac{Q}{C} = 0,$$

which is the same equation we had in Section 30–5 with the addition of the voltage drop *IR* across the resistor. Since $I = -dQ/dt$, as we saw in Section 30–5, this equation becomes

$$L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q = 0. \qquad \textbf{(30–17)}$$

This second-order differential equation in the variable *Q* has precisely the same form as that for a damped harmonic oscillator, Eq. 14–15:

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = 0.$$

Hence we can analyze our *LRC* circuit in the same way as for damped harmonic motion, Section 14–7. Our system may undergo damped oscillations, curve A in Fig. 30–14 (underdamped system), or it may be critically damped (curve B), or overdamped (curve C), depending on the relative values of *R*, *L*, and *C*. Using the results of Section 14–7, with *m* replaced by *L*, *b* by *R*, and *k* by C^{-1}, we find that the system will be underdamped when $b^2 < 4mk$ (page 415) or

$$R^2 < \frac{4L}{C}, \qquad \text{[underdamped, oscillating]}$$

and overdamped for $R^2 > 4L/C$. Critical damping (curve B in Fig. 30–14) occurs when $R^2 = 4L/C$. If *R* is smaller than $\sqrt{4L/C}$, the angular frequency, ω', will be

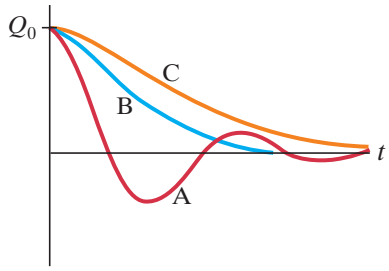
$$\omega' = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \qquad \textbf{(30–18)}$$

(compare to Eq. 14–18). And the charge *Q* as a function of time will be

$$Q = Q_0 e^{-\frac{R}{2L}t}\cos(\omega' t + \phi) \qquad \textbf{(30–19)}$$

(compare to Eq. 14–19) where ϕ is a phase constant which depends on the initial conditions, and is zero if $Q = Q_0$ at $t = 0$.

FIGURE 30–14 Charge *Q* on the capacitor in an *LRC* circuit as a function of time: curve A is for underdamped oscillation $(R^2 < 4L/C)$, curve B is for critically damped $(R^2 = 4L/C)$, and curve C is for overdamped $(R^2 > 4L/C)$.



Oscillators are an important element in many electronic devices: radios and television sets use them for tuning stations, and so do cell phones, plus many other electronic devices use oscillators for synchronizing timing. Because some resistance is always present, electrical oscillators generally need a periodic input of power to compensate for the energy converted to thermal energy in the resistance.

EXAMPLE 30–8 **Damped oscillations.** At $t = 0$, a 40-mH inductor is placed in series with a resistance $R = 3.0\,\Omega$ and a charged capacitor $C = 4.8\,\mu\text{F}$. (*a*) Show that this circuit will oscillate. (*b*) Determine the frequency. (*c*) What is the time required for the charge amplitude to drop to half its starting value? (*d*) What value of R will make the circuit nonoscillating?

APPROACH We first check R^2 vs. $4L/C$; then use Eqs. 30–18 and 30–19.

SOLUTION (*a*) In order to oscillate, the circuit must be underdamped, so we must have $R^2 < 4L/C$. Since $R^2 = 9.0\,\Omega^2$ and $4L/C = 4(0.040\,\text{H})/(4.8 \times 10^{-6}\,\text{F}) = 3.3 \times 10^4\,\Omega^2$, so R^2 is a lot less than $4L/C$, and the circuit will oscillate with very small damping.

(*b*) We use Eq. 30–18:

$$f' = \frac{\omega'}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} = 3.6 \times 10^2\,\text{Hz}.$$

(*c*) From Eq. 30–19, the amplitude will be half when

$$e^{-\frac{R}{2L}t} = \tfrac{1}{2}$$

or

$$t = \frac{2L}{R}\ln 2 = 18\,\text{ms}.$$

(*d*) To make the circuit critically damped or overdamped, we must use the criterion $R^2 \geq 4L/C = 3.3 \times 10^4\,\Omega^2$. Hence we must have $R \geq 180\,\Omega$.

30–7 AC Circuits and Reactance

We have previously discussed circuits that contain combinations of resistor, capacitor, and inductor, but only when they are connected to a dc source of emf or to zero voltage. Now we discuss these circuit elements when they are connected to a source of alternating voltage that produces an alternating current (ac).

First we examine, one at a time, how a resistor, a capacitor, and an inductor behave when connected to a source of alternating voltage, represented by the symbol



[alternating voltage]

which produces a sinusoidal voltage of frequency f. We assume (arbitrarily, for clarity) in each case that the emf gives rise to a current

$$I = I_0 \cos 2\pi f t = I_0 \cos \omega t \qquad \textbf{(30–20)}$$

where t is time and I_0 is the peak current. Remember (Section 25–7, Eqs. 25–9) that

$$V_{\text{rms}} = V_0/\sqrt{2} \quad \text{and} \quad I_{\text{rms}} = I_0/\sqrt{2}.$$

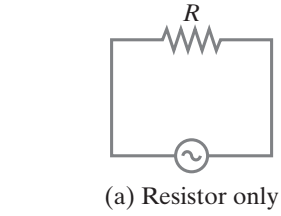
Resistor

When an ac source is connected to a resistor as in Fig. 30–15a, the current increases and decreases with the alternating voltage according to Ohm's law

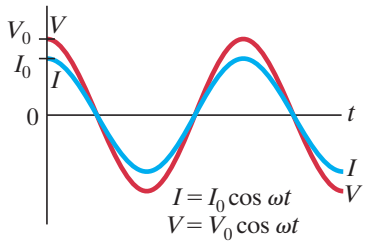
$$V = IR = I_0 R \cos \omega t = V_0 \cos \omega t$$

where $V_0 = I_0 R$ is the peak voltage. Figure 30–15b shows the voltage (red curve) and the current (blue curve) as a function of time. Because the current is zero when the voltage is zero and the current reaches a peak when the voltage does, we say that the current and voltage are **in phase**. Energy is transformed into heat (Section 25–7), at an average rate

$$\overline{P} = \overline{IV} = I_{\text{rms}}^2 R = V_{\text{rms}}^2/R.$$



(a) Resistor only



(b)

FIGURE 30–15 (a) Resistor connected to an ac source. (b) Current (blue curve) is in phase with the voltage (red) across a resistor.

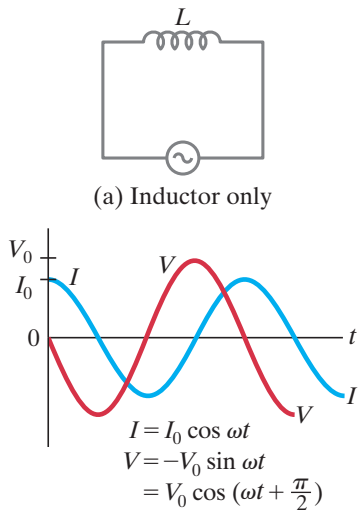


FIGURE 30–16 (a) Inductor connected to an ac source. (b) Current (blue curve) lags voltage (red curve) by a quarter cycle or $90° = \pi/2$ radians.

Inductor

In Fig. 30–16a an inductor of inductance L (symbol —ᔓᔓᔓ—) is connected to the ac source. We ignore any resistance it might have (it is usually small). The voltage applied to the inductor will be equal to the "back" emf generated in the inductor by the changing current as given by Eq. 30–5. This is because the sum of the electric potential changes around any closed circuit must add up to zero, according to Kirchhoff's rule. Thus (Eq. 30–5)

$$V - L\frac{dI}{dt} = 0$$

and inserting Eq. 30–20, $I = I_0 \cos \omega t$, we have (recall from Chapter 10 that 360° corresponds to 2π radians),

$$V = L\frac{dI}{dt} = -\omega L I_0 \sin \omega t. \qquad \textbf{(30–21)}$$

Using the identity $\sin\theta = -\cos(\theta + 90°)$, or in radians, $\sin\theta = -\cos\left(\theta + \frac{\pi}{2}\right)$, we can write

$$V = \omega L I_0 \cos\left(\omega t + \frac{\pi}{2}\right) = V_0 \cos\left(\omega t + \frac{\pi}{2}\right) \qquad \textbf{(30–22a)}$$

where we *must use radians* because ω has units of rad/s, and

$$V_0 = I_0 \omega L \qquad \textbf{(30–22b)}$$

is the peak voltage. The current I and voltage V as a function of time are graphed in Fig. 30–16b for a "pure" inductor. It is clear from this graph, as well as from Eqs. 30–22, that the current and voltage are out of phase by a quarter cycle, which is equivalent to $\frac{\pi}{2}$ radians or 90°. We see from the graph that

current lags the voltage by 90° $\left(\frac{\pi}{2}\right.$ radians$\left.\right)$ in an inductor.

That is, the current in an inductor reaches its peaks a quarter cycle later than the voltage does. Alternatively, we can say that the voltage leads the current by 90°.

Because the current and voltage in an inductor are out of phase by 90° $(\pi/2$ radians$)$, the product IV (= power) is as often positive as it is negative (Fig. 30–16b). So no energy is transformed in an inductor on the average; and no energy is dissipated as thermal energy.

Just as a resistor impedes the flow of charge, so too an inductor impedes the flow of charge in an alternating current due to the back emf produced. For a resistor R, the peak current and peak voltage are related by $V_0 = I_0 R$. We can write a similar relation for an inductor:

$$V_0 = I_0 X_L \qquad \left[\begin{array}{c}\text{maximum or rms values,}\\ \text{not at any instant}\end{array}\right] \qquad \textbf{(30–23a)}$$

where, from Eq. 30–22b (and using $\omega = 2\pi f$ where f is the frequency of the ac),

$$X_L = \omega L = 2\pi f L. \qquad \textbf{(30–23b)}$$

The term X_L is called the **inductive reactance** of the inductor, and has units of ohms. X_L is larger for higher frequencies f and larger inductance L. The greater X_L is, the more it impedes the flow of charge and the smaller the current. (Reactance and resistance are referred to collectively as *impedance*, as we discuss shortly.)

Equation 30–23a is valid for peak values I_0 and V_0; it is also valid for rms values, $V_{\text{rms}} = I_{\text{rms}} X_L$. Because the peak values of current and voltage are not reached at the same time, Eq. 30–23a is *not valid at a particular instant*, as is the case for a resistor $(V = IR)$. It simply is an equation that relates the maximum (or rms) values of V and I.

⚠ CAUTION

$V_0 = I_0 X_L$ is not valid at any moment in time

EXAMPLE 30–9 **Reactance of a coil.** A coil has a resistance $R = 1.00\ \Omega$ and an inductance of 0.300 H. Determine the current in the coil if (*a*) 120-V dc is applied to it, (*b*) 120-V ac (rms) at 60.0 Hz is applied.

APPROACH When the voltage is dc, there is no inductive reactance $(X_L = 2\pi f L = 0$ since $f = 0)$, so we apply Ohm's law for the resistance. When the voltage is ac, we calculate the reactance X_L and then use Eq. 30–23a.

SOLUTION (*a*) With dc, we have no X_L so we simply apply Ohm's law:

$$I = \frac{V}{R} = \frac{120\text{ V}}{1.00\ \Omega} = 120\text{ A}.$$

(*b*) The inductive reactance is

$$X_L = 2\pi f L = (6.283)\left(60.0\text{ s}^{-1}\right)(0.300\text{ H}) = 113\ \Omega.$$

A useful result here is that for normal U.S. 60-Hz electricity, $\omega = 2\pi f = 377$ rad/s. In comparison to this, the 1.0-Ω resistance can be ignored. Thus,

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{X_L} = \frac{120\text{ V}}{113\ \Omega} = 1.06\text{ A}.$$

NOTE It might be tempting to say that the total impedance is 113 Ω + 1 Ω = 114 Ω. This might imply that about 1% of the voltage drop is across the resistor, or about 1 V; and that across the inductance is 119 V. Although the 1 V across the resistor is correct, the other statements are not true because of the phase difference in an inductor. This will be discussed in the next Section.

Capacitor

When a capacitor is connected to a battery, the capacitor plates quickly acquire equal and opposite charges; but no steady current flows in the circuit. A capacitor prevents the flow of a dc current. But if a capacitor is connected to an alternating source of voltage, as in Fig. 30–17a, an alternating current will flow continuously. This can happen because when the ac voltage is first turned on, charge begins to flow and one plate acquires a negative charge and the other a positive charge. But when the voltage reverses itself, the charges flow in the opposite direction. Thus, for an alternating applied voltage, an ac current is present in the circuit continuously.

Let us look at this in more detail. By Kirchhoff's loop rule, the applied source voltage must equal the voltage V across the capacitor (Eq. 24–1, $Q = CV$) at any moment:

$$V = \frac{Q}{C}$$

where C is the capacitance and Q is the charge on the capacitor plates. The current I at any instant (given as $I = I_0 \cos\omega t$, Eq. 30–20) is

$$I = \frac{dQ}{dt} = I_0 \cos\omega t.$$

Hence the charge Q on the plates at any instant t is given by

$$Q = \int_0^t dQ = \int_0^t I_0 \cos\omega t\, dt = \frac{I_0}{\omega}\sin\omega t.$$

Then the voltage across the capacitor is

$$V = \frac{Q}{C} = I_0\left(\frac{1}{\omega C}\right)\sin\omega t.$$

Using the trigonometric identity $\sin\theta = \cos(90° - \theta) = \cos(\theta - 90°)$, we can rewrite this as (changing 90° to $\frac{\pi}{2}$ radians which must be used in calculations)

$$V = I_0\left(\frac{1}{\omega C}\right)\cos\left(\omega t - \frac{\pi}{2}\right) = V_0 \cos\left(\omega t - \frac{\pi}{2}\right) \qquad \textbf{(30–24a)}$$

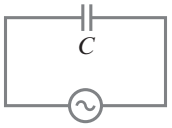
where

$$V_0 = I_0\left(\frac{1}{\omega C}\right) \qquad \textbf{(30–24b)}$$

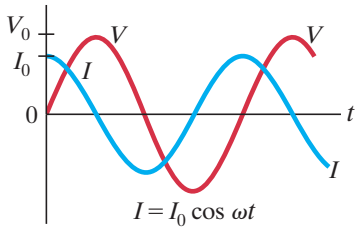
is the peak voltage. The current $I\ (= I_0 \cos\omega t)$ and voltage V (Eq. 30–24a) across the capacitor are graphed in Fig. 30–17b. It is clear from this graph, as well as a comparison of Eq. 30–24a with Eq. 30–20, that the current and voltage are out of phase by a quarter cycle or 90°, or $\frac{\pi}{2}$ radians:

current leads the voltage by 90° $\left(\frac{\pi}{2}\text{ rad}\right)$ for a capacitor.

Alternatively we can say that the voltage lags the current by 90° or $\pi/2$ radians. This is the opposite of what happens for an inductor.



(a) Capacitor only



$I = I_0 \cos\omega t$
$V = V_0 \sin\omega t$
$= V_0 \cos(\omega t - \frac{\pi}{2})$

(b)

FIGURE 30–17 (a) Capacitor connected to an ac source. (b) Current leads voltage by a quarter cycle, or 90° = $\pi/2$ radians.

CAUTION

Capacitors and inductors do not dissipate energy, only resistance does

Because the current and voltage are out of phase by 90° in a capacitor ($\pi/2$ radians), the average power dissipated is zero, just as for an inductor. Energy from the source is fed to the capacitor, where it is stored in the electric field between its plates. As the field decreases, the energy returns to the source. Thus *only a resistance will dissipate energy* as thermal energy in an ac circuit.

A relationship between the applied voltage and the current in a capacitor can be written just as for an inductance:

$$V_0 = I_0 X_C \qquad \left[\begin{array}{c}\text{maximum or rms values,}\\ \text{not at any instant}\end{array}\right] \qquad \textbf{(30–25a)}$$

where X_C is the **capacitive reactance** of the capacitor, and has units of ohms; X_C is given by (see Eq. 30–24b):

$$X_C = \frac{1}{\omega C} = \frac{1}{2\pi f C}. \qquad \textbf{(30–25b)}$$

When frequency f and/or capacitance C are smaller, X_C is larger and thus impedes the flow of charge more. That is, when X_C is larger, the current is smaller (Eq. 30–25a). In the next Section we use the term **impedance** to represent reactances and resistance.

Equation 30–25a relates the *peak* values of V and I, or the *rms* values ($V_{\text{rms}} = I_{\text{rms}} X_C$). But it is *not* valid at a particular instant because I and V are not in phase.

Note from Eq. 30–25b that for dc conditions, $\omega = 2\pi f = 0$ and X_C becomes infinite. This is as it should be, since a capacitor does not pass dc current. Also, note that the reactance of an inductor increases with frequency, but the reactance of a capacitor decreases with frequency.

EXAMPLE 30–10 **Capacitor reactance.** What is the rms current in the simple capacitor circuit of Fig. 30–17a (previous page) if $C = 1.00\,\mu\text{F}$ and $V_{\text{rms}} = 120\,\text{V}$? Calculate (*a*) for $f = 60.0\,\text{Hz}$, and then (*b*) for $f = 6.00 \times 10^5\,\text{Hz}$.

APPROACH We find the reactance using Eq. 30–25b, and solve for current in the equivalent form of Ohm's law, Eq. 30–25a.

SOLUTION (*a*) We have

$$X_C = \frac{1}{2\pi f C} = \frac{1}{(6.28)\left(60.0\,\text{s}^{-1}\right)\left(1.00 \times 10^{-6}\,\text{F}\right)} = 2.65\,k\Omega.$$

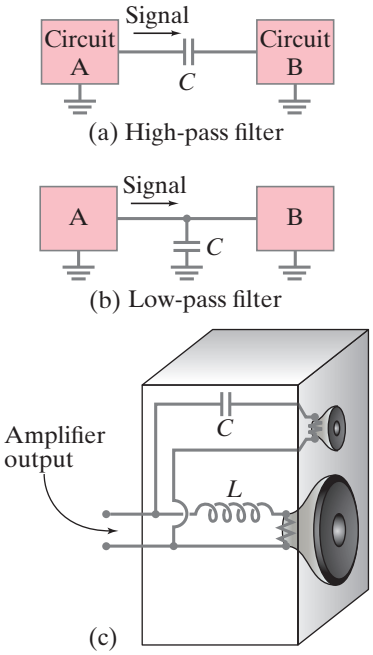
The rms current is (Eq. 30–25a):

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{X_C} = \frac{120\,\text{V}}{2.65 \times 10^3\,\Omega} = 45.2\,\text{mA}.$$

(*b*) For $f = 6.00 \times 10^5\,\text{Hz}$, X_C will be $0.265\,\Omega$ and $I_{\text{rms}} = 452\,\text{A}$, vastly larger!

NOTE The dependence of X_L on f is dramatic. For high frequencies, the capacitive reactance is very small.

FIGURE 30–18 (a) and (b) Two common uses for a capacitor as a filter. (c) Simple loudspeaker cross-over.



PHYSICS APPLIED

Capacitors as filters

Two common applications of capacitors are illustrated in Fig. 30–18a and b. In Fig. 30–18a, circuit A is said to be capacitively coupled to circuit B. The purpose of the capacitor is to prevent a dc voltage from passing from A to B but allowing an ac signal to pass relatively unimpeded (if C is sufficiently large, Eq. 30–25b). The capacitor in Fig. 30–18a is called a **high-pass filter** because it allows high-frequency ac to pass easily, but not dc and very little of low-frequency ac.

In Fig. 30–18b, the capacitor passes ac to ground, but not dc. In this case, a dc voltage can be maintained between circuits A and B, but an ac signal leaving A passes to ground instead of into B. Thus the capacitor in Fig. 30–18b acts as a **low-pass filter** when a constant dc voltage is required. Any high-frequency variation in voltage will pass to ground instead of into circuit B. (Very low-frequency ac will also be able to reach circuit B, at least in part.)

PHYSICS APPLIED

Loudspeaker cross-over

Loudspeakers having separate "woofer" (low-frequency speaker) and "tweeter" (high-frequency speaker) may use a simple "cross-over" that consists of a capacitor in the tweeter circuit to impede low-frequency signals, and an inductor in the woofer circuit to impede high-frequency signals $(X_L = 2\pi f L)$. Hence mainly low-frequency sounds reach and are emitted by the woofer. See Fig. 30–18c.

EXERCISE F At what frequency is the reactance of a 1.0-μF capacitor equal to $500\,\Omega$? (*a*) 320 Hz, (*b*) 500 Hz, (*c*) 640 Hz, (*d*) 2000 Hz, (*e*) 4000 Hz.

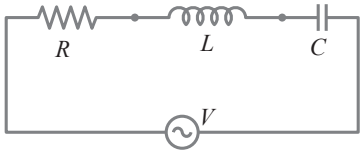


FIGURE 30–19 An *LRC* circuit.

30–8 *LRC* Series AC Circuit; Phasor Diagrams

Let us examine a circuit containing all three elements in series: a resistor R, an inductor L, and a capacitor C, Fig. 30–19. If a given circuit contains only two of these elements, we can still use the results of this Section by setting $R = 0$, $X_L = 0$, or $X_C = 0$, as needed. We let V_R, V_L, and V_C represent the voltage across each element at a *given instant* in time; and V_{R0}, V_{L0}, and V_{C0} represent the *maximum* (peak) values of these voltages. The voltage across each of the elements will follow the phase relations we discussed in the previous Section. At any instant the voltage V supplied by the source will be, by Kirchhoff's loop rule,

$$V = V_R + V_L + V_C. \qquad \textbf{(30–26a)}$$

Because the various voltages are not in phase, they do not reach their peak values at the same time, so the peak voltage of the source V_0 will *not* equal $V_{R0} + V_{L0} + V_{C0}$. The current in an *LRC* circuit at any instant is the same at all points in the circuit (charge does not pile up in the wires). Thus the currents in each element are in phase with each other, even though the voltages are not. We choose our origin in time $(t = 0)$ so that the current I at any time t is (as in Eq. 30–20)

$$I = I_0 \cos \omega t.$$

AC circuits are complicated to analyze. One approach† is to use a sort of vector device known as a **phasor diagram**. Arrows (treated like vectors) are drawn in an xy coordinate system to represent each voltage. The length of each arrow represents the magnitude of the peak voltage across each element:

$$V_{R0} = I_0 R, \qquad V_{L0} = I_0 X_L, \qquad \text{and} \qquad V_{C0} = I_0 X_C.$$

V_{R0} is in phase with the current and is initially $(t = 0)$ drawn along the positive x axis, as is the current (Fig. 30–20a). V_{L0} leads the current by 90°, so it leads V_{R0} by 90° and is initially drawn along the positive y axis. V_{C0} lags the current by 90°, so V_{C0} is drawn initially along the negative y axis, Fig. 30–20a.

If we let the vector diagram rotate counterclockwise at frequency f, we get the diagram shown in Fig. 30–20b; after a time, t, each arrow has rotated through an angle ωt. Then the *projections of each arrow onto the x axis represent the voltages across each element at the instant t*, as can be seen in Fig. 30–20c. For example $I = I_0 \cos \omega t$. [Compare Eqs. 30–22a and 30–24a with Fig. 30–20c to confirm the validity of the phasor diagram.]

The sum of the projections of the three voltage vectors represents the *instantaneous* voltage across the whole circuit, V, at a given instant. Therefore, the vector sum of these vectors will be the vector that represents the peak source voltage, V_0, as shown in Fig. 30–21 where it is seen that V_0 makes an angle ϕ with I_0 and V_{R0}. As time passes, V_0 rotates with the other vectors, so the instantaneous voltage V (projection of V_0 on the x axis) is (see Fig. 30–21)

$$V = V_0 \cos(\omega t + \phi). \qquad \textbf{(30–26b)}$$

The voltage V across the whole circuit must equal the source voltage (Fig. 30–19). Thus the voltage from the source is out of phase with the current by an angle ϕ. As a check, note in Eq. 30–26b and Fig. 30–21 that if $R = X_C = 0$, then $\phi = \pi/2$, and V_0 would lead the current by 90°, as it must for an inductor alone. Similarly, if $R = L = 0$, $\phi = -\pi/2$ and V_0 would lag the current by 90°, as it must for a capacitor alone.

†We could instead do our analysis by rewriting Eq. 30–26a as a differential equation (setting $V_C = Q/C$, $V_R = IR = (dQ/dt)R$, and $V_L = L\, dI/dt = Ld^2Q/dt^2$ and trying to solve the differential equation. The differential equation would look like Eq. 14–21 in Section 14–8 (on forced vibrations), and would be solved in the same way. Phasor diagrams are mathematically simpler, and also give us some physical insight.

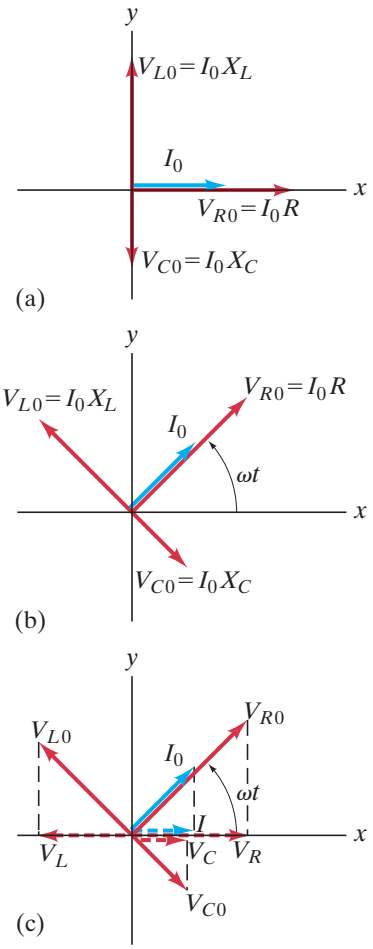
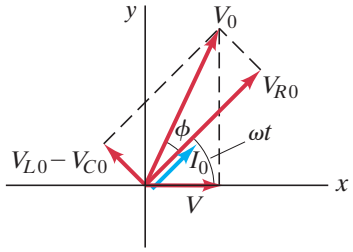


FIGURE 30–20 Phasor diagram for a series *LRC* circuit (voltages red, current blue). (a) At $t = 0$, (b) a time t later. (c) Projections (dashed) on the x axis which give I, V_R, V_C, and V_L at time t.

FIGURE 30–21 Phasor diagram for a series *LRC* circuit showing the sum vector, V_0.



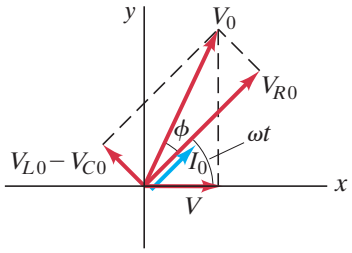


FIGURE 30–21 (Repeated.) Phasor diagram for a series *LRC* circuit showing the sum vector, V_0.

From this analysis we can now determine the total **impedance** Z of the circuit, which is defined in analogy to resistance and reactance as

$$V_{\text{rms}} = I_{\text{rms}} Z, \quad \text{or} \quad V_0 = I_0 Z, \tag{30–27}$$

where as usual $V_{\text{rms}} = V_0/\sqrt{2}$ and $I_{\text{rms}} = I_0/\sqrt{2}$.

From Fig. 30–21 we see, using the Pythagorean theorem (V_0 is the hypotenuse of a right triangle), that

$$V_0 = \sqrt{V_{R0}^2 + (V_{L0} - V_{C0})^2}$$
$$= I_0\sqrt{R^2 + (X_L - X_C)^2}.$$

Thus, using Eq. 30–27, the total impedance Z is (see also Eqs. 30–23b and 30–25b)

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \tag{30–28a}$$

$$= \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}. \tag{30–28b}$$

Also from Fig. 30–21, we can find the phase angle ϕ between voltage and current:

$$\tan\phi = \frac{V_{L0} - V_{C0}}{V_{R0}} = \frac{I_0(X_L - X_C)}{I_0 R} = \frac{X_L - X_C}{R}, \tag{30–29a}$$

and

$$\cos\phi = \frac{V_{R0}}{V_0} = \frac{I_0 R}{I_0 Z} = \frac{R}{Z}. \tag{30–29b}$$

Figure 30–21 was drawn for the case $X_L > X_C$, and the current *lags* the source voltage by ϕ. When the reverse is true, $X_L < X_C$, then ϕ in Eqs. 30–29 is less than zero, and the current *leads* the source voltage.

We saw earlier that power is dissipated only by a resistance; none is dissipated by inductance or capacitance. Therefore, the average power $\overline{P} = I_{\text{rms}}^2 R$. But from Eq. 30–29b, $R = Z\cos\phi$. Therefore

$$\overline{P} = I_{\text{rms}}^2 Z\cos\phi = I_{\text{rms}} V_{\text{rms}} \cos\phi. \tag{30–30}$$

The factor $\cos\phi$ is referred to as the **power factor** of the circuit. For a pure resistor, $\cos\phi = 1$ and $\overline{P} = I_{\text{rms}} V_{\text{rms}}$. For a capacitor or inductor alone, $\phi = -\pi/2$ or $+\pi/2$ respectively, so $\cos\phi = 0$ and no power is dissipated.

EXAMPLE 30–11 ***LRC* series circuit.** Suppose $R = 25.0\,\Omega$, $L = 30.0\,\text{mH}$, and $C = 12.0\,\mu\text{F}$ in a series *LRC* circuit, Fig. 30–19, and they are connected in series to a 90.0-V ac (rms) 500-Hz source. Calculate (*a*) the rms current in the circuit, (*b*) the voltmeter readings (rms) across each element, (*c*) the phase angle ϕ, and (*d*) the power dissipated in the circuit.

APPROACH To obtain the current we need to determine the impedance (Eq. 30–28b), and then use $I_{\text{rms}} = V_{\text{rms}}/Z$. Voltage drops across each element are found using Ohm's law or equivalent for each element: $V_R = IR$, $V_L = IX_L$, and $V_C = IX_C$.

SOLUTION (*a*) First, we find the reactance of the inductor and capacitor at $f = 500\,\text{Hz} = 500\,\text{s}^{-1}$:

$$X_L = 2\pi fL = 94.2\,\Omega, \qquad X_C = \frac{1}{2\pi fC} = 26.5\,\Omega.$$

Then the total impedance is

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{(25.0\,\Omega)^2 + (94.2\,\Omega - 26.5\,\Omega)^2} = 72.2\,\Omega.$$

From the impedance version of Ohm's law, Eq. 30–27,

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{Z} = \frac{90.0\,\text{V}}{72.2\,\Omega} = 1.25\,\text{A}.$$

(*b*) The rms voltage across each element is

$$(V_R)_{\text{rms}} = I_{\text{rms}} R = (1.25\,\text{A})(25.0\,\Omega) = 31.2\,\text{V}$$
$$(V_L)_{\text{rms}} = I_{\text{rms}} X_L = (1.25\,\text{A})(94.2\,\Omega) = 118\,\text{V}$$
$$(V_C)_{\text{rms}} = I_{\text{rms}} X_C = (1.25\,\text{A})(26.5\,\Omega) = 33.1\,\text{V}.$$

NOTE These voltages do *not* add up to the source voltage, 90.0 V (rms). Indeed, the rms voltage across the inductance (118 V) *exceeds* the source voltage. This can happen because the different voltages are out of phase with each other: so, at any instant the capacitor's voltage might be negative which compensates for a large positive inductor voltage. The rms voltages, however, are always positive by definition. Although the rms voltages need not add up to the source voltage, the instantaneous voltages at any time must add up to the source voltage at that instant.

CAUTION
Individual peak or rms voltages do NOT add up to source voltage (due to phase differences)

(*c*) The phase angle ϕ is given by Eq. 30–29b,

$$\cos\phi = \frac{R}{Z} = \frac{25.0\,\Omega}{72.2\,\Omega} = 0.346,$$

so $\phi = 1.22\,\text{rad} \approx 70°$. Note that ϕ is positive because $X_L > X_C$ in this case, so $V_{L0} > V_{C0}$ as in Fig. 30–21.

(*d*) $\overline{P} = I_{\text{rms}} V_{\text{rms}} \cos\phi = (1.25\,\text{A})(90.0\,\text{V})(25.0\,\Omega/72.2\,\Omega) = 39.0\,\text{W}$.

LRC circuits with elements in parallel are treated in Problems 89, 91, and 96.

30–9 Resonance in AC Circuits

The rms current in an *LRC* series circuit is given by (see Eqs. 30–27 and 30–28b):

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{Z} = \frac{V_{\text{rms}}}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}. \qquad \textbf{(30–31)}$$

Because the reactance of inductors and capacitors depends on the frequency $f\,(=\omega/2\pi)$ of the source, the current in an *LRC* circuit will depend on frequency. From Eq. 30–31 we can see that the current will be maximum at a frequency that satisfies

$$\left(\omega L - \frac{1}{\omega C}\right) = 0.$$

We solve this for ω and call the solution ω_0:

$$\omega_0 = \sqrt{\frac{1}{LC}}. \qquad \text{[resonance]} \quad \textbf{(30–32a)}$$

When $\omega = \omega_0$, the circuit is in **resonance**, and

$$f_0 = \omega_0/2\pi = \frac{1}{2\pi\sqrt{LC}} \qquad \textbf{(30–32b)}$$

is the **resonant frequency** of the circuit. At this frequency, $X_C = X_L$, so the impedance is purely resistive (and $\cos\phi = 1$). A graph of I_{rms} versus ω is shown in Fig. 30–22 for particular values of R, L, and C. For small R compared to X_L and X_C, the resonance peak will be higher and sharper. When R is very small, the circuit approaches the pure *LC* (oscillator) circuit we discussed in Section 30–5. When R is large compared to X_L and X_C, the resonance curve is relatively flat—there is little frequency dependence.

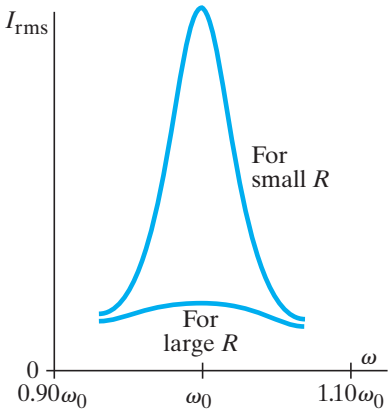


FIGURE 30–22 Current in *LRC* circuit as a function of angular frequency, ω, showing resonance peak at $\omega = \omega_0 = \sqrt{1/LC}$.

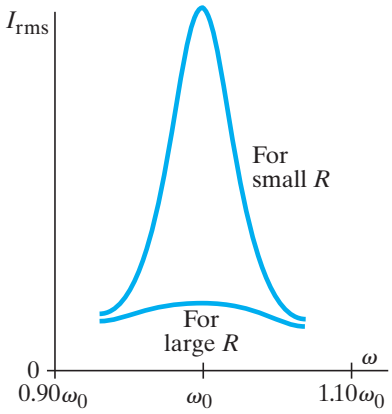


FIGURE 30–22 (Repeated.) Current in LRC circuit as a function of angular frequency, ω, showing resonance peak at $\omega = \omega_0 = \sqrt{1/LC}$.

This electrical resonance is analogous to mechanical resonance, discussed in Chapter 14. The energy transferred to the system by the source is a maximum at resonance whether it is electrical resonance, the oscillation of a spring, or pushing a child on a swing (Section 14–8). That this is true in the electrical case can be seen from Eq. 30–30; at resonance, $\cos\phi = 1$, and power $\overline{P}$ is a maximum. A graph of power versus frequency peaks like that for the current, Fig. 30–22.

Electric resonance is used in many circuits. Radio and TV sets, for example, use resonant circuits for tuning in a station. Many frequencies reach the circuit from the antenna, but a significant current flows only for those at or near the resonant frequency. Either L or C is variable in the resonant tuning circuit, so that different stations can be tuned in.

30–10 Impedance Matching

It is common to connect one electric circuit to a second circuit. For example, a TV antenna is connected to a TV receiver; an amplifier is connected to a loudspeaker; electrodes for an electrocardiogram are connected to a recorder. Maximum power is transferred from one to the other, with a minimum of loss, when the output impedance of the one device matches the input impedance of the second.

To show why, we consider simple circuits that contain only resistance. In Fig. 30–23 the source in circuit 1 could represent the signal from an antenna or a laboratory probe, and R_1 represents its resistance including internal resistance of the source. R_1 is called the output impedance (or resistance) of circuit 1. The output of circuit 1 is across the terminals a and b which are connected to the input of circuit 2, which may be very complicated. We let R_2 be the equivalent "input resistance" of circuit 2.

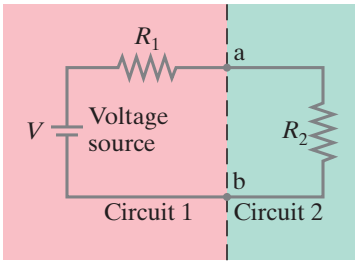


FIGURE 30–23 Output of the circuit on the left is input to the circuit on the right.

The power delivered to circuit 2 is $P = I^2R_2$ where $I = V/(R_1 + R_2)$ is the current that passes through both circuits. Thus

$$P = I^2R_2 = \frac{V^2R_2}{(R_1 + R_2)^2}.$$

If the resistance of the source is R_1, what value should R_2 have so that the maximum power is transferred to circuit 2? To determine a maximum, we take the derivative of P with respect to R_2 and set it equal to zero, which gives

$$V^2\left[\frac{1}{(R_1 + R_2)^2} - \frac{2R_2}{(R_1 + R_2)^3}\right] = V^2\left[\frac{R_1 + R_2 - 2R_2}{(R_1 + R_2)^3}\right] = 0$$

or

$$R_2 = R_1.$$

Thus, the maximum power is transmitted when the *output impedance* of one device *equals the input impedance* of the second. This is called **impedance matching**.

In an ac circuit that contains capacitors and inductors, the different phases are important and the analysis is more complicated. However, the same result holds: to maximize power transfer it is important to match impedances $(Z_2 = Z_1)$.

In addition, it is possible to seriously distort a signal if impedances do not match closely, and has in some cases led to meaningless or erroneous experimental results.

⚠ CAUTION

Erroneous results can occur if impedances don't match well

*30–11 Three-Phase AC

Transmission lines typically consist of four wires, rather than two. One of these wires is the ground; the remaining three are used to transmit three-phase ac power which is a superposition of three ac voltages 120° out of phase with each other:

$$V_1 = V_0 \sin \omega t$$
$$V_2 = V_0 \sin(\omega t + 2\pi/3)$$
$$V_3 = V_0 \sin(\omega t + 4\pi/3).$$

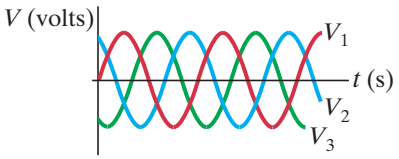


FIGURE 30–24 The three voltages, (red, green, blue) out of phase by 120° $(= \frac{2}{3}\pi$ radians$)$, in a three-phase power line.

(See Fig. 30–24.) Why is three-phase power used? We saw in Fig. 25–22 that single-phase ac (i.e., the voltage V_1 by itself) delivers power to the load in pulses. A much smoother flow of power can be delivered using three-phase power. Suppose that each of the three voltages making up the three-phase source is hooked up to a resistor R. Then the power delivered is:

$$P = \frac{1}{R}(V_1^2 + V_2^2 + V_3^2).$$

You can show (Problem 67) that this power is a constant equal to $3V_0^2/2R$, which is three times the rms power delivered by a single-phase source. This smooth flow of power makes electrical equipment run smoothly. Although houses use single-phase ac power, most industrial-grade machinery is wired for three-phase power.

EXAMPLE 30–12 **Three-phase circuit.** In a three-phase circuit, 266 V (rms) exists between line 1 and ground. What is the rms voltage between lines 2 and 3?

SOLUTION We are given $V_{\text{rms}} = V_0/\sqrt{2} = 266\,\text{V}$. Hence $V_0 = 376\,\text{V}$. Now $V_3 - V_2 = V_0[\sin(\omega t + 4\pi/3) - \sin(\omega t + 2\pi/3)] = 2V_0 \sin\frac{1}{2}(\frac{2\pi}{3})\cos\frac{1}{2}(2\omega t)$ where we used the identity: $\sin A - \sin B = 2\left[\sin\frac{1}{2}(A - B)\right]\left[\cos\frac{1}{2}(A + B)\right]$. The rms voltage is

$$(V_3 - V_2)_{\text{rms}} = \frac{1}{\sqrt{2}} 2V_0 \sin\frac{\pi}{3} = \sqrt{2}(376\,\text{V})(0.866) = 460\,\text{V (rms)}.$$

Summary

A changing current in a coil of wire will induce an emf in a second coil placed nearby. The **mutual inductance**, M, is defined as the proportionality constant between the induced emf $\mathcal{E}_2$ in the second coil and the time rate of change of current in the first:

$$\mathcal{E}_2 = -M\, dI_1/dt. \qquad \textbf{(30–3b)}$$

We can also write M as

$$M = \frac{N_2\Phi_{21}}{I_1} \qquad \textbf{(30–1)}$$

where Φ_{21} is the magnetic flux through coil 2 with N_2 loops, produced by the current I_1 in another coil (coil 1).

Within a single coil, a changing current induces an opposing emf, $\mathcal{E}$, so a coil has a **self-inductance** L defined by

$$\mathcal{E} = -L\, dI/dt. \qquad \textbf{(30–5)}$$

This induced emf acts as an *impedance* to the flow of an alternating current. We can also write L as

$$L = N\frac{\Phi_B}{I} \qquad \textbf{(30–4)}$$

where Φ_B is the flux through the inductance when a current I flows in its N loops.

When the current in an inductance L is I, the energy stored in the inductance is given by

$$U = \tfrac{1}{2}LI^2. \qquad \textbf{(30–6)}$$

This energy can be thought of as being stored in the magnetic field of the inductor. The energy density u in any magnetic field B is given by

$$u = \frac{1}{2}\frac{B^2}{\mu_0}, \qquad \textbf{(30–7)}$$

where μ_0 is replaced by μ if a ferromagnetic material is present.

When an inductance L and resistor R are connected in series to a constant dc source of emf, V_0, the current rises according to an exponential of the form

$$I = \frac{V_0}{R}(1 - e^{-t/\tau}), \qquad \textbf{(30–9)}$$

where

$$\tau = L/R \qquad \textbf{(30–10)}$$

is the **time constant**. The current eventually levels out at $I = I_0 = V_0/R$. If the battery is suddenly switched out of the ***LR* circuit**, and the circuit remains complete, the current drops exponentially, $I = I_0 e^{-t/\tau}$, with the same time constant τ.

The current in a pure ***LC* circuit** (or charge on the capacitor) would oscillate sinusoidally. The energy too would oscillate back and forth between electric and magnetic, from the capacitor to the inductor, and back again. If such a circuit has resistance (*LRC*), and the capacitor at some instant is charged, it can undergo damped oscillations or exhibit critically damped or overdamped behavior.

Capacitance and inductance offer *impedance* to the flow of alternating current just as resistance does. This impedance is referred to as **reactance**, X, and is defined (as for resistors) as the proportionality constant between voltage and current (either the rms or peak values). Across an inductor,

$$V_0 = I_0 X_L, \qquad \textbf{(30–23a)}$$

and across a capacitor,

$$V_0 = I_0 X_C. \qquad \textbf{(30–25a)}$$

The reactance of an inductor increases with frequency:

$$X_L = \omega L. \qquad \textbf{(30–23b)}$$

where $\omega = 2\pi f$ and f is the frequency of the ac. The reactance of a capacitor decreases with frequency:

$$X_C = \frac{1}{\omega C}. \qquad \textbf{(30–25b)}$$

Whereas the current through a resistor is always in phase with the voltage across it, this is not true for inductors and capacitors: in an inductor, the current lags the voltage by 90° or $\pi/2$ radians, and in a capacitor the current leads the voltage by 90° or $\pi/2$ radians.

In an ac ***LRC* series circuit**, the total **impedance** Z is defined by the equivalent of $V = IR$ for resistance: namely $V_0 = I_0 Z$ or $V_{\text{rms}} = I_{\text{rms}} Z$. The impedance Z is related to R, C, and L by

$$Z = \sqrt{R^2 + (X_L - X_C)^2}. \qquad \textbf{(30–28a)}$$

The current in the circuit lags (or leads) the source voltage by an angle ϕ given by $\cos\phi = R/Z$. Only the resistor in an ac *LRC* circuit dissipates energy, and at a rate

$$\overline{P} = I_{\text{rms}}^2 Z \cos\phi \qquad \textbf{(30–30)}$$

where the factor $\cos\phi$ is referred to as the **power factor**.

An *LRC* series circuit **resonates** at a frequency given by

$$\omega_0 = \sqrt{\frac{1}{LC}}. \qquad \textbf{(30–32a)}$$

The rms current in an *LRC* circuit is largest (as is the power transferred) when the applied voltage has a frequency equal to $f_0\ (= \omega_0/2\pi)$. The lower the resistance R, the higher and sharper the resonance peak.

Questions

1. How would you arrange two flat circular coils so that their mutual inductance was (*a*) greatest, (*b*) least (without separating them by a great distance)? Explain.
2. The second coil of N_2 turns in Fig. 30–2 is moved so it is near the end of the solenoid. How does this affect the mutual inductance?
3. Would two coils with mutual inductance also have self-inductance? Explain.
4. Is the energy density inside a solenoid greatest near the ends of the solenoid or near its center?
5. If you are given a fixed length of wire, how would you shape it to obtain the greatest self-inductance? The least?
6. Does the emf of the battery in Fig. 30–6a affect the time needed for the *LR* circuit to reach (*a*) a given fraction of its maximum possible current, (*b*) a given value of current? Explain.
7. A circuit with large inductive time constant carries a steady current. If a switch is opened, there can be a very large (and sometimes dangerous) spark or "arcing over." Explain. [*Hint*: Problem 25 may help.]
8. At the instant the battery is connected into the *LR* circuit of Fig. 30–6a, the emf in the inductor has its maximum value even though the current is zero. Explain.
9. What keeps an *LC* circuit oscillating even after the capacitor has discharged completely?
10. In an *LRC* circuit, can the rms voltage across (*a*) an inductor, (*b*) a capacitor, be greater than the rms voltage of the ac source? Explain.
11. When an ac generator is connected to an *LRC* circuit, where does the energy come from ultimately? Where does it go? How do the values of L, C, and R affect the energy supplied by the generator?
12. In an ac *LRC* circuit, if $X_L > X_C$, the circuit is said to be predominantly "inductive." And if $X_C > X_L$, the circuit is said to be predominantly "capacitive." Discuss the reasons for these terms. In particular, do they say anything about the relative values of L and C at a given frequency?
13. Do the results of Section 30–8 approach the proper expected results when ω approaches zero? What are the expected results?
14. Under what conditions is the impedance in an *LRC* circuit a minimum?
15. Is it possible for the instantaneous power output of an ac generator connected to an *LRC* circuit ever to be negative? Explain.
16. In an ac *LRC* circuit, does the power factor, $\cos\phi$, depend on frequency? Does the power dissipated depend on frequency?
17. Describe briefly how the frequency of the source emf affects the impedance of (*a*) a pure resistance, (*b*) a pure capacitance, (*c*) a pure inductance, (*d*) an *LRC* circuit near resonance (R small), (*e*) an *LRC* circuit far from resonance (R small).
18. Discuss the response of an *LRC* circuit as $R \to 0$ when the frequency is (*a*) at resonance, (*b*) near resonance, (*c*) far from resonance. Is there energy dissipation in each case? Discuss the transformations of energy that occur in each case.
19. An *LRC* resonant circuit with low resistance is often called an *oscillator* circuit. What is it that oscillates?
20. Compare the oscillations of an *LRC* circuit to the vibration of a mass m on a spring. What do L and C correspond to in the mechanical system?
21. Describe how to make the impedance in an *LRC* circuit a minimum.

MisConceptual Questions

1. How would you orient two flat circular coils so that their mutual inductance was greatest?
(*a*) Have their "faces" be parallel to each other.
(*b*) Have their "faces" be at 45° relative to each other.
(*c*) Have their "faces" be perpendicular to each other.
(*d*) The orientation of the faces does not affect their mutual inductance.

2. If you have 1.0 m of wire, how would you shape it to obtain the greatest self-inductance?
(*a*) Leave it straight.
(*b*) Make one single loop.
(*c*) Make two loops.
(*d*) Make many tiny loops.
(*e*) It won't matter.

3. A circuit with a large inductive time constant carries a steady current. When a switch is opened, causing the current to be interrupted, what happens?
(*a*) The current drops to zero in a very short time with no other effects.
(*b*) A current in the opposite direction of the original current is induced in the circuit.
(*c*) A (possibly large) emf is induced which can lead to a spark or "arcing over."
(*d*) The inductive reactance increases due to the sudden current decrease.
[*Hint*: Problem 25 may be helpful.]

4. What would increase the period of an oscillating LC circuit the most?
(*a*) Double the inductance.
(*b*) Double the capacitance.
(*c*) Double both the inductance and the capacitance.
(*d*) Double the maximum charge on the capacitor.
(*e*) Double the maximum current through the inductor.

5. How does the resistance affect the period of an oscillating LC circuit with low resistance?
(*a*) The smaller the resistance, the larger the period.
(*b*) The smaller the resistance, the smaller the period.
(*c*) The resistance does not affect the period.
(*d*) There is a specific non-zero resistance at which the period is minimized.
(*e*) There is a specific non-zero resistance at which the period is maximized.

6. Which of the following actions will *always* increase the rms current in an LRC series ac circuit?
(*a*) Increase the source voltage.
(*b*) Increase the source frequency.
(*c*) Increase the resistance in the circuit.
(*d*) Increase the capacitance in the circuit.
(*e*) Increase the inductance in the circuit.
(*f*) None of the above.

7. In an oscillating LC circuit, what oscillates?
(*a*) The amount of charge on the capacitor.
(*b*) The current in the inductor.
(*c*) The voltage across the capacitor.
(*d*) The voltage across the inductor.
(*e*) All of the above quantities.

8. Is the ac current in the inductor always the same as the current in the resistor of the LRC circuit of Fig. 30–19?
(*a*) No.
(*b*) Yes, but only when $X_C = X_L$.
(*c*) Yes, but only if $R = 0$.
(*d*) Yes, always.
(*e*) It depends.

9. A series LRC circuit has been connected to an ac source for a long time. The average power is given by Eq. 30–30. The average power
(*a*) is transformed into heat by the resistor.
(*b*) goes into energy stored by the magnetic field in the inductor.
(*c*) goes into energy stored by the electric field in the capacitor.
(*d*) (*a*), (*b*), and (*c*) are all true.
(*e*) Only (*b*) and (*c*) are true.

10. In an LRC circuit, what must be in phase with the input voltage?
(*a*) The voltage across the resistor.
(*b*) The voltage across the inductor.
(*c*) The voltage across the capacitor.
(*d*) The current through the resistor.
(*e*) All of the above.
(*f*) None of the above.

11. Which of the following is true about all series ac circuits?
(*a*) The voltage across any circuit element is a maximum when the current is a maximum in that circuit element.
(*b*) The current at any point in the circuit is always the same as the current at any other point in the circuit.
(*c*) The current in the circuit is a maximum when the source ac voltage is a maximum.
(*d*) Resistors, capacitors, and inductors can all change the phase of the current.

12. In an LRC circuit with ac source, the peak source voltage V_0 equals
(*a*) $V_{R0} + V_{L0} + V_{C0}$.
(*b*) $V_{R0} + V_{L0} - V_{C0}$.
(*c*) $V_{L0} - V_{C0}$.
(*d*) None of the above.

13. Under what circumstance is the impedance in an LRC circuit a minimum?
(*a*) When $C = 0$.
(*b*) When $L = 0$.
(*c*) At resonance.

Problems

30–1 Mutual Inductance

1. (II) Determine the mutual inductance per unit length between two long solenoids, one inside the other, whose radii are r_1 and r_2 $(r_2 < r_1)$ and whose turns per unit length are n_1 and n_2.

2. (II) A small thin coil with N_2 loops, each of area A_2, is placed inside a long solenoid, near its center. The solenoid has N_1 loops in its length ℓ and has area A_1. Determine the mutual inductance as a function of θ, the angle between the plane of the small coil and the axis of the solenoid.

3. (II) A 2.44-m-long coil containing 265 loops is wound on an iron core (average $\mu = 1850\mu_0$) along with a second coil of 115 loops. The loops of each coil have a radius of 2.00 cm. (*a*) Determine the mutual inductance M. (*b*) If the current in the first coil drops uniformly from 12.0 A to zero in 98.0 ms, determine the emf induced in the second coil.

4. (III) A long straight wire and a small rectangular wire loop lie in the same plane, Fig. 30–25. Determine the mutual inductance in terms of ℓ_1, ℓ_2, and w. Assume the wire is very long compared to ℓ_1, ℓ_2, and w, and that the rest of its circuit is very far away compared to ℓ_1, ℓ_2, and w.

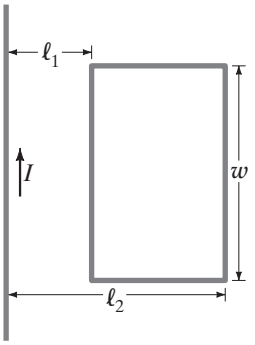


FIGURE 30–25
Problem 4.

30–2 Self-Inductance; Inductors

5. (I) Determine the inductance L of a 0.60-m-long air-filled solenoid 2.9 cm in diameter containing 7500 loops.

6. (I) How many turns of wire would be required to make a 130-mH inductor out of a 33.0-cm-long air-filled coil solenoid with a diameter of 4.2 cm?

7. (I) What is the inductance of a coil if the coil produces an emf of 2.20 V when the current in it changes from −28.0 mA to +25.0 mA in 12.0 ms?

8. (II) An air-filled cylindrical inductor has 3200 turns, and it is 2.5 cm in diameter and 28.2 cm long. (*a*) What is its inductance? (*b*) How many turns would you need to generate the same inductance if the core were iron-filled instead? Assume the magnetic permeability of iron is about 1200 times that of free space.

9. (II) A coil has 3.25-Ω resistance and 440-mH inductance. If the current is 3.00 A and is increasing at a rate of 3.15 A/s, what is the potential difference across the coil at this moment?

10. (II) If the outer conductor of a coaxial cable has radius 3.0 mm, what should be the radius of the inner conductor so that the inductance per unit length does not exceed 75 nH per meter?

11. (II) The wire of a tightly wound solenoid is unwound and used to make another tightly wound solenoid of 2.2 times the diameter. By what factor does the inductance change?

12. (II) Ignoring any mutual inductance, what is the equivalent inductance of two inductors connected (*a*) in series, (*b*) in parallel?

13. (III) A toroid has a rectangular cross section as shown in Fig. 30–26. Show that the self-inductance is

$$L = \frac{\mu_0 N^2 h}{2\pi} \ln \frac{r_2}{r_1}$$

where N is the total number of turns and r_1, r_2, and h are the dimensions shown in Fig. 30–26. [*Hint*: Use Ampère's law to get B as a function of r inside the toroid, and integrate.]

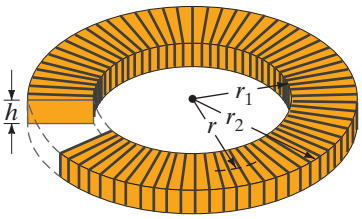


FIGURE 30–26
Problems 13 and 18. A toroid of rectangular cross section, with N turns carrying a current I.

30–3 Magnetic Energy Storage

14. (I) The magnetic field inside an air-filled solenoid 38.0 cm long and 2.10 cm in diameter is 0.720 T. Approximately how much energy is stored in this field?

15. (I) Typical large values for electric and magnetic fields attained in laboratories are about 1.0×10^4 V/m and 2.0 T. (*a*) Determine the energy density for each field and calculate the ratio of electric to magnetic energy. (*b*) What magnitude electric field would be needed to produce the same energy density as the 2.0-T magnetic field?

16. (II) What is the energy density at the center of a circular loop of wire carrying a 19.0-A current if the radius of the loop is 28.0 cm?

17. (II) Calculate the magnetic and electric energy densities at the surface of a 3.0-mm-diameter copper wire carrying an 18-A current. [*Hint*: Do not ignore resistance of the wire.]

18. (II) For the toroid of Fig. 30–26, determine the energy density in the magnetic field as a function of r $(r_1 < r < r_2)$ and integrate this over the volume to obtain the total energy stored in the toroid, which carries a current I in each of its N loops.

19. (II) Determine the total energy stored per unit length in the magnetic field between the coaxial cylinders of a coaxial cable (Fig. 30–5) by using Eq. 30–7 for the energy density and integrating over the volume.

20. (II) A long straight wire of radius r carries current I uniformly distributed across its cross-sectional area. Find the magnetic energy stored per unit length in the interior of this wire.

30–4 *LR* Circuits

21. (II) It takes 2.85 ms for the current in an LR circuit to increase from zero to 0.75 its maximum value. Determine (*a*) the time constant of the circuit, (*b*) the resistance of the circuit if $L = 31.0$ mH.

22. (II) How many time constants does it take for the potential difference across the resistor in an LR circuit like that in Fig. 30–7 to drop to 2.0% of its original value, after the switch is moved to the upper position, removing V_0 from the circuit?

23. (II) (*a*) Determine the energy stored in the inductor L as a function of time for the LR circuit of Fig. 30–6a. (*b*) After how many time constants does the stored energy reach 99.9% of its maximum value?

24. (II) In the circuit of Fig. 30–27, determine the current in each resistor (I_1, I_2, I_3) at the moment (*a*) the switch is closed, (*b*) a long time after the switch is closed. After the switch has been closed for a long time, and then reopened, what is each current (*c*) just after it is opened, (*d*) after a long time?

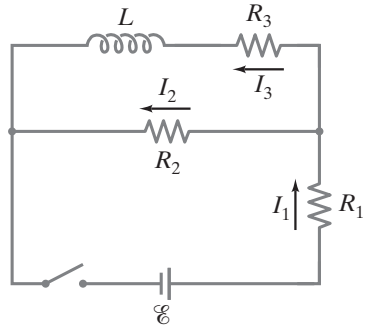


FIGURE 30–27 Problem 24.

25. (II) (*a*) In Fig. 30–28, assume that the switch has been in position A for sufficient time so that a steady current $I_0 = V_0/R$ flows through the resistor R. At time $t = 0$, the switch is quickly switched to position B and the current decays through resistor R' (which is much greater than R) according to $I = I_0 e^{-t/\tau'}$. Show that the maximum emf $\mathscr{E}_{\max}$ induced in the inductor during this time period is $(R'/R)V_0$. (*b*) If $R' = 45R$ and $V_0 = 145\,\text{V}$, determine $\mathscr{E}_{\max}$. [When a mechanical switch is opened, a high-resistance air gap is created, which is modeled as R' here. This Problem illustrates why **high-voltage sparking** can occur if a current-carrying inductor is suddenly cut off from its power source. The very high voltage can produce an electric field great enough to ionize atoms of air, which emit light when electrons recombine with the ions.]

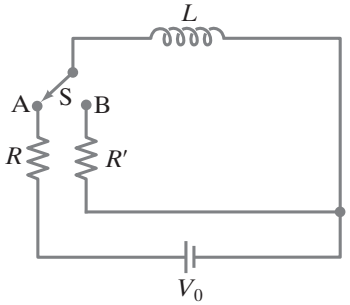


FIGURE 30–28 Problem 25.

26. (II) For the LR circuit of Fig. 30–6a determine dI/dt at $t = 0$ (when the battery is connected) and show that if I continued to increase at this rate, it would reach its maximum value in one time constant.

27. (III) Two tightly wound solenoids have the same length and circular cross-sectional area. But solenoid 1 uses wire that is 1.5 times as thick as solenoid 2. (*a*) What is the ratio of their inductances? (*b*) What is the ratio of their inductive time constants? (Assume the only resistance in the circuits is that of the wire itself.)

30–5 *LC* Circuits and Oscillations

28. (II) (*a*) If the initial conditions of an LC circuit are $I = I_0$ and $Q = 0$ at $t = 0$, write Q as a function of time. (*b*) Practically, how could you set up these initial conditions?

29. (II) A 475-pF capacitor is charged to 135 V and then quickly connected to a 175-mH inductor. Determine (*a*) the frequency of oscillation, (*b*) the peak value of the current, and (*c*) the maximum energy stored in the magnetic field of the inductor.

30. (II) At $t = 0$, let $Q = Q_0$, and $I = 0$ in an LC circuit. (*a*) The first time when the energy becomes shared equally by the inductor and the capacitor, what is the charge on the capacitor? (*b*) How much time has elapsed (in terms of the period T)?

30–6 *LC* Oscillations with Resistance

31. (II) A damped LC circuit loses 3.5% of its electromagnetic energy per cycle to thermal energy. If $L = 85\,\text{mH}$ and $C = 1.00\,\mu\text{F}$, what is the value of R?

32. (II) In an oscillating LRC circuit, how much time does it take for the energy stored in the fields of the capacitor and inductor to fall to 65% of the initial value? (See Fig. 30–13; assume $R \ll \sqrt{4L/C}$.)

33. (III) How much resistance must be added to a pure LC circuit ($L = 350\,\text{mH}$, $C = 1600\,\text{pF}$) to change the oscillator's frequency by 0.25%? Will it be increased or decreased?

30–7 AC Circuits; Reactance

34. (I) At what frequency will a 32.0-mH inductor have a reactance of 760 Ω?

35. (I) What is the reactance of an 8.4-μF capacitor at a frequency of (*a*) 60.0 Hz, (*b*) 1.00 MHz?

36. (I) Calculate the reactance of, and rms current in, a 36.0-mH radio coil connected to a 350-V (rms) 33.3-kHz ac line. Ignore resistance.

37. (II) An inductance coil operates at 240 V (rms) and 60.0 Hz. It draws 10.8 A (rms). What is the coil's inductance?

38. (II) A resistor R is in parallel with a capacitor C, and this parallel combination is in series with a resistor R'. If connected to an ac voltage source of frequency ω, what is the equivalent impedance of this circuit at the two extremes in frequency (*a*) $\omega = 0$, and (*b*) $\omega = \infty$?

39. (II) What is the inductance L of the primary of a transformer whose input is 110 V rms at 60.0 Hz and the rms current drawn is 3.8 A? Assume no current in the secondary.

40. (II) (*a*) What is the reactance of a well-insulated 0.030-μF capacitor connected to a 1.6-kV (rms) 720-Hz line? (*b*) What will be the peak value of the current?

41. (II) A capacitor is placed in parallel with some device, B, as in Fig. 30–18b, to filter out stray high-frequency signals, but to allow ordinary 60.0-Hz ac to pass through with little loss. Suppose that circuit B in Fig. 30–18b is a resistance $R = 530\,\Omega$ connected to ground, and that $C = 0.35\,\mu\text{F}$. Calculate the ratio of the capacitor's current amplitude to the incoming current's amplitude if the incoming current has a frequency of (*a*) 60.0 Hz; (*b*) 60.0 kHz.

42. (II) A current $I = 1.80\cos 377t$ (I in amps, t in seconds, and the argument of the cosine is in radians) flows in a series LR circuit in which $L = 3.45\,\text{mH}$ and $R = 1.35\,\text{k}\Omega$. What is the average power dissipation?

30–8 *LRC* Series AC Circuit; Phasor Diagrams

43. (I) A 10.0-kΩ resistor is in series with a 34.0-mH inductor and an ac source. Calculate the impedance of the circuit if the source frequency is (*a*) 55.0 Hz; (*b*) 55.0 kHz.

44. (I) A 75-Ω resistor and a 5.8-μF capacitor are connected in series to an ac source. Calculate the impedance of the circuit if the source frequency is (*a*) 60.0 Hz; (*b*) 6.0 MHz.

45. (II) For a 120-V, 60-Hz voltage, a current of 70 mA passing through the body for 1.0 s could be lethal. What must be the impedance of the body for this to occur?

46. (II) Determine the resistance of a coil if its impedance is 275 Ω and its reactance is 145 Ω.

47. (II) A 1.50-kΩ resistor in series with a 370-mH inductor is driven by an ac power supply. At what frequency is the impedance double that of the impedance at 60.0 Hz?

48. (II) An ac voltage source is connected in series with a 2.0-μF capacitor and a 750-Ω resistor. Using a digital ac voltmeter, the voltage source is measured to be 4.0 V rms, and the voltages across the resistor and across the capacitor are found to be 3.0 V rms and 2.7 V rms, respectively. Determine the frequency of the ac voltage source. Why is the voltage measured across the voltage source not equal to the sum of the voltages measured across the resistor and across the capacitor?

49. (II) Determine the total impedance, phase angle, and rms current in an *LRC* circuit connected to a 10.0-kHz, 625-V (rms) source if $L = 32.0\,\text{mH}$, $R = 8.70\,\text{k}\Omega$, and $C = 6250\,\text{pF}$.

50. (II) (*a*) What is the rms current in a series *LR* circuit when a 60.0-Hz, 120-V rms ac voltage is applied, where $R = 965\,\Omega$ and $L = 255\,\text{mH}$? (*b*) What is the phase angle between voltage and current? (*c*) How much power is dissipated? (*d*) What are the rms voltage readings across *R* and *L*?

51. (II) A 35-mH inductor with 2.0-Ω resistance is connected in series to a 21-μF capacitor and a 60.0-Hz, 45-V (rms) source. Calculate (*a*) the rms current, (*b*) the phase angle, and (*c*) the power dissipated in this circuit.

52. (II) A 25-mH coil whose resistance is 0.80 Ω is connected to a capacitor *C* and a 420-Hz source voltage. If the current and voltage are to be in phase, what value must *C* have?

53. (II) A 75-W incandescent lightbulb is designed to operate with an applied ac voltage of 120 V rms. The bulb is placed in series with an inductor *L*, and this series combination is then connected to a 60.0-Hz 240-V rms voltage source. For the bulb to operate properly, determine the required value for *L*. Assume the bulb has resistance *R* and negligible inductance.

54. (II) In the *LRC* circuit of Fig. 30–19, suppose $I = I_0 \sin\omega t$ (not cosine as in the text) and $V = V_0 \sin(\omega t + \phi)$. Determine the instantaneous power dissipated in the circuit from $P = IV$ using these equations and show that on the average, $\overline{P} = \frac{1}{2} V_0 I_0 \cos\phi$, which is in accord with Eq. 30–30.

55. (II) An *LRC* series circuit with $R = 120\,\Omega$, $L = 25\,\text{mH}$, and $C = 2.0\,\mu\text{F}$ is powered by an ac voltage source of peak voltage $V_0 = 340\,\text{V}$ and frequency $f = 660\,\text{Hz}$. (*a*) Determine the peak current that flows in this circuit. (*b*) Determine the phase angle of the source voltage relative to the current. (*c*) Determine the peak voltage across *R* and its phase angle relative to the source voltage. (*d*) Determine the peak voltage across *L* and its phase angle relative to the source voltage. (*e*) Determine the peak voltage across *C* and its phase angle relative to the source voltage.

56. (II) An *LR* circuit can be used as a "phase shifter." Assume that an "input" source voltage $V = V_0 \sin(2\pi f t + \phi)$ is connected across a series combination of an inductor $L = 55\,\text{mH}$ and resistor *R*. The "output" of this circuit is taken across the resistor. If $V_0 = 24\,\text{V}$ and $f = 225\,\text{Hz}$, determine the value of *R* so that the output voltage V_R lags the input voltage *V* by 25°. Compare (as a ratio) the peak output voltage with V_0.

30–9 Resonance in AC Circuits

57. (I) A 3400-pF capacitor is connected in series to a 26.0-μH coil of resistance 2.00 Ω. What is the resonant frequency of this circuit?

58. (I) What is the resonant frequency of the *LRC* circuit of Example 30–11? At what rate is energy taken from the generator, on the average, at this frequency?

59. (II) An *LRC* circuit has $L = 4.15\,\text{mH}$ and $R = 3.80\,\text{k}\Omega$. (*a*) What value must *C* have to produce resonance at 38.0 kHz? (*b*) What will be the maximum current at resonance if the peak external voltage is 136 V?

60. (II) The frequency of the ac voltage source (peak voltage V_0) in an *LRC* circuit is tuned to the circuit's resonant frequency $f_0 = 1/(2\pi\sqrt{LC})$. (*a*) Show that the peak voltage across the capacitor is $V_{C0} = V_0\, T_0/(2\pi\tau)$, where $T_0 (= 1/f_0)$ is the period of the resonant frequency and $\tau = RC$ is the time constant for charging the capacitor *C* through a resistor *R*. (*b*) Define $\beta = T_0/(2\pi\tau)$ so that $V_{C0} = \beta V_0$. Then β is the "amplification" of the source voltage across the capacitor. If a particular *LRC* circuit contains a 2.0-nF capacitor and has a resonant frequency of 5.0 kHz, what value of *R* will yield $\beta = 125$?

61. (II) Capacitors made from piezoelectric materials are commonly used as sound transducers ("loudspeakers"). They often require a large operating voltage. One method for providing the required voltage is to include the loudspeaker as part of an *LRC* circuit as shown in Fig. 30–29, where the loudspeaker is modeled electrically as the capacitance $C = 1.0\,\text{nF}$. Take $R = 35\,\Omega$ and $L = 55\,\text{mH}$. (*a*) What is the resonant frequency f_0 for this circuit? (*b*) If the voltage source has peak amplitude $V_0 = 2.0\,\text{V}$ at frequency $f = f_0$, find the peak voltage V_{C0} across the loudspeaker (i.e., the capacitor *C*). (*c*) Determine the ratio V_{C0}/V_0.

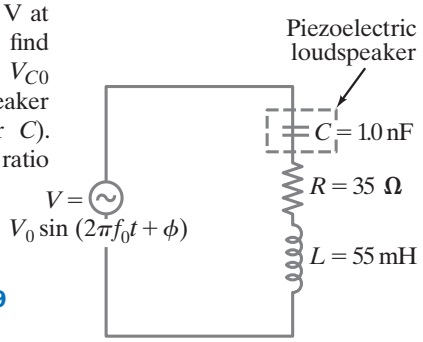


FIGURE 30–29
Problem 61.

62. (II) (*a*) Show that oscillation of charge *Q* on the capacitor of an *LRC* circuit has amplitude

$$Q_0 = \frac{V_0}{\sqrt{(\omega R)^2 + \left(\omega^2 L - \frac{1}{C}\right)^2}}.$$

(*b*) At what angular frequency, ω', will Q_0 be a maximum? (*c*) Compare to a forced damped harmonic oscillator (Chapter 14), and discuss. What do *L* and *C* correspond to in the mechanical system?

63. (II) A resonant circuit using a 220-nF capacitor is to resonate at 35.0 kHz. The air-core inductor is to be a solenoid with closely packed coils made from 12.0 m of insulated wire 1.1 mm in diameter. How many loops should the inductor contain?

64. (III) (*a*) Determine a formula for the average power $\overline{P}$ dissipated in an *LRC* circuit in terms of L, R, C, ω, and V_0. (*b*) At what frequency is the power a maximum? (*c*) Find an approximate formula for the width of the resonance peak in average power, $\Delta\omega$, which is the difference in the two (angular) frequencies where $\overline{P}$ has half its maximum value. Assume a sharp peak.

30–10 Impedance Matching

65. (II) The output of an electrocardiogram amplifier has an impedance of 45 kΩ. It is to be connected to an 8.0-Ω loudspeaker through a transformer. What should be the turns ratio of the transformer?

*30–11 Three-Phase AC

***66.** (II) Show that the power delivered by a three-phase ac source equals a constant $P = 3V_0^2/2R$, by combining the four equations in Section 30–11.

General Problems

67. At $t = 0$, the current through a 60.0-mH inductor is 50.0 mA and is increasing at the rate of 78.0 mA/s. What is the initial energy stored in the inductor, and how long does it take for the energy to increase by a factor of 8.0 from the initial value?

68. At time $t = 0$, the switch in the circuit shown in Fig. 30–30 is closed. After a sufficiently long time, steady currents I_1, I_2, and I_3 flow through resistors R_1, R_2, and R_3, respectively. Determine these three currents.

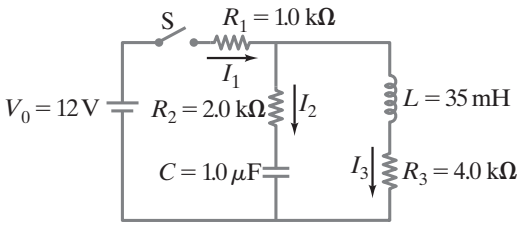


FIGURE 30–30 Problem 68.

69. (*a*) Show that the self-inductance L of a toroid (Fig. 30–31) of radius r_0 containing N loops each of diameter d is

$$L \approx \frac{\mu_0 N^2 d^2}{8 r_0}$$

if $r_0 \gg d$. Assume the field is uniform inside the toroid; is this actually true? (*b*) Is this result consistent with L for a solenoid? Should it be? (*c*) Calculate the inductance L of a large toroid if the diameter of the coils is 2.0 cm and the diameter of the whole ring is 66 cm. Assume the field inside the toroid is uniform. There are a total of 660 loops of wire.

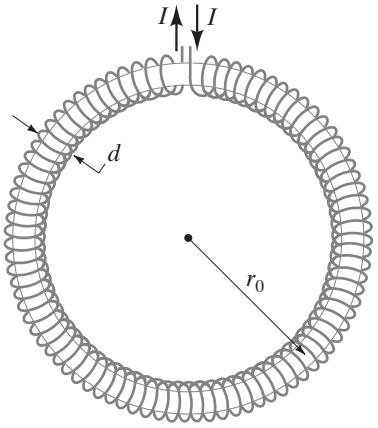


FIGURE 30–31 A toroid. Problem 69.

70. A pair of straight parallel thin wires, such as a lamp cord, each of radius r, are a distance ℓ apart and carry current to a circuit some distance away. Ignoring the field within each wire, show that the inductance per unit length is $(\mu_0/\pi)\ln[(\ell - r)/r]$.

71. Assuming the Earth's magnetic field averages about 0.50×10^{-4} T near the surface of the Earth, estimate the total energy stored in this field in the first 5.0 km above the Earth's surface.

72. (*a*) For an underdamped *LRC* circuit, determine a formula for the energy $U = U_E + U_B$ stored in the electric and magnetic fields as a function of time. Give answer in terms of the initial charge Q_0 on the capacitor. (*b*) Show how dU/dt is related to the rate energy is transformed in the resistor, I^2R.

73. A physics professor wants to demonstrate the large size of the henry unit. On the outside of a 12-cm-diameter plastic hollow tube, she wants to wind an air-filled solenoid with self-inductance of 1.0 H using copper wire with a 0.81-mm diameter. The solenoid is to be tightly wound with each turn touching its neighbor (the wire has a thin insulating layer on its surface so the neighboring turns are not in electrical contact). How long will the plastic tube need to be and what length of copper wire will be required? What will be the resistance of this solenoid?

74. The circuit shown in Fig. 30–32a can integrate (in the calculus sense) the input voltage V_{in}, if the time constant L/R is large compared with the time during which V_{in} varies. Explain how this *integrator* works and sketch its output for the square wave signal input shown in Fig. 30–32b. [*Hint*: Write Kirchhoff's loop rule for the circuit. Multiply each term in this differential equation (in I) by a factor $e^{Rt/L}$ to make it easier to integrate.]

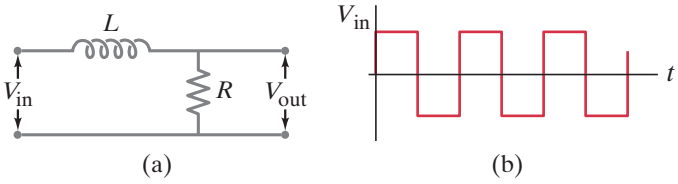


FIGURE 30–32 Problem 74.

75. Suppose circuit B in Fig. 30–18a consists of a resistance $R = 550\ \Omega$. The filter capacitor has capacitance $C = 1.6\ \mu\text{F}$. Will this capacitor act to eliminate 60.0-Hz ac but pass a high-frequency signal of frequency 6.0 kHz? To check this, determine the rms voltage drop across R for a 130-mV rms signal of frequency (*a*) 60.0 Hz; (*b*) 6.0 kHz.

76. In a series *LRC* circuit, the inductance is 33 mH, the capacitance is 55 nF, and the resistance is 1.50 kΩ. At what frequencies is the power factor equal to 0.24?

77. In our analysis of a series LRC circuit, Fig. 30–19, suppose we chose $V = V_0 \sin \omega t$. (*a*) Construct a phasor diagram, like that of Fig. 30–21, for this case. (*b*) Write a formula for the current I, defining all terms.

78. An ac voltage source $V = V_0 \sin(\omega t + 90°)$ is connected across an inductor L and current $I = I_0 \sin(\omega t)$ flows in this circuit. Note that the current and source voltage are 90° out of phase. (*a*) Directly calculate the average power delivered by the source over one period T of its sinusoidal cycle via the integral $\overline{P} = \int_0^T VI\, dt/T$. (*b*) Apply the relation $\overline{P} = I_{\text{rms}} V_{\text{rms}} \cos\phi$ to this circuit and show that the answer you obtain is consistent with that found in part (*a*). Comment on your results.

79. A circuit contains two elements, but it is not known if they are L, R, or C. The current in this circuit when connected to a 120-V rms 60.0-Hz source is 5.6 A and lags the voltage by 55°. What are the two elements and what are their values?

80. (*a*) What is the rms current in an RC circuit if $R = 6.40\,\text{k}\Omega$, $C = 1.80\,\mu\text{F}$, and the rms applied voltage is 120 V at 60.0 Hz? (*b*) What is the phase angle between voltage and current? (*c*) What is the power dissipated by the circuit? (*d*) What are the voltmeter readings across R and C?

81. An inductance coil draws 2.2 A dc when connected to a 45-V battery. When connected to a 60.0-Hz 120-V (rms) source, the current drawn is 3.8 A (rms). Determine the inductance and resistance of the coil.

82. The ***Q*-value** of a resonance circuit can be defined as the ratio of the voltage across the capacitor (or inductor) to the voltage across the resistor, at resonance. The larger the Q factor, the sharper the resonance curve will be and the sharper the tuning. (*a*) Show that the Q factor is given by the equation $Q = (1/R)\sqrt{L/C}$. (*b*) At a resonant frequency $f_0 = 1.0\,\text{MHz}$, what must be the value of L and R to produce a Q factor of 550? Assume that $C = 0.010\,\mu\text{F}$.

83. Show that the fraction of electromagnetic energy lost (to thermal energy) per cycle in a lightly damped $(R^2 \ll 4L/C)$ LRC circuit is approximately

$$\frac{\Delta U}{U} = \frac{2\pi R}{L\omega} = \frac{2\pi}{Q}.$$

The quantity Q can be defined as $Q = L\omega/R$, and is called the *Q-value*, or *quality factor*, of the circuit and is a measure of the damping present. A high Q-value means smaller damping and less energy input required to maintain oscillations.

84. In some experiments, very *tiny distances* or spaces ($\approx$ nm) can be measured by using capacitance. Consider forming an LC circuit using a parallel-plate capacitor with plate area A, and a known inductance L. (*a*) If charge is found to oscillate in this circuit at frequency $f = \omega/2\pi$ when the capacitor plates are separated by distance x, show that $x = 4\pi^2 A\epsilon_0 f^2 L$. (*b*) When the plate separation is changed by Δx, the circuit's oscillation frequency will change by Δf. Show that $\Delta x/x \approx 2(\Delta f/f)$. (*c*) If f is on the order of 1 MHz and can be measured to a precision of $\Delta f = 1$ Hz, with what percent accuracy can x be determined? Assume fringing effects at the capacitor's edges can be neglected.

85. An electronic device needs to be protected against sudden surges in current. In particular, after the power is turned on the current should rise to no more than 7.5 mA in the first 75 μs. The device has resistance 120 Ω and is designed to operate at 33 mA. How would you protect this device?

86. A voltage $V = 0.95 \sin 688t$ is applied to an LRC circuit (V is in volts, t is in seconds, and the argument of the sine is in radians) which has $L = 22.0\,\text{mH}$, $R = 23.2\,\text{k}\Omega$, and $C = 0.42\,\mu\text{F}$. (*a*) What is the impedance and phase angle? (*b*) How much power is dissipated in the circuit? (*c*) What is the rms current and voltage across each element?

87. **Filter circuit.** Figure 30–33 shows a simple filter circuit designed to pass dc voltages with minimal attenuation and to remove, as much as possible, any ac components (such as 60-Hz line voltage that could cause hum in an audio system, for example). Assume $V_{\text{in}} = V_1 + V_2$ where V_1 is dc and $V_2 = V_{20} \sin \omega t$, and that any resistance is very small. (*a*) Determine the current through the capacitor: give amplitude and phase (assume $R = 0$ and $X_L > X_C$). (*b*) Show that the ac component of the output voltage, $V_{2\,\text{out}}$, equals $(Q/C) - V_1$, where Q is the charge on the capacitor at any instant, and determine the amplitude and phase of $V_{2\,\text{out}}$. (*c*) Show that the attenuation of the ac voltage is greatest when $X_C \ll X_L$, and calculate the ratio of the output to input ac voltage in this case. (*d*) Compare the dc output voltage to input voltage.

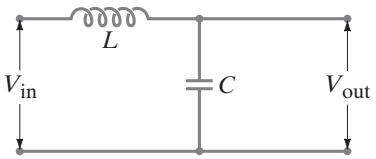


FIGURE 30–33
Problems 87 and 88.

88. Show that if the inductor L in the filter circuit of Fig. 30–33 (Problem 87) is replaced by a large resistor R, there will still be significant attenuation of the ac voltage and little attenuation of the dc voltage if the input dc voltage is high and the current (and power) are low.

89. **Parallel *LRC* Circuit.** A resistor R, capacitor C, and inductor L are connected in parallel across an ac generator as shown in Fig. 30–34. The source emf is $V = V_0 \sin \omega t$. Determine the current as a function of time (including amplitude and phase): (*a*) in the resistor, (*b*) in the inductor, (*c*) in the capacitor. (*d*) What is the total current leaving the source? (Give amplitude I_0 and phase.) (*e*) Determine the impedance Z defined as $Z = V_0/I_0$. (*f*) What is the power factor?

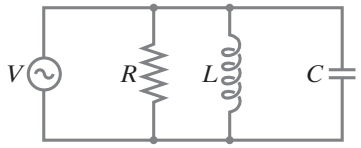


FIGURE 30–34
Problem 89.

90. Suppose a series LRC circuit has two resistors, R_1 and R_2, two capacitors, C_1 and C_2, and two inductors, L_1 and L_2, all in series. Calculate the total impedance of the circuit.

91. For the circuit shown in Fig. 30–35, show that if the condition $R_1 R_2 = L/C$ is satisfied then the potential difference between points a and b is zero for all frequencies.

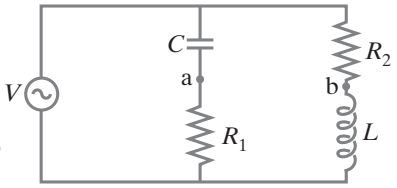


FIGURE 30–35
Problem 91.

92. In a certain LRC series circuit, when the ac voltage source has a particular frequency f, the peak voltage across the inductor is 5.0 times greater than the peak voltage across the capacitor. Determine f in terms of the resonant frequency f_0 of this circuit.

93. In a *plasma globe*, a hollow glass sphere is filled with low-pressure gas and a small spherical metal electrode is located at its center. Assume an ac voltage source of peak voltage V_0 and frequency f is applied between the metal sphere and the ground, and that a person is touching the outer surface of the globe with a fingertip, whose approximate area is $1.0\,\text{cm}^2$. The equivalent circuit for this situation is shown in Fig. 30–36, where R_G and R_P are the resistances of the gas and the person, respectively, and C is the capacitance formed by the gas, glass, and finger. (*a*) Determine C assuming it is a parallel-plate capacitor. The conductive gas and the person's fingertip form the opposing plates of area $A = 1.0\,\text{cm}^2$. The plates are separated by glass (dielectric constant $K = 5.0$) of thickness $d = 2.0\,\text{mm}$. (*b*) In a typical plasma globe, $f = 12\,\text{kHz}$. Determine the reactance X_C of C at this frequency in MΩ. (*c*) The voltage may be $V_0 = 2500\,\text{V}$. With this high voltage, the dielectric strength of the gas is exceeded and the gas becomes ionized. In this "plasma" state, the gas emits light ("sparks") and is highly conductive so that $R_G \ll X_C$. Assuming also that $R_P \ll X_C$, estimate the peak current that flows in the given circuit. Is this level of current dangerous? (*d*) If the plasma globe operated at $f = 1.0\,\text{MHz}$, estimate the peak current that would flow in the given circuit. Is this level of current dangerous?

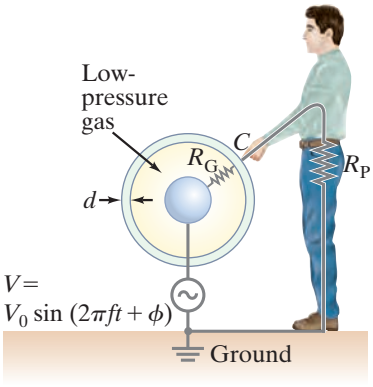


FIGURE 30–36
Problem 93.

94. You have a small electromagnet that consumes 350 W from a residential circuit operating at 120 V rms at 60.0 Hz. Using your ac multimeter, you determine that the unit draws 5.0 A rms. What are the values of the inductance and the internal resistance?

95. An inductor L in series with a resistor R, driven by a sinusoidal voltage source, responds as described by the following differential equation:

$$V_0 \sin\omega t = L\frac{dI}{dt} + RI.$$

Show that a current of the form $I = I_0 \sin(\omega t - \phi)$ flows through the circuit by direct substitution into the differential equation. Determine the amplitude of the current (I_0) and the phase difference ϕ between the current and the voltage source.

96. Series and parallel *LRC*. For the circuit shown in Fig. 30–37, $V = V_0 \sin\omega t$. Determine an equation for the current in each element of the circuit, as well as the total impedance. [*Hint*: Try a trial solution of the form $I = I_0 \sin(\omega t + \phi)$ for the current leaving the source.]

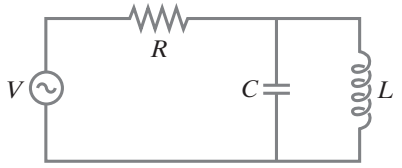


FIGURE 30–37
Problem 96.

97. The *RC* circuit shown in Fig. 30–38 is a **high-pass filter** because it passes high-frequency ac signals with less attenuation than low-frequency ac signals. (*a*) Show that the voltage gain is $A = V_{\text{out}}/V_{\text{in}} = 2\pi fRC/(4\pi^2 f^2 R^2 C^2 + 1)^{\frac{1}{2}}$. (*b*) Discuss the behavior of the gain A for $f \to 0$ and $f \to \infty$.

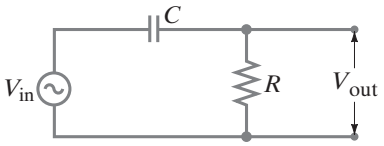


FIGURE 30–38
Problem 97.

98. The *RC* circuit shown in Fig. 30–39 is a **low-pass filter** because it passes low-frequency ac signals with less attenuation than high-frequency ac signals. (*a*) Show that the voltage gain is $A = V_{\text{out}}/V_{\text{in}} = 1/(4\pi^2 f^2 R^2 C^2 + 1)^{\frac{1}{2}}$. (*b*) Discuss the behavior of the gain A for $f \to 0$ and $f \to \infty$.

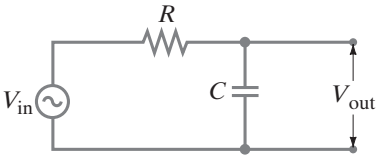


FIGURE 30–39
Problem 98.

99. To detect vehicles at traffic lights, wire loops with dimensions on the order of 2 m are often buried horizontally under roadways. Assume the self-inductance of such a coil is $L = 5.0\,\text{mH}$ and that it is part of an *LRC* circuit as shown in Fig. 30–40 with $C = 0.10\,\mu\text{F}$ and $R = 38\,\Omega$. The ac voltage has frequency f and rms voltage V_{rms}. (*a*) The frequency f is chosen to match the resonant frequency f_0 of the circuit. Find f_0 and determine what the rms voltage $(V_R)_{\text{rms}}$ across the resistor will be when $f = f_0$. (*b*) Assume that f, C, and R never change, but that, when a car is located above the buried coil, the coil's self-inductance decreases by 10% (due to induced eddy currents in the car's metal parts). Determine by what factor the voltage $(V_R)_{\text{rms}}$ decreases in the presence of a car in comparison to no car above the loop and thus how it detects the presence of a car. (*c*) Describe how the eddy currents induced in the car reduce L. [*Hint*: Recall Eq. 30–4, the definition of inductance.]

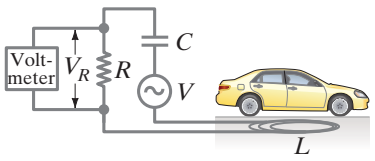


FIGURE 30–40
Problem 99.

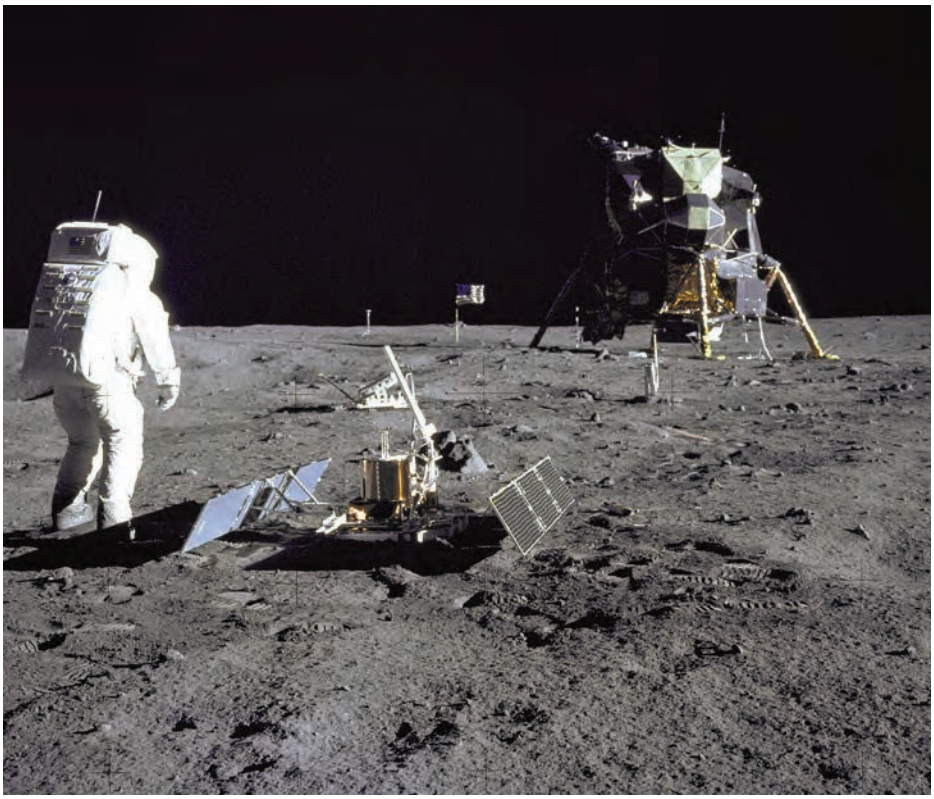
ANSWERS TO EXERCISES

A: (*b*).
B: (*a*).
C: From Eq. 30–5, L has dimensions VT/A so (L/R) has dimensions $(VT/A)/(V/A) = T$.
D: (*d*).
E: (*c*).
F: (*a*).

Wireless technology is all around us: radio and television, cell phones, wi-fi, Bluetooth, and all wireless communication. These devices communicate by electromagnetic waves traveling through space. Wireless devices are applications of Marconi's development of long-distance transmission of information more than a century ago.

In this photo we see the first humans to land on the Moon. In the background is a television camera that sent live moving images through empty space to Earth where it was shown live.

We will see in this Chapter that Maxwell predicted the existence of EM waves from his famous equations. Maxwell's equations themselves are a magnificent summary of electromagnetism. We will also see that EM waves carry energy and momentum, and that light itself is an electromagnetic wave.



CHAPTER 31

Maxwell's Equations and Electromagnetic Waves

CONTENTS

31 – 1 Changing Electric Fields Produce Magnetic Fields; Displacement Current
31 – 2 Gauss's Law for Magnetism
31 – 3 Maxwell's Equations
31 – 4 Production of Electromagnetic Waves
31 – 5 Electromagnetic Waves, and Their Speed, Derived from Maxwell's Equations
31 – 6 Light as an Electromagnetic Wave and the Electromagnetic Spectrum
31 – 7 Measuring the Speed of Light
31 – 8 Energy in EM Waves; the Poynting Vector
31 – 9 Radiation Pressure
31–10 Radio and Television; Wireless Communication

CHAPTER-OPENING QUESTION—Guess now!

Which of the following best describes the difference between radio waves and X-rays?

(a) X-rays are radiation whereas radio waves are electromagnetic waves.
(b) Both can be thought of as electromagnetic waves. They differ only in wavelength and frequency.
(c) X-rays are pure energy. Radio waves are made of fields, not energy.
(d) Radio waves come from electric currents in an antenna. X-rays are not related to electric charge.
(e) X-rays are made up of particles called photons whereas radio waves are oscillations in space.

The culmination of electromagnetic theory in the nineteenth century was the prediction, and the experimental verification, that waves of electromagnetic fields could travel through space. This achievement opened a whole new world of communication: first the wireless telegraph, then radio, television, and radar, and more recently cell phones, remote-control devices, wi-fi, GPS, Bluetooth, and all satellite communication. Most important was the spectacular prediction that light is an electromagnetic wave.

The theoretical prediction of electromagnetic waves was the work of the Scottish physicist James Clerk Maxwell (1831–1879; Fig. 31–1), who unified, in one magnificent theory, all the phenomena of electricity and magnetism.

The development of electromagnetic theory in the early part of the nineteenth century by Oersted, Ampère, and others was not actually done in terms of electric and magnetic fields. The idea of the field was introduced somewhat later by Faraday, and was not generally used until Maxwell showed that all electric and magnetic phenomena could be described using only four equations involving electric and magnetic fields. These equations, known as **Maxwell's equations**, are the basic equations for all electromagnetism. They are fundamental in the same sense that Newton's three laws of motion and the law of universal gravitation are for mechanics. In a sense, they are even more fundamental, because they are consistent with the theory of relativity (Chapter 36), whereas Newton's laws are not. Because all of electromagnetism is contained in this set of four equations, Maxwell's equations are considered one of the great triumphs of human intellect.

Before discussing Maxwell's equations and electromagnetic waves, we need to discuss a major prediction of Maxwell's (Section 31–1), and also Gauss's law for magnetism (Section 31–2).



FIGURE 31–1 James Clerk Maxwell (1831–1879).

31–1 Changing Electric Fields Produce Magnetic Fields; Displacement Current

Ampère's Law

Recall from Chapters 27 and 28 that a magnetic field is produced by an electric current, as was discovered by Oersted, and the mathematic relation is given by Ampère's law (Eq. 28–3):

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_{\text{encl}}.$$

Is it possible that magnetic fields could be produced in another way as well? Maxwell argued that if a changing magnetic field produces an electric field, as given by Faraday's law (Section 29–7), then the reverse might be true as well: **a changing electric field will produce a magnetic field**. This was an *hypothesis* by Maxwell, and it turned out to be correct. It is based on the idea of *symmetry* in nature. Indeed, the size of the effect in most cases is so small that Maxwell recognized it would be difficult to detect it experimentally.

To back up the idea that a changing electric field might produce a magnetic field, we use an indirect argument that goes something like this. According to Ampère's law (Section 28–4), we divide any closed path you choose into short segments $d\vec{\boldsymbol{\ell}}$, take the dot product of each $d\vec{\boldsymbol{\ell}}$ with the magnetic field $\vec{\mathbf{B}}$ at that segment, and sum (integrate) all these products over the complete closed path. That sum will then equal μ_0 times the total current I that passes through a surface bounded by the path of the line integral. When we applied Ampère's law to the field around a straight wire (Section 28–4), we imagined the current as passing through the flat circular area enclosed by our circular loop. That area is the surface 1 shown in Fig. 31–2. However, we could just as well imagine using the sack-shaped surface 2 in Fig. 31–2 as the surface for Ampère's law because the same current I passes through it.

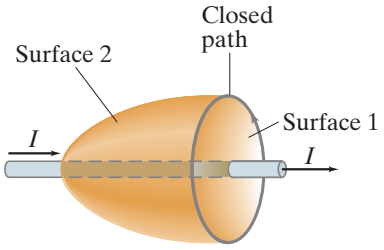
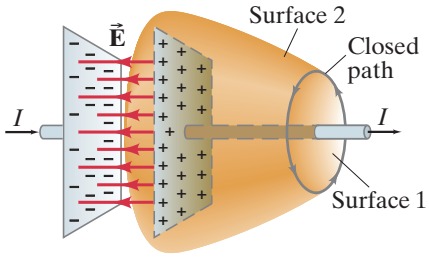


FIGURE 31–2 Ampère's law applied to two different surfaces bounded by the same closed path.

Now consider the closed circular path for the situation of Fig. 31–3, where a capacitor is being discharged, and two imaginary surfaces are shown. Ampère's law works for surface 1 (current I passes through surface 1), but it does not work for surface 2 because no current passes through surface 2. There is a magnetic field around the wire, so the left side of Ampère's law, $\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$, is not zero around the circular closed path; yet no current flows through surface 2, so the right side of Ampère's law *is* zero for surface 2. We seem to have a contradiction of Ampère's law.

There is a magnetic field present in Fig. 31–3, however, only if charge is flowing to or away from the capacitor plates. The changing charge on the plates means that the electric field between the plates is changing in time.

FIGURE 31–3 A capacitor discharging. A conduction current passes through surface 1, but no conduction current passes through the sack-like surface 2. An extra term is needed in Ampère's law.



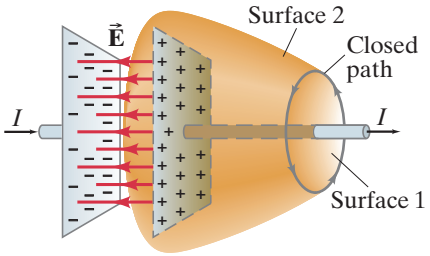


FIGURE 31–3 (Repeated.)
A capacitor discharging. A conduction current passes through surface 1, but no conduction current passes through the sack-like surface 2. An extra term is needed in Ampère's law.

Maxwell resolved the problem of no current through surface 2 in Fig. 31–3 by proposing that the changing electric field between the plates is *equivalent to* an electric current. He called it **displacement current** I_D. (An ordinary current I is then called a **conduction current**.)

Let us see what this displacement current term should be by determining it for the changing electric field between the capacitor plates in Fig. 31–3. The charge Q on a capacitor of capacitance C is $Q = CV$ where V is the potential difference between the plates (Eq. 24–1). Also recall that $V = Ed$ (Eq. 23–4) where d is the (small) separation of the plates and E is the (uniform) electric field magnitude between them, if we ignore any fringing of the field. Also, for a parallel-plate capacitor, $C = \epsilon_0 A/d$, where A is the area of each plate (Eq. 24–2). We combine these to obtain:

$$Q = CV = \left(\epsilon_0 \frac{A}{d}\right)(Ed) = \epsilon_0 AE.$$

If the charge on each plate changes at a rate dQ/dt, the electric field changes at a proportional rate. That is, by differentiating this expression for Q, we have:

$$\frac{dQ}{dt} = \epsilon_0 A \frac{dE}{dt}.$$

But dQ/dt is also the current I flowing into or out of the capacitor, so

$$I = \frac{dQ}{dt} = \epsilon_0 A \frac{dE}{dt} = \epsilon_0 \frac{d\Phi_E}{dt}$$

where $\Phi_E = EA$ is the **electric flux** through a surface enclosed by the integration path, which we now take as surface 2 in Fig. 31–3. In order to make Ampère's law work for surface 2 in Fig. 31–3, as well as for surface 1 (where current I flows), we therefore write:

AMPÈRE'S LAW (general form)

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I_{\text{encl}} + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}. \qquad \textbf{(31–1)}$$

This equation is the general form of **Ampère's law**,[†] and embodies Maxwell's idea that a magnetic field can be caused not only by an ordinary electric current, but also by a changing electric field or changing electric flux. Although we arrived at it for a special case, Eq. 31–1 has proved valid in general. The last term on the right in Eq. 31–1 is usually very small, and not easy to measure experimentally.

EXAMPLE 31–1 **Maxwell's extra term: charging capacitor.** A 30-pF air-gap capacitor has circular plates of area $A = 100\ \text{cm}^2$. It is charged by a 70-V battery through a 2.0-Ω resistor. At the instant the battery is connected, $t = 0$, the electric field between the plates is changing most rapidly. At this instant, calculate (*a*) the current in this RC circuit, and (*b*) the rate of change of electric field between the plates. (*c*) Determine the magnetic field induced between the plates. Assume $\vec{\mathbf{E}}$ is uniform between the plates at any instant and is zero at all points beyond the edges of the plates. (*d*) Determine $\vec{\mathbf{B}}$ beyond the edges of the plates.

APPROACH In Section 26–5 we discussed *RC* circuits and saw that the charge on a capacitor being charged, as a function of time, is

$$Q = CV_0\left(1 - e^{-t/RC}\right),$$

where V_0 is the voltage of the battery. To find the current at $t = 0$, we need to differentiate this and substitute the values $V_0 = 70\ \text{V}$, $C = 30\ \text{pF}$, $R = 2.0\ \Omega$.

SOLUTION (*a*) We take the derivative of Q and evaluate it at $t = 0$:

$$\left.\frac{dQ}{dt}\right|_{t=0} = \left.\frac{CV_0}{RC} e^{-t/RC}\right|_{t=0} = \frac{V_0}{R} = \frac{70\ \text{V}}{2.0\ \Omega} = 35\ \text{A}.$$

This is the rate at which charge accumulates on the capacitor and equals the current flowing in the circuit at $t = 0$.

[†]Actually, there is a third term on the right for the case when a magnetic field is produced by magnetized materials. This can be accounted for by changing μ_0 to μ but we will mainly be interested in cases where no magnetic material is present. In the presence of a dielectric, ϵ_0 is replaced by $\epsilon = K\epsilon_0$ (see Section 24–5).

(*b*) The electric field between two closely spaced conductors is given by (Eq. 21–8)

$$E = \frac{\sigma}{\epsilon_0} = \frac{Q/A}{\epsilon_0}$$

as we saw in Chapter 21 (see Example 21–14). Hence at $t = 0$

$$\frac{dE}{dt} = \frac{dQ/dt}{\epsilon_0 A} = \frac{35\ \text{A}}{(8.85\times10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)(1.0\times10^{-2}\ \text{m}^2)}$$
$$= 4.0\times10^{14}\ \text{V/m}\cdot\text{s}.$$

(*c*) Because of *symmetry*, we assume the lines of $\vec{\mathbf{B}}$ are circles perpendicular to $\vec{\mathbf{E}}$, as shown in Fig. 31–4a. This is the same symmetry we saw for the opposite situation of a changing magnetic field producing an electric field (Section 29–7, see Fig. 29–28c). To determine the magnitude of B between the capacitor plates we apply Ampère's law, Eq. 31–1, with the current $I_{\text{encl}} = 0$:

$$\oint \vec{\mathbf{B}}\cdot d\vec{\boldsymbol{\ell}} = \mu_0\epsilon_0\frac{d\Phi_E}{dt}.$$

We choose our path to be a circle of radius r, centered at the center of the plate and thus following a magnetic field line such as the one shown in Fig. 31–4. For $r \le r_0$ (the radius of plate) the flux through a circle of radius r is $E(\pi r^2)$ since E is assumed uniform between the plates at any moment. So from Ampère's law we have

$$B(2\pi r) = \mu_0\epsilon_0\frac{d}{dt}(\pi r^2 E)$$
$$= \mu_0\epsilon_0\pi r^2\frac{dE}{dt}.$$

Hence

$$B = \frac{\mu_0\epsilon_0}{2}r\frac{dE}{dt} \qquad [r \le r_0]$$

and increases linearly in r out to the edge, $r = r_0$.

(*d*) We have assumed $\vec{\mathbf{E}} \approx 0$ for $r > r_0$, so essentially all the electric flux is contained within the plates $(\text{area} = 4\pi r_0^2)$ and $\Phi_E = E\pi r_0^2$ is a constant for any $r > r_0$. Ampère's law gives

$$B(2\pi r) = \mu_0\epsilon_0\frac{d}{dt}(\pi r_0^2 E)$$
$$= \mu_0\epsilon_0\pi r_0^2\frac{dE}{dt}$$

or

$$B = \frac{\mu_0\epsilon_0 r_0^2}{2r}\frac{dE}{dt}. \qquad [r \ge r_0]$$

Thus B *decreases* with distance r from the center of the plates. See Fig. 31–4b.

NOTE B has its maximum value at $r = r_0$ which, from either relation above (using $r_0 = \sqrt{A/\pi} = 5.6\,\text{cm}$), when evaluated at $t = 0$, is

$$B_{\text{max}} = \frac{\mu_0\epsilon_0 r_0}{2}\frac{dE}{dt}$$
$$= \tfrac{1}{2}(4\pi\times10^{-7}\ \text{T}\cdot\text{m/A})(8.85\times10^{-12}\ \text{C}^2/\text{N}\cdot\text{m}^2)(5.6\times10^{-2}\ \text{m})(4.0\times10^{14}\ \text{V/m}\cdot\text{s})$$
$$= 1.2\times10^{-4}\ \text{T}.$$

This is a small field even at $t = 0$, and decreases very quickly because $dE/dt = (dQ/dt)/\epsilon_0 A = (V_0/R\epsilon_0 A)e^{-t/RC}$ (see part *a* and *b* above) where $RC = 6.0\times10^{-11}\,\text{s}$, so Maxwell's extra term in Ampère's law, which initially is on the order of Earth's magnetic field, is so brief it is not so easy to measure.

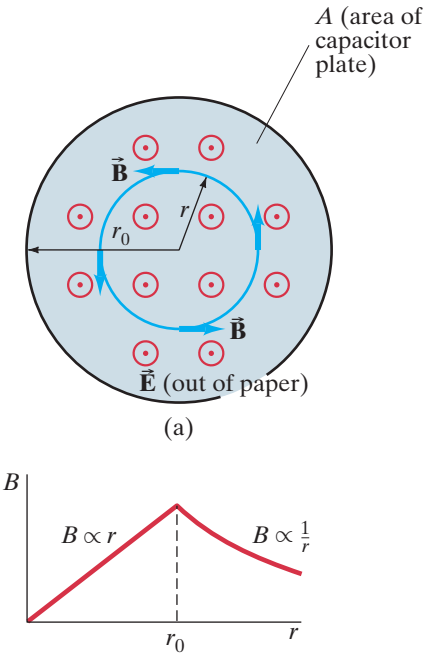


FIGURE 31–4 (Example 31–1.) (a) A circular plate of a parallel-plate capacitor. $\vec{\mathbf{E}}$ between plates points out toward viewer; lines of $\vec{\mathbf{B}}$ are circles. (b) Graph of B vs. r.

We can also consider the magnetic field B outside the capacitor plates of Example 31–1 in terms of the current I in the circuit. The electric field between the plates is $E = \sigma/\epsilon_0 = Q/(\epsilon_0 A)$, as we saw in part *b*, so $dE/dt = I/(\epsilon_0 A)$.

Hence B for $r > r_0$ is, with $A = \pi r_0^2$,

$$B = \frac{\mu_0 \epsilon_0 r_0^2}{2r} \frac{dE}{dt} = \frac{\mu_0 \epsilon_0 r_0^2}{2r} \frac{I}{\epsilon_0 \pi r_0^2} = \frac{\mu_0 I}{2\pi r}.$$

This is the same formula for the field that surrounds a wire (Eq. 28–1). Thus the B field outside the capacitor (Example 31–1) is the same as B outside the wire connected to it carrying current I. In other words, the magnetic field produced by the changing electric field between the plates is the same as that produced by the current in the wire.

Displacement Current

Maxwell interpreted the second term on the right in Eq. 31–1 as being *equivalent* to an electric current. He called it a **displacement current**, I_D. An ordinary current I is then called a **conduction current**. Ampère's law can then be written

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 (I + I_D)_{\text{encl}} \tag{31–2}$$

where

Displacement current

$$I_D = \epsilon_0 \frac{d\Phi_E}{dt}. \tag{31–3}$$

The term "displacement current" was based on an old discarded theory. Don't let it confuse you: I_D does not represent a flow of electric charge,[†] nor is there a displacement.

31–2 Gauss's Law for Magnetism

We are almost in a position to state Maxwell's equations, but first we need to discuss the magnetic equivalent of Gauss's law. As we saw in Chapter 29, for a magnetic field $\vec{\mathbf{B}}$ the *magnetic flux* Φ_B through a surface is defined as

$$\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}$$

where the integral is over the area of either an open or a closed surface. The magnetic flux through a closed surface—that is, a surface which completely encloses a volume—is written

$$\Phi_B = \oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}.$$

In the electric case, we saw in Section 22–2 on Gauss's law that the electric flux Φ_E through a closed surface is equal to the total net charge Q enclosed by the surface, divided by ϵ_0 (Eq. 22–4):

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q_{\text{encl}}}{\epsilon_0}.$$

This relation is Gauss's law for electricity.

We can write a similar relation for the magnetic flux. We have seen (Section 27–1) that in spite of intense searches, no isolated magnetic poles (monopoles)—the magnetic equivalent of single electric charges—have ever been observed. Hence, **Gauss's law for magnetism** is

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0. \tag{31–4}$$

In terms of magnetic field lines, this relation tells us that as many lines enter the enclosed volume as leave it. If, indeed, magnetic monopoles do not exist, then there are no "sources" or "sinks" for magnetic field lines to start or stop on, unlike electric field lines that start on positive charges and end on negative charges.

[†]The interpretation of the changing electric field as a current does fit in well with our discussion in Chapter 30 where we saw that an alternating current can be said to pass through a capacitor (although charge doesn't). It also means that Kirchhoff's junction rule will be valid even at a capacitor plate: conduction current flows into the plate, but no conduction current flows out of the plate—instead a "displacement current" flows out of one plate (toward the other plate).

Magnetic field lines must then be continuous. Even for a bar magnet, a magnetic field $\vec{\mathbf{B}}$ exists inside as well as outside the magnetic material, and the lines of $\vec{\mathbf{B}}$ are closed loops as shown in Fig. 31–5.

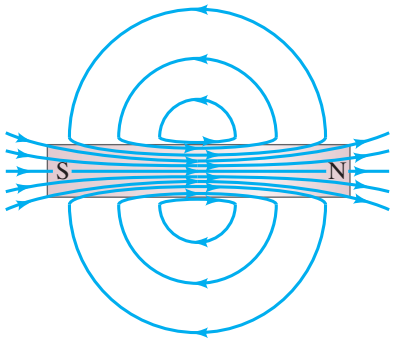


FIGURE 31–5 Magnetic field lines for a bar magnet. All the lines curve around, as shown for a few, and are continuous.

31–3 Maxwell's Equations

With the extension of Ampère's law given by Eq. 31–1, plus Gauss's law for magnetism (Eq. 31–4), we are now ready to state all four of Maxwell's equations. We have seen them all before in the past ten Chapters. In the absence of dielectric or magnetic materials, **Maxwell's equations** are:

MAXWELL'S EQUATIONS

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q}{\epsilon_0} \qquad \textbf{(31–5a)}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0 \qquad \textbf{(31–5b)}$$

$$\oint \vec{\mathbf{E}} \cdot d\vec{\ell} = -\frac{d\Phi_B}{dt} \qquad \textbf{(31–5c)}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\ell} = \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}. \qquad \textbf{(31–5d)}$$

The first two of Maxwell's equations are the same as Gauss's law for electricity (Chapter 22, Eq. 22–4) and Gauss's law for magnetism (Section 31–2, Eq. 31–4). The third is the general form of Faraday's law (Chapter 29, Eq. 29–8) and the fourth is Ampère's law as modified by Maxwell (Eq. 31–1). The first two equations have integrals over a closed surface on the left, and the second two have line integrals around a closed path. The quantities on the right of each of these equations, Q, I, and the fluxes, are those enclosed by, or through, a chosen surface (related to the integral on the left). (We drop the subscripts "encl" for simplicity.)

They can be summarized in words: (1) a generalized form of Coulomb's law relating electric field to its sources, electric charges; (2) the same for the magnetic field, except that if there are no magnetic monopoles, magnetic field lines are continuous—they do not begin or end (as electric field lines do on charges); (3) an electric field is produced by a changing magnetic field; (4) a magnetic field is produced by an electric current or by a changing electric field.

Maxwell's equations are the basic equations for all electromagnetism, and are as fundamental as Newton's three laws of motion and the law of universal gravitation. Maxwell's equations can also be written in differential form; see Appendix F.

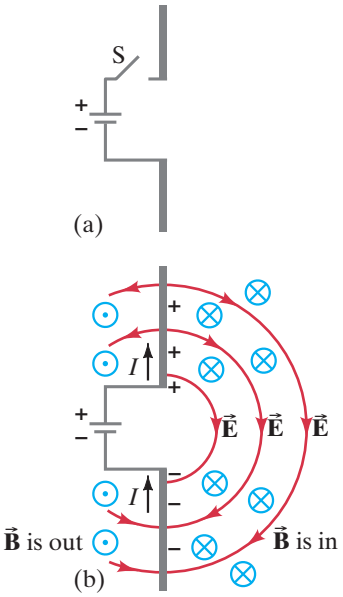
In earlier Chapters, we have seen that we can treat electric and magnetic fields separately if they do not vary in time. But we cannot treat them independently if they change in time because a changing magnetic field produces an electric field; and a changing electric field produces a magnetic field. An important outcome of these relations is the production of electromagnetic waves.

31–4 Production of Electromagnetic Waves

According to Maxwell, a magnetic field will be produced in empty space if there is a changing electric field. From this, Maxwell derived another startling conclusion. If a changing magnetic field produces an electric field, that electric field is itself changing. This changing electric field will, in turn, produce a magnetic field, which will be changing, and so it too will produce a changing electric field; and so on. Maxwell found that the net result of these interacting changing fields was a *wave* of electric and magnetic fields that can propagate (travel) through space! We now examine, in a simplified way, how such **electromagnetic waves** can be produced.

Consider two conducting rods that will serve as an "antenna" (Fig. 31–6a). Suppose these two rods are connected by a switch to the opposite terminals of a battery. When the switch is closed, the upper rod quickly becomes positively charged and the lower one negatively charged. Electric field lines are formed as indicated by the red lines in Fig. 31–6b. While the charges are flowing, a current exists whose direction is indicated by the black arrows. A magnetic field is therefore produced near the antenna. The magnetic field lines encircle the rod-like antenna and therefore, in Fig. 31–6b, $\vec{\mathbf{B}}$ points into the page ($\otimes$) on the right and out of the page ($\odot$) on the left.

FIGURE 31–6 Fields produced by charge flowing into conductors. It takes time for the $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ fields to travel outward to distant points. The fields are shown to the right of the antenna, but they move out in all directions, symmetrically about the (vertical) antenna.



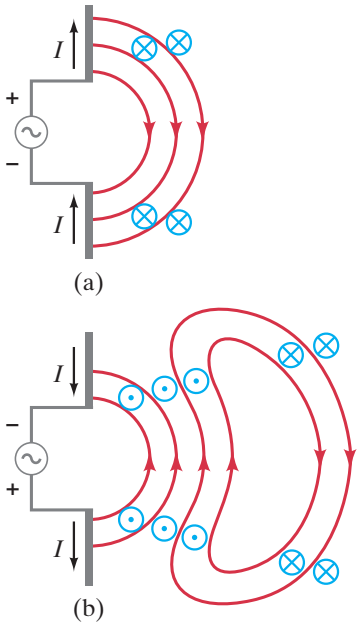


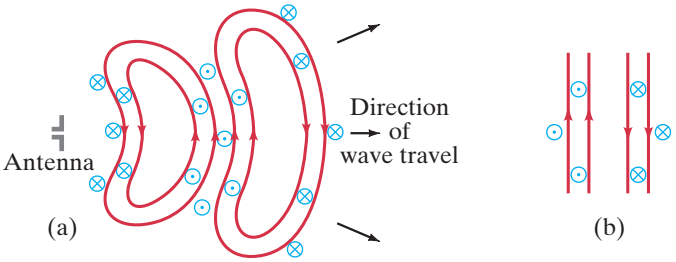
FIGURE 31–7 Sequence showing electric and magnetic fields that spread outward from oscillating charges on two conductors (the antenna) connected to an ac source (see the text).

Now we ask: how far out do these electric and magnetic fields extend? In the static case, the fields would extend outward indefinitely far. However, when the switch in Fig. 31–6 is closed, the fields quickly appear nearby, but it takes time for them to reach distant points. Both electric and magnetic fields store energy, and this energy cannot be transferred to distant points at infinite speed.

Now we look at the situation of Fig. 31–7 where our antenna is connected to an ac generator. In Fig. 31–7a, the connection has just been completed. Charge starts building up and fields form just as in Fig. 31–6. The + and − signs in Fig. 31–7a indicate the net charge on each rod at a given instant. The black arrows indicate the direction of the current. The electric field is represented by the red lines in the plane of the page; and the magnetic field, according to the right-hand rule, is into ($\otimes$) or out of ($\odot$) the page, in blue. In Fig. 31–7b, the voltage of the ac generator has reversed in direction; the current is reversed and the new magnetic field is in the opposite direction. Because the new fields have changed direction, the old lines fold back to connect up to some of the new lines and form closed loops as shown.[†] The old fields, however, don't suddenly disappear; they are on their way to distant points. Indeed, because a changing magnetic field produces an electric field, and a changing electric field produces a magnetic field, this combination of changing electric and magnetic fields moving outward is self-supporting, no longer depending on the antenna charges.

The fields not far from the antenna, referred to as the *near field*, become quite complicated, but in this book we are not so interested in them. We are instead mainly interested in the fields far from the antenna (they are generally what we detect), which we refer to as the **radiation field**, or *far field*. The electric field lines form loops, as shown in Fig. 31–8, and continue moving outward. The magnetic field lines also form closed loops, but are not shown because they are perpendicular to the page. Although the lines are shown only on the right of the source, fields also travel in other directions. The field strengths are greatest in directions perpendicular to the oscillating charges; and they drop to zero along the direction of oscillation—above and below the antenna in Fig. 31–8.

FIGURE 31–8 (a) The radiation fields (far from the antenna) produced by a sinusoidal signal on the antenna. The red closed loops represent electric field lines. The magnetic field lines, perpendicular to the page and represented by blue $\otimes$ and $\odot$, also form closed loops. (b) Very far from the antenna the wave fronts (field lines) are essentially flat over a fairly large area, and are referred to as *plane waves*.



The magnitudes of both $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ in the radiation field are found to decrease with distance as $1/r$. (Compare this to the static electric field given by Coulomb's law where $\vec{\mathbf{E}}$ decreases as $1/r^2$.) The energy carried by the electromagnetic wave is proportional (as for any wave, Chapter 15) to the square of the amplitude, E^2 or B^2, as will be discussed further in Section 31–8. So the intensity of the wave decreases as $1/r^2$. Thus energy carried by EM waves follows the **inverse square law** just as for sound waves (Eqs. 15–8).

Several things about the radiation field can be noted from Fig. 31–8. First, *the electric and magnetic fields at any point are perpendicular to each other, and perpendicular to the direction of wave travel*. Second, we can see that the fields alternate in direction ($\vec{\mathbf{B}}$ is into the page at some points and out of the page at others; $\vec{\mathbf{E}}$ points up at some points and down at others). Thus, the field strengths vary from a maximum

[†]We are considering waves traveling through empty space. There are no charges for lines of $\vec{\mathbf{E}}$ to start or stop on, so they form closed loops. Magnetic field lines always form closed loops.

in one direction, to zero, to a maximum in the other direction. The electric and magnetic fields are "in phase": that is, they each are zero at the same points and reach their maximum at the same points in space. Finally, very far from the antenna (Fig. 31–8b) the field lines are quite flat over a reasonably large area, and the waves are referred to as **plane waves**.

If the source voltage varies sinusoidally, then the electric and magnetic field strengths in the radiation field will also vary sinusoidally. The sinusoidal character of the waves is diagrammed in Fig. 31–9, which displays the field directions and magnitudes plotted as a function of position along the direction of wave travel. Notice that $\vec{\mathbf{B}}$ and $\vec{\mathbf{E}}$ are perpendicular to each other and to the direction of travel (= the direction of the wave velocity $\vec{\mathbf{v}}$). The direction of $\vec{\mathbf{v}}$ can be had from a right-hand rule using $\vec{\mathbf{E}} \times \vec{\mathbf{B}}$: fingers along $\vec{\mathbf{E}}$, then along $\vec{\mathbf{B}}$, gives $\vec{\mathbf{v}}$ along thumb.

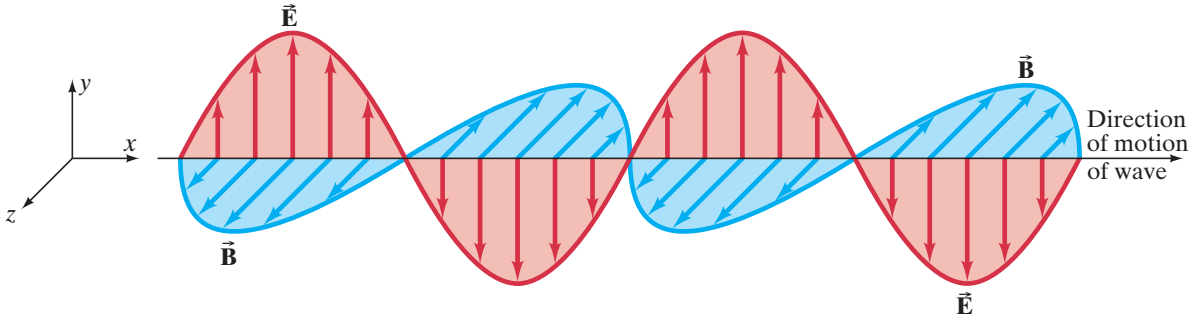


FIGURE 31–9 Electric and magnetic field strengths in an electromagnetic wave. $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ are at right angles to each other. The entire pattern moves in a direction perpendicular to both $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$.

We call these waves electromagnetic (EM) waves. They are *transverse* waves because the amplitude is perpendicular to the direction of wave travel. However, EM waves are always waves of *fields*, not of matter (like waves on water or a rope). Because they are fields, EM waves can propagate in empty space.

As we have seen, EM waves are produced by electric charges that are oscillating, and hence are undergoing *acceleration*. In fact, we can say in general that

accelerating electric charges give rise to electromagnetic waves.

Electromagnetic waves can be produced in other ways as well, requiring description at the atomic and nuclear levels, as we will discuss later.

EXERCISE A At a particular instant in time, a wave has its electric field pointing north and its magnetic field pointing vertically up. In which direction is the wave traveling? (*a*) South, (*b*) west, (*c*) east, (*d*) down, (*e*) not enough information.

31–5 Electromagnetic Waves, and Their Speed, Derived from Maxwell's Equations

Let us now examine how the existence of EM waves follows from Maxwell's equations. We will see that Maxwell's prediction of the existence of EM waves was startling. Equally startling was the speed at which they were predicted to travel.

We begin by considering a region of free space, where there are *no charges or conduction currents*—that is, far from the source so that the wave fronts (the field lines in Fig. 31–8b) are essentially flat over a reasonable area. We call them **plane waves**, as we saw, because at any instant $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ are uniform over a reasonably large plane perpendicular to the direction of propagation. We choose a coordinate system, so that the wave is traveling in the x direction with velocity $\vec{\mathbf{v}} = v\hat{\mathbf{i}}$, with $\vec{\mathbf{E}}$ parallel to the y axis and $\vec{\mathbf{B}}$ parallel to the z axis, as in Fig. 31–9.

Maxwell's equations in empty space, with $Q = I = 0$, become

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = 0 \qquad \textbf{(31–6a)}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0 \qquad \textbf{(31–6b)}$$

$$\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{d\Phi_B}{dt} \qquad \textbf{(31–6c)}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}. \qquad \textbf{(31–6d)}$$

Notice the beautiful *symmetry* of these equations. The term on the right in the last equation ("displacement current term") is essential for this symmetry. It is also essential if electromagnetic waves are to be produced, as we will now see.

If the wave is sinusoidal with wavelength λ and frequency f, then, as we saw in Chapter 15, Section 15–4, such a traveling wave moving to the right along the x axis can be written as

$$\begin{aligned} E &= E_y = E_0 \sin(kx - \omega t) \\ B &= B_z = B_0 \sin(kx - \omega t) \end{aligned} \qquad \textbf{(31–7)}$$

where

$$k = \frac{2\pi}{\lambda}, \qquad \omega = 2\pi f, \qquad \text{and} \qquad f\lambda = \frac{\omega}{k} = v, \qquad \textbf{(31–8)}$$

with v being the speed of the wave. For a wave moving to the left, the minus sign becomes plus, $(kx + \omega t)$, as in Eq. 15–13. Although visualizing the wave as sinusoidal is helpful, we will not have to assume so in most of what follows.

Consider now a small rectangle in the plane of the electric field as shown in Fig. 31–10. This rectangle has a finite height Δy, and a very thin width which we take to be the infinitesimal distance dx. First we show that $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{v}}$ are in the orientation shown by applying Lenz's law to this rectangular loop. The changing magnetic flux through this loop is related to the electric field around the loop by Faraday's law (Maxwell's third equation, Eq. 31–6c). For the case shown, B through the loop is decreasing in time as the wave moves to the right. So the electric field must be in a direction to oppose this change. This means E must be greater on the right side of the loop than on the left, as shown (so it could produce a counterclockwise current whose magnetic field would act to oppose the decrease in Φ_B—even though there is no current). This brief argument shows that the orientations of $\vec{\mathbf{E}}$, $\vec{\mathbf{B}}$, and $\vec{\mathbf{v}}$ are in the correct relation as shown. That is, $\vec{\mathbf{v}}$ is in the direction of $\vec{\mathbf{E}} \times \vec{\mathbf{B}}$.

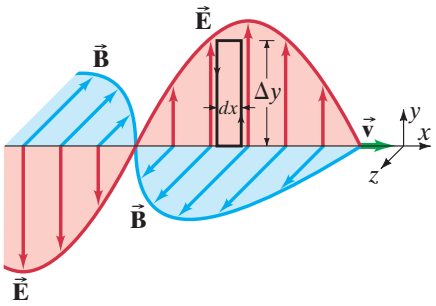


FIGURE 31–10 Applying Faraday's law to the rectangle $(\Delta y)(dx)$.

Let us again apply Faraday's law, Maxwell's third equation (Eq. 31–6c), in a more detailed way to the same rectangle of height Δy and width dx shown in Fig. 31–10. First we consider $\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}}$ in Eq. 31–6c. Along the short top and bottom sections of length dx, $\vec{\mathbf{E}}$ is perpendicular to $d\vec{\boldsymbol{\ell}}$, so $\vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = 0$. Along the vertical sides, we let E be the electric field along the left side, and on the right side where it will be slightly larger, it is $E + dE$. Thus, if we take our loop counterclockwise,

$$\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = (E + dE)\,\Delta y - E\,\Delta y = dE\,\Delta y.$$

For the right side of Faraday's law, $-d\Phi_B/dt$, the magnetic flux through the loop changes as

$$\frac{d\Phi_B}{dt} = \frac{dB}{dt} dx\,\Delta y,$$

since the area of the loop, $(dx)(\Delta y)$, is not changing. Thus, Faraday's law gives us

$$dE\,\Delta y = -\frac{dB}{dt} dx\,\Delta y$$

or

$$\frac{dE}{dx} = -\frac{dB}{dt}.$$

Actually, both E and B are functions of position x and time t. We should therefore

rewrite this equation using partial derivatives:

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t} \tag{31–9}$$

where $\partial E/\partial x$ means the derivative of E with respect to x while t is held fixed, and $\partial B/\partial t$ is the derivative of B with respect to t while x is kept fixed.

We can obtain another important relation between E and B in addition to Eq. 31–9. To do so, we consider now a small rectangle in the plane of $\vec{\mathbf{B}}$, whose length and width are Δz and dx as shown in Fig. 31–11. To this rectangular loop we apply Maxwell's fourth equation (the extension of Ampère's law) in empty space, Eq. 31–6d:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}.$$

Along the short sides (dx), the dot product $\vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}}$ is zero since $\vec{\mathbf{B}}$ is perpendicular to $d\vec{\boldsymbol{\ell}}$. Along the longer sides (Δz), we let B be the magnetic field along the left side of length Δz, and $B + dB$ be the (slightly larger) field along the right side. We again integrate counterclockwise, so

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = B\,\Delta z - (B + dB)\,\Delta z = -dB\,\Delta z.$$

The right side of Maxwell's fourth equation, Eq. 31–6d, is

$$\mu_0 \epsilon_0 \frac{d\Phi_E}{dt} = \mu_0 \epsilon_0 \frac{dE}{dt} dx\,\Delta z,$$

where the area $dx\,\Delta y$ is not changing. Equating these last two expressions, we obtain

$$-dB\,\Delta z = \mu_0 \epsilon_0 \frac{dE}{dt} dx\,\Delta z$$

or

$$\frac{\partial B}{\partial x} = -\mu_0 \epsilon_0 \frac{\partial E}{\partial t} \tag{31–10}$$

where we have replaced dB/dx and dE/dt by the proper partial derivatives as before.

We can use Eqs. 31–9 and 31–10 to obtain a relation between the magnitudes of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$, and the speed v. Let E and B be given by Eqs. 31–7 as a function of x and t. When we apply Eq. 31–9, taking the derivatives $\partial E/\partial x$ and $\partial B/\partial t$ as given by Eqs. 31–7, we obtain

$$kE_0 \cos(kx - \omega t) = \omega B_0 \cos(kx - \omega t)$$

so

$$\frac{E_0}{B_0} = \frac{\omega}{k} = v,$$

since $v = \omega/k$ (see Eq. 31–8 or 15–12). Since E and B are in phase, we see that E and B are related by

$$\frac{E}{B} = v \tag{31–11}$$

at any point in space, where v is the velocity of the wave.

Now we apply Eq. 31–10 to the sinusoidal fields (Eqs. 31–7), taking partial derivatives, and obtain

$$kB_0 \cos(kx - \omega t) = \mu_0 \epsilon_0 \omega E_0 \cos(kx - \omega t)$$

or

$$\frac{B_0}{E_0} = \frac{\mu_0 \epsilon_0 \omega}{k} = \mu_0 \epsilon_0 v.$$

We just saw that $B_0/E_0 = 1/v$, so

$$\mu_0 \epsilon_0 v = \frac{1}{v}.$$

Solving for v gives us an astonishing result:

$$v = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = c, \tag{31–12}$$

Speed of light

where c is the special symbol for the speed of electromagnetic waves in free space.

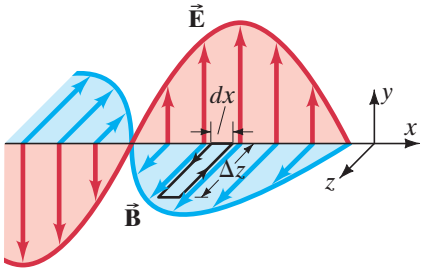


FIGURE 31–11 Applying Maxwell's fourth equation to the rectangle $(\Delta z)(dx)$.

We see that c is a constant, independent of the wavelength or frequency. If we put in values for ϵ_0 and μ_0 we find

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = \frac{1}{\sqrt{(8.85 \times 10^{-12}\,\mathrm{C^2/N \cdot m^2})(4\pi \times 10^{-7}\,\mathrm{T \cdot m/A})}}$$
$$= 3.00 \times 10^8\,\mathrm{m/s}.$$

This is a remarkable result. For this is precisely equal to the measured speed of light!

EXAMPLE 31–2 Determining $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ in EM waves. Assume a 60.0-Hz EM wave is a sinusoidal wave propagating in the z direction with $\vec{\mathbf{E}}$ pointing in the x direction, and $E_0 = 2.00\,\mathrm{V/m}$. Write vector expressions for $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ as functions of position and time.

APPROACH We find λ from $\lambda f = v = c$. Then we use Fig. 31–9 and Eqs. 31–7 and 31–8 for the mathematical form of traveling electric and magnetic fields of an EM wave.

SOLUTION The wavelength is

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8\,\mathrm{m/s}}{60.0\,\mathrm{s^{-1}}} = 5.00 \times 10^6\,\mathrm{m}.$$

From Eq. 31–8 we have

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{5.00 \times 10^6\,\mathrm{m}} = 1.26 \times 10^{-6}\,\mathrm{m^{-1}}$$
$$\omega = 2\pi f = 2\pi(60.0\,\mathrm{Hz}) = 377\,\mathrm{rad/s}.$$

From Eq. 31–11 with $v = c$, we find that

$$B_0 = \frac{E_0}{c} = \frac{2.00\,\mathrm{V/m}}{3.00 \times 10^8\,\mathrm{m/s}} = 6.67 \times 10^{-9}\,\mathrm{T}.$$

The direction of propagation is that of $\vec{\mathbf{E}} \times \vec{\mathbf{B}}$, as in Fig. 31–9. With $\vec{\mathbf{E}}$ pointing in the x direction, and the wave propagating in the $+z$ direction, $\vec{\mathbf{B}}$ must point in the y direction. Using Eqs. 31–7 we find:

$$\vec{\mathbf{E}} = \hat{\mathbf{i}}(2.00\,\mathrm{V/m})\sin\left[(1.26 \times 10^{-6}\,\mathrm{m^{-1}})z - (377\,\mathrm{rad/s})t\right]$$
$$\vec{\mathbf{B}} = \hat{\mathbf{j}}(6.67 \times 10^{-9}\,\mathrm{T})\sin\left[(1.26 \times 10^{-6}\,\mathrm{m^{-1}})z - (377\,\mathrm{rad/s})t\right]$$

*Derivation of Speed of Light (General)

We can derive the speed of EM waves without having to assume sinusoidal waves by combining Eqs. 31–9 and 31–10 as follows. We take the derivative with respect to t of Eq. 31–10

$$\frac{\partial^2 B}{\partial t\,\partial x} = -\mu_0 \epsilon_0 \frac{\partial^2 E}{\partial t^2}.$$

We next take the derivative of Eq. 31–9 with respect to x:

$$\frac{\partial^2 E}{\partial x^2} = -\frac{\partial^2 B}{\partial t\,\partial x}.$$

Since $\partial^2 B/\partial t\,\partial x$ appears in both these relations, we obtain

$$\frac{\partial^2 E}{\partial t^2} = \frac{1}{\mu_0 \epsilon_0}\frac{\partial^2 E}{\partial x^2}. \qquad \textbf{(31–13a)}$$

By taking the derivatives of Eqs. 31–9 (with respect to t) and 31–10 (with respect to x) we obtain the same relation for B:

$$\frac{\partial^2 B}{\partial t^2} = \frac{1}{\mu_0 \epsilon_0}\frac{\partial^2 B}{\partial x^2}. \qquad \textbf{(31–13b)}$$

Both of Eqs. 31–13 have the form of the **wave equation** for a plane wave traveling

in the x direction, as discussed in Section 15–5 (Eq. 15–16):

$$\frac{\partial^2 D}{\partial t^2} = v^2 \frac{\partial^2 D}{\partial x^2},$$

where D stands for any type of displacement. We see that the velocity v for Eqs. 31–13 is given by

$$v^2 = \frac{1}{\mu_0 \epsilon_0}$$

in agreement with Eq. 31–12. Thus we see that a natural outcome of Maxwell's equations is that E and B obey the wave equation for waves traveling with speed $v = 1/\sqrt{\mu_0 \epsilon_0}$. It was on this basis that Maxwell predicted the existence of electromagnetic waves and predicted their speed.

31–6 Light as an Electromagnetic Wave and the Electromagnetic Spectrum

Maxwell's prediction that EM waves should exist was startling. Equally remarkable was the speed at which EM waves were predicted to travel in empty space (Eqs. 31–11 and 31–12),

$$c = \frac{E}{B} = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3.00 \times 10^8 \text{ m/s},$$

Speed of light

the same as the measured speed of light in vacuum.

Light had been shown some 60 years before Maxwell's work to behave like a wave (we'll discuss this in Chapter 34). But nobody knew what kind of wave it was. What is it that is oscillating in a light wave? Maxwell, on the basis of the calculated speed of EM waves, argued that light must be an electromagnetic wave. This idea soon came to be generally accepted by scientists, but not fully until after EM waves were experimentally detected. EM waves were first generated and detected experimentally by Heinrich Hertz (1857–1894) in 1887, eight years after Maxwell's death. Hertz used a spark-gap apparatus in which charge was made to rush back and forth for a short time, generating waves whose frequency was about 10^9 Hz. He detected them some distance away using a loop of wire in which an emf was induced when a changing magnetic field passed through. These waves were later shown to travel at the speed of light, 3.00×10^8 m/s, and to exhibit all the characteristics of light such as reflection, refraction, and interference. The only difference was that their frequencies were not visible to the human eye. Hertz's experiment was a strong confirmation of Maxwell's theory.

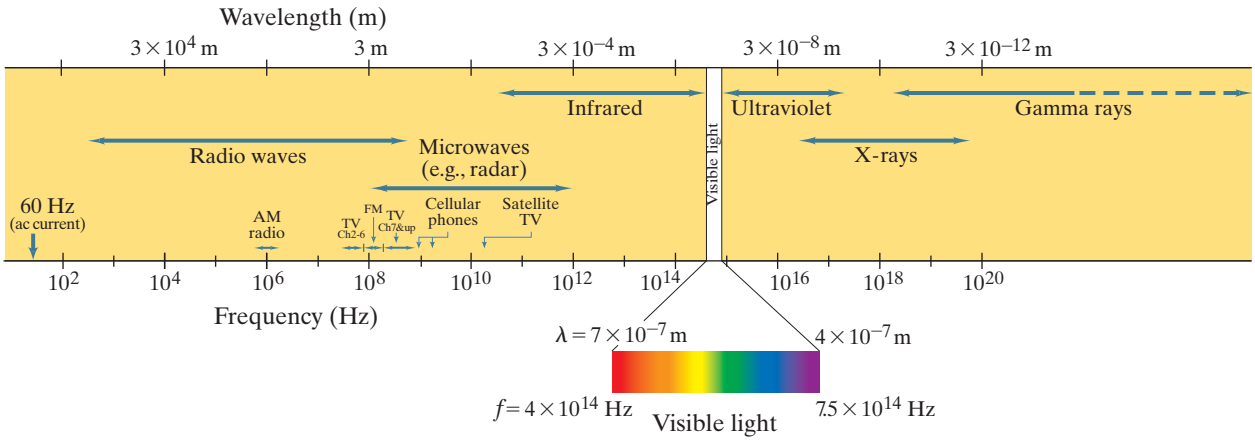
The wavelengths of visible light were measured in the first decade of the nineteenth century, long before anyone imagined that light was an electromagnetic wave. The wavelengths were found to lie between 4.0×10^{-7} m and 7.5×10^{-7} m, or 400 nm to 750 nm $(1 \text{ nm} = 10^{-9} \text{ m})$. The frequencies of visible light can be found using Eq. 15–1 or 31–8, which we rewrite here:

$$c = \lambda f, \qquad \textbf{(31–14)}$$

where f and λ are the frequency and wavelength, respectively, of the wave. Here, c is the speed of light, 3.00×10^8 m/s; it gets the special symbol c because of its universality for all EM waves in "free space" (meaning empty space = a vacuum). Equation 31–14 tells us that the frequencies of visible light are between 4.0×10^{14} Hz and 7.5×10^{14} Hz. (Recall that 1 Hz = 1 cycle per second = 1 s^{-1}.)

But visible light is only one kind of EM wave. As we have seen, Hertz produced EM waves of much lower frequency, about 10^9 Hz. These are now called **radio waves**, because frequencies in this range are used to transmit radio and TV signals. Electromagnetic waves, or EM radiation as we sometimes call it, have been produced or detected over a wide range of frequencies. They are usually categorized as shown in Fig. 31–12, which is known as the **electromagnetic spectrum**.

FIGURE 31–12 Electromagnetic spectrum.



Radio waves and microwaves can be produced in the laboratory using electronic equipment (Fig. 31–7). Higher-frequency waves are very difficult to produce electronically. These and other types of EM waves are produced in natural processes, as emission from atoms, molecules, and nuclei (more on this later). EM waves can be produced by the acceleration of electrons or other charged particles, such as electrons in the antenna of Fig. 31–7. X-rays have very short wavelengths (and very high frequencies), and they are produced (Chapter 35) when fast-moving electrons are rapidly decelerated upon striking a metal target. Even the visible light emitted by an ordinary incandescent bulb is due to electrons undergoing acceleration within the hot filament.

EXERCISE B Return to the Chapter-Opening Question, page 898, and answer it again now. Try to explain why you may have answered differently the first time.

We will meet various types of EM waves later. However, it is worth mentioning here that infrared (IR) radiation (EM waves whose frequency is just less than that of visible light) is mainly responsible for the heating effect on Earth due to the Sun. The Sun emits not only visible light but substantial amounts of IR and UV (ultraviolet) as well. The molecules of our skin tend to "resonate" at infrared frequencies, so it is these that are preferentially absorbed and thus warm us up. We humans experience EM waves differently, depending on their wavelengths: Our eyes detect wavelengths between about 4×10^{-7} m and 7.5×10^{-7} m (visible light), whereas our skin "detects" longer wavelengths (IR). Many EM wavelengths we don't detect directly at all.

CAUTION
Sound and EM waves are different

Light and other electromagnetic waves travel at a speed of 3×10^8 m/s. Compare this to sound, which travels (see Chapter 16) at a speed of about 300 m/s in air, a million times slower; or to typical freeway speeds of a car, 30 m/s (100 km/h, or 60 mi/h), 10 million times slower than light. They differ also in that sound is a longitudinal wave, whereas EM waves are transverse. Finally EM waves differ from sound waves in another big way: sound waves travel in a medium such as air, and involve motion of air molecules; EM waves do not involve any material—only fields, and they can travel in empty space.

EXAMPLE 31–3 **Wavelengths of EM waves.** Calculate the wavelength (a) of a 60-Hz EM wave, (b) of a 93.3-MHz FM radio wave, and (c) of a beam of visible red light from a laser at frequency 4.74×10^{14} Hz.

APPROACH All of these waves are electromagnetic waves, so their speed is $c = 3.00 \times 10^8$ m/s. We solve for λ in Eq. 31–14: $\lambda = c/f$.

SOLUTION (a) $$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{60 \text{ s}^{-1}} = 5.0 \times 10^6 \text{ m},$$

or 5000 km. 60 Hz is the frequency of ac current in the United States, and, as we see here, one wavelength stretches all the way across the continental USA.

(b) $$\lambda = \frac{3.00 \times 10^8 \text{ m/s}}{93.3 \times 10^6 \text{ s}^{-1}} = 3.22 \text{ m}.$$

The length of an effective FM radio antenna is often about half this $\left(\frac{1}{2}\lambda\right)$, or $1\frac{1}{2}$ m.

(c) $$\lambda = \frac{3.00 \times 10^8 \text{ m/s}}{4.74 \times 10^{14} \text{ s}^{-1}} = 6.33 \times 10^{-7} \text{ m} \ (= 633 \text{ nm}).$$

EXERCISE C What are the frequencies of (a) an 80-m-wavelength radio wave, and (b) an X-ray of wavelength 5.5×10^{-11} m?

EXAMPLE 31–4 **ESTIMATE** **Cell phone antenna.** The antenna of a cell phone is often $\frac{1}{4}$ wavelength long. A particular cell phone has an 8.5-cm-long straight rod for its antenna. Estimate the operating frequency of this phone.

APPROACH The basic equation relating wave speed, wavelength, and frequency is $c = \lambda f$. The wavelength λ equals four times the antenna's length.

SOLUTION The antenna is $\frac{1}{4}\lambda$ long, so $\lambda = 4(8.5 \text{ cm}) = 34 \text{ cm} = 0.34 \text{ m}$. Then $f = c/\lambda = (3.0 \times 10^8 \text{ m/s})/(0.34 \text{ m}) = 8.8 \times 10^8 \text{ Hz} = 880 \text{ MHz}$.

NOTE Radio antennas are not always straight conductors. The conductor may be a round loop to save space. See Fig. 31–22b.

EXERCISE D How long should a $\frac{1}{4}\lambda$ antenna be for an aircraft radio operating at 165 MHz?

Electromagnetic waves can travel along transmission lines as well as in empty space. When a source of emf is connected to a transmission line—be it two parallel wires or a coaxial cable (Figs. 31–13 and 28–12)—the electric field within the wire is not set up immediately at all points along the wires. This is based on the same argument we used in Section 31–4 with reference to Fig. 31–7. Indeed, it can be shown that if the wires are separated by empty space or air, the electrical signal travels along the wires at the speed $c = 3.0 \times 10^8$ m/s. For example, when you flip a light switch, the light actually goes on a tiny fraction of a second later. If the wires are in a medium whose electric permittivity is ϵ and magnetic permeability is μ (Sections 24–5 and 28–10, respectively), the speed is not given by Eq. 31–12, but by

$$v = \frac{1}{\sqrt{\epsilon\mu}}$$

instead.

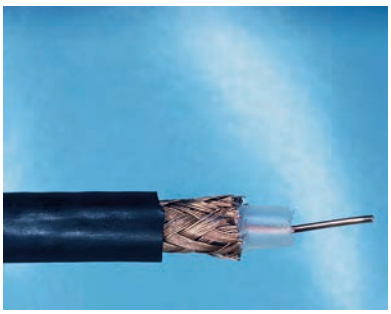


FIGURE 31–13
Coaxial cable.

EXAMPLE 31–5 **ESTIMATE** **Phone call time lag.** You make a telephone call from New York to a friend in London. Estimate how long it will take the electrical signal generated by your voice to reach London, assuming the signal is (*a*) carried on a telephone cable under the Atlantic Ocean, and (*b*) sent via satellite 36,000 km above the ocean. Would there be a noticeable delay in either case?

APPROACH The signal is carried (*a*) on a telephone cable [wire or optical fiber (Section 32–7)], or (*b*) in the air via satellite. In either case it is an electromagnetic wave. Electronics as well as the cable slow things down, but as a rough estimate we take the speed to be $c = 3.0 \times 10^8\,\text{m/s}$.

SOLUTION The distance from New York to London is about 5000 km.

(*a*) The time delay via the cable is $t = d/c \approx (5 \times 10^6\,\text{m})/(3.0 \times 10^8\,\text{m/s}) = 0.017\,\text{s}$.

(*b*) Via satellite the time would be longer because communications satellites, which are usually geosynchronous (Example 6–6), move at a height of 36,000 km. The signal would have to go up to the satellite and back down, or 72,000 km, but the actual distance the signal would travel would be a little more than this as the signal would go up and down on a diagonal (5000 km New York to London, small compared to the distance up to the satellite). Thus $t = d/c \approx (7.2 \times 10^7\,\text{m})/(3 \times 10^8\,\text{m/s}) \approx 0.24\,\text{s}$, one way. Both directions $\approx \frac{1}{2}\,\text{s}$.

NOTE When the signal travels via underwater cable, which nearly all intercontinental calls are today, there is barely a hint of a delay and conversations are normal. If the phone call is sent via satellite, the delay would be noticeable. The length of time between the end of when you speak and your friend receives it and replies, and then you hear the reply, would be about a half second beyond the normal time in a conversation, as we just calculated. This would be enough to be noticeable, and you would have to adjust for it so you don't start talking again while your friend's reply is on the way back to you.

EXERCISE E If you are on the phone via satellite to someone only 100 km away, would you notice the same effect discussed in the NOTE above?

EXERCISE F If your voice traveled as a sound wave, how long would it take to go from New York to London?

31–7 Measuring the Speed of Light

Galileo attempted to measure the speed of light by trying to measure the time required for light to travel a known distance between two hilltops. He stationed an assistant on one hilltop and himself on another, and ordered the assistant to lift the cover from a lamp the instant he saw a flash from Galileo's lamp. Galileo measured the time between the flash of his lamp and when he received the light from his assistant's lamp. The time was so short that Galileo concluded it merely represented human reaction time, and that the speed of light must be extremely high.

The first successful determination that the speed of light is finite was made by the Danish astronomer Ole Roemer (1644–1710). Roemer had noted that the carefully measured orbital period of Io, a moon of Jupiter with an average period of 42.5 h, varied slightly, depending on the relative position of Earth and Jupiter. He attributed this variation in the apparent period to the change in distance between the Earth and Jupiter during one of Io's periods, and the time it took light to travel the extra distance. Roemer concluded that the speed of light—though great—is finite.

Since then a number of techniques have been used to measure the speed of light. Among the most important were those carried out by the American Albert A. Michelson (1852–1931).

Michelson used the rotating mirror apparatus diagrammed in Fig. 31–14 for a series of high-precision experiments carried out from 1880 to the 1920s. Light from a source would hit one face of a rotating eight-sided mirror. The reflected light traveled to a stationary mirror a large distance away and back again as shown. If the rotating mirror was turning at just the right rate, the returning beam of light would reflect from the face of one of the mirrors into a small telescope through which the observer looked. If the speed of rotation was only slightly different, the beam would be deflected to one side and would not be seen by the observer. From the required speed of the rotating mirror and the carefully measured distance to the stationary mirror, the speed of light could be calculated. In the 1920s, Michelson set up the rotating mirror on the top of Mt. Wilson in southern California and the stationary mirror on Mt. San Antonio (Mt. Baldy) 35 km away. He later measured the speed of light in vacuum using a long evacuated tube.

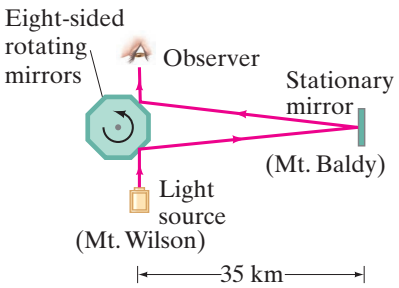


FIGURE 31–14 Michelson's speed-of-light apparatus (not to scale).

The most recent measurements of the speed of light, c, in vacuum yielded

$$c = 2.99792458 \times 10^8 \text{ m/s}.$$

In 1983, international organizations decided to *define* this value as the speed of light. (See Section 1–4, where we also saw that as a consequence, the length of the standard meter became more precise.) We usually round off c to

$$c = 3.00 \times 10^8 \text{ m/s}$$

when extremely precise results are not required. In air, the speed is only slightly less.

31–8 Energy in EM Waves; the Poynting Vector

Electromagnetic waves carry energy from one region of space to another. This energy is associated with the moving electric and magnetic fields. In Section 24–4, we saw that the energy density u_E (J/m^3) stored in an electric field E is $u_E = \frac{1}{2}\epsilon_0 E^2$ (Eq. 24–6). The energy density stored in a magnetic field B, as we discussed in Section 30–3, is given by $u_B = \frac{1}{2}B^2/\mu_0$ (Eq. 30–7). Thus, the total energy stored per unit volume in a region of space where there is an electromagnetic wave is

$$u = u_E + u_B = \frac{1}{2}\epsilon_0 E^2 + \frac{1}{2}\frac{B^2}{\mu_0}. \qquad \textbf{(31–15)}$$

In this equation, E and B represent the electric and magnetic field strengths of the wave at any instant in a small region of space. We can write Eq. 31–15 in terms of the E field alone, using Eqs. 31–11 $(B = E/c)$ and 31–12 $\left(c = 1/\sqrt{\epsilon_0\mu_0}\right)$ to obtain

$$u = \frac{1}{2}\epsilon_0 E^2 + \frac{1}{2}\frac{\epsilon_0\mu_0 E^2}{\mu_0} = \epsilon_0 E^2. \qquad \textbf{(31–16a)}$$

Note here that the energy density associated with the B field equals that due to the E field, and each contributes half to the total energy. We can also write the energy density in terms of the B field only (using Eq. 31–11):

$$u = \epsilon_0 E^2 = \epsilon_0 c^2 B^2 = \frac{B^2}{\mu_0}, \qquad \textbf{(31–16b)}$$

or in one term containing both E and B,

$$u = \epsilon_0 E^2 = \epsilon_0 EcB = \frac{\epsilon_0 EB}{\sqrt{\epsilon_0\mu_0}}$$

or

$$u = \sqrt{\frac{\epsilon_0}{\mu_0}}\,EB. \qquad \textbf{(31–16c)}$$

Equations 31–16 give the energy density of EM waves in any region of space at any instant.

Now let us determine the energy the wave transports per unit time per unit area. This is given by a vector $\vec{\mathbf{S}}$, which is called the **Poynting vector**.[†] The units of $\vec{\mathbf{S}}$ are $\mathrm{W/m^2}$. The direction of $\vec{\mathbf{S}}$ is the direction in which the energy is transported, which is the direction in which the wave is moving.

Let us imagine the wave is passing through an area A perpendicular to the x axis as shown in Fig. 31–15. In a short time dt, the wave moves to the right a distance $dx = c\,dt$ where c is the wave speed. The energy that passes through A in the time dt is the energy that occupies the volume $dV = A\,dx = Ac\,dt$. The energy density u is $u = \epsilon_0 E^2$ where E is the electric field in this volume at the given instant.

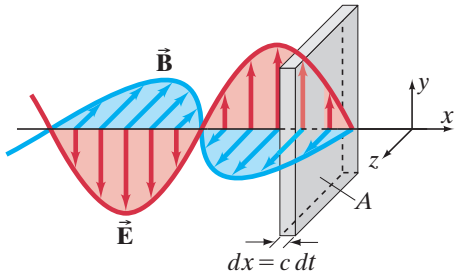


FIGURE 31–15 Electromagnetic waves carrying energy through area A.

So the total energy dU contained in this volume dV is the energy density u times the volume: $dU = u\,dV = (\epsilon_0 E^2)(Ac\,dt)$. Therefore S, the energy crossing the area A per time dt, is

$$S = \frac{1}{A}\frac{dU}{dt} = \epsilon_0 cE^2. \qquad \textbf{(31–17)}$$

Since $E = cB$ and $c = 1/\sqrt{\epsilon_0\mu_0}$, this can also be written:

$$S = \epsilon_0 cE^2 = \frac{cB^2}{\mu_0} = \frac{EB}{\mu_0}.$$

The direction of $\vec{\mathbf{S}}$ is along $\vec{\mathbf{v}}$, perpendicular to $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$, so the Poynting vector $\vec{\mathbf{S}}$ can be written

$$\vec{\mathbf{S}} = \frac{1}{\mu_0}(\vec{\mathbf{E}} \times \vec{\mathbf{B}}). \qquad \textbf{(31–18)}$$

Equation 31–17 or 31–18 gives the energy transported per unit area per unit time at any *instant*. We often want to know the *average* over an extended period of time since the frequencies are usually so high we don't detect the rapid time variation. If E and B are sinusoidal, then $\overline{E^2} = E_0^2/2$, just as for electric currents and voltages (Section 25–7), where E_0 is the *maximum* value of E. Thus we can write for the magnitude of the Poynting vector, on the average,

$$\overline{S} = \frac{1}{2}\epsilon_0 cE_0^2 = \frac{1}{2}\frac{c}{\mu_0}B_0^2 = \frac{E_0 B_0}{2\mu_0}, \qquad \textbf{(31–19a)}$$

where B_0 is the maximum value of B. This time averaged value of $\vec{\mathbf{S}}$ is the **intensity**, defined as the average power transferred across unit area (Section 15–3). We can also write for the average value of S:

$$\overline{S} = \frac{E_{\mathrm{rms}} B_{\mathrm{rms}}}{\mu_0} \qquad \textbf{(31–19b)}$$

where E_{rms} and B_{rms} are the rms values $\left(E_{\mathrm{rms}} = \sqrt{\overline{E^2}} = E_0/\sqrt{2}\right.$ and $\left.B_{\mathrm{rms}} = \sqrt{\overline{E^2}} = B_0/\sqrt{2}\right)$.

EXAMPLE 31–6 ***E* and *B* from the Sun.** Radiation from the Sun reaches the Earth (above the atmosphere) with an intensity of about $1350\ \mathrm{J/s \cdot m^2} = 1350\ \mathrm{W/m^2}$. Assume that this is a single EM wave, and calculate the maximum values of E and B.

APPROACH We solve Eq. 31–19a $\left(\overline{S} = \frac{1}{2}\epsilon_0 cE_0^2\right)$ for E_0 in terms of $\overline{S}$ and use $\overline{S} = 1350\ \mathrm{J/s \cdot m^2}$.

[†]After J. H. Poynting (1852–1914).

SOLUTION $E_0 = \sqrt{\frac{2\bar{S}}{\epsilon_0 c}} = \sqrt{\frac{2(1350\,\text{J/s}\cdot\text{m}^2)}{(8.85\times10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)(3.00\times10^8\,\text{m/s})}}$

$= 1.01\times10^3\,\text{V/m}.$

From Eq. 31–11, $B = E/c$, so

$$B_0 = \frac{E_0}{c} = \frac{1.01\times10^3\,\text{V/m}}{3.00\times10^8\,\text{m/s}} = 3.37\times10^{-6}\,\text{T}.$$

NOTE Although B has a small numerical value compared to E (because of the way the different units for E and B are defined), B contributes the same energy to the wave as E does, as we saw earlier (Eqs. 31–15 and 16).

CAUTION

E and B have very different values (due to how units are defined), but E and B contribute equal *energy*

31–9 Radiation Pressure

If electromagnetic waves carry energy, then we would expect them to also carry linear momentum. When an electromagnetic wave encounters the surface of an object, a force will be exerted on the surface as a result of the momentum transfer ($F = dp/dt$), just as when a moving object strikes a surface. The force per unit area exerted by the waves is called **radiation pressure**, and its existence was predicted by Maxwell. He showed that if a beam of EM radiation (light, for example) is completely absorbed by an object, then the momentum transferred is

$$\Delta p = \frac{\Delta U}{c} \qquad \begin{bmatrix}\text{radiation}\\ \text{fully}\\ \text{absorbed}\end{bmatrix} \qquad \textbf{(31–20a)}$$

where ΔU is the energy absorbed by the object in a time Δt, and c is the speed of light.† If instead, the radiation is fully reflected (suppose the object is a mirror), then the momentum transferred is twice as great, just as when a ball bounces elastically off a surface:

$$\Delta p = \frac{2\,\Delta U}{c}. \qquad \begin{bmatrix}\text{radiation}\\ \text{fully}\\ \text{reflected}\end{bmatrix} \qquad \textbf{(31–20b)}$$

If a surface absorbs some of the energy, and reflects some of it, then $\Delta p = a\,\Delta U/c$, where a has a value between 1 and 2.

Using Newton's second law we can calculate the force and the pressure P exerted by radiation on an object from the change in momentum p. The force F is given by

$$F = \frac{dp}{dt}.$$

The average rate that energy is delivered to the object is related to the Poynting vector by (Eq. 31–17)

$$\frac{dU}{dt} = \bar{S}A,$$

where A is the cross-sectional area of the object which intercepts the radiation. The radiation pressure P (assuming full absorption) is given by (see Eq. 31–20a)

$$P = \frac{F}{A} = \frac{1}{A}\frac{dp}{dt} = \frac{1}{Ac}\frac{dU}{dt} = \frac{\bar{S}}{c}. \qquad \begin{bmatrix}\text{fully}\\ \text{absorbed}\end{bmatrix} \qquad \textbf{(31–21a)}$$

If the light is fully reflected, the pressure is twice as great (Eq. 31–20b):

$$P = \frac{2\bar{S}}{c}. \qquad \begin{bmatrix}\text{fully}\\ \text{reflected}\end{bmatrix} \qquad \textbf{(31–21b)}$$

†Very roughly, if we think of light as particles (and we do—see Chapter 37), the force that would be needed to bring such a particle moving at speed c to "rest" (i.e., absorption) is $F = \Delta p/\Delta t$. But F is also related to energy by Eq. 8–7, $F = \Delta U/\Delta x$, so $\Delta p = F\,\Delta t = \Delta U/(\Delta x/\Delta t) = \Delta U/c$ where we identify $(\Delta x/\Delta t)$ as the speed of light c.

EXAMPLE 31–7 **ESTIMATE** **Solar pressure.** Radiation from the Sun that reaches the Earth's surface (after passing through the atmosphere) transports energy at a rate of about $1000\,\text{W/m}^2$. Estimate the pressure and force exerted by the Sun on your outstretched hand.

APPROACH The radiation is partially reflected and partially absorbed, so let us estimate simply $P = \bar{S}/c$.

SOLUTION $$P \approx \frac{\bar{S}}{c} = \frac{1000\,\text{W/m}^2}{3\times10^8\,\text{m/s}} \approx 3\times10^{-6}\,\text{N/m}^2.$$

An estimate of the area of your outstretched hand might be about 10 cm by 20 cm, so $A = 0.02\,\text{m}^2$. Then the force is

$$F = PA \approx (3\times10^{-6}\,\text{N/m}^2)(0.02\,\text{m}^2) \approx 6\times10^{-8}\,\text{N}.$$

NOTE These numbers are tiny. The force of gravity on your hand, for comparison, is maybe a half pound, or with $m = 0.2\,\text{kg}$, $mg \approx (0.2\,\text{kg})(9.8\,\text{m/s}^2) \approx 2\,\text{N}$. The radiation pressure on your hand is imperceptible compared to gravity.

EXAMPLE 31–8 **ESTIMATE** **A solar sail.** Proposals have been made to use the radiation pressure from the Sun to help propel spacecraft around the solar system. (*a*) About how much force would be applied on a $1\,\text{km}\times1\,\text{km}$ highly reflective sail when about the same distance from the Sun as the Earth is? (*b*) By how much would this increase the speed of a 5000-kg spacecraft in one year? (*c*) If the spacecraft started from rest, about how far would it travel in a year?

APPROACH (*a*) Pressure P is force per unit area, so $F = PA$. We use the estimate of Example 31–7, doubling it for a reflecting surface $P = 2\bar{S}/c$. (*b*) We find the acceleration from Newton's second law, and assume it is constant, and then find the speed from $v = v_0 + at$. (*c*) The distance traveled is given by $x = \frac{1}{2}at^2$.

SOLUTION (*a*) Doubling the result of Example 31–7, we get a solar pressure that is about $2\bar{S}/c \approx 10^{-5}\,\text{N/m}^2$, rounding off. Then the force is $F \approx PA = (10^{-5}\,\text{N/m}^2)(10^3\,\text{m})(10^3\,\text{m}) \approx 10\,\text{N}$.

(*b*) The acceleration is $a \approx F/m \approx (10\,\text{N})/(5000\,\text{kg}) \approx 2\times10^{-3}\,\text{m/s}^2$. One year has $(365\,\text{days})(24\,\text{h/day})(3600\,\text{s/h}) \approx 3\times10^7\,\text{s}$. The speed increase is $v - v_0 = at \approx (2\times10^{-3}\,\text{m/s}^2)(3\times10^7\,\text{s}) \approx 6\times10^4\,\text{m/s}$ ($\approx 200{,}000\,\text{km/h!}$).

(*c*) Starting from rest, this acceleration would result in a distance traveled of about $d = \frac{1}{2}at^2 \approx \frac{1}{2}(2\times10^{-3}\,\text{m/s}^2)(3\times10^7\,\text{s})^2 \approx 10^{12}\,\text{m}$ in a year, about seven times the Sun–Earth distance. This result would apply if the spacecraft was far from the Earth so the Earth's gravitational force is small compared to 10 N. Also gravity due to the Sun and planets has been ignored, but in reality would have to be considered.

NOTE A large sail providing a small force over a long time could result in a lot of motion.

PHYSICS APPLIED
Optical tweezers (move cell parts, DNA elasticity)

Although you cannot directly feel the effects of radiation pressure, the phenomenon is quite dramatic when applied to atoms irradiated by a finely focused laser beam. An atom has a mass on the order of $10^{-27}\,\text{kg}$, and a laser beam can deliver energy at a rate of $1000\,\text{W/m}^2$. This is the same intensity used in Example 31–7, but here a radiation pressure of $10^{-6}\,\text{N/m}^2$ would be very significant on a molecule whose mass might be 10^{-26} to $10^{-23}\,\text{kg}$. It is possible to move atoms and molecules around by steering them with a laser beam, in a device called **optical tweezers**. Optical tweezers have some remarkable applications. They are of great interest to biologists, especially since optical tweezers can manipulate live microorganisms, and components within a cell, without damaging them. Optical tweezers have been used to measure the elastic properties of DNA by pulling each end of the molecule with such a laser "tweezers."

31–10 Radio and Television; Wireless Communication

PHYSICS APPLIED
Wireless transmission

Electromagnetic waves offer the possibility of transmitting information over long distances. Among the first to realize this and put it into practice was Guglielmo Marconi (1874–1937) who, in the 1890s, invented and developed wireless communication. With it, messages could be sent at the speed of light without the use of wires. The first signals were merely long and short pulses that could be translated into words by a code, such as the "dots" and "dashes" of Morse code (Section 23–9): they were digital wireless, believe it or not. In 1895 Marconi sent wireless signals a kilometer or two in Italy. By 1901 he had sent test signals 3000 km across the ocean from Newfoundland, Canada, to Cornwall, England (Fig. 31–16). In 1903 he sent the first practical commercial messages from Cape Cod, Massachusetts, to England: the London *Times* printed news items sent from its New York correspondent. (1903 was also the year of the first powered airplane flight by the Wright brothers. The hallmarks of the modern age—wireless communication and flight—date from the same year.)

Our modern world of wireless communication, including radio, television, cordless phones, cell phones, Bluetooth, wi-fi, and satellite communication, are based on Marconi's pioneering work.

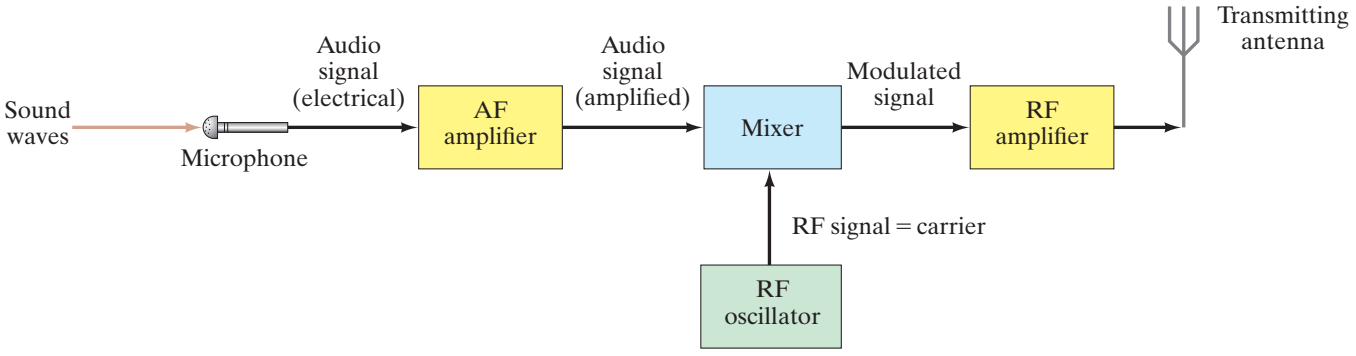
The early twentieth century also saw the development of vacuum tubes. Out of this early work radio and television were born. We now discuss briefly (1) how radio and TV signals are transmitted, and (2) how they are received at home.

The process by which a radio station transmits information (words and music) is outlined in Fig. 31–17. The audio (sound) information is changed into an electrical signal of the same frequencies by, say, a microphone, or an electronic reader (laser for CDs and DVDs). This electrical signal is called an audio frequency (AF) signal, because the frequencies are in the audio range (20 to 20,000 Hz). The signal is amplified electronically and is then mixed with a radio-frequency (RF) signal which is its **carrier frequency**. Each station has its own carrier frequency, and we use it to tune into different stations. AM radio stations have carrier frequencies from about 530 kHz to 1700 kHz. For example, "710 on your dial" means a station whose carrier frequency is 710 kHz. FM radio stations have much higher carrier frequencies, between 88 MHz and 108 MHz. The carrier frequencies for broadcast TV stations in the United States lie between 54 MHz and 72 MHz, between 76 MHz and 88 MHz, between 174 MHz and 216 MHz, and between 470 MHz and 608 MHz.[†] Today's digital broadcasting (see Sections 23–9 and 23–10) uses the same frequencies as the pre-2009 analog transmission.



FIGURE 31–16 Guglielmo Marconi (1874–1937), on the left, receiving signals in Cornwall, 1901.

FIGURE 31–17 Block diagram of a radio transmitter.



[†]The band 608-614 MHz is reserved for radio astronomy. All TV stations have a bandwidth of 6 MHz.

PHYSICS APPLIED
AM and FM

The mixing of the audio and carrier frequencies is done in two ways. In **amplitude modulation** (AM), the amplitude of the high-frequency carrier wave is made to vary in proportion to the amplitude of the audio signal, as shown in Fig. 31–18. It is called "amplitude modulation" because the *amplitude* of the carrier is altered ("modulate" means to change or alter). In **frequency modulation** (FM), the *frequency* of the carrier wave is made to change in proportion to the audio signal's amplitude, as shown in Fig. 31–19. In both cases, the mixed signal is amplified further and sent to the transmitting antenna (Fig. 31–17), where the complex mixture of frequencies is sent out in the form of EM waves. In digital communication, the signal is put into a digital form (Section 23–9) which modulates the carrier.

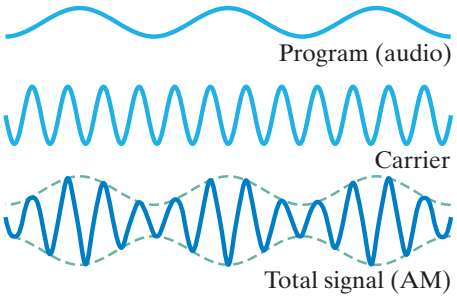


FIGURE 31–18 In amplitude modulation (AM), the amplitude of the carrier signal is made to vary in proportion to the audio signal's amplitude.

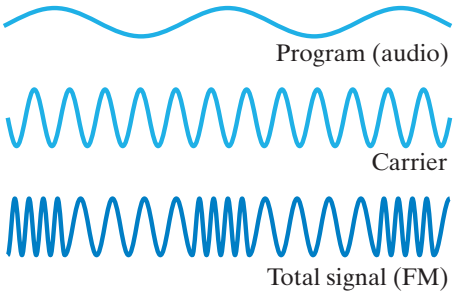


FIGURE 31–19 In frequency modulation (FM), the frequency of the carrier signal is made to change in proportion to the audio signal's amplitude. This method is used by FM radio and television.

Digital television transmission (Sections 23–9 and 23–10) works in a similar way, usually using a form of AM mixed with carrier frequencies.

PHYSICS APPLIED
Radio and TV receivers

Now let us look at the other end of the process, the reception of radio and TV programs at home. A simple radio receiver is diagrammed in Fig. 31–20.

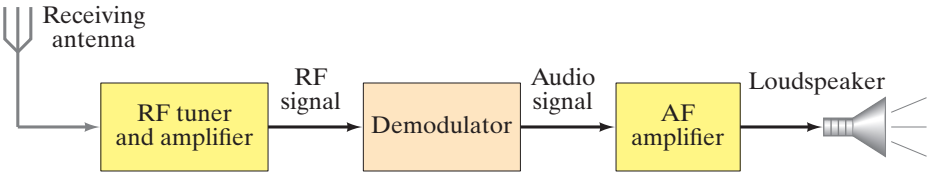
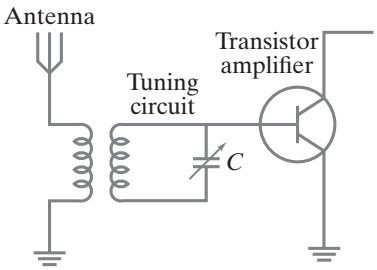


FIGURE 31–20 Block diagram of a simple radio receiver.

The EM waves sent out by all stations are received by the antenna. The signals the antenna detects and sends to the receiver are very small and contain frequencies from many different stations. The receiver uses a resonant LC circuit (Section 30–9) to select out a particular RF frequency (actually a narrow range of frequencies) corresponding to a particular station. A simple way of tuning a station is shown in Fig. 31–21. A particular station is "tuned in" by adjusting C and/or L so that the resonant frequency of the circuit $f_0 = 1/(2\pi\sqrt{LC})$, Eq. 30–32b, equals that of the station's carrier frequency. The signal, containing both audio and carrier frequencies, next goes to the *demodulator*, or *detector* (Fig. 31–19), where "demodulation" takes place—that is, the audio signal is separated from the RF carrier frequency. The audio signal is amplified and sent to a loudspeaker or headphones.

FIGURE 31–21 Simple tuning stage of a radio.



Modern receivers have more stages than those shown. Various means are used to increase the sensitivity and selectivity (ability to detect weak signals and distinguish them from other stations), and to minimize distortion of the original signal.[†]

A television receiver does similar things to both the audio and the video signals. The audio signal is amplified and goes to the loudspeaker, and the video signal goes finally to the monitor screen, such as an LCD (Sections 23–10 and 34–8).

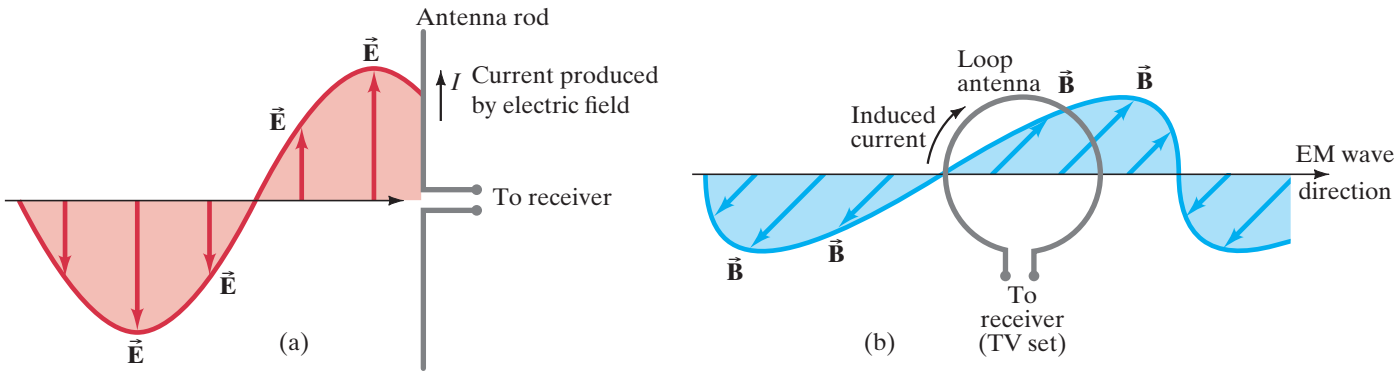


FIGURE 31–22 Receiving antennas. (a) Electric field of EM wave produces a current in an antenna consisting of straight wire or rods. (b) The moving and changing magnetic field induces an emf and current in a loop antenna.

One kind of receiving **antenna** consists of one or more conducting rods; the electric field in the EM waves exerts a force on the electrons in the conductor, causing them to move back and forth at the frequencies of the waves (Fig. 31–22a). A second type of antenna consists of a coil of wire which detects the magnetic field of the wave: the changing B field induces an emf in the coil (Fig. 31–22b). A satellite dish (Fig. 31–23) consists of a parabolic reflector that focuses the EM waves onto a small antenna (sometimes called a "horn"), similar to a concave mirror telescope.

FIGURE 31–23 A satellite dish.



EXAMPLE 31–9 **Tuning a station.** Calculate the transmitting wavelength of an FM radio station that transmits at 100.1 MHz.

APPROACH Radio is transmitted as an EM wave, so the speed is $c = 3.0 \times 10^8$ m/s. The wavelength is found from Eq. 31–14, $\lambda = c/f$.

SOLUTION The carrier frequency is $f = 100.1\text{ MHz} \approx 1.0 \times 10^8\text{ s}^{-1}$, so

$$\lambda = \frac{c}{f} = \frac{(3.0 \times 10^8\text{ m/s})}{(1.0 \times 10^8\text{ s}^{-1})} = 3.0\text{ m}.$$

NOTE The wavelengths of other FM signals (88 MHz to 108 MHz) are close to the 3.0-m wavelength of this station. FM antennas are typically 1.5 m long, or about a half wavelength. This length is chosen so that the antenna reacts in a resonant fashion and thus is more sensitive to FM frequencies. AM radio antennas would have to be very long and impractical to be either $\frac{1}{2}\lambda$ or $\frac{1}{4}\lambda$.

[†]For *FM stereo broadcasting*, two signals are carried by the carrier wave. One signal contains frequencies up to about 15 kHz, which includes most audio frequencies. The other signal includes the same range of frequencies, but 19 kHz is added to it. A stereo receiver subtracts this 19,000-Hz signal and distributes the two signals to the left and right channels. The first signal consists of the sum of left and right channels (L + R), so monophonic radios (one speaker) detect all the sound. The second signal is the difference between left and right (L − R). Hence a stereo receiver must add and subtract the two signals to get pure left and right signals for each channel.

Other EM Wave Communications

PHYSICS APPLIED
Cell phones, remote controls, cable TV, satellite TV and radio

The various regions of the radio-wave spectrum are assigned by governmental agencies for various purposes. Besides radio and television, there are "bands" assigned for use by ships, airplanes, police, military, amateurs, satellites and space, and radar. Cell phones (also called mobile phones), for example, are complete radio transmitters and receivers. In the U.S., the first cell phone technology was analog. The first digital service in the U.S. was CDMA, functioning on bands in the vicinity of 800 MHz and 1900 MHz (=1.9 GHz). Later, some U.S. providers followed Europe, Asia, and much of the rest of the world in using an international standard called GSM (Global System for Mobile Communication), mainly on 900-MHz and 1800-MHz bands. GSM phones require a SIM card. A 700-MHz band that used to carry TV broadcast channels (no longer used) is now available for cell phones. More recent technology, like 4G LTE and 5G use a SIM card, and use the same bands plus newly assigned bands around 24 GHz, 28 GHz, 39 GHz, and others, plus in the 600 MHz region where other TV broadcast channels were. Radio-controlled toys (cars, sailboats, robotic animals, etc.) can use various frequencies from 27 MHz to 75 MHz. Automobile remote entry (keyless) may operate around 300 MHz or 400 MHz.

Cable TV channels are carried as electromagnetic waves along a coaxial cable (Figs. 28–12 and 31–13) rather than being broadcast and received through the "air." The channels are in the same part of the EM spectrum, hundreds of MHz, but some are at frequencies not available for TV broadcast. Digital satellite TV and radio are carried in the microwave portion of the spectrum (12 to 18 GHz and 2.3 GHz, respectively).



FIGURE 31–24 The first person on the Moon, Neil Armstrong, July 20, 1969, pointed out "One small step for a man, one giant leap for mankind."

Wireless from the Moon

In 1969, astronauts first landed on the Moon. It was shown live on television (Fig. 31–24). The transmitting TV camera can be seen in the Chapter-Opening photo, page 898. At that time, someone pointed out that Columbus and other early navigators could have imagined that humans might one day reach the Moon. But they would never have believed possible that moving images could be sent from the Moon to the Earth through empty space.

Summary

James Clerk Maxwell synthesized an elegant theory in which all electric and magnetic phenomena could be described using four equations, now called **Maxwell's equations**. They are based on earlier ideas, but Maxwell added one more—that a changing electric field produces a magnetic field. Maxwell's equations are

$$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q}{\epsilon_0} \qquad \textbf{(31–5a)}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0 \qquad \textbf{(31–5b)}$$

$$\oint \vec{\mathbf{E}} \cdot d\vec{\ell} = -\frac{d\Phi_B}{dt} \qquad \textbf{(31–5c)}$$

$$\oint \vec{\mathbf{B}} \cdot d\vec{\ell} = \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}. \qquad \textbf{(31–5d)}$$

The first two are Gauss's laws for electricity and for magnetism; the other two are Faraday's law and Ampère's law (as extended by Maxwell).

Maxwell's theory predicted that transverse **electromagnetic (EM) waves** would be produced by accelerating electric charges, and these waves would propagate (move) through space at the speed of light:

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3.00 \times 10^8 \,\text{m/s}. \qquad \textbf{(31–12)}$$

The wavelength λ and frequency f of EM waves are related to their speed c by

$$c = \lambda f, \qquad \textbf{(31–14)}$$

just as for other waves.

The oscillating electric and magnetic fields in an EM wave are perpendicular to each other and to the direction of propagation. These EM waves are waves of fields, not matter, and can propagate in empty space.

After EM waves were experimentally detected it became generally accepted that light is an EM wave. The **electromagnetic spectrum** includes EM waves of a wide variety of wavelengths, from microwaves and radio waves to visible light to X-rays and gamma rays, all of which travel through space at a speed $c = 3.00 \times 10^8$ m/s.

The energy carried by EM waves can be described by the **Poynting vector**

$$\vec{\mathbf{S}} = \frac{1}{\mu_0} \vec{\mathbf{E}} \times \vec{\mathbf{B}} \qquad \textbf{(31–18)}$$

which gives the rate energy is carried across unit area per unit time when the electric and magnetic fields in an EM wave in free space are $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$.

EM waves carry momentum and exert a **radiation pressure** proportional to the intensity S of the wave.

Radio, TV, cell phone, and other wireless signals are transmitted through space in the radio-wave or microwave part of the EM spectrum.

Questions

1. An electric field $\vec{\mathbf{E}}$ points away from you, and its magnitude is increasing. Will the induced magnetic field be clockwise or counterclockwise? What if $\vec{\mathbf{E}}$ points toward you and is decreasing?
2. What is the direction of the displacement current in Fig. 31–3? (*Note*: The capacitor is discharging.)
3. Why is it that the magnetic field of a displacement current in a capacitor is so much harder to detect than the magnetic field of a conduction current?
4. Are there any good reasons for calling the term $\mu_0 \epsilon_0 d\Phi_E/dt$ in Eq. 31–1 as being due to a "current"? Explain.
5. The electric field in an EM wave traveling north oscillates in an east–west plane. Describe the direction of the magnetic field vector in this wave. Explain.
6. Is sound an electromagnetic wave? If not, what kind of wave is it?
7. Can EM waves travel through a perfect vacuum? Can sound waves?
8. When you flip a light switch on, does the light go on immediately? Explain.
9. Are the wavelengths of radio and television signals longer or shorter than those detectable by the human eye?
10. What does the wavelength calculated in Example 31–2 tell you about the phase of a 60-Hz ac current that starts at a power plant as compared to its phase at a house 200 km away?
11. When you connect two loudspeakers to the output of an audio system, should you be sure the lead-in wires are equal in length to avoid a time lag between speakers? Explain.
12. In the electromagnetic spectrum, what type of EM wave would have a wavelength of 10^3 km; 1 km; 1 m; 1 cm; 1 mm; 1 μm?
13. Can radio waves have the same frequencies as sound waves (20 Hz–20,000 Hz)?
14. Discuss how cordless telephones make use of EM waves. What about cell phones?
15. If a radio transmitter has a vertical antenna, should a receiver's antenna (rod type) be vertical or horizontal to obtain best reception?
16. The carrier frequencies of FM broadcasts are much higher than for AM broadcasts. On the basis of what you learned about diffraction in Chapter 15, explain why AM signals can be detected more readily than FM signals behind low hills or buildings.
17. A lost person may signal by switching a flashlight on and off using Morse code. This is actually a modulated EM wave. Is it AM or FM? What is the frequency of the carrier, approximately?

MisConceptual Questions

1. The radius of an atom is on the order of 10^{-10} m. In comparison, the wavelength of visible light is
   (*a*) much smaller.
   (*b*) about the same size.
   (*c*) much larger.
2. Which of the following travel at the same speed as light?
   (*a*) Radio waves.
   (*b*) Microwaves.
   (*c*) Radar.
   (*d*) Ultrasonic waves.
   (*e*) Infrared radiation.
   (*f*) Cell phone signals.
   (*g*) Gamma rays.
   (*h*) X-rays.
3. In empty space, which quantity is always larger for X-ray radiation than for a radio wave?
   (*a*) Amplitude.
   (*b*) Wavelength.
   (*c*) Frequency.
   (*d*) Speed.
4. If electrons in a vertical wire vibrate up and down 1000 times per second, they will create an electromagnetic wave having
   (*a*) a wavelength of 1000 m.
   (*b*) a frequency of 1000 Hz.
   (*c*) a speed of 1000 m/s.
   (*d*) an amplitude of 1000 m.
5. If the Earth–Sun distance were doubled, the intensity of radiation from the Sun that reaches the Earth's surface would
   (*a*) quadruple.
   (*b*) double.
   (*c*) drop to $\frac{1}{2}$.
   (*d*) drop to $\frac{1}{4}$.
6. An electromagnetic wave is traveling vertically downward. At a certain moment the electric field points west. In which direction does the magnetic field point at this instant?
   (*a*) North.
   (*b*) South.
   (*c*) East.
   (*d*) West.
   (*e*) Up.
   (*f*) Down.
   (*g*) Either (*a*) or (*b*).
   (*h*) Either (*c*) or (*d*).
   (*i*) Either (*e*) or (*f*).
7. If the intensity of an electromagnetic wave doubles,
   (*a*) the electric field must also double.
   (*b*) the magnetic field must also double.
   (*c*) both the magnetic field and the electric field must increase by a factor of $\sqrt{2}$.
   (*d*) Any of the above.

8. Maxwell's four equations can be used as a starting point to find
   (*a*) the electric field produced by a stationary point charge.
   (*b*) the electric field produced by a moving point charge.
   (*c*) the gravitational field of a charged object.
   (*d*) electromagnetic waves produced by charges and currents.
   (*e*) the magnetic field produced by a current carrying wire.
9. Which will produce an EM wave?
   (*a*) A steady electric current.
   (*b*) An alternating current.
   (*c*) A proton in uniform circular motion.
   (*d*) None of the above.
10. If all else is the same, for which surface would the radiation pressure from light be the greatest?
   (*a*) A black surface.
   (*b*) A gray surface.
   (*c*) A yellow surface.
   (*d*) A white surface.
   (*e*) All experience the same radiation pressure, because they are exposed to the same light.
11. Starting in 2009, TV stations in the U.S. switched to digital signals. [See Sections 31–10, 23–9, 23–10.] To watch today's digital broadcast TV, could you use a pre-2009 TV antenna meant for analog? Explain.
   (*a*) No; analog antennas do not receive digital signals.
   (*b*) No; digital signals are broadcast at different frequencies, so you need a different antenna.
   (*c*) Yes; digital signals are broadcast with the same carrier frequencies, so your old antenna will be fine.
   (*d*) No; you cannot receive digital signals through an antenna and need to switch to cable or satellite.
12. Which of the following do not produce magnetic fields?
   (*a*) Stationary objects that have electric charge.
   (*b*) Magnets.
   (*c*) Steady electric currents.
   (*d*) Constant electric fields.
   (*e*) Changing electric fields.

Problems

31–1 $\vec{B}$ Produced by Changing $\vec{E}$

1. (I) Determine the rate at which the electric field changes near the center between the round plates of a capacitor, 5.0 cm in diameter, if the plates are spaced 1.1 mm apart and the voltage across them is changing at a rate of 120 V/s.
2. (I) Calculate the displacement current I_D between the square plates, 5.8 cm on a side, of a capacitor if the electric field is changing at a rate of $2.6 \times 10^6\ \text{V/m} \cdot \text{s}$.
3. (II) At a given instant, a 2.4-A current flows in the wires connected to a parallel-plate capacitor. What is the rate at which the electric field is changing between the plates if the square plates are 1.60 cm on a side?
4. (II) A 2200-nF capacitor with circular parallel plates 2.0 cm in diameter is accumulating charge at the rate of 38.0 mC/s at some instant in time. (*a*) What will be the induced magnetic field strength 10.0 cm radially outward from the center of the plates? (*b*) What will be the value of the magnetic field after the capacitor is fully charged?
5. (II) Show that the displacement current through a parallel-plate capacitor can be written $I_D = C\,dV/dt$, where V is the voltage across the capacitor at any instant.
6. (II) Suppose an air-gap capacitor has circular plates of radius $r = 2.5$ cm and separation $d = 1.6$ mm. A 68.0-Hz emf, $\mathscr{E} = \mathscr{E}_0 \cos \omega t$, is applied to the capacitor. The maximum displacement current is 35 μA. Determine (*a*) the maximum conduction current I, (*b*) the value of $\mathscr{E}_0$, (*c*) the maximum value of $d\Phi_E/dt$ between the plates. Neglect fringing.
7. (III) Suppose that a circular parallel-plate capacitor has radius $r_0 = 3.0$ cm and plate separation $d = 5.0$ mm. A sinusoidal potential difference $V = V_0 \sin(2\pi ft)$ is applied across the plates, where $V_0 = 180$ V and $f = 60$ Hz. (*a*) In the region between the plates, show that the magnitude of the induced magnetic field is given by $B = B_0(r) \cos(2\pi ft)$, where $B_0(r)$ is a function of the radial distance r from the capacitor's central axis. (*b*) Determine the expression for the amplitude $B_0(r)$ of this time-dependent (sinusoidal) field when $r \le r_0$, and when $r > r_0$. (*c*) Plot $B_0(r)$ in tesla for the range $0 \le r \le 10$ cm.

31–5 EM Waves

8. (I) If the electric field in an EM wave has a peak magnitude of 0.45×10^{-4} V/m, what is the peak magnitude of the magnetic field strength?
9. (I) If the magnetic field in a traveling EM wave has a peak magnitude of 16.5 nT, what is the peak magnitude of the electric field?
10. (I) In an EM wave traveling west, the B field oscillates up and down vertically and has a frequency of 85.0 kHz and an rms strength of 7.75×10^{-9} T. Determine the frequency and rms strength of the electric field. What is the direction of the electric field oscillations?
11. (I) How long does it take light to reach us from the Sun, 1.50×10^8 km away?
12. (II) How long should it take the voices of astronauts on the Moon to reach the Earth? Explain in detail.
13. (II) The electric field of a plane EM wave is given by $E_x = E_0 \cos(kz + \omega t)$, $E_y = E_z = 0$. Determine (*a*) the direction of propagation and (*b*) the magnitude and direction of $\vec{\mathbf{B}}$.
14. (III) Consider two possible candidates $E(x, t)$ as solutions of the wave equation for an EM wave's electric field. Let A and α be constants. Show that (*a*) $E(x, t) = Ae^{-\alpha(x-vt)^2}$ satisfies the wave equation, and that (*b*) $E(x, t) = Ae^{-(\alpha x^2 - vt)}$ does not satisfy the wave equation.

31–6 Electromagnetic Spectrum

15. (I) (*a*) What is the wavelength of a 35.75×10^9-Hz radar signal? (*b*) What is the frequency of an X-ray with wavelength 0.12 nm?
16. (I) An EM wave has frequency 2.65×10^{14} Hz. What is its wavelength, and how would we classify it?
17. (I) Electromagnetic waves and sound waves can have the same frequency. (*a*) What is the wavelength of a 1.00-kHz electromagnetic wave? (*b*) What is the wavelength of a 1.00-kHz sound wave? (The speed of sound in air is 341 m/s.) (*c*) Can you hear a 1.00-kHz electromagnetic wave?

18. (II) A light-year is a measure of distance (not time). How many meters does light travel in a year?

19. (II) Our nearest star (other than the Sun) is 4.2 light-years away. That is, it takes 4.2 years for the light it emits to reach Earth. How far away is it in meters?

20. (II) How long would it take a message sent as radio waves from Earth to reach Mars when Mars is (*a*) nearest Earth, (*b*) farthest from Earth? Assume that Mars and Earth are in the same plane and that their orbits around the Sun are circles (Mars is 230×10^6 km from the Sun).

21. (II) Pulsed lasers used for science and medicine produce very brief bursts of electromagnetic energy. If the laser light wavelength is 1062 nm (Neodymium–YAG laser), and the pulse lasts for 32 picoseconds, how many wavelengths are found within the laser pulse? How brief would the pulse need to be to fit only one wavelength?

22. (II) An electromagnetic wave has an electric field given by

$$\vec{\mathbf{E}} = \hat{\mathbf{i}}(275\ \text{V/m})\sin\left[(0.053\ \text{m}^{-1})z - (1.6 \times 10^7\ \text{rad/s})t\right].$$

(*a*) What are the wavelength and frequency of the wave? (*b*) Write down an expression for the magnetic field.

31–7 Speed of Light

23. (II) What is the minimum angular speed at which Michelson's eight-sided mirror would have had to rotate to reflect light into an observer's eye by succeeding mirror faces (1/8 of a revolution, Fig. 31–14)?

24. (II) A student wants to scale down Michelson's light-speed experiment to a size that will fit in one room. An eight-sided mirror is available, and the stationary mirror can be mounted 16 m from the rotating mirror. If the arrangement is otherwise as shown in Fig. 31–14, at what minimum rate would the mirror have to rotate? Does this seem feasible?

31–8 Energy in EM Waves; Poynting Vector

25. (I) The $\vec{\mathbf{E}}$ field in an EM wave has a peak of 23.5 mV/m. What is the average rate at which this wave carries energy across unit area per unit time?

26. (II) The magnetic field in a traveling EM wave has an rms strength of 22.5 nT. How long does it take to deliver 415 J of energy to 1.00 cm^2 of a wall that it hits perpendicularly?

27. (II) How much energy is transported across a 1.00-cm^2 area per hour by an EM wave whose E field has an rms strength of 25.8 mV/m?

28. (II) A spherically spreading EM wave comes from a 2100-W source. At a distance of 5.0 m, what is the intensity, and what is the rms value of the electric field?

29. (II) If the amplitude of the B field of an EM wave is 1.8×10^{-7} T, (*a*) what is the amplitude of the E field? (*b*) What is the average power transported across unit area by the EM wave?

30. (II) What is the average energy contained in a 1.00-m^3 volume near the Earth's surface due to radiant energy from the Sun? See Example 31–6.

31. (II) An 18.5-mW laser puts out a narrow beam 2.40 mm in diameter. What are the rms values of E and B in the beam?

32. (II) Estimate the average power output of the Sun, given that about 1350 W/m^2 reaches the upper atmosphere of the Earth.

33. (II) A high-energy pulsed laser emits a 1.0-ns-long pulse of average power 1.5×10^{11} W. The beam is nearly a cylinder 1.8×10^{-3} m in radius. Determine (*a*) the energy delivered in each pulse, and (*b*) the rms value of the electric field.

34. (III) (*a*) When a circular parallel-plate capacitor is being charged as in Example 31–1, show that the Poynting vector $\vec{\mathbf{S}}$ points radially inward toward the center of the capacitor, parallel to the plates. (*b*) Integrate $\vec{\mathbf{S}}$ over the cylindrical boundary of the capacitor gap to show that the rate at which energy enters the capacitor is equal to the rate at which electrostatic energy is being stored in the electric field of the capacitor (Section 24–4). Ignore fringing of $\vec{\mathbf{E}}$.

35. (III) The Arecibo radio telescope in Puerto Rico can detect a radio wave with an intensity as low as 1×10^{-23} W/m^2. Consider a "best-case" scenario for communication with extraterrestrials: suppose an advanced civilization a distance x away from Earth is able to transform the entire power output of a Sun-like star completely into a radio-wave signal which is transmitted uniformly in all directions. (*a*) In order for Arecibo to detect this radio signal, what is the maximum value for x in light-years ($1\ \text{ly} \approx 10^{16}$ m)? (*b*) How does this maximum value compare with the 100,000-ly size of our Milky Way galaxy? The intensity of sunlight at Earth's orbital distance from the Sun is 1350 W/m^2. [*Hint*: Assume the inverse square law.]

31–9 Radiation Pressure

36. (II) Estimate the radiation pressure due to a bulb emitting 5.0 W of EM radiation on a highly absorbing surface at a distance of 8.0 cm from the center of the bulb. Estimate the force exerted on your fingertip if you place it at this point.

37. (II) Laser light can be focused (at best) to a spot with a radius r equal to its wavelength λ. Suppose a 1.0-W beam of green laser light ($\lambda = 5 \times 10^{-7}$ m) forms such a spot and illuminates a cylindrical object of radius r and length r (Fig. 31–25). Estimate (*a*) the radiation pressure and force on the object, and (*b*) its acceleration, if its density equals that of water and it absorbs all the radiation. [This order-of-magnitude calculation convinced researchers of the feasibility of "optical tweezers," page 916.]

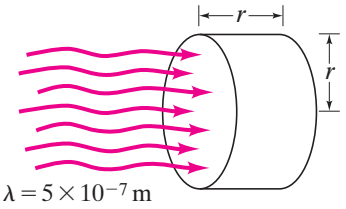


FIGURE 31–25
Problem 37.

38. (II) What size should the solar panel on a satellite orbiting Jupiter be if it is to collect the same amount of radiation from the Sun as a 1.0-m^2 solar panel on a satellite orbiting Earth? [*Hint*: Assume the inverse square law.]

39. (II) Suppose you have a car with a 100-hp engine. How large a solar panel would you need to replace the engine with solar power? Assume that the solar panels can utilize 20% of the maximum solar energy that reaches the Earth's surface ($1000\ \text{W/m}^2$). Explain why or why not this is practical.

31–10 Radio, TV

40. (I) What is the range of wavelengths for (*a*) FM radio (88 MHz to 108 MHz) and (*b*) AM radio (535 kHz to 1700 kHz)?

41. (I) Estimate the wavelength for a 1.9-GHz cell phone transmitter.

42. (I) Compare 1030 on the AM dial to 103.1 on FM. Which has the longer wavelength, and by what factor is it larger?

43. (I) What are the wavelengths for two TV channels that broadcast at 54.0 MHz (Channel 2) and 692 MHz (Channel 51)?

44. (I) The variable capacitor in the tuner of an AM radio has a capacitance of 1800 pF when the radio is tuned to a station at 550 kHz. What must the capacitance be for a station near the other end of the dial, 1610 kHz?

45. (I) The oscillator of a 98.3-MHz FM station has an inductance of 2.8 μH. What value must the capacitance be?

46. (II) A certain FM radio tuning circuit has a fixed capacitor $C = 680$ pF. Tuning is done by a variable inductance. What range of values must the inductance have to tune stations from 88 MHz to 108 MHz?

47. (II) An amateur radio operator wishes to build a receiver that can tune a range from 14.0 MHz to 15.0 MHz. A variable capacitor has a minimum capacitance of 95 pF. (*a*) What is the required value of the inductance? (*b*) What is the maximum capacitance used on the variable capacitor?

48. (II) A satellite beams microwave radiation with a power of 16 kW toward the Earth's surface, 550 km away. When the beam strikes Earth, its circular diameter is about 1500 m. Find the rms electric field strength of the beam.

49. (III) A 1.40-m-long FM antenna is oriented parallel to the electric field of an EM wave. How large must the electric field be to produce a 1.00-mV (rms) voltage between the ends of the antenna? What is the rate of energy transport per m^2?

General Problems

50. Who will hear the voice of a singer first: a person in the balcony 50.0 m away from the stage (see Fig. 31–26), or a person 1800 km away at home whose ear is next to the radio listening to a live broadcast? Roughly how much sooner? Assume the microphone is a few centimeters from the singer and the temperature is 20°C.

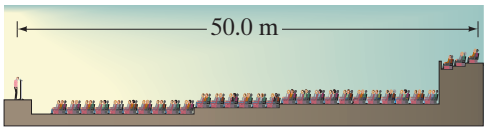


FIGURE 31–26 Problem 50.

51. A global positioning system (GPS) functions by determining the travel times for EM waves from various satellites to a moving GPS receiver on Earth (car or hiker). If the receiver is to detect a change in the receiver's position on the order of 3 m, what is the associated change in travel time (in ns) that must be measured?

52. Light is emitted from an incandescent lightbulb filament in wave-train bursts about 10^{-8} s in duration. What is the length in space of such wave trains?

53. The voice from an astronaut on the Moon (Fig. 31–27) was beamed to an auditorium full of people on Earth. (*a*) If you were standing 32 m from the loudspeaker on Earth, what was the total time lag between when you heard the sound and when the sound originally entered a microphone on the Moon? (*b*) Explain whether the microphone was inside the space helmet, or outside, and why.



FIGURE 31–27 Problem 53.

54. Radio-controlled clocks throughout the United States receive a radio signal from a transmitter in Fort Collins, Colorado, that accurately (within a microsecond) marks the beginning of each minute. A slight amount of time is added by a clock at any location to correct for the signal travel time to the clock from Fort Collins. Assuming Fort Collins is no more than 3000 km from any point in the U.S., what is the longest travel-time delay?

55. If the Sun were to disappear or radically change its output, how long would it take for us on Earth to learn about it?

56. *Cosmic microwave background* (CMB) radiation fills space with an average energy density of about 4×10^{-14} J/m^3. (*a*) Find the rms value of the electric field associated with this radiation. (*b*) How far from a 6.8-kW radio transmitter emitting uniformly in all directions would you find a comparable value?

57. What are E_0 and B_0 at a point 2.50 m from a light source whose output is 5.0 W? Assume the bulb emits radiation of a single frequency uniformly in all directions.

58. Estimate the rms electric field in the sunlight that hits Mars, knowing that the Earth receives about 1350 W/m^2 and that Mars is 1.52 times farther from the Sun (on average) than is the Earth.

59. The average intensity of a particular TV station's signal is 1.0×10^{-13} W/m^2 when it arrives at a 33-cm-diameter satellite TV antenna. (*a*) Calculate the total energy received by the antenna during 3.0 hours of viewing this station's programs. (*b*) Estimate the amplitudes of the E and B fields of the EM wave.

60. What length antenna would be appropriate for a portable device that could receive satellite TV?

61. A radio station is allowed to broadcast at an average power not to exceed 25 kW. If an electric field amplitude of 0.020 V/m is considered to be acceptable for receiving the radio transmission, estimate how many kilometers away you might be able to detect this station.

62. How practical is *solar power* for various devices? Assume that on a sunny day, sunlight has an intensity of 1000 W/m^2 at the surface of Earth and that a solar-cell panel can convert 20% of that sunlight into electric power. Calculate the area A of solar panel needed to power (*a*) a calculator that consumes 50 mW, (*b*) a hair dryer that consumes 1875 W, (*c*) a car that would require 120 hp. (*d*) In each case, would the area A be small enough to be mounted on the device itself, or in the case of (*b*) on the roof of a house?

63. A point source emits light energy uniformly in all directions at an average rate P_0 with a single frequency f. Show that the peak electric field in the wave is given by

$$E_0 = \sqrt{\frac{\mu_0 c P_0}{2\pi r^2}}.$$

[*Hint*: The surface area of a sphere is $4\pi r^2$.]

64. Estimate how long an AM antenna would have to be if it were (*a*) $\frac{1}{2}\lambda$ or (*b*) $\frac{1}{4}\lambda$. AM radio is roughly 1 MHz (530 kHz to 1.7 MHz).

65. A powerful laser portrayed in a movie provides a 3-mm-diameter beam of green light with a power of 3 W. A good agent inside a spacecraft aims the laser beam at an enemy astronaut hovering outside. The mass of the enemy astronaut is 120 kg and the spacecraft 185,000 kg. (*a*) Determine the "radiation-pressure" force exerted on the enemy by the laser beam assuming her suit is perfectly reflecting. (*b*) If the enemy is 30 m from the spacecraft's center of mass, estimate the gravitational force the spacecraft exerts on the enemy. (*c*) Which of the two forces is larger, and by what factor?

66. Show that displacement current, $\epsilon_0\,(d\Phi_E/dt)$, has the SI units of amperes.

67. When estimating the distance to lightning, you can time how long after the light flash is seen until the thunder is heard. Sound travels 343 m/s, so a good estimate is to allow 3.0 seconds per kilometer. There is actually a small mistake with this method because the time for the light to travel is not accounted for. How big is this mistake (%) for a lightning flash 2.0 km away?

68. What is the maximum power level of a radio station so as to avoid electrical breakdown of air at a distance of 0.75 m from the transmitting antenna? Assume the antenna is a point source. Air breaks down in an electric field of about 3×10^6 V/m.

69. In free space ("vacuum"), where the net charge and current flow is zero, the speed of an EM wave is given by $v = 1/\sqrt{\epsilon_0\mu_0}$. If, instead, an EM wave travels in a nonconducting ("dielectric") material with dielectric constant K, then $v = 1/\sqrt{K\epsilon_0\mu_0}$. For frequencies corresponding to the visible spectrum (near 5×10^{14} Hz), the dielectric constant of water is 1.77. Predict the speed of light in water and compare this value (as a percentage) with the speed of light in a vacuum.

70. Why do some **microwave ovens** rotate the food container? Suppose the metal walls form a rectangular cavity of dimensions 37 cm × 37 cm × 20 cm. When 2.45-GHz microwaves are continuously introduced into this cavity, reflection of incident waves from the walls sets up standing waves with nodes at the walls. Along the 37-cm dimension of the oven, how many nodes exist (excluding the nodes at the wall) and what is the distance between adjacent nodes? (No heating occurs at these nodes, so rotating the food container gives more even heating of the food.)

71. Imagine that a steady current I flows in a straight cylindrical wire of radius R_0 and resistivity ρ. (*a*) If the current is then changed at a rate dI/dt, show that a displacement current I_D exists in the wire of magnitude $\epsilon_0\rho(dI/dt)$. (*b*) If the current in a copper wire is changed at the rate of 1.0 A/ms, determine the magnitude of I_D. (*c*) Determine the magnitude of the magnetic field B_D created by I_D at the surface of a copper wire with $R_0 = 1.0$ mm. Compare (as a ratio) B_D with the field created at the surface of the wire by a steady current of 1.0 A.

72. At a given instant in time, a traveling EM wave is noted to have its maximum magnetic field pointing west and its maximum electric field pointing south. In which direction is the wave traveling? If the rate of energy flow is 520 W/m^2, what are the maximum values for the two fields?

73. The electric and magnetic fields of a certain EM wave in free space are given by $\vec{\mathbf{E}} = E_0 \sin(kx - \omega t)\,\hat{\mathbf{j}} + E_0 \cos(kx - \omega t)\,\hat{\mathbf{k}}$ and $\vec{\mathbf{B}} = B_0 \cos(kx - \omega t)\,\hat{\mathbf{j}} - B_0 \sin(kx - \omega t)\,\hat{\mathbf{k}}$. (*a*) Show that $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ are perpendicular to each other at all times. (*b*) For this wave, $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ are in a plane parallel to the yz plane. Show that the wave moves in a direction perpendicular to both $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$. (*c*) At any arbitrary choice of position x and time t, show that the magnitudes of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ always equal E_0 and B_0, respectively. (*d*) At $x = 0$, draw the orientation of $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ in the yz plane at $t = 0$. Then qualitatively describe the motion of these vectors in the yz plane as time increases. [Note: The EM wave in this Problem is "circularly polarized."]

74. The radiation pressure (Section 31–9) created by electromagnetic waves might someday be used to power spacecraft through the use of a "solar sail," Example 31–8. (*a*) Assuming total reflection, what would be the pressure on a solar sail located at the same distance from the Sun as the Earth (where $I = 1350$ W/m^2)? (*b*) Suppose the sail material has a mass density of 1 g/m^2. What would be the acceleration of the sail due to solar radiation pressure? (*c*) A realistic solar sail would have a payload. How big a sail would you need to accelerate a 100-kg payload at 1×10^{-3} m/s^2?

75. Suppose a 25-kW radio station emits EM waves uniformly in all directions. (*a*) How much energy per second crosses a 1.0-m^2 area 1.0 km from the transmitting antenna? (*b*) What is the rms magnitude of the $\vec{\mathbf{E}}$ field at this point, assuming the station is operating at full power? What is the rms voltage induced in a 1.0-m-long vertical car antenna (*c*) 1.0 km away, (*d*) 50 km away?

76. How large an emf (rms) will be generated in an antenna that consists of a circular coil 2.2 cm in diameter having 280 turns of wire, when an EM wave of frequency 810 kHz transporting energy at an average rate of 1.0×10^{-4} W/m^2 passes through it? [*Hint*: You can use Eq. 29–4 (for a generator) because that equation can be applied to an observer moving *with* the generator's coil as it rotates at $\omega = 2\pi f$ with f the frequency of the magnetic field.]

77. 9.0 km from a radio station's transmitting antenna, the amplitude of the electric field is 0.12 V/m. What is the average power output of the radio station?

78. Suppose that a right-moving EM wave overlaps with a left-moving EM wave so that, in a certain region of space, the total electric field in the y direction and magnetic field in the z direction are given by $E_y = E_0 \sin(kx - \omega t) + E_0 \sin(kx + \omega t)$ and $B_z = B_0 \sin(kx - \omega t) - B_0 \sin(kx + \omega t)$. (*a*) Find the mathematical expression that represents the standing electric and magnetic waves in the y and z directions, respectively. (*b*) Determine the Poynting vector and find the x locations at which it is zero at all times.

79. The electric field of an EM wave pulse traveling along the x axis in free space is given by $E_y = E_0 \exp\left[-\alpha^2x^2 - \beta^2t^2 + 2\alpha\beta xt\right]$, where E_0, α, and β are positive constants. (*a*) Is the pulse moving in the $+x$ or $-x$ direction? (*b*) Express β in terms of α and c (speed of light in free space). (*c*) Determine the expression for the magnetic field of this EM wave.

ANSWERS TO EXERCISES

A: (*c*).
B: (*b*).
C: (*a*) 3.8×10^6 Hz; (*b*) 5.5×10^{18} Hz.
D: 45 cm.
E: Yes; the signal still travels 72,000 km.
F: Over 4 hours.

Reflection from still water, as from a glass mirror, can be analyzed using the ray model of light.

Is this picture right side up? Or upside down? How can you tell? What are the clues? Notice the people and position of the Sun. Ray diagrams, which we will learn to draw in this Chapter, can provide the answer. See Example 32–3.

In this first Chapter on light and optics, we use the ray model of light to understand the formation of images by plane mirrors, and by curved (spherical) mirrors. We also begin our study of refraction—how light rays bend when they go from one medium to another—which prepares us for our study in the next Chapter of lenses, which are the crucial part of so many optical instruments.



CHAPTER 32

Light: Reflection and Refraction

CONTENTS

32–1 The Ray Model of Light
32–2 Reflection; Image Formation by a Plane Mirror
32–3 Formation of Images by Spherical Mirrors
32–4 Seeing Yourself in a Magnifying Mirror (Concave)
32–5 Convex (Rearview) Mirrors
32–6 Index of Refraction
32–7 Refraction: Snell's Law
32–8 Visible Spectrum and Dispersion
32–9 Total Internal Reflection; Fiber Optics
*32–10 Refraction at a Spherical Surface

CHAPTER-OPENING QUESTION—Guess now!

A 2.0-m-tall person is standing 2.0 m from a flat vertical mirror staring at her image. What minimum height must the mirror itself have if the person is to see her entire body, from the top of her head to her feet?

(a) 0.50 m.
(b) 1.0 m.
(c) 1.5 m.
(d) 2.0 m.
(e) 2.5 m.

The sense of sight is extremely important to us, for it provides us with a large part of our information about the world. How do we see? What is the something called *light* that enters our eyes and causes the sensation of sight? How does light behave so that we can see everything that we do? We saw in Chapter 31 that light can be considered a form of electromagnetic radiation. We now examine the subject of light in detail in the next four Chapters.

We see an object in one of two ways: (1) the object may be a *source* of light, such as a lightbulb, a flame, or a star, in which case we see the light emitted directly from the source; or, more commonly, (2) we see an object by light *reflected* from it. In the latter case, the light may have originated from the Sun, from artificial lights, or from a campfire. An understanding of how objects *emit* light was not achieved until the 1920s, and will be discussed in Chapter 37. How light is *reflected* from objects was understood much earlier, and will be discussed here, in Section 32–2.

32–1 The Ray Model of Light

A great deal of evidence suggests that *light travels in straight lines* under a wide variety of circumstances.[†] For example, a source of light like the Sun (which at its great distance from us is nearly a "point source") casts distinct shadows, and the beam from a laser pointer appears to be a straight line. In fact, we infer the positions of objects in our environment by assuming that light moves from the object to our eyes in straight-line paths. Our orientation to the physical world is based on this assumption.

This reasonable assumption is the basis of the **ray model** of light. This model assumes that light travels in straight-line paths called **light rays**. Actually, a ray is an idealization; it is meant to represent an extremely narrow beam of light. When we see an object, according to the ray model, light reaches our eyes from each point on the object. Although light rays leave each point in many different directions, normally only a small bundle of these rays can enter the pupil of an observer's eye, as shown in Fig. 32–1. If the person's head moves to one side, a different bundle of rays will enter the eye from each point.

We saw in Chapter 31 that light can be considered as an electromagnetic wave. Although the ray model of light does not deal with this aspect of light (we discuss the wave nature of light in Chapters 34 and 35), the ray model has been very successful in describing many aspects of light such as reflection, refraction, and the formation of images by mirrors and lenses. Because these explanations involve straight-line rays at various angles, this subject is referred to as **geometric optics**.

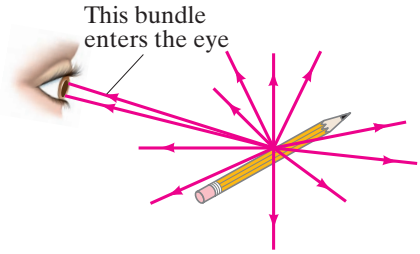


FIGURE 32–1 Light rays come from each single point on an object. A small bundle of rays leaving one point is shown entering a person's eye. This is a fundamental diagram. It will help us understand how we see images made by mirrors and lenses. Rays entering the eye are diverging. Our eyes would be useless if they could not focus diverging rays!

32–2 Reflection; Image Formation by a Plane Mirror

When light strikes the surface of an object, some of the light is reflected. The rest can be absorbed by the object (and transformed to thermal energy) or, if the object is transparent like glass or water, part can be transmitted through. For a very smooth shiny object such as a silvered mirror, over 95% of the light may be reflected.

When a narrow beam of light strikes a flat surface (Fig. 32–2), we define the **angle of incidence**, θ_i, to be the angle an incident ray makes with the normal (perpendicular) to the surface, and the **angle of reflection**, θ_r, to be the angle the reflected ray makes with the normal. It is found that the *incident and reflected rays lie in the same plane with the normal to the surface*, and that

the angle of reflection equals the angle of incidence, $\theta_r = \theta_i$.

LAW OF REFLECTION

This is the **law of reflection**, and it is depicted in Fig. 32–2. It was known to the ancient Greeks and was mentioned by Dante in his great masterpiece, *The Divine Comedy*, written around 1300.[‡] You can confirm it yourself by shining a narrow flashlight beam or a laser pointer at a mirror in a darkened room.

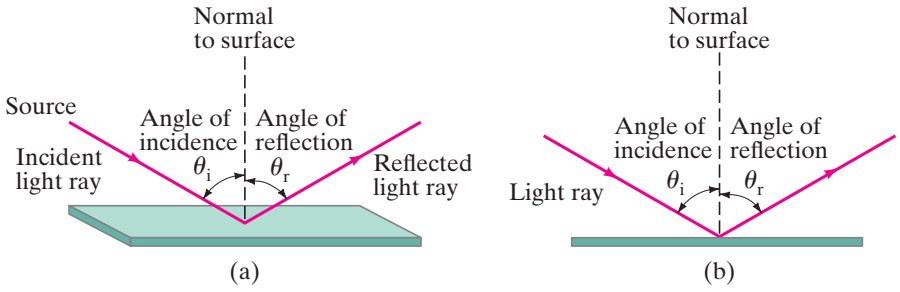


FIGURE 32–2 Law of reflection: (a) shows a 3-D view of an incident ray being reflected at the top of a flat surface; (b) shows a side or "end-on" view, which we will usually use because of its clarity.

[†]In ignoring the wave properties of light we must be careful that when the light rays pass by objects or through apertures, these must be large compared to the wavelength of the light (so the wave phenomena of interference and diffraction, as discussed in Chapter 15, can be ignored), and we ignore what happens to the light at the edges of objects until we get to Chapters 34 and 35.

[‡]There can be science even in great poetry. Dante considered this symmetry of reflection as testimony to harmony in the universe. (Ref.: Part 2, Chapter 15.)

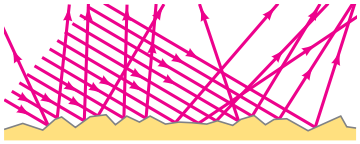


FIGURE 32–3 Diffuse reflection from a rough surface.

When light is incident upon a rough surface, even microscopically rough such as this page, it is reflected in many directions, as shown in Fig. 32–3. This is called **diffuse reflection**. The law of reflection still holds, however, at each small section of the surface. Because of diffuse reflection in all directions, an ordinary object can be seen at many different angles by the light reflected from it. When you move your head to the side, different reflected rays reach your eye from each point on the object (such as this page), Fig. 32–4a. Let us compare diffuse reflection to reflection from a mirror, which is known as **specular reflection**. ("Speculum" is Latin for mirror.) When a narrow beam of light shines on a mirror, the light will not reach your eye unless your eye is positioned at just the right place where the law of reflection is satisfied, as shown in Fig. 32–4b. This is what gives rise to the special image-forming properties of mirrors.

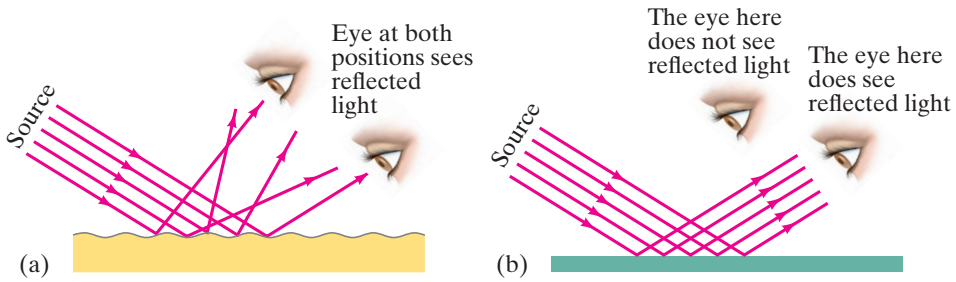


FIGURE 32–4 A narrow beam of light shines on (a) white paper, and (b) a mirror. In part (a), you can see with your eye the white light reflected at various positions because of diffuse reflection. But in part (b), you see the reflected light only when your eye is placed correctly $(\theta_r = \theta_i)$. Mirror reflection is also known as specular reflection. (Galileo, using similar arguments, showed that the Moon must have a rough surface rather than a highly polished surface like a mirror, as some people thought.)

EXAMPLE 32–1 **Reflection from flat mirrors.** Two flat mirrors are perpendicular to each other. An incoming beam of light makes an angle of 15° with the first mirror as shown in Fig. 32–5a. What angle will the outgoing beam make with the second mirror?

APPROACH We sketch the path of the beam as it reflects off the two mirrors, and draw the two normals to the mirrors for the two reflections. We use geometry and the law of reflection to find the various angles.

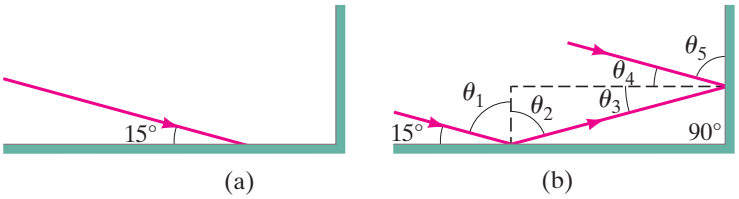


FIGURE 32–5 Example 32–1.

SOLUTION In Fig. 32–5b, $\theta_1 + 15° = 90°$, so $\theta_1 = 75°$; by the law of reflection $\theta_2 = \theta_1 = 75°$ too. Using the fact that the sum of the three angles of a triangle is always 180°, and noting that the two normals to the two mirrors are perpendicular to each other, we have $\theta_2 + \theta_3 + 90° = 180°$. Thus $\theta_3 = 180° - 90° - 75° = 15°$. By the law of reflection, $\theta_4 = \theta_3 = 15°$, so $\theta_5 = 75°$ is the angle the reflected ray makes with the second mirror surface.

NOTE The outgoing ray is parallel to the incoming ray. Reflectors on bicycles, cars, and other applications use this principle. Light reflects back toward the source.

When you look straight into a mirror, you see what appears to be yourself as well as various objects around and behind you, Fig. 32–6. Your face and the other objects look as if they are in front of you, beyond the mirror. But what you see in the mirror is an **image** of the objects, including yourself, that are in front of the mirror. Also, you don't see yourself as others see you, because left and right appear reversed in the image. That dimple on your left cheek seems to be on the right cheek of the "person" in the mirror.

A **plane mirror** is one with a smooth flat reflecting surface. Figure 32–7 shows how an image is formed by a plane mirror according to the ray model. We are viewing the mirror, on edge, in the diagram of Fig. 32–7, and the rays are shown reflecting from the front surface. (Good mirrors are generally made by putting a highly reflective metallic coating on one surface of a very flat piece of glass.) Rays from two different points on an object (the bottle on the left in Fig. 32–7) are shown: two rays are shown leaving from a point on the top of the bottle, and two more from a point on the bottom. Rays leave each point on the object going in many directions (as in Fig. 32–1), but only those that enclose the bundle of rays that enter the eye from each of the two points are shown. Each set of diverging rays that reflect from the mirror and enter the eye *appears to come from a single point* behind the mirror, called the **image point**, as shown by the dashed lines. That is, our eyes and brain interpret any rays that enter an eye as having traveled straight-line paths, like in Fig. 32–1. The point from which each bundle of rays seems to come is one point on the image. For each point on the object, there is a corresponding image point. (This analysis of how a plane mirror forms an image was published by Kepler in 1604.)



FIGURE 32–6 When you look in a mirror, you see an image of yourself and objects around you. You don't see yourself as others see you, because left and right appear reversed in the image. (Note photographer is visible in the mirror. Why?)

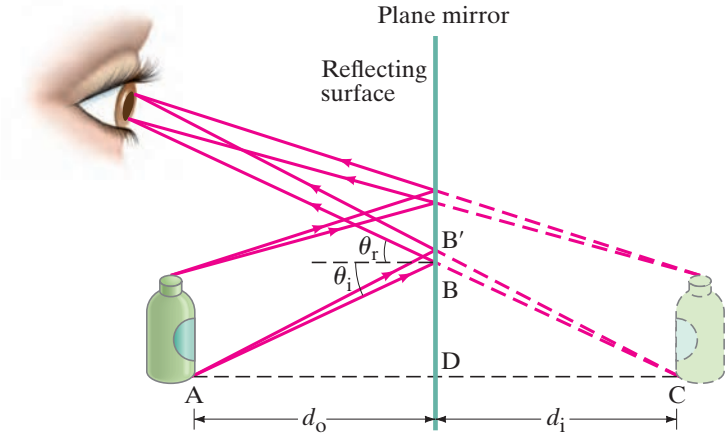


FIGURE 32–7 Formation of a virtual image by a plane (flat) mirror. Only the bundle of rays from the top and bottom of the object which reach the eye is shown. (We will usually show images using dashed lines or paler colors.)

Let us concentrate on the two rays that leave point A on the object in Fig. 32–7, and strike the mirror at points B and B′. We use geometry now, for the rays at B. The angles ADB and CDB are right angles; and because of the law of reflection, $\theta_i = \theta_r$ at point B. Therefore, by geometry, angles ABD and CBD are also equal. The two triangles ABD and CBD are thus congruent, and the length AD = CD. That is, the image appears as far behind the mirror as the object is in front. The **image distance**, d_i (perpendicular distance from mirror to image, Fig. 32–7), equals the **object distance**, d_o (perpendicular distance from object to mirror). That is

$$d_i = d_o. \qquad \text{[plane mirror]}$$

From the geometry, we also can see that the height of the image is the same as that of the object.

The light rays do not actually pass through the image location itself in Fig. 32–7. (Note where the red lines are drawn dashed to show they are our projections, not rays.) The image would not appear on paper or film placed at the location of the image. Therefore, the image in a plane mirror is called a **virtual image**. This is to distinguish it from a **real image** in which the light *does* pass through the image and which therefore could appear on a white surface, or on film or on an electronic sensor placed at the image position. Our eyes can see both real and virtual images, as long as the diverging rays enter our pupils. We will see that curved mirrors and lenses can form real images, as well as virtual. A movie projector lens, for example, produces a real image that is visible on the screen.

⚠ CAUTION

Light rays do not actually pass through image in a plane mirror

EXAMPLE 32–2 **How tall must a full-length mirror be?** A woman 1.60 m tall stands in front of a vertical plane mirror. What is the minimum height of the mirror, and how high must its lower edge be above the floor, if she is to be able to see her whole body? Assume that her eyes are 10 cm below the top of her head.

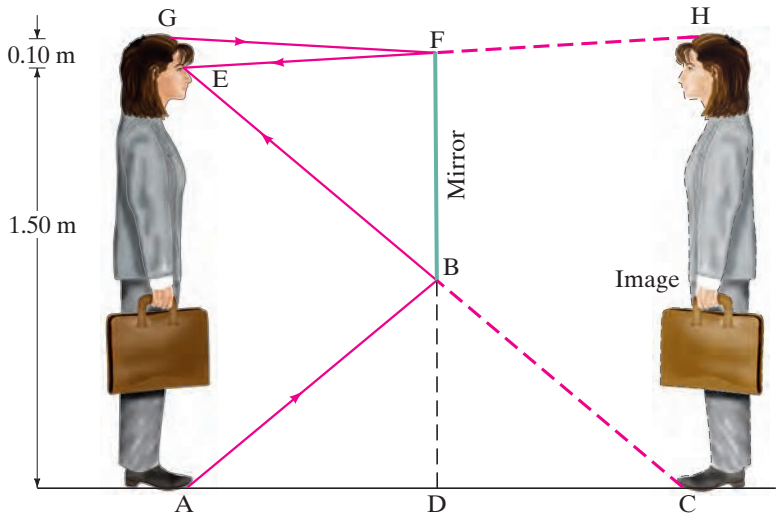


FIGURE 32–8 Seeing oneself in a mirror. Example 32–2.

APPROACH For her to see her whole body, as in Fig. 32–8, light rays from the top of her head (point G) and from the bottom of her foot (A) must reflect from the mirror and enter her eye. We don't show two rays diverging from each point as we did in Fig. 32–7, where we wanted to find where the image is. Now that we know the image is the same distance behind a plane mirror as the object is in front, we only need to show one ray leaving point G (top of head) and one ray leaving point A (her toe), and then use geometry.

PHYSICS APPLIED
How tall a mirror do you need to see a reflection of your entire self?

SOLUTION First consider the ray that leaves her foot at A, reflects at B, and enters the eye at E. The mirror needs to extend no lower than B. The angle of reflection equals the angle of incidence, so the height BD is half of the height AE. Because $AE = 1.60\,\text{m} - 0.10\,\text{m} = 1.50\,\text{m}$, then $BD = 0.75\,\text{m}$. So the mirror's bottom edge must be 0.75 m above the floor. Similarly, if the woman is to see the top of her head, the top edge of the mirror only needs to reach point F, which is 5 cm below the top of her head (half of $GE = 10\,\text{cm}$). Thus, $DF = 1.55\,\text{m}$, and the mirror needs to have a vertical height of only $(1.55\,\text{m} - 0.75\,\text{m}) = 0.80\,\text{m}$.

NOTE We see that a mirror, if positioned at the correct height, as in Fig. 32–8, need be only half as tall as a person for that person to be able to see all of himself or herself.

EXERCISE A Does the result of Example 32–2 depend on your distance from the mirror? (Try it and see, it's fun.)

EXERCISE B Return to the Chapter-Opening Question, page 926, and answer it again now. Try to explain why you may have answered differently the first time.

CONCEPTUAL EXAMPLE 32–3 **Is the photo upside down?** Close examination of the photograph on the first page of this Chapter, page 926, reveals that in the top portion the Sun is seen clearly, whereas in the lower portion the Sun is partially blocked by the tree branches. Show why the reflection is not the same as the real scene by drawing a sketch of this situation, showing the Sun, the camera, the branch, and two rays going from the Sun to the camera, one ray direct and one reflected. Is the photograph right side up?

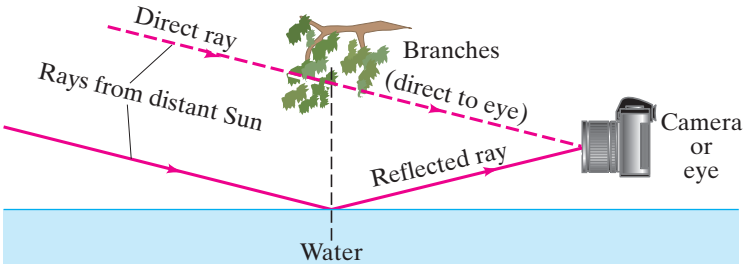


FIGURE 32–9 Example 32–3, analyzing the photo on page 926 that opens this Chapter: is it right side up, or not?

RESPONSE Figure 32–9 is drawn assuming the photo on page 926 is upside down. The Sun is so far away, its light rays arrive parallel (see Figs. 32–12 and 32–13), and the Sun's rays as seen directly come through the branches (dashed red ray). The unblocked view of the Sun would be the reflection: the ray traveling below the branches would reflect off the water and into the camera. This works—it is what we see in the photo if we turn it upside down. But a diagram assuming the photo on page 926 is right side up (Sun seen directly is less high above the horizon than is the reflection) won't work (Problem 6). The photo on page 926 is upside down.

Also, what about the people in the photo? Try to draw a diagram showing why they don't appear in the reflection. [*Hint*: Assume they are not sitting at the edge of the pool, but back from the edge.] Then try to draw a diagram of the reverse (i.e., assume the photo is right side up so the people are visible only in the reflection). Reflected images are not perfect replicas when different planes (distances) are involved.

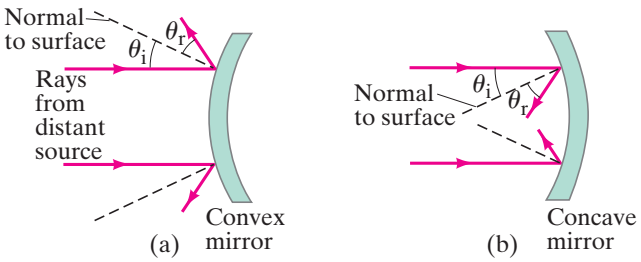


FIGURE 32–10 Mirrors with convex and concave spherical surfaces. Note that $\theta_r = \theta_i$ for each ray. (The black dashed lines are perpendicular to the mirror surface at each point shown).

FIGURE 32–11 (a) A concave cosmetic mirror gives a magnified image. (b) A convex mirror in a store reduces image size and so includes a wide field of view.


(a)


(b)

32–3 Formation of Images by Spherical Mirrors

Reflecting surfaces can also be *curved*, usually *spherical*, which means they form a section of a sphere. A **spherical mirror** is called **convex** if the reflection takes place on the outer surface of the spherical shape so that the center of the mirror surface bulges out toward the viewer, Fig. 32–10a. A mirror is called **concave** if the reflecting surface is on the inner surface of the center of the sphere so that the mirror surface curves away from the viewer (like a "cave"), Fig. 32–10b. Concave mirrors are used as shaving or cosmetic mirrors (**magnifying mirrors**), Fig. 32–11a, because they magnify. Convex mirrors are sometimes used on cars and trucks (rearview mirrors) and in shops (to watch for theft), because they take in a wide field of view, Fig. 32–11b.

Focal Point and Focal Length

To see how spherical mirrors form images, we first consider an object that is very far from a concave mirror. For a distant object, as shown in Fig. 32–12, the rays from each point on the object that strike the mirror will be nearly parallel. *For an object infinitely far away* (the Sun and stars approach this), *the rays would be precisely parallel.*

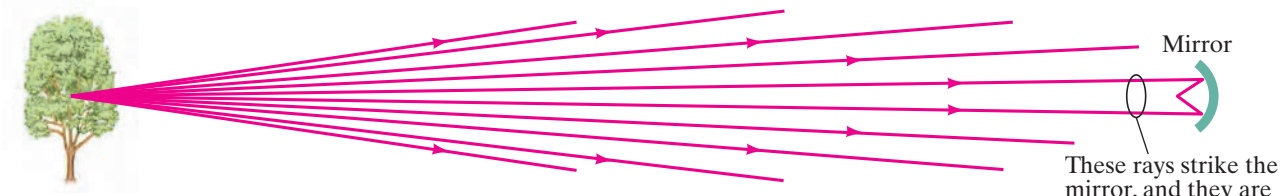


FIGURE 32–12 If the object's distance is large compared to the size of the mirror (or lens), the rays arrive nearly parallel. They are parallel for an object at infinity (∞).

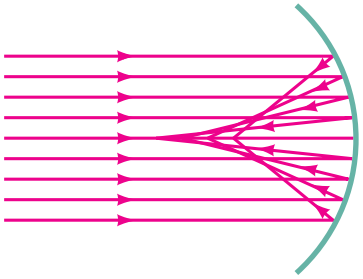


FIGURE 32–13 Parallel rays striking a concave spherical mirror do not intersect (or focus) at precisely a single point. (This "defect" is referred to as "spherical aberration.")

Now consider such parallel rays falling on a concave mirror as in Fig. 32–13. The law of reflection holds for each of these rays at the point each strikes the mirror. As can be seen, they are not all brought to a single point. In order to form a sharp image, the rays must come to a point. Thus a spherical mirror will not make as sharp an image as a plane mirror will. However, as we show below, if the mirror is small compared to its radius of curvature, so that a reflected ray makes only a *small angle* with the incident ray (2θ in Fig. 32–14), then the rays will cross each other at very nearly a single point, or **focus**. In the case shown in Fig. 32–14, the incoming rays are parallel to the **principal axis**, which is defined as the straight line perpendicular to the curved surface at its center (line CA in Fig. 32–14). The point F, where incident parallel rays come to a focus after reflection, is called the **focal point** of the mirror. The distance between F and the center of the mirror, length FA, is called the **focal length**, f, of the mirror. The focal point is also the *image point for an object infinitely far away* along the principal axis. The image of the Sun, for example, would be at F.

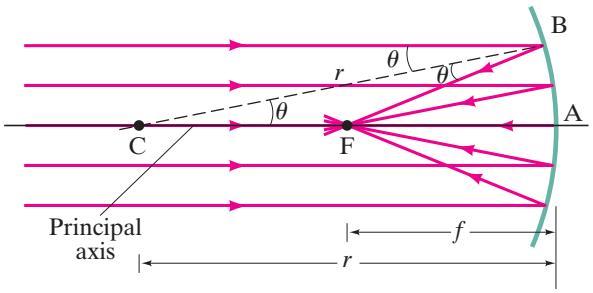


FIGURE 32–14 Rays parallel to the principal axis of a concave spherical mirror come to a focus at F, the focal point, as long as the mirror is small in width as compared to its radius of curvature, r, so that the rays are "paraxial"—that is, make only small angles with the horizontal axis. The smaller the angles, the better the approximation.

Now we will show, for a mirror whose reflecting surface is small compared to its radius of curvature, that the rays very nearly meet at a common point, F, and we will also determine the focal length f. In this approximation, we consider only rays that make a small angle with the principal axis; such rays are called **paraxial rays**, and their angles are exaggerated in Fig. 32–14 to make the labels clear. First we consider a ray that strikes the mirror at B in Fig. 32–14. The point C is the center of curvature of the mirror (the center of the sphere of which the mirror is a part). So the dashed line CB is equal to r, the radius of curvature, and CB is normal to the mirror's surface at B. The incoming ray that hits the mirror at B makes an angle θ with this normal, and hence the reflected ray, BF, also makes an angle θ with the normal (law of reflection). The angle BCF is also θ, as shown (parallel lines—see Fig. A–1 in Appendix A). The triangle CBF is isosceles because two of its angles are equal. Thus length CF = FB. We assume the mirror surface's height is small compared to the mirror's radius of curvature, so the angles are small,[†] and the length FB is nearly equal to length FA. In this approximation, FA = FC. But FA = f, the focal length, and CA = 2 × FA = r. Thus the focal length is half the radius of curvature:

Focal length of spherical mirror

$$f = \frac{r}{2}. \qquad \text{[spherical mirror]} \quad \textbf{(32–1)}$$

We assumed only that the angle θ was small, so this result applies for all other incident paraxial rays. Thus all paraxial rays pass through the same very tiny point F, the focal point.

Since it is only approximately true that the rays come to a perfect focus at F, the more curved the mirror, the worse the approximation (Fig. 32–13) and the more blurred the image. This "defect" of spherical mirrors is called **spherical aberration**; we will discuss it more with regard to lenses in Chapter 33. A **parabolic reflector**, on the other hand, will reflect the rays to a perfect focus. However, because parabolic shapes are much harder to make and thus much more expensive, spherical mirrors are used for most purposes. (Many astronomical telescopes use parabolic reflectors, as do TV satellite dish antennas which concentrate radio waves to nearly a point, Fig. 31–23.) We consider here only spherical mirrors and we will assume that they are small compared to their radius of curvature so that the image is sharp and Eq. 32–1 holds.

[†]Just to give an idea, the error in this assumption rises to 1% at $\theta \approx 15° = 0.25$ rad if we use $\sin\theta$ or $\tan\theta$ as our guide: $\tan\theta \approx \sin\theta \approx \theta$ in radians (0.259 vs. 0.262, page A-5 in Appendix A).

Image Formation—Ray Diagrams

We saw that for an object at infinity, the image is located at the focal point of a concave spherical mirror, where $f = r/2$. But where does the image lie for an object not at infinity? First consider the object shown as an arrow in Fig. 32–15a, which is placed between F and C at point O (O for object). Let us determine where the image will be for a given point O′ at the top of the object, by finding the point where rays drawn from the tip of the arrow converge after reflecting from the mirror. To do this we can draw several rays and make sure these reflect from the mirror such that the angle of reflection equals the angle of incidence.

(a) Ray 1 goes out from O′ parallel to the axis and reflects through F.

(b) Ray 2 goes through F and then reflects back parallel to the axis.

(c) Ray 3 is perpendicular to mirror, and so must reflect back on itself and go through C (center of curvature).

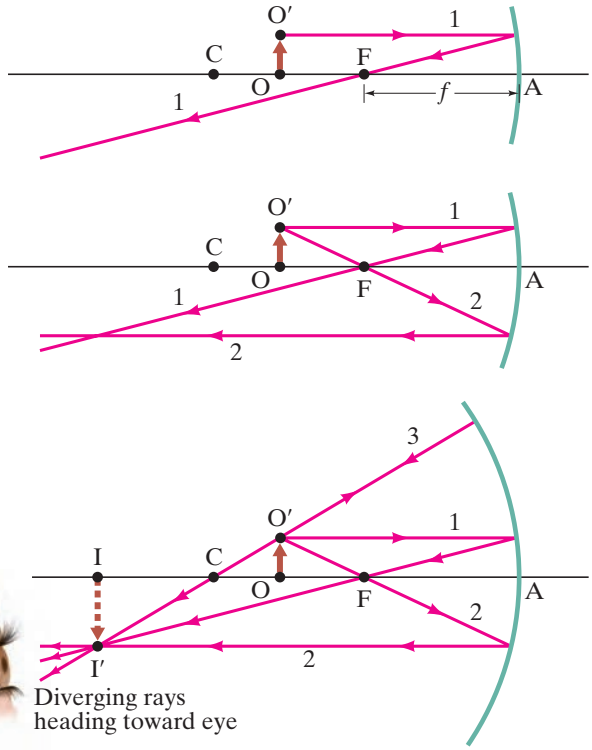


FIGURE 32–15 Rays leave point O′ on the object (a brown arrow). Shown are the three most useful rays for determining where the image I′ is formed. The image is shown as a dashed brown arrow. [Note that our mirror is not small compared to f, so the rays shown are not actually paraxial and our diagram will not give the precise position of the image.]

Many rays could be drawn leaving any point on an object, but determining the image position is faster if we deal with three particular rays. These are the rays labeled 1, 2, and 3 in Fig. 32–15 and we draw them leaving object point O′ as follows:

Ray 1 leaving O′ is drawn parallel to the axis; therefore after reflection it must pass along a line through F, Fig. 32–15a (just as parallel rays did in Fig. 32–14).

Ray 2 leaves O′ and is made to pass through F (Fig. 32–15b); therefore it must reflect so it is parallel to the axis. (In reverse, a parallel ray passes through F.)

Ray 3 is drawn along a radius of the spherical surface (Fig. 32–15c) and is perpendicular to the mirror, so it is reflected back on itself and passes through C, the center of curvature.

RAY DIAGRAM
Finding the image position for a curved mirror

All three rays leave a single point O′ on the object. After reflection from a (small) mirror, the point at which these rays cross is the image point I′. All other rays from the same object point will also pass through this image point. To find the image point for any object point, only these three types of rays need to be drawn. Only two of these rays are needed, but the third serves as a check.

PROBLEM SOLVING
Image point is where reflected rays intersect

We have shown the image point in Fig. 32–15 only for a single point on the object. Other points on the object are imaged nearby. For instance, the bottom of the arrow, on the principal axis at point O, is imaged on the axis at point I. So a complete image of the object is formed (dashed arrow in Fig. 32–15c). Because the light actually passes through the image, this is a **real image** that will appear on a white surface or film placed there. This can be compared to the virtual image formed by a plane mirror (Fig. 32–7: the light does not pass through that image).

The image in Fig. 32–15 can be seen by the eye only when the eye is placed to the left of the image, so that rays *diverging* from a point on the image (as point I′) enter the eye as shown in Fig. 32–15c. A normal eye can focus parallel or *diverging* rays (not converging), as in normal viewing: see the important Fig. 32–1.

CAUTION
Only diverging rays can be focused by the eye

Mirror Equation and Magnification

Image points can be determined, roughly, by drawing the three rays as just described, Fig. 32–15. But it is difficult to draw small angles for the "paraxial" rays as we assumed. For more accurate results, we now derive an equation that gives the image distance if the object distance and radius of curvature of the mirror are known. To do this, we refer to Fig. 32–16. The **object distance**, d_o, is the distance of the object (point O) from the center of the mirror. The **image distance**, d_i, is the distance of the image (point I) from the center of the mirror. The height of the object OO′ is called h_o and the height of the image, II′, is h_i. Two rays leaving O′ are shown: O′FBI′, which is ray 2 in Fig. 32–15, plus ray O′AI′, which is a fourth type of ray that reflects at the center of the mirror (equal angles θ) and in many cases can be useful to find an image point (labeled 4 in Fig. 32–16).

RAY DIAGRAM
Ray 4 to help find image

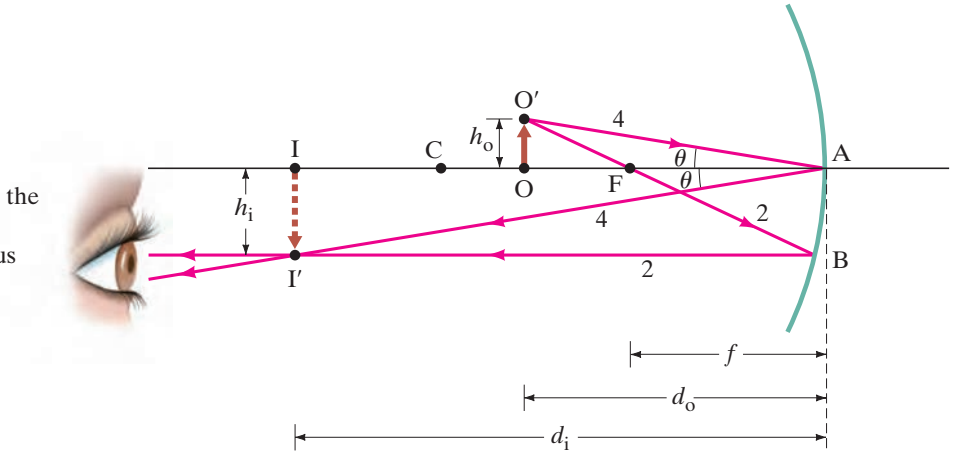


FIGURE 32–16 Diagram for deriving the mirror equation. For the derivation, we assume the mirror size is small compared to its radius of curvature.

The ray O′AI′ obeys the law of reflection, so the two right triangles O′AO and I′AI are similar. Therefore, we have

$$\frac{h_o}{h_i} = \frac{d_o}{d_i}.$$

For the other ray shown, O′FBI′, we use the fact that for a mirror small compared to its radius, the triangles O′FO and AFB are similar (because the angles at F are equal and we use the approximation $AB = h_i$) and obtain

$$\frac{h_o}{h_i} = \frac{OF}{FA} = \frac{d_o - f}{f},$$

where we used the fact that $FA = f$, the focal length of the mirror.

The left sides of the two preceding expressions are the same, so we can equate the right sides:

$$\frac{d_o}{d_i} = \frac{d_o - f}{f}.$$

We now divide both sides by d_o and rearrange to obtain

MIRROR EQUATION

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}. \qquad \textbf{(32–2)}$$

This is the equation we were seeking. It is called the **mirror equation** and relates the object and image distances to the focal length f (where $f = r/2$).

The mirror equation also holds for a plane mirror: the focal length is $f = r/2 = \infty$ (Eq. 32–1) because $r = \infty$ for a flat surface, and Eq. 32–2 gives $d_i = -d_o$, as it should in accord with Fig. 32–7 and the discussion on page 929; the sign convention on the next page gives us the minus sign.

⚠ CAUTION
Ray diagrams are not precise

Note that the image distance cannot be obtained accurately by measuring on a ray diagram like Fig. 32–16, because our diagram violates the paraxial ray assumption (we draw rays at steeper angles to make them clearly visible).

The **lateral magnification**, m, of a mirror is defined as the height of the image divided by the height of the object. From our first set of similar triangles in Fig. 32–16, or the first equation on page 934, we can write:

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}. \qquad \textbf{(32–3)}$$

The minus sign in Eq. 32–3 is inserted as a convention. Indeed, we must be careful about the signs of all quantities in Eqs. 32–2 and 32–3. Sign conventions are chosen so as to give the correct locations and orientations of images, as predicted by ray diagrams. The **sign conventions** we use are:

PROBLEM SOLVING
Sign conventions for mirrors

1. the image height h_i is positive if the image is upright, and negative if inverted, relative to the object (assuming h_o is taken as positive);
2. d_i or d_o is positive if image or object is in front of the mirror (as in Fig. 32–16); if either image or object is behind the mirror, the corresponding distance is negative. [An example of $d_i < 0$ can be seen in Fig. 32–17, on the next page.][†]

Thus the lateral magnification (Eq. 32–3) is positive for an upright image and negative for an inverted image (upside down). In everyday life, what we usually mean by magnification is not lateral magnification but rather *angular* magnification, which we discuss shortly.

Concave Mirror Examples

EXAMPLE 32–4 **Image in a concave mirror.** A 1.50-cm-high object is placed 20.0 cm from a concave mirror with radius of curvature 30.0 cm. Determine (*a*) the position of the image, and (*b*) its size.

APPROACH The ray diagram is basically the same as Fig. 32–16, since the object is between F and C. We determine the focal length from the radius of curvature (Eq. 32–1), $f = r/2 = 15.0$ cm. The position and size of the image are found from Eqs. 32–2 and 32–3.

SOLUTION Referring to Fig. 32–16, we have CA $= r = 30.0$ cm, FA $= f = 15.0$ cm, and OA $= d_o = 20.0$ cm.

(*a*) We start with the mirror equation, Eq. 32–2, rearranging it:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o} = \frac{1}{15.0\text{ cm}} - \frac{1}{20.0\text{ cm}} = 0.0167\text{ cm}^{-1}.$$

CAUTION
Remember to take the reciprocal

So $d_i = 1/(0.0167\text{ cm}^{-1}) = 60.0$ cm. Because d_i is positive, the image is 60.0 cm in front of the mirror, on the same side as the object.

(*b*) From Eq. 32–3, the lateral magnification is

$$m = -\frac{d_i}{d_o} = -\frac{60.0\text{ cm}}{20.0\text{ cm}} = -3.00.$$

The image is 3.0 times larger than the object, and its height is

$$h_i = mh_o = (-3.00)(1.5\text{ cm}) = -4.5\text{ cm}.$$

The minus sign reminds us that the image is inverted, as shown in Fig. 32–16.

NOTE When an object is farther from a concave mirror than the focal point, we can see from Fig. 32–15 or 32–16 that the image is always inverted and real.

CONCEPTUAL EXAMPLE 32–5 **Reversible rays.** If the object in Example 32–4 is placed instead where the image is (see Fig. 32–16), where will the new image be?

RESPONSE The mirror equation is *symmetric* in d_o and d_i. Thus the new image will be where the old object was. Indeed, in Fig. 32–16 we need only reverse the direction of the rays to get our new situation.

[†] d_o is always positive for a real object; d_o can be negative only if the object is itself an image formed by another mirror or lens—see Example 33–6, next Chapter.

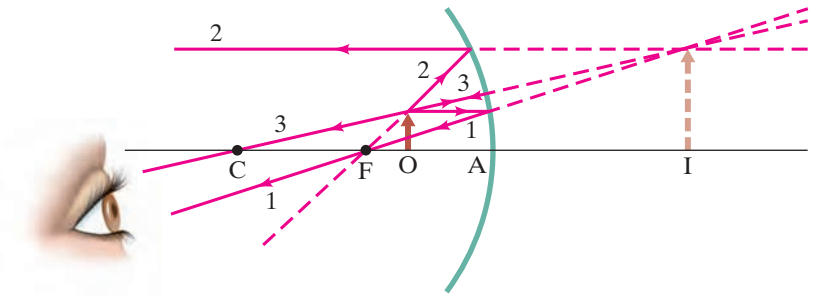


FIGURE 32–17 Object placed within the focal point F. The image (dashed arrow) is *behind* the mirror and is *virtual*, Example 32–6. [Note that the vertical scale (height of object = 1.0 cm) is different from the horizontal (OA = 10.0 cm) for ease of drawing, and reduces the precision of the drawing.]

EXAMPLE 32–6 **Object closer to concave mirror than focal point.** A 1.00-cm-high object is placed 10.0 cm from a concave mirror whose radius of curvature is 30.0 cm. (*a*) Draw a ray diagram to locate (approximately) the position of the image. (*b*) Determine the position of the image and the lateral magnification analytically.

APPROACH We draw the ray diagram using the rays as in Fig. 32–15, page 933. An analytic solution uses Eqs. 32–1, 32–2, and 32–3.

SOLUTION (*a*) Since $f = r/2 = 15.0\,\text{cm}$, the object $(d_o = 10.0\,\text{cm})$ is between the mirror and the focal point. We draw the three rays as described earlier (Fig. 32–15); they are shown leaving the tip of the object in Fig. 32–17. Ray 1 leaves the tip of our object heading toward the mirror parallel to the axis, and reflects through F. Ray 2 cannot head toward F because it would not strike the mirror; so ray 2 must point as if it started at F (dashed line in Fig. 32–17) and heads to the mirror, and then is reflected parallel to the principal axis. Ray 3 is perpendicular to the mirror and reflects back on itself. The rays reflected from the mirror diverge and so never meet at a point. They appear to be coming from a point behind the mirror (dashed red rays) just as for a flat mirror, Fig. 32–7. This point where the dashed rays meet is the image point of the tip of the arrow. The image is thus *behind* the mirror and is *virtual*.

(*b*) We use Eq. 32–2 to find d_i when $d_o = 10.0\,\text{cm}$:

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o} = \frac{1}{15.0\,\text{cm}} - \frac{1}{10.0\,\text{cm}} = \frac{2-3}{30.0\,\text{cm}} = -\frac{1}{30.0\,\text{cm}}.$$

Therefore, $d_i = -30.0\,\text{cm}$. The minus sign means the image is behind the mirror, which our diagram also showed us. The lateral magnification is $m = -d_i/d_o = -(-30.0\,\text{cm})/(10.0\,\text{cm}) = +3.00$. So the image is 3.00 times larger than the object. The plus sign indicates that the image is upright (same as object), which is consistent with the ray diagram, Fig. 32–17.

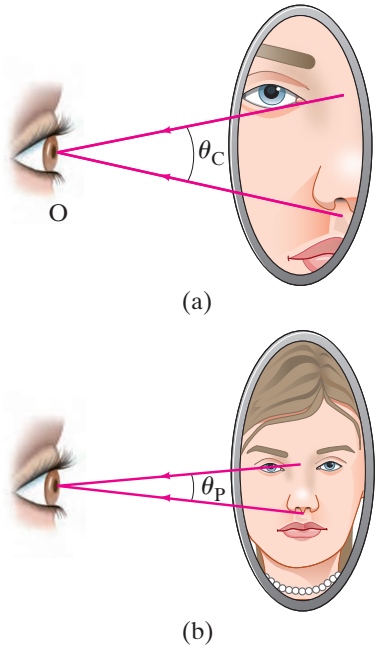


FIGURE 32–18 Looking at your face (a) in a magnifying mirror, and (b) in a plane flat mirror. The eye on the left is yours (object O), looking at *your* image in the mirror.

32–4 Seeing Yourself in a Magnifying Mirror (Concave)

To understand how a concave mirror acts as a magnifying mirror, we start with Fig. 32–17 and assume the object O is *you*, or your eye and part of your face. When a makeup or shaving mirror claims to be 5×, for example, the mirror is supposed to make an image that looks 5 times larger than when you look in an ordinary plane mirror. In Fig. 32–18 we are looking at two such mirrors. The concave mirror in (a) shows less of your face than the plane mirror in (b). But the part of your face you can see in (a) looks enlarged and you can see more detail of your eyes, cheeks, and nose. To specify the magnification, we compare the angles that some feature in your image, such as your nose, subtends at your eye. These angles are labeled θ_C (C for concave) and θ_P (P for plane) in Fig. 32–18. Their ratio is defined as the **angular magnification** M:

$$M = \frac{\theta_C}{\theta_P},$$

and is roughly about 2× in Fig. 32–18. This M corresponds to what we see, and usually mean in everyday life, vs. the m of Eq. 32–3 which is different. [M compares two different images; m compares an object to an image (distances or heights).]

Let us make a calculation of M using Fig. 32–17, but forget the eye shown on the left. Instead, we assume your eye and face are the object labeled O. Let us assume that the mirror has focal length $f = 15.0\,\text{cm}$ and your eye is at $d_o = 10.0\,\text{cm}$,

PHYSICS APPLIED
Magnifying mirror: Seeing yourself upright and magnified in a concave mirror

just as in Example 32–6, where we found the image distance to be $d_i = -30.0$ cm and the lateral magnification $m = 3.0$. Let the height of the object h_o (in cm) be the height of your nose or maybe your eye. To find θ_C in Fig. 32–18a, we assume the angle is small so $\theta_C \approx$ (height of image)/(distance eye to image). The height of the image is $h_i = mh_o = 3.0h_o$ (Eq. 32–3), and the distance from your eye to the image (see Fig. 32–17) is your eye's distance to the mirror ($d_o = 10.0$ cm) plus the distance from the mirror to the image in Fig. 32–17, $|d_i| = 30.0$ cm. Thus (cm units omitted)

$$\theta_C = \frac{h_i}{d_o + |d_i|} = \frac{3.0h_o}{10.0 + 30.0} = \left(\frac{3}{40}\right)h_o.$$

The angle θ_P for the plane mirror in Fig. 32–18b is easier. The image height equals the object height, and the distance from your eye to the image is $d_o + |d_i| = 2d_o$ because the image distance equals the object distance for a plane mirror (Fig. 32–7). Thus

$$\theta_P = \frac{h_i}{2d_o} = \frac{h_o}{20.0} = \left(\frac{1}{20}\right)h_o.$$

Then

$$M = \frac{\theta_C}{\theta_P} = \frac{(3/40)h_o}{(1/20)h_o} = \frac{(3)(20)}{40} = 1.5\times.$$

To get a larger magnification, let us move our eye, which is object O in Fig. 32–17 (not the eye shown), very close to the focal point F, say $d_o = 14.0$ cm. To find d_i we use Eq. 32–2

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o} = \frac{1}{15.0\text{ cm}} - \frac{1}{14.0\text{ cm}} = \frac{14.0\text{ cm} - 15\text{ cm}}{(15.0\text{ cm})(14.0\text{ cm})} = -\frac{1.0}{210\text{ cm}}.$$

Hence $d_i = -210$ cm and the lateral magnification is $m = h_i/h_o = -d_i/d_o =$ (210 cm)/(14.0 cm) = 15. The image height is $h_i = mh_o = 15h_o$ so when $d_o = 14.0$ cm,

$$\theta_C = \frac{15h_o}{d_o + |d_i|} = \frac{15h_o}{14.0 + 210} = 0.067h_o,$$

where we omitted units (all cm). Looking in a plane mirror from a distance of $d_o = 14.0$ cm means

$$\theta_P = \frac{h_o}{2d_o} = \left(\frac{1}{28}\right)h_o = 0.036h_o.$$

Hence

$$M = \frac{\theta_C}{\theta_P} = \frac{0.067h_o}{0.036h_o} = 1.9\times,$$

or almost 2×. If you move your eye (point O) closer and closer to F in Fig. 32–17, the angular magnification approaches 2.0×. That is in fact the best sharp image you can get from any spherical concave mirror.

What happens if you move your eye further from the mirror, so it is between F and C (center of mirror's curvature) as in Fig. 32–16? (Try it and see.) Your eye (and face) is the object at point O. If you set $d_o = 20.0$ cm, or any value between $d_o = f$ and $d_o = r = 2f$, your image will be real, *inverted*, and *behind* you (see Fig. 32–16 or use Eq. 32–2). In the ray diagram, the rays entering your eye are *converging* (imagine your eye just below point O in Fig. 32–16). But a normal eye is designed to focus *diverging* rays, as in reality (Fig. 32–1). Nonetheless, if you experiment with a concave mirror, as you place your eye farther back from the focal point ($d_o > f$), at first you will see an *upright* slightly blurry image of your face with $M > 2\times$. As you increase the distance between your eye and the mirror, the image gets larger but blurrier.† You maybe can reach 5× but the image is quite blurry. (10×, 20× is unrealistic!) As the eye-mirror distance, d_o, gets close to the mirror's radius of curvature (point C in Fig. 32–16) the image becomes one big blur. There seems to be a paradox when the object (your face) is between F and C. The image is *inverted* and *behind* you, as we just saw, yet you still *see* an upright image of yourself, though blurred. This can be explained: the converging rays from the mirror are intercepted by your eye but will *not* be focused by the eye's lens to a point on the retina, but instead as a small blob or circle (like "circles of confusion" discussed in the next Chapter).

Finally, what happens when your eye is beyond point C, the mirror's center of curvature. You can get a sharp image of yourself once again, but upside down. And it will seem small now because the mirror is rather far from you. This inverted image of yourself will be in front of you and readily visible. Draw the ray diagram, or see Fig. 32–50b (Problem 31).

†Farsighted eyes do better and focus slightly converging rays (see Fig. 33–27).

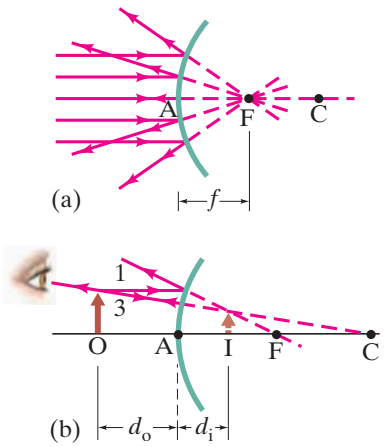


FIGURE 32–19 Convex mirror: (a) the focal point is at F, behind the mirror; (b) the image I of the object at O is virtual, upright, and smaller than the object. [Not to scale for Example 32–7.]

32–5 Convex (Rearview) Mirrors

The analysis used for concave mirrors can be applied to **convex** mirrors. Even the mirror equation (Eq. 32–2) holds for a convex mirror. Figure 32–19a shows parallel rays falling on a convex mirror. Again spherical aberration is significant (Fig. 32–13), unless we assume the mirror is small compared to its radius of curvature. The rays reflected from the mirror diverge, but seem to come from point F *behind* the mirror, Fig. 32–19a. This is the **focal point**, and its distance from the center of the mirror (point A) is the **focal length**, f. The equation $f = r/2$ is valid also for a convex mirror. We see that an object at infinity (parallel rays) produces a virtual image in a convex mirror. Indeed, no matter where the object is placed on the reflecting side of a convex mirror, the image will be virtual and upright as seen in Fig. 32–19b. To find the image we draw rays 1 and 3 according to the rules used before (page 933), as shown in Fig. 32–19b. Note that although rays 1 and 3 don't actually pass through points F and C, the line along which each is drawn does (shown dashed).

The mirror equation, Eq. 32–2, holds for convex mirrors but the focal length f and radius of curvature must be considered negative. The proof is left for you to do as Problem 23.

PHYSICS APPLIED
Convex rearview mirror

FIGURE 32–20 Example 32–7.



EXAMPLE 32–7 **Convex rearview mirror.** An external rearview car mirror is convex with a radius of curvature of 0.80 m (Fig. 32–20). Determine the location of the image and its lateral magnification for a truck 10.0 m from the mirror.

APPROACH The ray diagram will be like Fig. 32–19b, but the large object distance $(d_o = 10.0\,\text{m})$ makes a precise drawing difficult. We have a convex mirror, so r is negative by convention.

SOLUTION The center of curvature of a convex mirror is behind the mirror, as is its focal point, so we set $r = -0.80\,\text{m}$ so that the focal length is $f = r/2 = -0.40\,\text{m}$. The object is in front of the mirror, $d_o = 10.0\,\text{m}$. Solving the mirror equation, Eq. 32–2, for $1/d_i$ gives

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o} = \frac{1}{-0.40\,\text{m}} - \frac{1}{10.0\,\text{m}} = \frac{-10.0 - 0.40}{4.0\,\text{m}} = -\frac{-10.4}{4.0\,\text{m}}.$$

Thus $d_i = -4.0\,\text{m}/10.4 = -0.38\,\text{m}$ or 38 cm behind the mirror. Equation 32–3 gives the lateral magnification

$$m = -\frac{d_i}{d_o} = -\frac{(-0.38\,\text{m})}{(10.0\,\text{m})} = +0.038.$$

CAUTION
How large does the object seem, and how far away?

How large does the truck *look* in the rearview mirror of Fig. 32–20? We again use *angular magnification*, $M = \theta_C/\theta_P$, as we did in Section 32–4, because we want to compare the size of the image in the convex rearview mirror to what it would be in a flat mirror. In the convex mirror, the angle that the image of the truck makes with your eye (note Figs. 32–18 and 32–19) is $\theta_C = h_i/(d_{eye} + |d_i|)$ where d_{eye} is the distance of your eye from the mirror, about 0.5 m for a mirror on the driver's side. In Example 32–7, $|d_i| = 0.38\,\text{m}$ and $h_i = mh_o = 0.038h_o$, so $\theta_C = 0.038h_o/(0.5\,\text{m} + 0.38\,\text{m}) = 0.043h_o$ where h_o is the height of the truck in meters. But if you looked in a *plane* mirror, the lateral magnification $m = 1$, so $h_i = h_o$, and $|d_i| = d_o = 10\,\text{m}$. So the angle you would view the truck is $\theta_P = h_o/(d_{eye} + d_o) = h_o/(0.5\,\text{m} + 10\,\text{m}) = 0.095h_o$. The angular magnification is

$$M = \frac{\theta_C}{\theta_P} = \frac{0.043h_o}{0.095h_o} = 0.45,$$

CAUTION
*This object is **closer** than it seems in a convex mirror (BEWARE)*

so the image in the convex mirror is about half what it would be in a normal plane mirror. The truck would thus look *twice as far away* as it really is!

If we consider the mirror on the passenger side, $d_{eye} \approx 1.5\,\text{m}$, you will get a smaller result (see Problem 96).

32–6 Index of Refraction

We saw in Chapter 31 that the speed of light in vacuum (like other EM waves) is

$$c = 2.99792458 \times 10^8 \text{ m/s},$$

which is usually rounded off to

$$3.00 \times 10^8 \text{ m/s}$$

when extremely precise results are not required. In air, the speed is only slightly less. In other transparent materials, such as glass and water, the speed is always less than that in vacuum. For example, in water light travels at about $\frac{3}{4}c$. The ratio of the speed of light in vacuum to the speed v in a given material is called the **index of refraction**, n, of that material:

$$n = \frac{c}{v}. \tag{32–4}$$

The index of refraction is never less than 1, and values for various materials are given in Table 32–1. For example, since $n = 1.33$ for water, the speed of light in water is

$$v = \frac{c}{n} = \frac{(3.00 \times 10^8 \text{ m/s})}{1.33} = 2.26 \times 10^8 \text{ m/s}.$$

As we shall see later, n varies somewhat with the wavelength of the light—except in vacuum—so a particular wavelength is specified in Table 32–1, that of yellow light with wavelength $\lambda = 589$ nm.

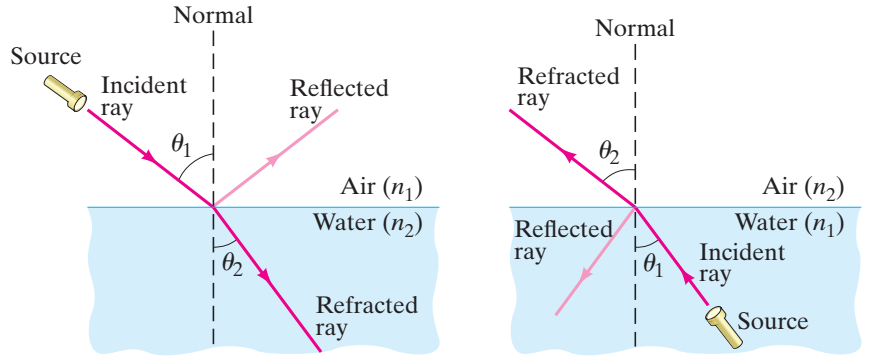
That light travels more slowly in matter than in vacuum can be explained at the atomic level as being due to the absorption and reemission of light by atoms and molecules of the material.

TABLE 32–1 Indices of Refraction (at $\lambda = 589$ nm)

Material	$n = \frac{c}{v}$
Vacuum	1.0000
Air (at STP)	1.0003
Water	1.33
Ethyl alcohol	1.36
Glass	
Fused quartz	1.46
Crown glass	1.52
Light flint	1.58
Plastic	
Acrylic, Lucite, CR-39	1.50
Polycarbonate	1.59
"High-index"	1.6–1.7
Sodium chloride	1.53
Diamond	2.42

32–7 Refraction: Snell's Law

When light passes from one transparent medium into another with a different index of refraction, some or all of the incident light is reflected at the boundary. The rest passes into the new medium. If a ray of light is incident at an angle to the surface (other than perpendicular), the ray changes direction as it enters the new medium. This change in direction, or bending, of the light ray is called **refraction**.



(a) $n_2 > n_1$: Ray bends toward ⊥ (b) $n_1 > n_2$: Ray bends away from ⊥

FIGURE 32–21 Refraction. (a) Light refracted when passing from air (n_1) into water (n_2): $n_1 < n_2$. (b) Light refracted when passing from water (n_1) into air (n_2): $n_1 > n_2$.

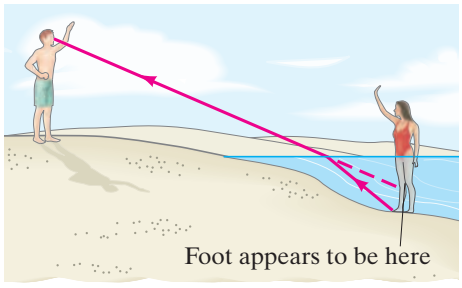
Figure 32–21a shows a ray passing from air into water. Angle θ_1 is the angle the incident ray makes with the normal (perpendicular) to the surface and is called the **angle of incidence**. Angle θ_2 is the **angle of refraction**, the angle the refracted ray makes with the normal to the surface. Notice that the ray bends toward the normal when entering the water. This is always the case when the ray enters a medium where the speed of light is *less* (and the index of refraction greater, Eq. 32–4). If light travels from one medium into a second where its speed is *greater*, the ray bends away from the normal; this is shown in Fig. 32–21b for a ray traveling from water to air.

⚠ CAUTION

Angles of incidence and refraction are measured from the perpendicular, not from the surface



(a)



(b)

FIGURE 32–22 (a) Photograph, and (b) ray diagram showing why a person's legs look shorter standing in water: a ray from the bather's foot to the observer's eye bends at the water's surface, and our brain interprets the light as traveling in a straight line, from higher up (dashed line).

FIGURE 32–23 A straw in water looks bent even when it isn't.



Refraction is responsible for optical illusions, such as a person standing in water appears to have shortened legs (Fig. 32–22). Rays leaving the person's foot are bent at the surface. The observer's brain assumes the rays to have traveled a straight-line path (dashed red line), and so the feet appear to be higher than they really are. Similarly, when you put a straw in water, it appears to be bent (Fig. 32–23). This also means that water is deeper than it appears.

Snell's Law

We can write an equation that relates the angle of refraction to the angle of incidence at an interface (Fig. 32–21), and we might expect that it would depend on the speed of light in the two media. In about 1621, Willebrord Snell (1591–1626) found just such a relationship experimentally. Known as **Snell's law**, it is written:

SNELL'S LAW (LAW OF REFRACTION)

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \qquad \textbf{(32–5)}$$

θ_1 is the angle of incidence and θ_2 is the angle of refraction; n_1 and n_2 are the respective indices of refraction in the materials. See Fig. 32–21. The incident and refracted rays lie in the same plane, which also includes the perpendicular to the surface. Snell's law is the **law of refraction**. (Snell's law was derived in Section 15–10 for water waves where Eq. 15–19 is just a combination of Eqs. 32–5 and 32–4, and we derive it again in Chapter 34 using the wave theory of light.)

Snell's law shows that if $n_2 > n_1$, then $\theta_2 < \theta_1$. Thus, if light enters a medium where n is greater (and its speed is less), the ray is bent toward the normal. And if $n_2 < n_1$, then $\theta_2 > \theta_1$, so the ray bends away from the normal. See Fig. 32–21.

EXERCISE C Light passes from a medium with $n = 1.3$ (water) into a medium with $n = 1.5$ (glass). Is the light bent toward or away from the perpendicular to the interface?

EXAMPLE 32–8 **Refraction through flat glass.** Light traveling in air strikes a flat piece of uniformly thick glass at an incident angle of 60.0°, as shown in Fig. 32–24. If the index of refraction of the glass is 1.50, (*a*) what is the angle of refraction θ_A in the glass; (*b*) what is the angle θ_B at which the ray emerges from the glass?

APPROACH We apply Snell's law twice: at the first surface, where the light enters the glass, and again at the second surface where it leaves the glass and enters the air.

SOLUTION (*a*) The incident ray is in air, so $n_1 = 1.00$ and $n_2 = 1.50$. Applying Snell's law where the light enters the glass $(\theta_1 = 60.0°,\ \theta_2 = \theta_A)$ gives

$$(1.00)\sin 60.0° = (1.50)\sin\theta_A$$

or

$$\sin\theta_A = \frac{1.00}{1.50}\sin 60.0° = 0.5774,$$

so $\theta_A = 35.3°$.

(*b*) Since the faces of the glass are parallel, the incident angle at the second surface is also θ_A (geometry), so $\sin\theta_A = 0.5774$. At this second interface, $n_1 = 1.50$ and $n_2 = 1.00$. Thus the ray re-enters the air at an angle θ_B given by

$$\sin\theta_B = \frac{1.50}{1.00}\sin\theta_A = 0.866,$$

and $\theta_B = 60.0°$. The direction of a light ray is thus unchanged by passing through a flat piece of glass of uniform thickness.

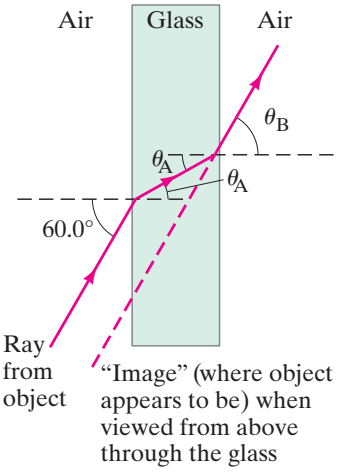


FIGURE 32–24 Light passing through a piece of glass (Example 32–8).

NOTE This result is valid for any angle of incidence. The ray is displaced slightly to one side, however. You can observe this by looking through a piece of glass (near its edge) at some object and then moving your head to the side slightly so that you see the object directly. It "jumps."

EXAMPLE 32–9 **Apparent depth of a pool.** A swimmer has dropped her goggles to the bottom of a pool at the shallow end, marked as 1.0 m deep. But the goggles don't appear to be that deep. Why? How deep do the goggles appear to be when you look straight down into the water?

APPROACH We draw a ray diagram showing two rays going upward at a small angle, heading toward your eye, from a point on the goggles on the bottom of the pool, and being refracted at the water's (flat) surface, Fig. 32–25. The two rays traveling upward from the goggles are refracted *away* from the normal as they exit the water, and so appear to be diverging from a point above the goggles (dashed lines), which is why the water appears to be less deep than it actually is. We are looking straight down, so all angles are small (but exaggerated in Fig. 32–25 for clarity).

SOLUTION To calculate the apparent depth d' (Fig. 32–25), given a real depth $d = 1.0\,\text{m}$, we use Snell's law with $n_1 = 1.33$ for water and $n_2 = 1.0$ for air:

$$\sin\theta_2 = n_1 \sin\theta_1 .$$

We are considering only small angles, so $\sin\theta \approx \tan\theta \approx \theta$, with θ in radians. So Snell's law becomes

$$\theta_2 \approx n_1\theta_1 .$$

From Fig. 32–25, we see that $\theta_2 \approx \tan\theta_2 = x/d'$ and $\theta_1 \approx \tan\theta_1 = x/d$. Putting these into Snell's law, $\theta_2 \approx n_1\theta_1$, we get

$$\frac{x}{d'} \approx n_1 \frac{x}{d}$$

or

$$d' \approx \frac{d}{n_1} = \frac{1.0\,\text{m}}{1.33} = 0.75\,\text{m}.$$

The pool seems only three-fourths as deep as it actually is.

NOTE Water in general is deeper than it looks—a useful safety guideline.

⚠ CAUTION (real life)
Water is deeper than it looks

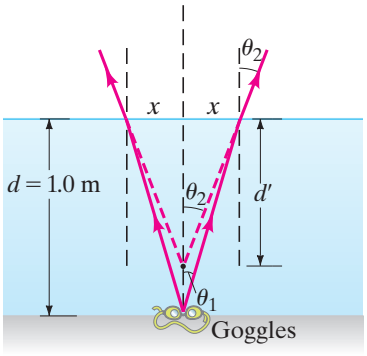


FIGURE 32–25 Example 32–9. Objects under water appear to be less deep than they actually are.

32–8 The Visible Spectrum and Dispersion

Two important properties of light are intensity (or brightness) and color. The **intensity** of light is the energy it carries per unit area per unit time (see Eqs. 31–18 and 31–19). The **color** of light, as we shall discuss in Chapter 34, is related to the frequency f or wavelength λ of the light. (Recall $\lambda f = c = 3.0 \times 10^8\,\text{m/s}$, Eq. 31–14.) Visible light—that to which our eyes are sensitive—consists of frequencies from about $4 \times 10^{14}\,\text{Hz}$ to $7.5 \times 10^{14}\,\text{Hz}$, corresponding to wavelengths in air of about 400 nm to 700 nm.[†] This is the **visible spectrum**, and within it lie the different colors from violet to red, as shown in Fig. 32–26.

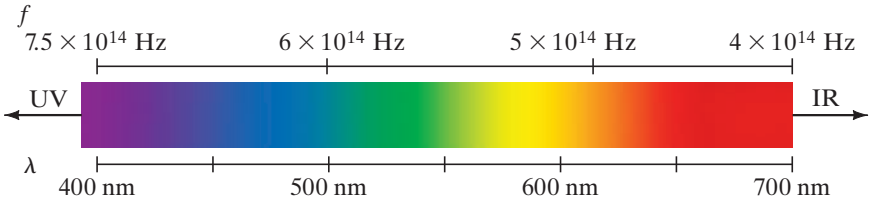


FIGURE 32–26 The spectrum of visible light, showing the range of frequencies and wavelengths in air for the various colors. Many colors, such as brown, do not appear in the spectrum; they are made from a mixture of wavelengths.

Light with wavelength (in air) shorter than 400 nm (= violet) is called **ultraviolet** (UV), and light with wavelength longer than 700 nm (= red) is called **infrared** (IR).[‡] Although human eyes are not sensitive to UV or IR, some types of photographic film and detectors do respond to them.

[†]Sometimes the angstrom (Å) unit is used when referring to light: $1\,\text{Å} = 1 \times 10^{-10}\,\text{m}$. Visible light has wavelengths in air of 4000 Å to 7000 Å.

[‡]The complete electromagnetic spectrum is illustrated in Fig. 31–12.

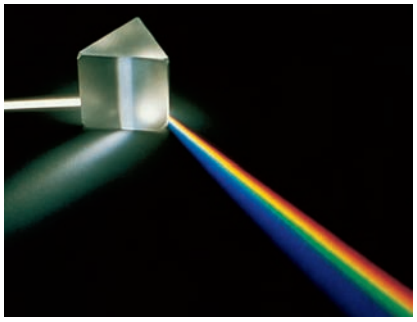


FIGURE 32–27 White light passing through a prism is broken down into its constituent colors.

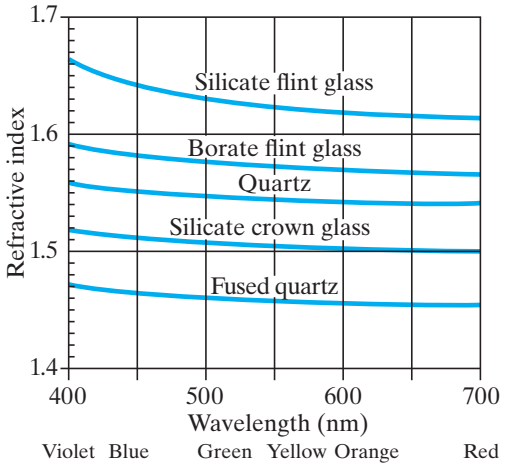
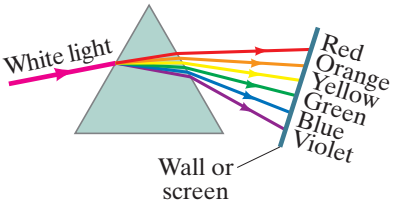


FIGURE 32–28 Index of refraction as a function of wavelength for various transparent solids.

FIGURE 32–29 White light dispersed by a prism into the visible spectrum.



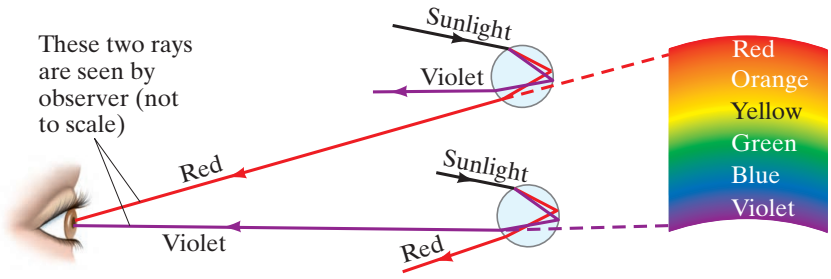
PHYSICS APPLIED

Rainbows

A prism can separate white light into a rainbow of colors, as shown in Fig. 32–27. This happens when the index of refraction of a material depends on the wavelength, as shown for several materials in Fig. 32–28. **White light** from the Sun is a mixture of all visible wavelengths, and when incident on a prism, as in Fig. 32–29, the different wavelengths are bent to varying degrees. Because the index of refraction is greater for the shorter wavelengths, violet light is bent the most and red the least, as shown in Fig. 32–29. This spreading of white light into the full spectrum is called **dispersion**.

Rainbows are a spectacular example of dispersion—by drops of water. You can see rainbows when you look at falling water droplets (rain) with the Sun behind you. Figure 32–30 shows how red and violet rays are bent by spherical water droplets and are reflected off the back surface of the droplet. Red is bent the least and so reaches the observer's eyes from droplets higher in the sky, as shown in Fig. 32–30a. Thus the top of the rainbow is red.

FIGURE 32–30 (a) Ray diagram explaining how a rainbow (b) is formed.



(a)



(b)

The visible spectrum, Fig. 32–26, does not show all the colors seen in nature. For example, there is no brown in Fig. 32–26. Many of the colors we see are a mixture of wavelengths. For practical purposes, most natural colors can be reproduced using three primary colors. They are red, green, and blue for direct source viewing such as TV and computer monitors. For inks used in printing, the primary colors are cyan (the blue color of the margin notes in this book), yellow, and magenta (the pinkish red color we use for light rays in ray diagrams).

For any wave, its velocity v is related to its wavelength λ and frequency f by $v = f\lambda$ (Eq. 15–1 or 31–14). When a mechanical wave, such as sound, travels from one material into another, the frequency of the wave does not change across the boundary. The reason is that at the boundary, a point (an atom) is forced to oscillate at that frequency, and so do adjacent points (atoms) beyond the boundary as the wave continues to move.

In the same way, an EM wave does not change frequency when it crosses a boundary. Thus if light travels from air into a material with index of refraction n, the wavelength becomes (recall Eq. 32–4):

$$\lambda_n = \frac{v}{f} = \frac{c}{nf} = \frac{\lambda}{n} \qquad \textbf{(32–6)}$$

where λ is the wavelength in vacuum or air and λ_n is the wavelength in the material with index of refraction n.

CONCEPTUAL EXAMPLE 32–10 **Observed color of light under water.** We said that color depends on wavelength. For example, for light of wavelength $\lambda_0 = 650\,\text{nm}$ in air, we see red. If we observe the same object when under water, it still looks red. But the wavelength in water λ_w is (Eq. 32–6) $\lambda_w = \lambda_0/n_w = 650\,\text{nm}/1.33 = 489\,\text{nm}$. Light with wavelength 489 nm in air would appear blue in air. Can you explain why the light appears red rather than blue when observed under water?

RESPONSE Today we have little doubt that it is our brains that express colors, based on the frequencies of light that strike the receptor cells within the retina (at the rear of the eyeball, as diagrammed in the next Chapter, Fig. 33–25). For objects under water, the water does nothing to change the frequency, but does change the wavelength to λ_0/n_w. When that light enters the eye, the frequency is still unchanged, but the speed is changed to c/n_{eye} where n_{eye} is the index of refraction of the fluid that fills the interior of the eye and is in contact with the retina. The wavelength of light that reaches the retina is $\lambda_{eye} = \lambda_0/n_{eye}$, and is the same whether the light enters from the air or from water. Another way of saying this is that color depends on the *frequency* of the light, which does not change.

32–9 Total Internal Reflection; Fiber Optics

When light passes from one material into a second material where the index of refraction is less (say, from water into air), the refracted light ray bends away from the normal, as for rays I and J in Fig. 32–31. At a particular incident angle, the angle of refraction will be 90°, and the refracted ray would skim the surface (ray K). The incident angle at which this occurs is called the **critical angle**, θ_C.

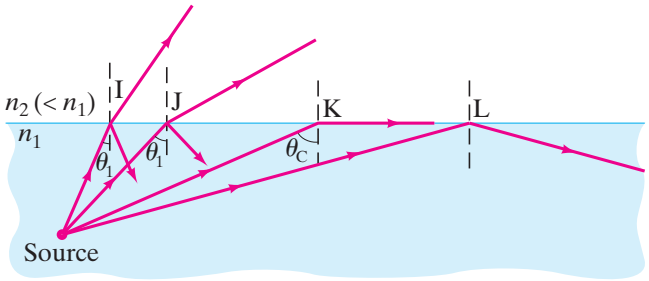


FIGURE 32–31 Since $n_2 < n_1$, light rays are totally internally reflected if the incident angle $\theta_1 > \theta_C$, as for ray L. If $\theta_1 < \theta_C$, as for rays I and J, only a part of the light is reflected, and the rest is refracted.

From Snell's law, the critical angle θ_C is given by

$$\sin\theta_C = \frac{n_2}{n_1}\sin 90^\circ = \frac{n_2}{n_1}. \qquad \textbf{(32–7)}$$

For any incident angle less than θ_C, there will be a refracted ray, although part of the light will also be reflected at the boundary. However, for incident angles θ_1 greater than θ_C, Snell's law would tell us that $\sin\theta_2$ $(= n_1 \sin\theta_1/n_2)$ would be greater than 1.00 when $n_2 < n_1$. Yet the sine of an angle can never be greater than 1.00. In this case there is no refracted ray at all, and *all of the light is reflected*, as for ray L in Fig. 32–31. This effect is called **total internal reflection**. Total internal reflection occurs only when light strikes a boundary where the medium beyond has a *lower* index of refraction.

⚠ CAUTION
Total internal reflection (occurs only if refractive index is smaller *beyond boundary)*

CONCEPTUAL EXAMPLE 32–11 **View up from under water.** Describe what a person would see who looked up at the world from beneath the perfectly smooth surface of a lake or swimming pool.

RESPONSE For an air–water interface, the critical angle is given by

$$\sin\theta_C = \frac{1.00}{1.33} = 0.750.$$

Therefore, $\theta_C = 49°$. Thus the person would see the outside world compressed into a circle whose edge makes a 49° angle with the vertical. Beyond this angle, the person would see reflections from the sides and bottom of the lake or pool (Fig. 32–32).

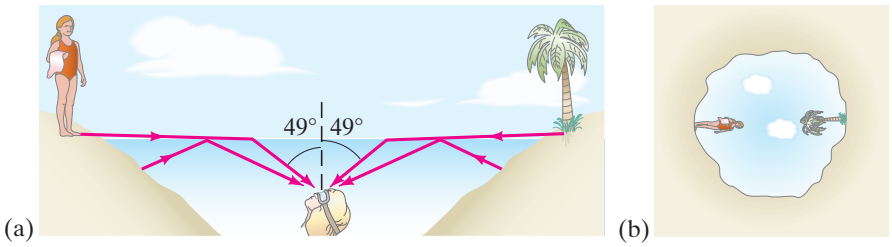


FIGURE 32–32 (a) Light rays entering submerged person's eye, and (b) view looking upward from beneath the water (the surface of the water must be very smooth). Example 32–11.

EXERCISE D Light traveling in air strikes a glass surface with $n = 1.48$. For what range of angles will total internal reflection occur?

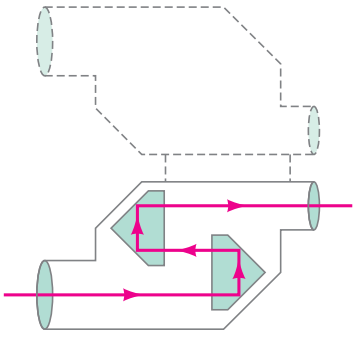
EXERCISE E Fill a sink with water. Place a friend's waterproof watch just below the surface with the watch's flat crystal parallel to the water surface. From above you can still see the watch reading. As you move your head to one side far enough what do you see, and why?

Diamonds achieve their brilliance (Fig. 32–33) from a combination of dispersion and total internal reflection. Because diamonds have a very high index of refraction of about 2.4, the critical angle for total internal reflection is only 25°.



FIGURE 32–33 Diamond.

FIGURE 32–34 Total internal reflection of light by prisms in binoculars.



The light dispersed into a spectrum inside the diamond therefore strikes many of the internal surfaces of the diamond before it strikes one at less than 25° and emerges. After many such reflections, the light has traveled far enough that the colors have become sufficiently separated to be seen individually and brilliantly by the eye after leaving the diamond.

Many optical instruments, such as binoculars, use total internal reflection within a prism to reflect light. The advantage is that very nearly 100% of the light is reflected, whereas even the best mirrors reflect somewhat less. Thus the image is brighter, especially after several reflections. For glass with $n = 1.50$, $\theta_C = 41.8°$. Therefore, 45° prisms will reflect all the light internally, if oriented as shown in the binoculars of Fig. 32–34.

EXERCISE F What would happen if we immersed the 45° glass prisms in Fig. 32–34 in water?

Fiber Optics; Medical Instruments

Total internal reflection is the principle behind **fiber optics**. Glass and plastic fibers as thin as a few micrometers in diameter are commonly used. A bundle of such slender transparent fibers is called a **light pipe** or **fiber-optic cable**. Light[†] can be transmitted along the fiber with almost no loss because of total internal reflection. Figure 32–35 shows how light traveling down a thin fiber makes only glancing collisions with the walls so that total internal reflection occurs. Even if the light pipe is bent gently into a complicated shape, the critical angle still won't be exceeded, so light is transmitted practically undiminished to the other end. Very small losses do occur, mainly by reflection at the ends and absorption within the fiber. The angles in Fig. 32–35 are very small in the thin fibers so the light travels nearly parallel to the walls; its speed is $c/n = c/1.5 \approx \frac{2}{3}c$ for glass fibers.[‡]

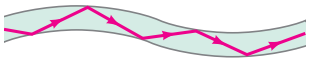


FIGURE 32–35 Light reflected totally at the interior surface of a glass or transparent plastic fiber.

Important applications of fiber-optic cables are in communications and medicine. They are used in place of wire to carry telephone calls, video signals, and computer data. The signal is a modulated light beam (a light beam whose intensity can be varied—see Figs. 31–18 and 31–19) and data is transmitted at a much higher rate and with less loss and less interference than an electrical signal in a copper wire. Fibers have been developed that can support several hundred separate wavelengths, each modulated to carry more than 500 gigabits $(5\times10^{11}\text{ bits})$ of information per second. That amounts to more than 50 terabits $(5\times10^{13}\text{ bits})$ per second for one hundred wavelengths.

PHYSICS APPLIED

Fiber optics in communications and medicine—bronchoscopes, colonoscopes, endoscopes

The use of fiber optics to transmit a clear picture is particularly useful in medicine, Fig. 32–36. For example, a patient's lungs can be examined by inserting a fiber-optic cable known as a **bronchoscope** through the mouth and down the bronchial tube. Light is sent down an outer set of fibers to illuminate the lungs. The reflected light returns up a central core set of fibers. Light directly in front of each fiber travels up that fiber. At the opposite end, a viewer sees a series of bright and dark spots, much like a TV screen—that is, a picture of what lies at the opposite end. Lenses are used at each end of the cable. The image may be viewed directly or on a monitor screen or film. The fibers must be optically insulated from one another, usually by a thin coating of material with index of refraction less than that of the fiber. The more fibers there are, and the thinner they are, the more detailed the picture. Such instruments, including bronchoscopes, colonoscopes (for viewing the colon), and **endoscopes** (stomach or other organs), are extremely useful for examining hard-to-reach places.

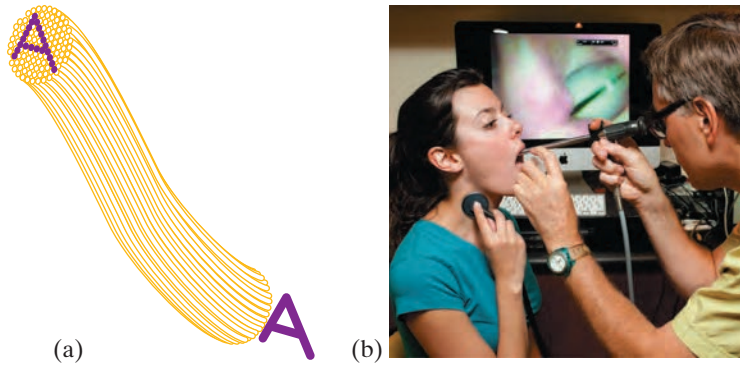


FIGURE 32–36 (a) How a fiber-optic image is made. (b) Example of a fiber-optic device inserted through the mouth to view the vocal cords, with image on screen.

[†]Fiber-optic devices use not only visible light but also infrared light, ultraviolet light, and microwaves.

[‡]A few years back, some Wall Street "traders" made money utilizing the finite speed of light, $\approx \frac{2}{3}c \approx 200\text{ km}$ per millisecond (ms). Actual transactions took place some distance from New York City. By intercepting large stock orders before they arrived, these techies could make their own orders that took advantage of their foreknowledge. It was "high-frequency trading."

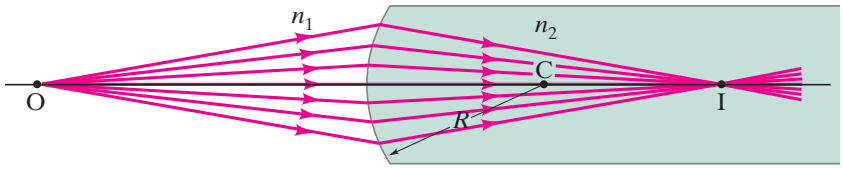


FIGURE 32–37 Rays from a point O on an object will be focused at a single image point I by a spherical boundary between two transparent materials $(n_2 > n_1)$, as long as the rays make small angles with the axis.

*32–10 Refraction at a Spherical Surface

We now examine the refraction of rays at the spherical surface of a transparent material. Such a surface could be one face of a lens or the cornea of the eye. To be general, let us consider an object which is located in a medium whose index of refraction is n_1 (it may or may not be air), and rays from each point on the object can enter a medium whose index of refraction is n_2. The radius of curvature of the spherical boundary is R, and its center of curvature is at point C, Fig. 32–37. We now show that all rays leaving a point O on the object will be focused at a single point I, the image point, if we consider only *paraxial* rays: rays that make a small angle with the axis. To do so, we consider a single ray that leaves point O as shown in Fig. 32–38, and reaches the spherical surface at point P.

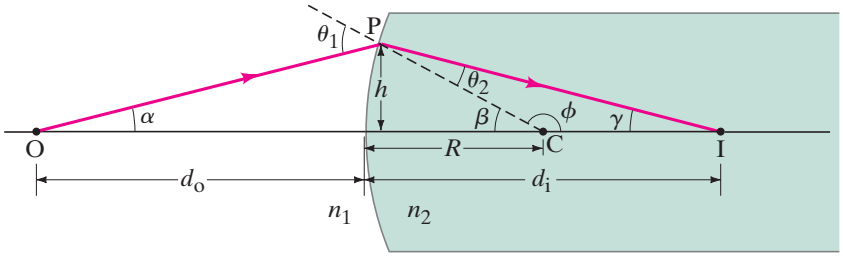


FIGURE 32–38 Diagram for showing that all paraxial rays from O focus at the same point I $(n_2 > n_1)$.

From Snell's law, Eq. 32–5, we have

$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$

We are assuming that angles θ_1, θ_2, α, β, and γ are small, so $\sin\theta \approx \theta$ (in radians), and Snell's law becomes, approximately,

$$n_1\theta_1 = n_2\theta_2.$$

Also, $\beta + \phi = 180°$ and $\theta_2 + \gamma + \phi = 180°$, so

$$\beta = \gamma + \theta_2,$$

or

$$\theta_2 = \beta - \gamma.$$

Similarly for triangle OPC,

$$\theta_1 = \alpha + \beta.$$

The θ_1 and θ_2 in these last two equations can be substituted into $n_1\theta_1 = n_2\theta_2$ to yield

$$n_1\alpha + n_2\gamma = (n_2 - n_1)\beta.$$

Since we are considering only the case of small angles, we can write, approximately,

$$\alpha = \frac{h}{d_o}, \qquad \beta = \frac{h}{R}, \qquad \gamma = \frac{h}{d_i},$$

where d_o and d_i are the object and image distances and h is the height as shown in Fig. 32–38. We substitute these into the previous equation, divide through

by h, and obtain

$$\frac{n_1}{d_o} + \frac{n_2}{d_i} = \frac{n_2 - n_1}{R}. \qquad \textbf{(32–8)}$$

For a given object distance d_o, this equation tells us d_i, the image distance, does not depend on the angle of a ray. Hence all paraxial rays meet at the same point I (within the accuracy of the approximation $\sin\theta \approx \theta$). This is true only for rays that make small angles with the axis and with each other, and is equivalent to assuming that the width of the refracting spherical surface is small compared to its radius of curvature. If this assumption is not true, the rays will not converge to a point; there will be spherical aberration, just as there is for a mirror (see Fig. 32–13), and the image will be blurry. (Spherical aberration will be discussed further in Section 33–10.)

We derived Eq. 32–8 using Fig. 32–38 for which the spherical surface is convex (as viewed by the incoming ray). It is also valid for a concave surface—as can be seen using Fig. 32–39—if we use the following sign conventions:

1. If the surface is convex (so the center of curvature C is on the side of the surface opposite to that from which the light comes), R is positive. If the surface is concave (C on the same side from which the light comes), R is negative.
2. The image distance, d_i, follows the same convention: positive if on the opposite side from where the light comes, negative if on the same side.
3. The object distance is positive if on the same side from which the light comes (this is the normal case, although when several surfaces bend the light it may not be so); otherwise it is negative.

For the case shown in Fig. 32–39 with a concave surface, both R and d_i are negative when used in Eq. 32–8. Note, in this case, that the image is virtual.

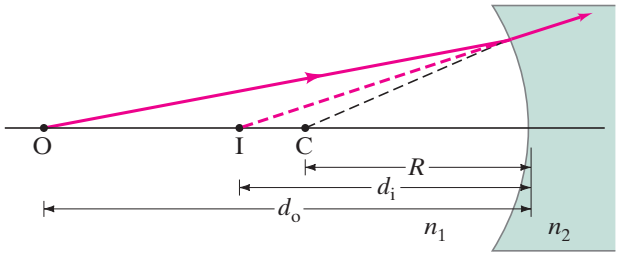


FIGURE 32–39 Rays from O refracted by a concave surface form a virtual image $(n_2 > n_1)$. Per our conventions, $R < 0$, $d_i < 0$, $d_o > 0$.

EXAMPLE 32–12 Apparent depth II. A person looks vertically down into a 1.0-m-deep pool. How deep does the water appear to be?

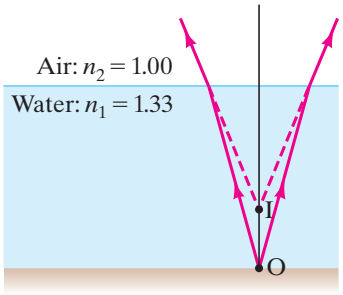
APPROACH Example 32–9 solved this problem using Snell's law. Here we use Eq. 32–8.

SOLUTION A ray diagram is shown in Fig. 32–40. Point O represents a point on the pool's bottom. The rays diverge and appear to come from point I, the image. We have $d_o = 1.0\,\text{m}$ and, for a flat surface, $R = \infty$. Then Eq. 32–8 becomes

$$\frac{1.33}{1.0\,\text{m}} + \frac{1.00}{d_i} = \frac{(1.00 - 1.33)}{\infty} = 0.$$

Hence $d_i = -(1.0\,\text{m})/(1.33) = -0.75\,\text{m}$. So the pool appears to be only three-fourths as deep as it actually is, the same result we found in Example 32–9. The minus sign tells us the image point I is on the same side of the surface as O, and the image is virtual. At angles other than vertical, this conclusion must be modified.

FIGURE 32–40 Example 32–12.



EXAMPLE 32–13 **A spherical "lens."** A point source of light is placed at a distance of 25.0 cm from the center of a glass sphere $(n = 1.5)$ of radius 10.0 cm, Fig. 32–41. Find the image of the source (= object).

APPROACH As shown in Fig. 32–41, there are two refractions, and we treat them successively, one at a time. The light rays from the source first refract from the convex glass surface facing the source. We analyze this first refraction, treating it as in Fig. 32–37, ignoring the back side of the sphere. Then we treat the back surface, using the first image as the object for the second surface. To use Eq. 32–8 we need to consider only paraxial rays, so we need to block light from entering too high or too low into the sphere (the vertical brown lines).

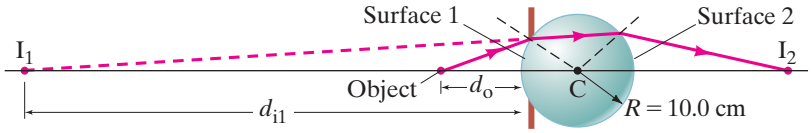


FIGURE 32–41 Example 32–13.

SOLUTION Using Eq. 32–8 (assuming paraxial rays) with $n_1 = 1.0$, $n_2 = 1.5$, $R = 10.0$ cm, and $d_o = 25.0\text{ cm} - 10.0\text{ cm} = 15.0$ cm, we solve for the image distance as formed at surface 1, d_{i1}:

$$\frac{1}{d_{i1}} = \frac{1}{n_2}\left(\frac{(n_2 - n_1)}{R} - \frac{n_1}{d_o}\right) = \frac{1}{1.5}\left(\frac{1.5 - 1.0}{10.0\text{ cm}} - \frac{1.0}{15.0\text{ cm}}\right) = -\frac{1}{90.0\text{ cm}}.$$

Thus, the image I_1 of the first refraction is located 90.0 cm *to the left* of the front surface, and is virtual. This image (I_1) now serves as the "object" for the refraction occurring at the back surface (surface 2) of the sphere. This surface is concave so $R = -10.0$ cm, and we consider a ray close to the axis. Then the object distance (from surface 2) is $d_{o2} = 90.0\text{ cm} + 2(10.0\text{ cm}) = 110.0$ cm, and Eq. 32–8 yields, with $n_1 = 1.5$, $n_2 = 1.0$,

$$\frac{1}{d_{i2}} = \frac{1}{1.0}\left(\frac{1.0 - 1.5}{-10.0\text{ cm}} - \frac{1.5}{110.0\text{ cm}}\right) = \frac{4.0}{110.0\text{ cm}}$$

so $d_{i2} = 110.0\text{ cm}/4.0 \approx 28$ cm. Thus, the final image is located a distance 28 cm from the back side of the sphere, or 38 cm from the center of the sphere.

Summary

Light appears to travel along straight-line paths, called **rays**, through uniform transparent materials including air and glass. When light reflects from a flat surface, the *angle of reflection equals the angle of incidence*. This **law of reflection** explains why mirrors can form **images**.

In a **plane mirror**, the image is virtual, upright, the same size as the object, and located as far behind the mirror as the object is in front.

A **spherical mirror** can be concave or convex. A **concave** spherical mirror focuses parallel rays of light (light from a very distant object) to a point called the **focal point**. The distance of this point from the mirror is the **focal length** f of the mirror and

$$f = \frac{r}{2} \qquad \textbf{(32–1)}$$

where r is the radius of curvature of the mirror.

Parallel rays falling on a **convex mirror** reflect from the mirror as if they diverged from a common point behind the mirror. The distance of this point from the mirror is the focal length and is considered negative for a convex mirror.

For a given object, the approximate position and size of the image formed by a mirror can be found by ray tracing. More precisely, the relation between image and object distances, d_i and d_o, and the focal length f, is given by the **mirror equation**:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}. \qquad \textbf{(32–2)}$$

The ratio of image height h_i to object height h_o, which equals the lateral magnification m of a spherical mirror, is

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}. \qquad \textbf{(32–3)}$$

If the rays that converge to form an image actually pass through the image, so the image would appear on a screen or film placed there, the image is said to be a **real image**. If the light rays do not actually pass through the image, the image is a **virtual image**.

The speed of light v depends on the **index of refraction**, n, of the material:

$$n = \frac{c}{v}, \qquad \textbf{(32–4)}$$

where c is the speed of light in vacuum.

When light passes from one transparent medium into another, the rays bend or refract. The **law of refraction** (**Snell's law**) states that

$$n_1 \sin\theta_1 = n_2 \sin\theta_2, \qquad \textbf{(32–5)}$$

where n_1 and θ_1 are the index of refraction and angle with the normal (perpendicular) to the surface for the incident ray, and n_2 and θ_2 are for the refracted ray.

When light of wavelength λ enters a medium with index of refraction n, the wavelength is reduced to

$$\lambda_n = \frac{\lambda}{n}. \qquad \textbf{(32–6)}$$

The frequency does not change.

The frequency or wavelength of light determines its color. The **visible spectrum** in air extends from about 400 nm (violet) to about 750 nm (red).

Glass prisms spread white light out into its constituent colors because the index of refraction varies with wavelength, a phenomenon known as **dispersion**.

When light rays reach the boundary of a material where the index of refraction decreases, the rays will be **totally internally reflected** if the incident angle, θ_1, is such that Snell's law would predict $\sin\theta_2 > 1$. This occurs if θ_1 exceeds the critical angle θ_C given by

$$\sin\theta_C = \frac{n_2}{n_1}. \qquad \textbf{(32–7)}$$

Questions

1. What would be the appearance of the Moon if it had (*a*) a rough surface; (*b*) a polished mirrorlike surface?
2. Archimedes is said to have burned the whole Roman fleet in the harbor of Syracuse, Italy, by focusing the rays of the Sun with a huge spherical mirror. Is this[†] reasonable?
3. What is the focal length of a plane mirror? What is the magnification of a plane mirror?
4. An object is placed along the principal axis of a spherical mirror. The magnification of the object is −3.0. Is the image real or virtual, inverted or upright? Is the mirror concave or convex? On which side of the mirror is the image located?
5. Using the rules for the three rays discussed with reference to Fig. 32–15, draw ray 2 for Fig. 32–19b.
6. Does the mirror equation, Eq. 32–2, hold for a plane mirror? Explain.
7. If a concave mirror produces a real image, is the image necessarily inverted?
8. How might you determine the speed of light in a solid, rectangular, transparent object?
9. When you look at the Moon's reflection from a ripply sea, it appears elongated (Fig. 32–42). Explain.
10. How can you "see" a round drop of water on a table even though the water is transparent and colorless?
11. What is the angle of refraction when a light ray is incident perpendicular to the boundary between two transparent materials?
12. When you look down into a swimming pool or a lake, are you likely to overestimate or underestimate its depth? Explain. How does the apparent depth vary with the viewing angle? (Use ray diagrams.)
13. Draw a ray diagram to show why a stick or straw looks bent when part of it is under water (Fig. 32–23).
14. Although a plane mirror appears to reverse left and right, it doesn't reverse up and down. Discuss why this happens, noting that front to back is also reversed. Also discuss what happens if, while standing, you look up vertically at a horizontal mirror on the ceiling.
15. When a wide beam of parallel light enters water at an angle, the beam broadens. Explain.
16. You look into an aquarium and view a fish inside. One ray of light from the fish as it exits the tank is shown in Fig. 32–43. The apparent position of the fish is also shown. In the drawing, indicate the approximate position of the *actual* fish. Briefly justify your answer.



FIGURE 32–42
Question 9.

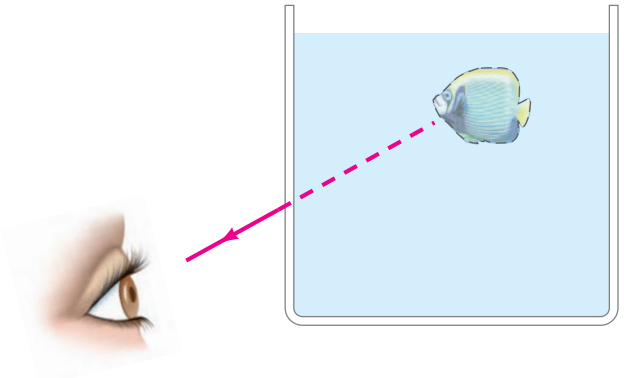


FIGURE 32–43 Question 16.

[†]Students at MIT did a feasibility study. See web.mit.edu/2.009/www/experiments/deathray/10_ArchimedesResult.html.

17. A child looks into a pool to see how deep it is. She then drops a small toy into the pool to help decide how deep the pool is. After this careful investigation, she decides it is safe to jump in—only to discover the water is over her head. What went wrong with her interpretation of her experiment?

18. What type of mirror is shown in Fig. 32–44? Explain.



FIGURE 32–44 Question 18 and Problem 15.

19. Can a light ray traveling in air be totally reflected when it strikes a smooth water surface if the incident angle is chosen correctly? Explain.

20. When you look up at an object in air from beneath the surface in a swimming pool, does the object appear to be the same size as when you see it directly in air? Explain. (See Fig. 32–32.)

21. Light rays from stars (including our Sun) always bend toward the vertical direction as they pass through the Earth's atmosphere. (*a*) Why does this make sense? (*b*) What can you conclude about the apparent positions of stars as viewed from Earth?

MisConceptual Questions

1. Suppose you are standing about 3 m in front of a mirror. You can see yourself just from the top of your head to your waist, where the bottom of the mirror cuts off the rest of your image. If you walk one step closer to the mirror
   (*a*) you will not be able to see any more of your image.
   (*b*) you will be able to see more of your image, below your waist.
   (*c*) you will see less of your image, with the cutoff rising to be above your waist.

2. When light traveling in air enters water at an angle,
   (*a*) it changes direction by bending away from the surface.
   (*b*) it slows down.
   (*c*) its wavelength changes.
   (*d*) Statements (*a*) and (*b*) are true but (*c*) is not.
   (*e*) Statements (*a*), (*b*), and (*c*) are all true.
   (*f*) None of these statements is true.

3. If you want to see yourself full-length in a mirror, what must be the *minimum* size of the mirror? (Try it and see to be sure.)
   (*a*) Your full height and body width.
   (*b*) Your full height and twice your body width.
   (*c*) Half of your height and half body width.
   (*d*) Half of your height and twice your body width.
   (*e*) It depends on how far from the mirror you stand.

4. You want to create a spotlight that will shine a bright beam of light with all of the light rays parallel to each other. You have a large concave spherical mirror and a small lightbulb. Where should you place the lightbulb?
   (*a*) At the focal point of the mirror.
   (*b*) At the radius of curvature of the mirror.
   (*c*) At any point, because all rays bouncing off the mirror will be parallel.
   (*d*) None of the above; you can't make parallel rays with a concave mirror.

5. Which of the following statements about the focal point of a concave spherical mirror is true?
   (*a*) Parallel rays incident on the mirror are reflected through the focal point.
   (*b*) The image produced by the mirror is at the focal point.
   (*c*) The location of the focal point depends on where the object is located.
   (*d*) None of the above.

6. When you look at a fish in a still stream from the bank, the fish appears shallower than it really is due to refraction. From directly above, it appears
   (*a*) deeper than it really is.
   (*b*) at its actual depth.
   (*c*) shallower than its real depth.
   (*d*) It depends on your height above the water.

7. Images are
   (*a*) always smaller than the object.
   (*b*) always inverted relative to the object.
   (*c*) always at the point where light rays from a point object would cross.
   (*d*) always the same size as the object.
   (*e*) None of the above.

8. If you shine a light through an optical fiber, why does it come out the end but not out the sides?
   (*a*) It does come out the sides, but this effect is not obvious because the sides are so much longer than the ends.
   (*b*) The sides are mirrored, so the light reflects.
   (*c*) Total internal reflection makes the light reflect from the sides.
   (*d*) The light flows along the length of the fiber, never touching the sides.

9. When moonlight strikes the surface of a calm lake, what happens to this light?
   (*a*) All of it reflects from the water surface back to the air.
   (*b*) Some of it reflects back to the air; some enters the water.
   (*c*) All of it enters the water.
   (*d*) All of it disappears via absorption by water molecules.

10. The colors in a rainbow are caused by
(*a*) the interaction of the light reflected from different raindrops.
(*b*) different amounts of absorption for light of different colors by the water in the raindrops.
(*c*) different amounts of refraction for light of different colors by the water in the raindrops.
(*d*) the downward motion of the raindrops.

11. An observer directly above two closely spaced lines on paper, views them through an angled piece of glass, Fig. 32–45. To the observer, the lines appear
(*a*) shifted to the right.
(*b*) shifted to the left.
(*c*) spaced farther apart.
(*d*) spaced closer together.
(*e*) the same as if the glass were not there.

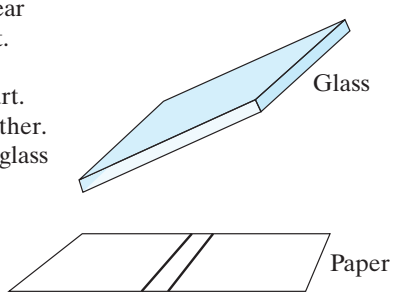


FIGURE 32–45
MisConceptual Question 11.

12. To shoot a swimming fish with an intense light beam from a *laser gun*, you should aim
(*a*) directly at the image.
(*b*) slightly above the image.
(*c*) slightly below the image.

13. Parallel light rays cross interfaces from medium 1 into medium 2 and then into medium 3 as shown in Fig. 32–46. What can we say about the relative sizes of the indices of refraction of these media?
(*a*) $n_1 > n_2 > n_3$.
(*b*) $n_3 > n_2 > n_1$.
(*c*) $n_2 > n_3 > n_1$.
(*d*) $n_1 > n_3 > n_2$.
(*e*) $n_2 > n_1 > n_3$.
(*f*) None of the above.

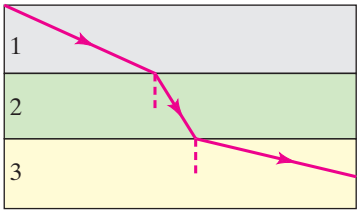


FIGURE 32–46
MisConceptual Question 13.

Problems

32–2 Reflection; Plane Mirrors

1. (I) When you look at yourself in a 60-cm-tall plane mirror, you see the same amount of your body whether you are close to the mirror or far away. (Try it and see.) Use ray diagrams to show why this should be true.

2. (I) Suppose that you want to take a photograph of yourself as you look at your image in a mirror 2.4 m away. For what distance should the camera lens be focused?

3. (II) Two plane mirrors meet at a 135° angle, Fig. 32–47. If light rays strike one mirror at 32° as shown, at what angle ϕ do they leave the second mirror?

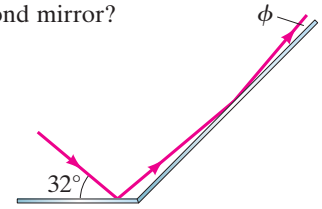


FIGURE 32–47
Problem 3.

4. (II) A person whose eyes are 1.64 m above the floor stands 2.60 m in front of a vertical plane mirror whose bottom edge is 38 cm above the floor, Fig. 32–48. What is the horizontal distance x, from the base of the wall supporting the mirror to the nearest point on the floor that can be seen reflected in the mirror?

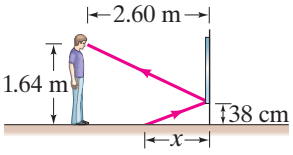


FIGURE 32–48
Problem 4.

5. (II) Show that if two plane mirrors meet at an angle ϕ, a single ray reflected successively from both mirrors is deflected through an angle of 2ϕ independent of the incident angle. Assume $\phi < 90°$ and that *only* two reflections, one from each mirror, take place.

6. (II) (*a*) Use a ray diagram, as we did in Example 32–3, to show that this Chapter's opening photograph, page 926, cannot be right-side up considering the position of the Sun. (*b*) Confirm this result by using ray diagrams involving the people shown.

7. (II) Suppose you are 96 cm from a plane mirror. What area of the mirror is used to reflect the rays entering one eye from a point on the tip of your nose if your pupil diameter is 4.5 mm?

8. (II) Stand up two plane mirrors so they form a 90.0° angle as in Fig. 32–49. When you look into this double mirror, you see yourself as others see you, instead of reversed as in a single mirror. Make a ray diagram to show how this occurs.

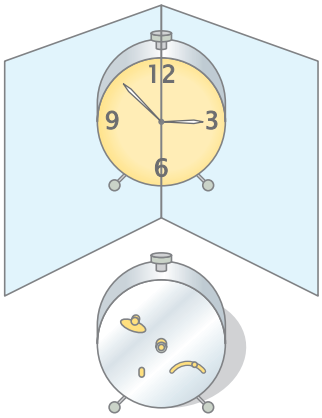


FIGURE 32–49
Problems 8 and 9.

9. (III) Suppose a third mirror is placed beneath the two shown in Fig. 32–49, so that all three are perpendicular to each other. (*a*) Show that for such a "corner reflector," any incident ray will return in its original direction after three reflections. (*b*) What happens if it makes only two reflections?

32–3 to 32–5 Spherical Mirrors

10. (I) A solar cooker, really a concave mirror pointed at the Sun, focuses the Sun's rays 17.2 cm in front of the mirror. What is the radius of the spherical surface from which the mirror was made?

11. (I) How far from a concave mirror (radius 30.0 cm) must an object be placed if its image is to be at infinity?

12. (II) The lateral magnification of a convex mirror is +0.75 for objects 3.2 m from the mirror. What is the focal length of this mirror?

13. (II) A small candle is 41 cm from a concave mirror having a radius of curvature of 24 cm. (*a*) What is the focal length of the mirror? (*b*) Where will the image of the candle be located? (*c*) Will the image be upright or inverted?

14. (II) You look at yourself in a shiny 8.4-cm-diameter Christmas tree ball. If your face is 25.0 cm away from the ball's front surface, where is your image? Is it real or virtual? Is it upright or inverted?

15. (II) A mirror at an amusement park shows an upright image of any person who stands 2.1 m in front of it. If the image height is three times the person's height, what is the radius of curvature of the mirror? (See Fig. 32–44.)

16. (II) A dentist wants a small mirror that, when 2.00 cm from a tooth, will produce a 3.0× upright image. What kind of mirror must be used and what must its radius of curvature be?

17. (II) A 4.2-cm-tall object is placed 26 cm in front of a spherical mirror. It is desired to produce a virtual image that is upright and 3.2 cm tall. (*a*) What type of mirror should be used? (*b*) Where is the image located? (*c*) What is the focal length of the mirror? (*d*) What is the radius of curvature of the mirror?

18. (II) You are standing 2.7 m from a convex security mirror in a store. You estimate the height of your image to be half what it would be in a plane mirror at the same place. Estimate the radius of curvature of the mirror.

19. (II) An object 4.0 mm high is placed 18 cm from a convex mirror of radius of curvature 18 cm. (*a*) Show by ray tracing that the image is virtual, and estimate the image distance. (*b*) Show that the (negative) image distance can be computed from Eq. 32–2 using a focal length of −9.0 cm. (*c*) Compute the image size, using Eq. 32–3.

20. (II) The image of a distant tree is virtual, very small, and 13.0 cm behind a curved mirror. What kind of mirror is it, and what is its radius of curvature?

21. (II) Use two techniques, (*a*) a ray diagram, and (*b*) the mirror equation, to show that the magnitude of the lateral magnification of a concave mirror is less than 1 if the object is beyond the center of curvature C $(d_o > r)$, and is greater than 1 if the object is within C $(d_o < r)$.

22. (II) Show, using a ray diagram, that the lateral magnification m of a convex mirror is $m = -d_i/d_o$, just as for a concave mirror. [*Hint*: Consider a ray from the top of the object that reflects at the center of the mirror.]

23. (II) Use ray diagrams to show that the mirror equation, Eq. 32–2, is valid for a convex mirror as long as f is considered negative.

24. (II) In Example 32–4, show that if the object is moved 10.0 cm farther from the concave mirror, the object's image size will equal the object's actual size. Stated as a multiple of the focal length, what is the object distance for this "actual-sized image" situation?

25. (II) (*a*) Where should an object be placed in front of a concave mirror so that it produces an image at the same location as the object? (*b*) Is the image real or virtual? (*c*) Is the image inverted or upright? (*d*) What is the lateral magnification of the image?

26. (II) Let the focal length of a convex mirror be written as $f = -|f|$. Show that the lateral magnification m of an object a distance d_o from this mirror is given by $m = |f|/(d_o + |f|)$. Based on this relation, explain why your nose looks bigger than the rest of your face when looking into a convex mirror.

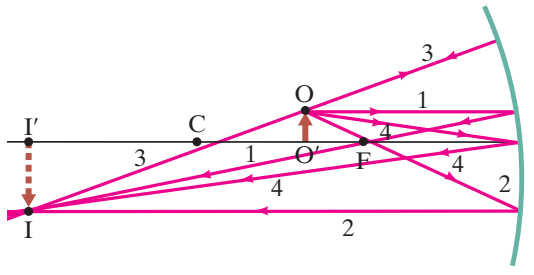
27. (II) A spherical mirror of focal length f produces an image of an object with lateral magnification m. (*a*) Show that the object is a distance $d_o = f\left(1 - \frac{1}{m}\right)$ from the reflecting side of the mirror. (*b*) Use the relation in part (*a*) to show that, no matter where an object is placed in front of a convex mirror, its image will have a lateral magnification in the range $0 \le m \le +1$.

28. (II) A concave mirror has focal length f. When an object is placed a distance $d_o > f$ from this mirror, a real image with lateral magnification m is formed. (*a*) Show that $m = f/(f - d_o)$. (*b*) Sketch m vs. d_o over the range $f < d_o < +\infty$ where $f = 0.45$ m. (*c*) For what value of d_o will the real image have the same size as the object? (*d*) To obtain a real image that is much larger than the object, in what general region should the object be placed relative to the mirror?

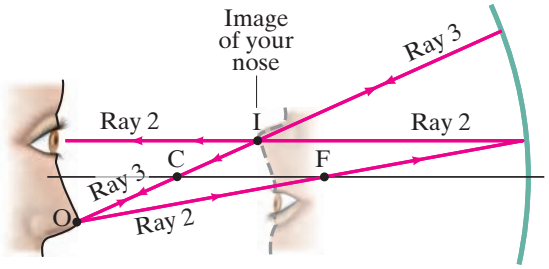
29. (II) When walking toward a concave mirror you notice that your image flips at a distance of 0.80 m from the mirror. What is the radius of curvature of the mirror? [*Hint*: Carefully examine Section 32–4.]

30. (II) A shaving or makeup mirror is designed to magnify your face by a factor of 1.8 (when compared to a flat mirror) when your face is placed 20.0 cm in front of it. (*a*) What type of mirror is it? (*b*) Describe the type of image that it makes of your face. (*c*) Calculate the required radius of curvature for the mirror.

31. (III) Two ray diagrams involving a concave mirror are shown in Figs. 32–50a and b. (*a*) and (*b*): you are looking into the mirror. The object is OO′, which includes your eye O and maybe a point on your nose, O′. Describe the situation for each diagram, where the image is, and what you see. In part (*a*), explain why what you actually see—a blurry image—is not inverted as shown. In part (*b*) note that rays moving left from I toward your eye are *diverging* so make a sharp image. (*c*) In Fig. 32–50a, suppose the object OO′, is *not* you, but is some ordinary object, such as an arrow, and your eye is on the far left, beyond the image I I′ so diverging rays enter your eye. Describe what you see in the mirror in detail and any other pertinent details.



(a)



(b)

FIGURE 32–50 Problem 31.

32. (III) An object is placed a distance r in front of a wall, where r exactly equals the radius of curvature of a certain concave mirror. At what distance from the wall should this mirror be placed so that a real image of the object is formed on the wall? What is the lateral magnification of the image?

33. (III) A short thin object (like a short length of wire) of length ℓ is placed along the axis of a spherical mirror (perpendicular to the glass surface). Show that its image has length $\ell' = m^2\ell$ so the **longitudinal magnification** is equal to $-m^2$ where m is the normal "lateral" magnification, Eq. 32–3. Why the minus sign? [*Hint*: Find the image positions for both ends of the wire, and assume ℓ is very small.]

32–6 Index of Refraction

34. (I) The speed of light in ice is 2.29×10^8 m/s. What is the index of refraction of ice?

35. (I) What is the speed of light in (*a*) ethyl alcohol, (*b*) lucite, (*c*) crown glass?

36. (II) The speed of light in a certain substance is 78% of its value in water. What is the index of refraction of that substance?

32–7 Refraction: Snell's Law

37. (I) A diver shines a narrow flashlight beam upward from beneath the water at a 32.8° angle to the vertical. At what angle does the light leave the water?

38. (I) A flashlight beam strikes the surface of a pane of glass ($n = 1.56$) at a 69° angle to the normal. What is the angle of refraction?

39. (I) Rays of the Sun are seen to make a 38.0° angle to the vertical beneath the water. At what angle above the horizon is the Sun?

40. (I) A light beam coming from an underwater spotlight exits the water at an angle of refraction equal to 52.0°. At what angle of incidence did it hit the air-water interface from below the surface?

41. (II) A beam of light in air strikes a slab of glass ($n = 1.62$) and is partially reflected and partially refracted. Determine the angle of incidence if the angle of reflection is twice the angle of refraction.

42. (II) A light beam strikes a 2.5-cm-thick piece of plastic with a refractive index of 1.62 at a 45° angle. The plastic is on top of a 3.8-cm-thick piece of glass for which $n = 1.47$. What is the distance D in Fig. 32–51?

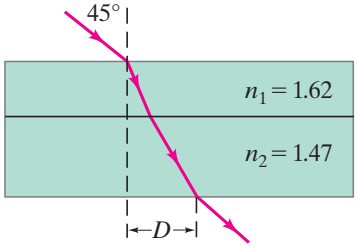


FIGURE 32–51
Problem 42.

43. (II) An aquarium filled with water has flat glass sides whose index of refraction is 1.51. A beam of light from outside the aquarium strikes the glass at a 43.5° angle to the perpendicular (Fig. 32–52). What is the angle of this light ray when it enters (*a*) the glass, and then (*b*) the water? (*c*) What would be the refracted angle if the ray entered the water directly?

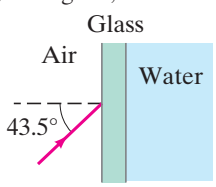


FIGURE 32–52
Problem 43.

44. (II) In searching the bottom of a pool at night, a watchman shines a narrow beam of light from his flashlight, 1.3 m above the water level, onto the surface of the water at a point 2.8 m from his foot at the edge of the pool (Fig. 32–53). Where does the spot of light hit the bottom of the pool which is 2.1 m deep? Measure from the bottom of the wall beneath his foot.

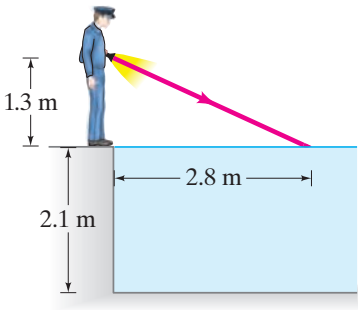


FIGURE 32–53
Problem 44.

45. (II) The block of glass ($n = 1.5$) shown in cross section in Fig. 32–54 is surrounded by air. A ray of light enters the block at its left-hand face with incident angle θ_1 and reemerges into the air from the right-hand face directed parallel to the block's base. Determine θ_1.

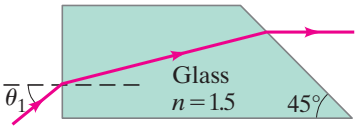


FIGURE 32–54 Problem 45.

46. (II) A laser beam of circular cross-section has diameter $d_1 = 3.0$ mm. It strikes a flat air–glass surface at an incident angle $\theta_1 = 22°$. The refracted beam in the glass ($n = 1.50$) is oval. Determine the length and width of the oval laser beam in the glass.

47. (II) A light ray is incident on an equilateral glass prism at a 45.0° angle to one face, Fig. 32–55. Calculate the angle at which the light ray emerges from the opposite face. Assume that $n = 1.58$.

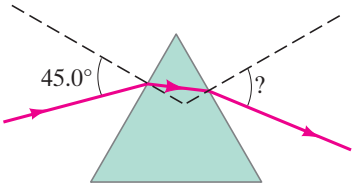


FIGURE 32–55 Problems 47 and 61.

48. (II) Show in general that for a light beam incident on a uniform layer of transparent material, as in Fig. 32–24, the direction of the emerging beam is parallel to the incident beam, independent of the incident angle θ. Assume the air on the two sides of the transparent material is the same.

49. (III) A light ray is incident on a flat piece of glass with index of refraction n as in Fig. 32–24. Show that if the incident angle θ is small, the emerging ray is displaced a distance $d = t\theta(n-1)/n$, where t is the thickness of the glass, θ is in radians, and d is the perpendicular distance between the incident ray and the (dashed) line of the emerging ray (Fig. 32–24).

32–8 Visible Spectrum; Dispersion

50. (I) By what percent is the speed of blue light (450 nm) less than the speed of red light (680 nm), in silicate flint glass (see Fig. 32–28)?

51. (I) A light beam strikes a piece of glass at a 55.00° incident angle. The beam contains two wavelengths, 450.0 nm and 700.0 nm, for which the index of refraction of the glass is 1.4831 and 1.4754, respectively. What is the angle between the two refracted beams?

52. (II) A parallel beam of light containing two wavelengths, $\lambda_1 = 461$ nm and $\lambda_2 = 656$ nm, enters the silicate flint glass of an equilateral prism as shown in Fig. 32–56. At what angle does each beam leave the prism (give angle with normal to the face)? See Fig. 32–28.

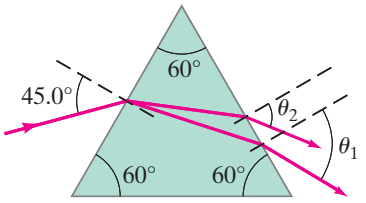


FIGURE 32–56 Problem 52.

53. (III) For visible light, the index of refraction n of glass is roughly 1.5, although this value varies by about 1% across the visible range. Consider a ray of white light incident from air at angle θ_1 onto a flat piece of glass. (*a*) Show that, upon entering the glass, the visible colors contained in this incident ray will be dispersed over a range of refracted angles given approximately by

$$\Delta\theta_2 \approx \frac{\sin\theta_1}{\sqrt{n^2 - \sin^2\theta_1}}\frac{\Delta n}{n}.$$

[*Hint*: For x in radians, $(d/dx)(\sin^{-1}x) = 1/\sqrt{1-x^2}$.] (*b*) If $\theta_1 = 0°$, what is $\Delta\theta_2$ in degrees? (*c*) If $\theta_1 = 90°$, what is $\Delta\theta_2$ in degrees?

32–9 Total Internal Reflection

54. (I) What is the critical angle for the interface between water and crown glass? To be internally reflected, the light must start in which material?

55. (I) The critical angle for a certain liquid–air surface is 52.6°. What is the index of refraction of the liquid?

56. (II) A beam of light is emitted in a pool of water from a depth of 65.0 cm. Where must it strike the air–water interface, relative to the spot directly above it, in order that the light does *not* exit the water?

57. (II) A ray of light, after entering a light fiber, reflects at an angle of 14.5° with the long axis of the fiber, as in Fig. 32–57. Calculate the distance along the axis of the fiber that the light ray travels between successive reflections off the sides of the fiber. Assume that the fiber has an index of refraction of 1.55 and is 1.60×10^{-4} m in diameter.



FIGURE 32–57 Problem 57.

58. (II) A beam of light is emitted 7.7 cm beneath the surface of a liquid and strikes the surface 7.2 cm from the point directly above the source. If total internal reflection occurs, what can you say about the index of refraction of the liquid?

59. (II) Two rays A and B travel down a cylindrical optical fiber of diameter $d = 75.0\,\mu$m, length $\ell = 1.0$ km, and index of refraction $n_1 = 1.465$. Ray A travels a straight path down the fiber's axis, whereas ray B propagates down the fiber by repeated reflections at the critical angle each time it impinges on the fiber's boundary. Determine the extra time Δt it takes for ray B to travel down the entire fiber in comparison with ray A (Fig. 32–58), assuming (*a*) the fiber is surrounded by air, (*b*) the fiber is surrounded by a cylindrical glass "cladding" with index of refraction $n_2 = 1.460$.

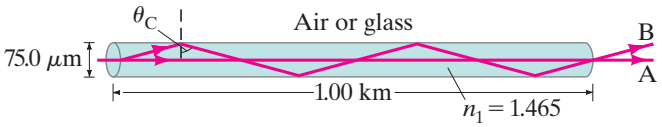


FIGURE 32–58 Problem 59.

60. (II) (*a*) What is the minimum index of refraction for a glass or plastic prism to be used in binoculars (Fig. 32–34) so that total internal reflection occurs at 45°? (*b*) Will binoculars work if their prisms (assume $n = 1.58$) are immersed in water? (*c*) What minimum n is needed if the prisms are immersed in water?

61. (III) Suppose a ray strikes the left face of the prism in Fig. 32–55 at 45.0° as shown, but is totally internally reflected at the opposite side. If the apex angle (at the top) is $\theta = 63.0°$, what can you say about the index of refraction of the prism?

62. (III) A beam of light enters the end of an optic fiber as shown in Fig. 32–59. (*a*) Show that we can guarantee total internal reflection at the side surface of the material (at point A), if the index of refraction is greater than about 1.42. In other words, regardless of the angle α, the light beam reflects back into the material at point A, assuming air outside. (*b*) What if the fiber is immersed in water?

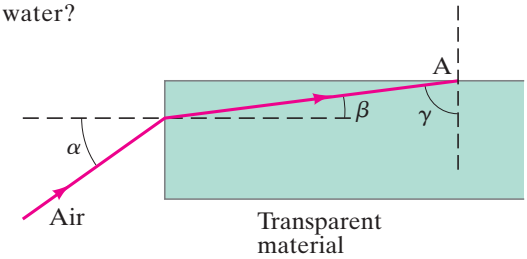


FIGURE 32–59 Problem 62.

*32–10 Refraction at Spherical Surface

***63.** (II) A 13.0-cm-thick plane piece of glass ($n = 1.54$) lies on the surface of a 12.0-cm-deep pool of water. How far below the top of the glass does the bottom of the pool seem, as viewed from directly above?

***64.** (II) A fish is swimming in water inside a thin spherical glass bowl of uniform thickness. Assuming the radius of curvature of the bowl is 32.0 cm, locate the image of the fish if the fish is located: (*a*) at the center of the bowl; (*b*) 20.0 cm from the side of the bowl between the observer and the center of the bowl. Assume the fish is small.

***65.** (III) In Section 32–8, we derived Eq. 32–8 for a convex spherical surface with $n_2 > n_1$. Using the same conventions and using diagrams similar to Fig. 32–38, show that Eq. 32–8 is valid also for (*a*) a convex spherical surface with $n_2 < n_1$, (*b*) a concave spherical surface with $n_2 > n_1$, and (*c*) a concave spherical surface with $n_2 < n_1$.

General Problems

66. Two identical concave mirrors are set facing each other 1.0 m apart. A small lightbulb is placed halfway between the mirrors. A small piece of paper placed just to the left of the bulb prevents light from the bulb from directly shining on the left mirror, but light reflected from the right mirror still reaches the left mirror. A good image of the bulb appears on the left side of the piece of paper. What is the focal length of the mirrors?

67. A pulse of light takes 2.63 ns (see Table 1–4) to travel 0.500 m in a certain material. Determine the material's index of refraction, and identify this material.

68. (*a*) A plane mirror can be considered a limiting case of a spherical mirror. Specify what this limit is. (*b*) Determine an equation that relates the image and object distances in this limit of a plane mirror. (*c*) Determine the magnification of a plane mirror in this same limit. (*d*) Are your results in parts (*b*) and (*c*) consistent with the discussion of Section 32–2 on plane mirrors?

69. The critical angle of a certain piece of plastic in air is $\theta_C = 35.8°$. What is the critical angle of the same plastic if it is immersed in water?

70. Two plane mirrors are facing each other 2.2 m apart as in Fig. 32–60. You stand 1.5 m away from one of these mirrors and look into it. You will see multiple images of yourself. (*a*) How far away from you are the first three images of yourself in the mirror in front of you? (*b*) Are these first three images facing toward you or away from you?

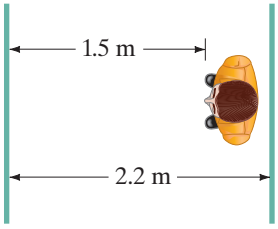


FIGURE 32–60 Problem 70.

71. We wish to determine the depth of a swimming pool filled with water by measuring the width ($x = 5.20$ m) and then noting that the bottom edge of the pool is just visible at an angle of 13.0° above the horizontal as shown in Fig. 32–61. Calculate the depth of the pool.

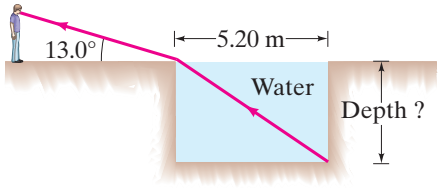


FIGURE 32–61 Problem 71.

72. A 1.80-m-tall person stands 4.20 m from a convex mirror and notices that he looks precisely half as tall as he does in a plane mirror placed at the same distance. What is the radius of curvature of the convex mirror? (Assume that $\sin\theta \approx \theta$.) [*Hint*: The viewing *angle* is half.]

73. Each student in a physics lab is assigned to find the location where a bright object may be placed in order that a concave mirror, with radius of curvature $r = 54$ cm, will produce an image three times the size of the object. Two students complete the assignment at different times using identical equipment, but when they compare notes later, they discover that their answers for the object distance are not the same. Explain why they do not necessarily need to repeat the lab, and justify your response with a calculation.

74. A slab of thickness D, whose two faces are parallel, has index of refraction n. A ray of light incident from air onto one face of the slab at incident angle θ_1 splits into two rays A and B. Ray A reflects directly back into the air, while B travels a total distance ℓ within the slab before reemerging from the slab's face a distance d from its point of entry (Fig. 32–62). (*a*) Derive expressions for ℓ and d in terms of D, n, and θ_1. (*b*) For normal incidence (i.e., $\theta_1 = 0°$) show that your expressions yield the expected values for ℓ and d.

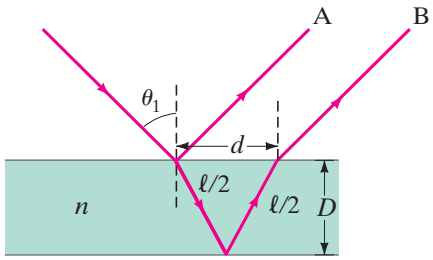


FIGURE 32–62 Problem 74.

75. A kaleidoscope makes symmetric patterns with two plane mirrors having a 60° angle between them as shown in Fig. 32–63. Draw the location of the images (some of them images of images) of the object placed midway between the mirrors.



FIGURE 32–63 Problem 75.

76. Sunlight is reflected off the Moon. How long does it take that light to reach us from the Moon?

77. The label on a laser says it produces light of wavelength 670 nm. The laser beam passes through a block of plastic for which $n = 1.57$. What is the wavelength of the light inside the plastic?

78. The index of refraction for a piece of plastic is 1.50 for light of wavelength 550 nm. A beam of white light strikes the plastic surface at a 45° incident angle. The angle between the violet light and the red light in the refracted beam is 0.24°. What is the percentage difference in the plastic's index of refraction from one end of the visible spectrum to the other?

79. When light passes through a prism, the angle that the refracted ray makes relative to the incident ray is called the deviation angle δ, Fig. 32–64. Show that this angle is a minimum when the ray passes through the prism symmetrically, perpendicular to the bisector of the apex angle ϕ, and show that the minimum deviation angle, δ_m, is related to the prism's index of refraction n by

$$n = \frac{\sin\frac{1}{2}(\phi + \delta_m)}{\sin\phi/2}.$$

[*Hint*: For θ in radians, $(d/d\theta)(\sin^{-1}\theta) = 1/\sqrt{1-\theta^2}$.]

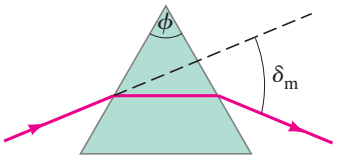


FIGURE 32–64 Problems 79 and 80.

80. If the apex angle of a prism is $\phi = 68°$ (see Fig. 32–63), what is the minimum incident angle for a ray if it is to emerge from the opposite side (i.e., not be totally internally reflected), given $n = 1.58$?

81. **Fermat's principle** states that "light travels between two points along the path that requires the least time, as compared to other nearby paths." From Fermat's principle derive (*a*) the law of reflection $(\theta_i = \theta_r)$ and (*b*) the law of refraction (Snell's law). [*Hint*: Choose two appropriate points so that a ray between them can undergo reflection or refraction. Draw a rough path for a ray between these points, and write down an expression of the time required for light to travel the arbitrary path chosen. Then take the derivative to find the minimum.]

82. A ray of light with wavelength λ is incident from air at precisely $60°$ $(=\theta)$ on a spherical water drop of radius r and index of refraction n (which depends on λ). When the ray reemerges into the air from the far side of the drop, it has been deflected an angle ϕ from its original direction as shown in Fig. 32–65. By how much does the value of ϕ for violet light $(n = 1.341)$ differ from the value for red light $(n = 1.330)$?

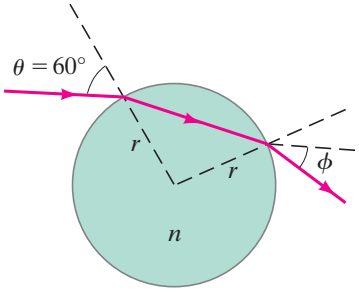


FIGURE 32–65
Problem 82.

83. A triangular prism made of crown glass $(n = 1.52)$ with base angles of 26.0° is surrounded by air. If parallel rays are incident normally on its base as shown in Fig. 32–66, what is the angle ϕ between the two emerging rays?

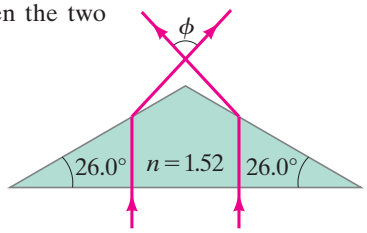


FIGURE 32–66
Problem 83.

84. Figure 32–67 shows a liquid-detecting prism device that might be used inside a washing machine. If no liquid covers the prism's hypotenuse, total internal reflection of the beam from the light source produces a large signal in the light sensor. If liquid covers the hypotenuse, some light escapes from the prism into the liquid and the light sensor's signal decreases. Thus a large signal from the light sensor indicates the absence of liquid in the reservoir. Determine the allowable range for the prism's index of refraction n.

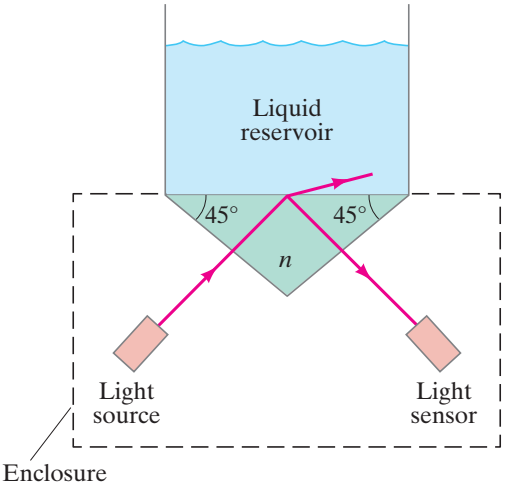


FIGURE 32–67 Problem 84.

85. Light from a laser (in air) strikes the exact center of one face of a solid glass cube $(n = 1.40)$ at an angle θ relative to the normal. The refracted beam travels inside the glass until it strikes an adjacent face of the cube. The original angle of incidence θ is such that no light exits the cube where the beam strikes the second face. What is the maximum value θ can have?

86. An optical fiber is a long transparent cylinder of diameter d and index of refraction n. If this fiber is bent sharply, some light hitting the side of the cylinder may escape rather than reflect back into the fiber (Fig. 32–68). What is the smallest radius r at a short bent section for which total internal reflection will be assured for light initially travelling parallel to the axis of the fiber?

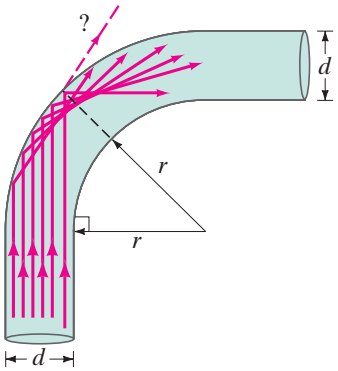


FIGURE 32–68 Problem 86.

87. An object is placed 21 cm from a certain mirror. The image is half the height of the object, inverted, and real. How far is the image from the mirror, and what is the radius of curvature of the mirror?

88. The end faces of a cylindrical glass rod $(n = 1.51)$ are perpendicular to the sides. Show that a light ray entering an end face at any angle will be totally internally reflected inside the rod when it strikes the sides. Assume the rod is in air. What if it were in water?

89. The index of refraction, n, of crown flint glass at different wavelengths (λ) of light are given in the Table below.

λ (nm)	1060	546.1	365.0	312.5
n	1.50586	1.51978	1.54251	1.5600

Make a graph of n versus λ. The variation in index of refraction with wavelength is given by the Cauchy equation $n = A + B/\lambda^2$. Make another graph of n versus $1/\lambda^2$ and determine the constants A and B for the glass by fitting the data with a straight line.

***90.** Suppose Fig. 32–37 shows a cylindrical rod whose end has a radius of curvature $R = 2.0$ cm, and the rod is immersed in water with index of refraction of 1.33. The rod has index of refraction 1.49. Find the location and height of the image of an object 2.0 mm high located 23 cm away from the rod.

91. A coin lies at the bottom of a 0.95-m-deep pool. If a viewer sees it at a 45° angle, where is the image of the coin, relative to the coin? [*Hint*: The image is found by tracing back to the intersection of two rays.]

92. Consider a ray of sunlight incident from air on a spherical raindrop of radius r and index of refraction n. Defining θ to be its incident angle, the ray then follows the path shown in Fig. 32–69, exiting the drop at a "scattering angle" ϕ compared with its original incoming direction. (*a*) Show that $\phi = 180° + 2\theta - 4\sin^{-1}(\sin\theta/n)$. (*b*) The parallel rays of sunlight illuminate a raindrop with rays of all possible incident angles from 0° to 90°. Plot ϕ vs. θ in the range $0° \le \theta \le 90°$, in 0.5° steps, assuming $n = 1.33$ as is appropriate for water at visible-light wavelengths. (*c*) From your plot, you should find that a fairly large fraction of the incident angles have nearly the same scattering angle. Approximately what fraction of the possible incident angles is within roughly 1° of $\phi = 139°$? [This subset of incident rays is what creates the *rainbow*. Wavelength-dependent variations in n cause the rainbow to form at slightly different ϕ for the various visible colors.]

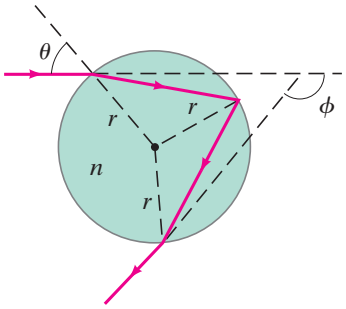


FIGURE 32–69 Problem 92.

93. The paint used on highway signs often contains small transparent spheres which provide nighttime illumination of the sign's lettering by retro-reflecting vehicle headlight beams. Consider a light ray from air incident on one such sphere of radius r and index of refraction n. Let θ be its incident angle, and let the ray follow the path shown in Fig. 32–70, so that the ray exits the sphere in the direction exactly antiparallel to its incoming direction. Considering only rays for which $\sin\theta$ can be approximated as θ, determine the required value for n.

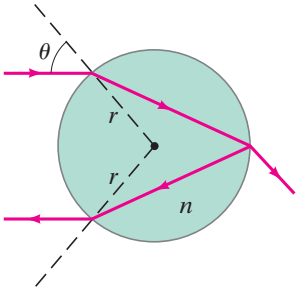


FIGURE 32–70 Problem 93.

94. You hold a small flat mirror 0.50 m in front of you and can see your reflection twice in that mirror because there is a full-length mirror 1.0 m behind you (Fig. 32–71). Determine the distance of each image from you.

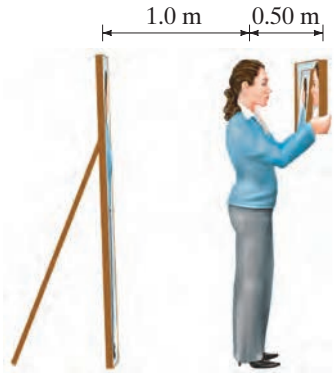


FIGURE 32–71 Problem 94.

95. (*a*) Did the person we see in Fig. 32–72 shoot the picture we are looking at? We see her in three different mirrors. Describe (*b*) what type of mirror each is, and (*c*) her position relative to the focal point and center of curvature.



FIGURE 32–72 Problem 95.

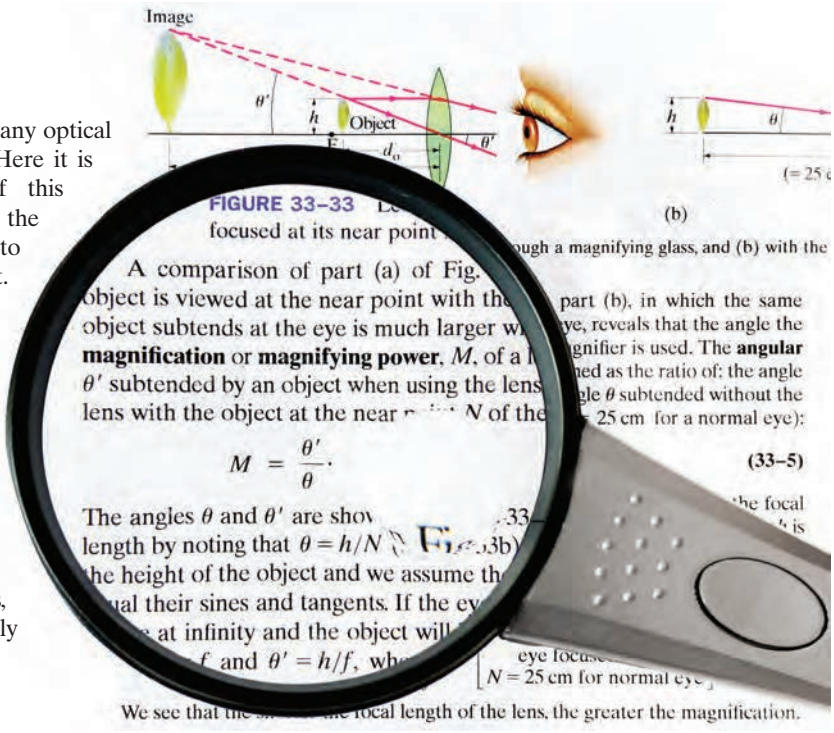
96. If the truck in Example 32–7 is viewed by the driver in the passenger side mirror (as in Fig. 32–20), d_{eye} would be about 1.5 m. What then would the angular magnification be?

ANSWERS TO EXERCISES

A: No.
B: (*b*).
C: Toward.
D: None.
E: The face of the watch becomes shiny. You can no longer see the numbers; total internal reflection.
F: No total internal reflection, $\theta_C > 45°$.

This magnifying glass is one of the many optical devices we discuss in this Chapter. Here it is magnifying part of page 979 of this Chapter, which describes how the magnifying glass works according to geometry and the ray model of light.

In this Chapter we examine thin lenses, seeing how they form images by refraction. We see how to determine image position as a function of object position and the focal length of the lens, using the ray model and geometry. We then examine optical devices including film and digital cameras, the human eye, eyeglasses, simple magnifiers, telescopes, microscopes, and finally optical aberrations.



CHAPTER 33

Lenses and Optical Instruments

CONTENTS

33 – 1 Thin Lenses; Ray Tracing and Focal Length
33 – 2 The Thin Lens Equation
33 – 3 Combinations of Lenses
33 – 4 Lensmaker's Equation
33 – 5 Cameras: Film and Digital
33 – 6 The Human Eye; Corrective Lenses
33 – 7 Magnifying Glass
33 – 8 Telescopes
33 – 9 Compound Microscope
33–10 Aberrations of Lenses and Mirrors

CHAPTER-OPENING QUESTION—Guess now!

A converging lens, like the type used in a magnifying glass,

(a) always produces a magnified image (image taller than the object).
(b) can also produce an image smaller than the object.
(c) always produces an upright image.
(d) can also produce an inverted image (upside down).
(e) None of these statements are true.

The laws of reflection and refraction are the basis for explaining the operation of many optical instruments. We now discuss and analyze simple lenses using the ray model of optics as in the previous Chapter. We then analyze a number of optical instruments, from the magnifying glass and the human eye to telescopes and microscopes. The importance of lenses is that they form images of objects—see Fig. 33–1.

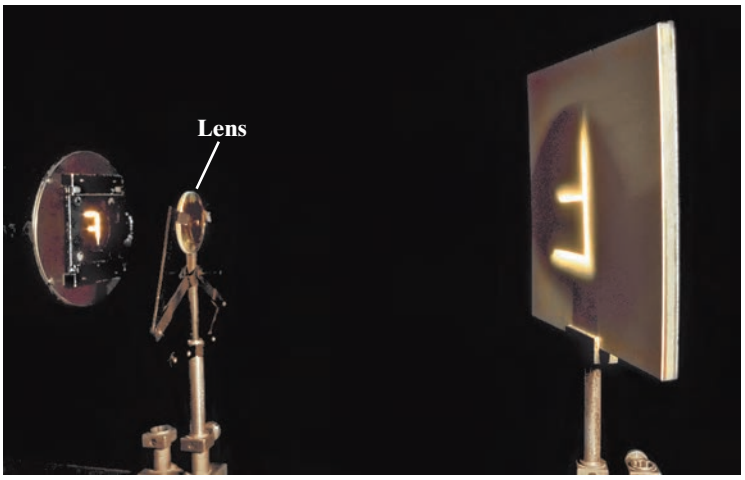


FIGURE 33–1 Converging lens (in holder) forms an image (large "F" on screen at right) of a bright object (illuminated "F" at the left).

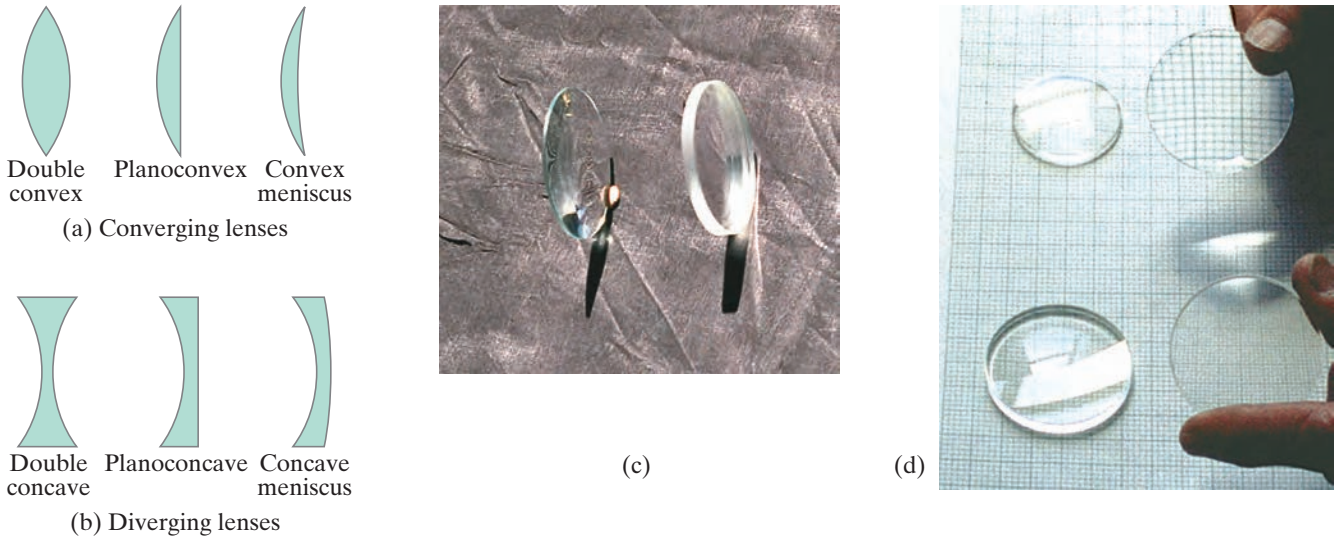


FIGURE 33–2 (a) Converging lenses and (b) diverging lenses, shown in cross section. Converging lenses are thicker at the center whereas diverging lenses are thicker at the edges. (c) Photo of a converging lens (on the left) and a diverging lens (right). (d) Converging lenses (top), and diverging lenses (below), lying flat, and raised off the paper to form images.

33–1 Thin Lenses; Ray Tracing and Focal Length

The most important simple optical device is the thin lens. The development of optical devices using lenses dates to the sixteenth and seventeenth centuries, although the earliest record of eyeglasses dates from the late thirteenth century. Today we find lenses in eyeglasses, cameras, magnifying glasses, telescopes, binoculars, microscopes, and medical instruments. A thin lens is usually circular, and its two faces are portions of a sphere. (Cylindrical faces are also possible, but we will concentrate on spherical.) The two faces can be concave, convex, or plane. Several types are shown in Fig. 33–2a and b in cross section.

Consider parallel rays striking the double convex lens shown in cross section in Fig. 33–3a. We assume the lens is made of transparent material such as glass or transparent plastic, with index of refraction greater than that of the air outside. The **axis** of a lens is a straight line passing through the center of the lens and perpendicular to its two surfaces (Fig. 33–3). From Snell's law, we can see that each ray in Fig. 33–3a is bent toward the axis when the ray enters the lens and again when it leaves the lens at the back surface. (Note the dashed lines indicating the normals to each surface for the top ray.) If rays parallel to the axis fall on a thin lens, they will be focused to a point called the **focal point**, F. This will not be precisely true for a lens with spherical surfaces. But it will be very nearly true—that is, parallel rays will be focused to a tiny region that is nearly a point—if the diameter of the lens is small compared to the radii of curvature of the two lens surfaces. This criterion is satisfied by a **thin lens**, one that is very thin compared to its diameter, and we consider only thin lenses here.

The rays from a point on a distant object are essentially parallel—see Fig. 32–12. Therefore we can say that *the focal point is the image point for an object at infinity on the lens axis*, Fig. 33–3a. Thus, the focal point of a lens can be found by locating the point where the Sun's rays (or those from some other distant object) are brought to a sharp image, Fig. 33–4. The distance of the focal point from the center of the lens is called the **focal length**, f, Fig. 33–3a. A lens can be turned around so that light can pass through it from the opposite side. The *focal length is the same on both sides*, as we shall see later, even if the curvatures of the two lens surfaces are different. If parallel rays fall on a lens at an angle, as in Fig. 33–3b, they focus at a point F_a. The plane (or surface) containing all focus points, such as F and F_a in Fig. 33–3b, is called the **focal plane** of the lens.

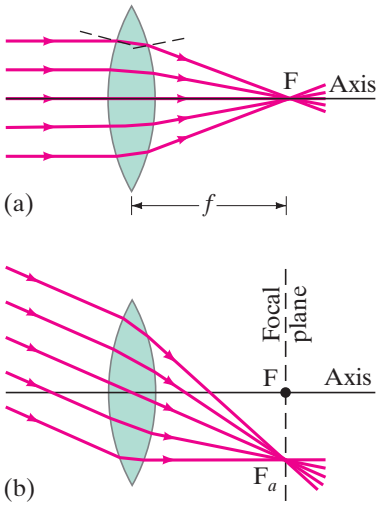


FIGURE 33–3 Parallel rays are brought to a focus by a converging thin lens.

FIGURE 33–4 Image of the Sun burning wood.



Any lens (in air) that is thicker in the center than at the edges will make parallel rays converge to a point, and is called a **converging lens** (see Fig. 33–2a). Lenses that are thinner in the center than at the edges (Fig. 33–2b) are called **diverging lenses** because they make parallel light diverge, as shown in Fig. 33–5.

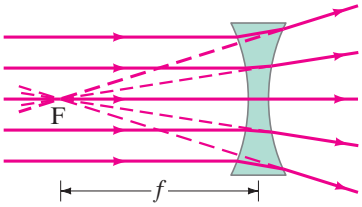


FIGURE 33–5 Diverging lens.

The focal point, F, of a diverging lens is defined as that point from which refracted rays, originating from parallel incident rays, seem to emerge as shown in Fig. 33–5. And the distance from F to the center of the lens is called the **focal length**, f, just as for a converging lens.

Optometrists and ophthalmologists, instead of using the focal length, use the reciprocal of the focal length to specify the strength of eyeglass (or contact) lenses. This is called the **power**, P, of a lens:

$$P = \frac{1}{f}. \qquad \textbf{(33–1)}$$

The unit for lens power is the **diopter** (D), which is an inverse meter: $1\,\text{D} = 1\,\text{m}^{-1}$. For example, a 20-cm-focal-length lens has a power $P = 1/(0.20\,\text{m}) = 5.0\,\text{D}$. We will mainly use the focal length, but we will refer again to the power of a lens when we discuss eyeglass lenses in Section 33–6.

The most important parameter of a lens is its focal length f, which is the same on both sides of the lens. For a converging lens, f can be measured by finding the image point for the Sun or other distant objects. Once f is known, the image position can be determined for any object. To find the image point by drawing rays would be difficult if we had to determine the refractive angles at the front surface of the lens and again at the back surface where the ray exits. We can save ourselves a lot of effort by making use of certain facts we already know, such as that a ray parallel to the axis of the lens passes (after refraction) through the focal point. To determine an image point, we can consider only the three rays indicated in Fig. 33–6, which uses an arrow (on the left) as the object, and a converging lens forming an image (dashed arrow) to the right. These rays, emanating from a single point on the object, are drawn as if the lens were infinitely thin, and we show only a single sharp bend at the center line of the lens instead of the refractions at each surface. These three rays are drawn as follows:

RAY DIAGRAM
Finding the image position formed by a thin lens

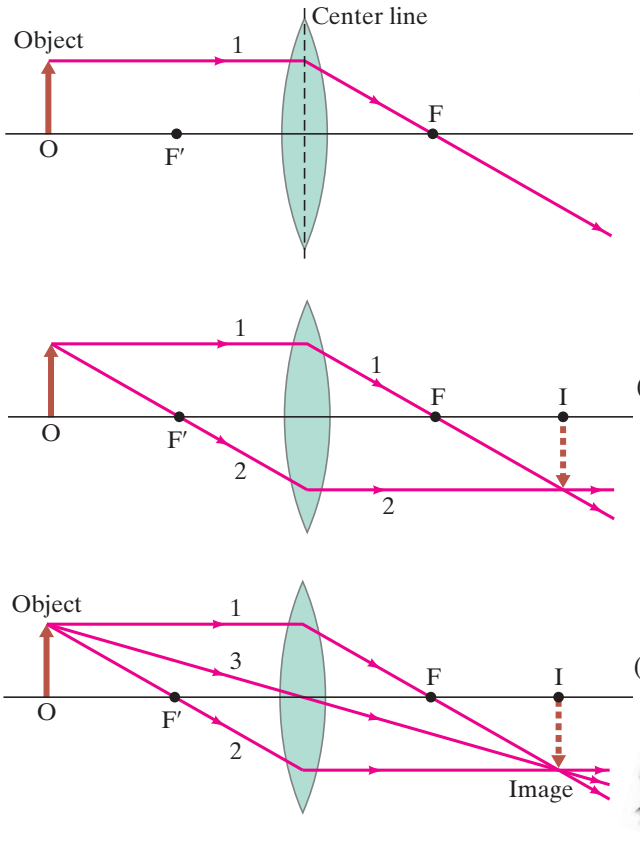
Ray 1 is drawn parallel to the axis, Fig. 33–6a; therefore it is refracted by the lens so that it passes along a line through the focal point F behind the lens. (See also Fig. 33–3a.)

Ray 2 is drawn to pass through the other focal point F′ (front side of lens in Fig. 33–6) and emerge from the lens parallel to the axis, Fig. 33–6b. (In reverse it would be a parallel ray going left and passing through F′.)

Ray 3 is directed toward the very center of the lens, where the two surfaces are essentially parallel to each other, Fig. 33–6c. This ray therefore emerges from the lens at the same angle as it entered. The ray would be displaced slightly to one side, as we saw in Example 32–8 (Fig. 32–24), but since we assume the lens is thin, we draw ray 3 straight through as shown.

The point where these three rays cross is the image point for that object point. Actually, any two of these rays will suffice to locate the image point, but drawing the third ray can serve as a check.

Using these three rays for one object point, we can find the image point for that point of the object (the top of the arrow in Fig. 33–6). The image points for all other points on the object can be found similarly to determine the complete



(a) Ray 1 leaves one point on object going parallel to the axis, then refracts through focal point behind the lens.

(b) Ray 2 passes through F′ in front of the lens; therefore it is parallel to the axis behind the lens.

(c) Ray 3 passes straight through the center of the lens (assumed very thin).

FIGURE 33–6 Finding the image by ray tracing for a converging lens. Rays are shown leaving one point on the object (an arrow). Shown are the three most useful rays, leaving the tip of the object, for determining where the image of that point is formed. (Note that the focal points F and F′ on either side of the lens are the same distance f from the center of the lens.)

image of the object. Because the rays actually pass through the image for the case shown in Fig. 33–6, it is a **real image** (see page 929). The image could be detected on film or by an electronic camera sensor, and actually be seen on a white surface or screen placed at the position of the image (Fig. 33–7).

CONCEPTUAL EXAMPLE 33–1 **Half-blocked lens.** What happens to the image of an object if the top half of a lens is covered by a piece of cardboard?

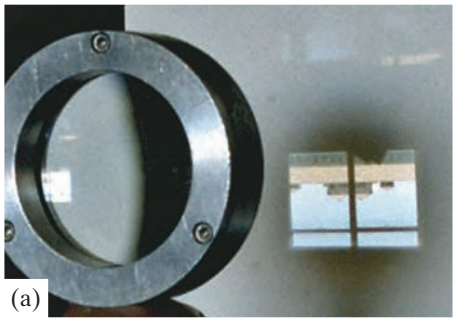
RESPONSE Let us look at the rays in Fig. 33–6. If the top half (or any half of the lens) is blocked, you might think that half the image is blocked. But in Fig. 33–6c, we see how the rays used to create the "top" of the image pass through both the top and the bottom of the lens. Only three of many rays are shown—many more rays pass through the lens, and they can form the image. You don't lose the image. But covering part of the lens cuts down on the total light received and reduces the brightness of the image.

NOTE If the lens is partially blocked by your thumb, you may notice an out of focus image of part of that thumb.

Seeing the Image

The image can also be seen directly by the eye when the eye is placed behind the image, as shown in Fig. 33–6c, so that some of the rays diverging from each point on the image can enter the eye. We can see a sharp image only for rays *diverging* from each point on the image, because we see normal objects when diverging rays from each point enter the eye, as shown in Fig. 32–1. A normal eye cannot focus converging rays; if your eye was positioned between points F and I in Fig. 33–6c, it would not see a clear image. (More about our eyes in Section 33–6.) Figure 33–7 shows an image seen (a) on a white surface and (b) directly by the eye (and a camera) placed behind the image. The eye can see both real and virtual images (see next page) as long as the eye is positioned so diverging rays enter it.

FIGURE 33–7 (a) A converging lens can form a real image (here of a distant building, upside down) on a white wall. (b) That same real image is also directly visible to the eye. [Figure 33–2d shows images (graph paper) seen by the eye made by both diverging and converging lenses.]



(a)



(b)

Diverging Lens

By drawing the same three rays emerging from a single object point, we can determine the image position formed by a diverging lens, as shown in Fig. 33–8. Note that ray 1 is drawn parallel to the axis, but does not pass through the focal point F′ behind the lens. Instead it seems to come (dashed line) from the focal point F in front of the lens. Ray 2 is directed toward F′ and is refracted parallel to the lens axis by the lens. Ray 3 passes directly through the center of the lens. The three refracted rays seem to emerge from a point on the left of the lens. This is the image point, I. Because the rays do not pass through the image, it is a **virtual image**. Note that the eye does not distinguish between real and virtual images—both are visible.

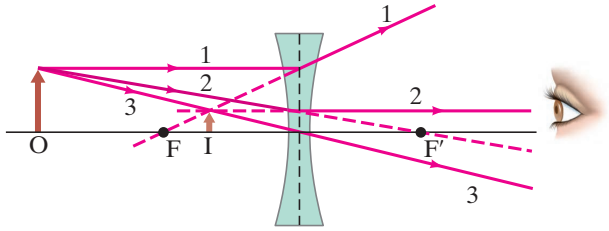


FIGURE 33–8 Finding the image by ray tracing for a diverging lens.

33–2 The Thin Lens Equation

We now derive an equation that relates the image distance to the object distance and the focal length of a thin lens. This equation will make the determination of image position quicker and more accurate than doing ray tracing. Let d_o be the object distance, the distance of the object from the center of the lens, and d_i be the image distance, the distance of the image from the center of the lens.

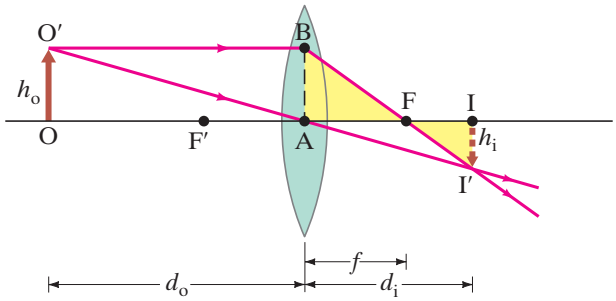


FIGURE 33–9 Deriving the lens equation for a converging lens.

Let h_o and h_i refer to the heights of the object and image. Consider the two rays shown in Fig. 33–9 for a converging lens, assumed to be very thin. The right triangles FI′I and FBA (highlighted in yellow) are similar because angle AFB equals angle IFI′. Therefore

$$\frac{h_i}{h_o} = \frac{d_i - f}{f},$$

since length AB $= h_o$. Triangles OAO′ and IAI′ are similar as well. Therefore,

$$\frac{h_i}{h_o} = \frac{d_i}{d_o}.$$

We equate the right sides of these two equations (the left sides are the same), and divide by d_i to obtain

$$\frac{1}{f} - \frac{1}{d_i} = \frac{1}{d_o}$$

or

THIN LENS EQUATION

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}. \qquad \textbf{(33–2)}$$

This is called the **thin lens equation**. It relates the image distance d_i to the object distance d_o and the focal length f. It is the most useful equation in geometric optics. (Interestingly, it is exactly the same as the mirror equation, Eq. 32–2.)

If the object is at infinity, then $1/d_o = 0$ and Eq. 33–2 gives $d_i = f$. Thus the focal length is the image distance for an object at infinity, as mentioned earlier.

We can derive the lens equation for a diverging lens using Fig. 33–10. Triangles IAI′ and OAO′ are similar; and triangles IFI′ and AFB are similar. Thus (noting that length AB $= h_o$)

$$\frac{h_i}{h_o} = \frac{d_i}{d_o}$$

and

$$\frac{h_i}{h_o} = \frac{f - d_i}{f}.$$

When we equate the right sides of these two equations and simplify, we obtain

$$\frac{1}{d_o} - \frac{1}{d_i} = -\frac{1}{f}.$$

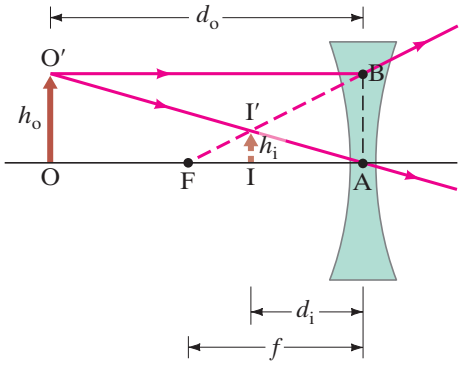


FIGURE 33–10 Deriving the lens equation for a diverging lens.

This equation becomes the same as Eq. 33–2 if we make f and d_i negative. That is, we take f to be *negative for a diverging lens*, and d_i negative when the image is on the same side of the lens as the light comes from. Thus Eq. 33–2 will be valid for both converging and diverging lenses, and for *all* situations, if we use the following **sign conventions**:

CAUTION
Focal length is negative for diverging lens

PROBLEM SOLVING
SIGN CONVENTIONS for lenses

1. The focal length is positive for converging lenses and negative for diverging lenses.
2. The object distance is positive if the object is on the side of the lens from which the light is coming. This is always the case for real objects; but when lenses are used in combination, it might not be so (see Example 33–6, where the object distance for the diverging lens is negative, and the object is "virtual" and behind the lens).
3. The image distance is positive if the image is on the opposite side of the lens from where the light is coming; if it is on the same side, d_i is negative. Equivalently, the image distance is positive for a real image (Fig. 33–9) and negative for a virtual image (Fig. 33–10).
4. The height of the image, h_i, is positive if the image is upright, and negative if the image is inverted relative to the object. (h_o is always taken as upright and positive.)

The **lateral magnification**, m, of a lens is defined as the ratio of the image height to object height, $m = h_i/h_o$. From Figs. 33–9 and 33–10 and the conventions just stated (for which we will need a minus sign), we have

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}. \qquad \textbf{(33–3)}$$

For an upright image the magnification is positive (as in Fig. 33–10), and for an inverted image the magnification is negative (Fig. 33–9).

From sign convention 1, it follows that the power (Eq. 33–1) of a converging lens, in diopters, is positive, whereas the power of a diverging lens is negative. A converging lens is sometimes referred to as a **positive lens**, and a diverging lens as a **negative lens**.

Diverging lenses (see Fig. 33–10) always produce an upright virtual image for any real object, no matter where that object is. Converging lenses can produce real (inverted) images as in Fig. 33–9, or virtual (upright) images, depending on object position, as we shall see.

PROBLEM SOLVING

Thin Lenses

1. Draw a **ray diagram**, as precise as possible, but even a rough one can serve as confirmation of analytic results. Choose one point on the object and draw at least two, or preferably three, of the easy-to-draw rays described in Figs. 33–6 and 33–8. The image point is where the rays intersect.
2. For analytic solutions, solve for unknowns in the **thin lens equation** (Eq. 33–2) and the **magnification equation** (Eq. 33–3). The thin lens equation involves reciprocals—don't forget to take the reciprocal.
3. Follow the **sign conventions** listed on page 963.
4. Check that your analytic answers are **consistent** with your ray diagram.

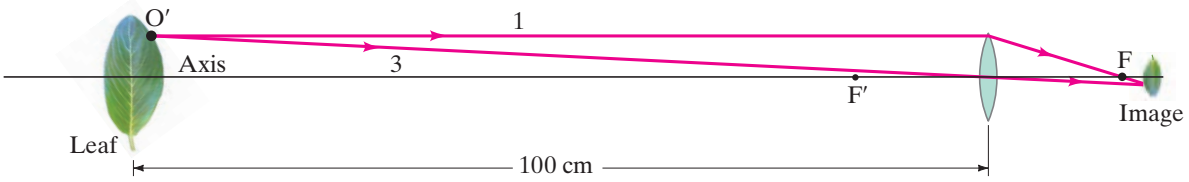


FIGURE 33–11 Example 33–2. (Not to scale.)

EXAMPLE 33–2 **Image formed by converging lens.** What is (*a*) the position, and (*b*) the size, of the image of a 7.6-cm-high leaf placed 1.00 m from a +50.0-mm-focal-length camera lens?

APPROACH We follow the steps of the Problem Solving Strategy explicitly.

SOLUTION

1. **Ray diagram**. Figure 33–11 is an approximate ray diagram, showing only rays 1 and 3 for a single point on the leaf. We see that the image ought to be a little behind the focal point F, to the right of the lens.
2. **Thin lens and magnification equations**. (*a*) We find the image position analytically using the thin lens equation, Eq. 33–2. The camera lens is converging, with $f = +5.00\,\text{cm}$, and $d_o = 100\,\text{cm}$, and so the thin lens equation gives
$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o} = \frac{1}{5.00\,\text{cm}} - \frac{1}{100\,\text{cm}} = \frac{20.0 - 1.0}{100\,\text{cm}} = \frac{19.0}{100\,\text{cm}}.$$
Then, taking the reciprocal,
$$d_i = \frac{100\,\text{cm}}{19.0} = 5.26\,\text{cm},$$
or 52.6 mm behind the lens.
(*b*) The magnification is
$$m = -\frac{d_i}{d_o} = -\frac{5.26\,\text{cm}}{100\,\text{cm}} = -0.0526,$$
so
$$h_i = mh_o = (-0.0526)(7.6\,\text{cm}) = -0.40\,\text{cm}.$$
The image is 4.0 mm high.
3. **Sign conventions**. The image distance d_i came out positive, so the image is behind the lens. The image height is $h_i = -0.40\,\text{cm}$; the minus sign means the image is inverted.
4. **Consistency**. The analytic results of steps 2 and 3 are consistent with the ray diagram, Fig. 33–11: the image is behind the lens and inverted.

NOTE Part (*a*) tells us that the image is 2.6 mm farther from the lens than the image for an object at infinity, which would equal the focal length, 50.0 mm. Indeed, when focusing a camera lens, the closer the object is to the camera, the farther the lens must be from the sensor or film.

EXERCISE A If the leaf (object) of Example 33–2 is moved farther from the lens, does the image move closer to or farther from the lens? (Don't calculate!)

EXAMPLE 33–3 **Object close to converging lens.** An object is placed 10 cm from a 15-cm-focal-length converging lens. Determine the image position and size (*a*) analytically, and (*b*) using a ray diagram.

APPROACH The object is within the focal point—closer to the lens than the focal point F as $d_o < f$. We first use Eqs. 33–2 and 33–3 to obtain an analytic solution, and then confirm with a ray diagram using the special rays 1, 2, and 3 for a single object point.

SOLUTION (*a*) Given $f = 15\,\text{cm}$ and $d_o = 10\,\text{cm}$, then

$$\frac{1}{d_i} = \frac{1}{15\,\text{cm}} - \frac{1}{10\,\text{cm}} = \frac{2-3}{30\,\text{cm}} = -\frac{1}{30\,\text{cm}},$$

and $d_i = -30\,\text{cm}$. (Remember to take the reciprocal!) Because d_i is negative, the image must be virtual and on the same side of the lens as the object (sign convention 3, page 963). The magnification

$$m = -\frac{d_i}{d_o} = -\frac{-30\,\text{cm}}{10\,\text{cm}} = 3.0.$$

The image is three times as large as the object and is upright. This lens is being used as a magnifying glass, which we discuss in more detail in Section 33–7.

(*b*) The ray diagram is shown in Fig. 33–12 and in accord with the results in part (*a*). We choose point O′ on the top of the object and draw ray 1. Ray 2, however, may take some thought: if we draw it heading toward F′, it is going the wrong way—so we have to draw it as if coming from F′ (and so dashed), striking the lens, and then going out parallel to the lens axis. We project it backward, with a dashed line, as we must do also for ray 1, in order to find where they cross. Ray 3 is drawn through the lens center, and it crosses the other two rays at the image point, I′.

NOTE From Fig. 33–12 we can see that, when an object is placed between a converging lens and its focal point, the image is virtual.

CAUTION
Don't forget to take the reciprocal

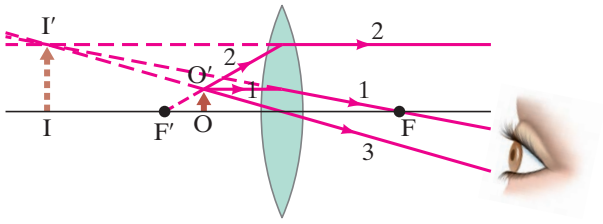


FIGURE 33–12 An object placed within the focal point of a converging lens produces a virtual image. Example 33–3.

EXERCISE B Return to the Chapter-Opening Question, page 958, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 33–4 **Diverging lens.** Where must a small insect be placed if a 25-cm-focal-length diverging lens is to form a virtual image 20 cm from the lens, on the same side as the object?

APPROACH The ray diagram is basically that of Fig. 33–10 because our lens here is diverging and our image is given as in front of the lens within the focal distance. (It would be a valuable exercise to draw the ray diagram to scale, precisely, now.) The insect's distance, d_o, can be calculated using the thin lens equation.

SOLUTION The lens is diverging, so f is negative: $f = -25\,\text{cm}$. The image distance must be negative too because the image is in front of the lens (sign conventions), so $d_i = -20\,\text{cm}$. The lens equation, Eq. 33–2, gives

$$\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_i} = -\frac{1}{25\,\text{cm}} + \frac{1}{20\,\text{cm}} = \frac{-4+5}{100\,\text{cm}} = \frac{1}{100\,\text{cm}}.$$

So the object must be 100 cm in front of the lens.

33–3 Combinations of Lenses

Many optical instruments use lenses in combination. When light passes through more than one lens, we can start the analysis by finding the image formed by the first lens as if it were alone. Then that image becomes the *object* for the second lens. Next we find the image formed by this second lens using the first image as object. This second image is the final image if there are only two lenses. The total magnification will be the product of the separate magnifications of each lens. Even if the second lens intercepts the light from the first lens before it forms an image, this technique still works.

EXAMPLE 33–5 **A two-lens system.** Two converging lenses, A and B, with focal lengths $f_A = 20.0\,\text{cm}$ and $f_B = 25.0\,\text{cm}$, are placed 80.0 cm apart, as shown in Fig. 33–13a. An object is placed 60.0 cm in front of the first lens as shown in Fig. 33–13b. Determine (*a*) the position, and (*b*) the magnification, of the final image formed by the combination of the two lenses.

APPROACH Starting at the tip of our object O, we draw rays 1, 2, and 3 for the first lens, A, and also a ray 4 which, after passing through lens A, acts for the second lens, B, as ray 3′ (through the center). We use primes now for the standard rays relative to lens B. Ray 2 for lens A exits parallel, and so is ray 1′ for lens B. To determine the position of the image I_A formed by lens A, we use Eq. 33–2 with $f_A = 20.0\,\text{cm}$ and $d_{oA} = 60.0\,\text{cm}$. The distance of I_A (lens A's image) from lens B is the object distance d_{oB} for lens B. The final image is found using the thin lens equation, this time with all distances relative to lens B. For (*b*) the magnifications are found from Eq. 33–3 for each lens in turn.

SOLUTION (*a*) The object is a distance $d_{oA} = +60.0\,\text{cm}$ from the first lens, A, and this lens forms an image whose position can be calculated using the thin lens equation:

$$\frac{1}{d_{iA}} = \frac{1}{f_A} - \frac{1}{d_{oA}} = \frac{1}{20.0\,\text{cm}} - \frac{1}{60.0\,\text{cm}} = \frac{3-1}{60.0\,\text{cm}} = \frac{1}{30.0\,\text{cm}}.$$

CAUTION
Object distance for second lens is **not** *equal to the image distance for first lens*

So the first image I_A is at $d_{iA} = 30.0\,\text{cm}$ behind the first lens. This image becomes the object for the second lens, B. It is a distance $d_{oB} = 80.0\,\text{cm} - 30.0\,\text{cm} = 50.0\,\text{cm}$ in front of lens B (Fig. 33–13b). The image formed by lens B, again using the thin lens equation, is at a distance d_{iB} from the lens B:

$$\frac{1}{d_{iB}} = \frac{1}{f_B} - \frac{1}{d_{oB}} = \frac{1}{25.0\,\text{cm}} - \frac{1}{50.0\,\text{cm}} = \frac{2-1}{50.0\,\text{cm}} = \frac{1}{50.0\,\text{cm}}.$$

Hence $d_{iB} = 50.0\,\text{cm}$ behind lens B. This is the final image—see Fig. 33–13b.

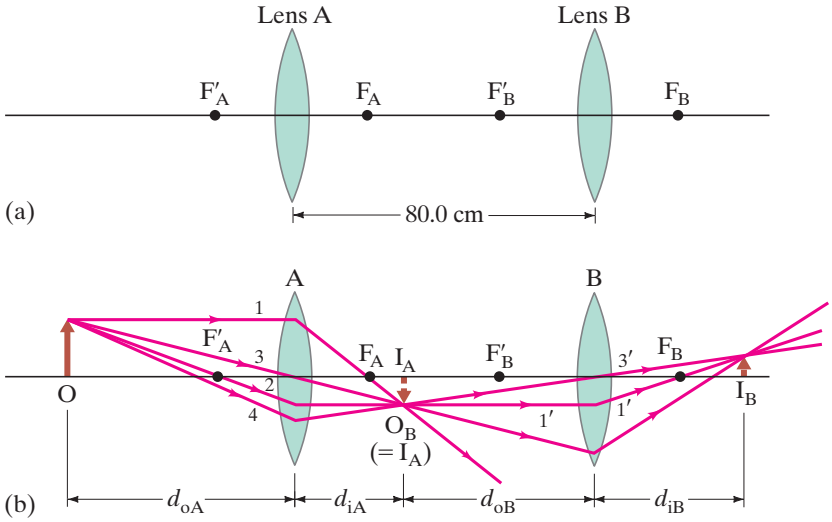


FIGURE 33–13 Two lenses, A and B, used in combination, Example 33–5. The small numbers refer to the easily drawn rays; those numbers have a prime (′) when referring to lens B.

(*b*) Lens A has a magnification (Eq. 33–3)

$$m_A = -\frac{d_{iA}}{d_{oA}} = -\frac{30.0\,\text{cm}}{60.0\,\text{cm}} = -0.500.$$

Thus, the first image is inverted and is half as high as the object (again Eq. 33–3):

$$h_{iA} = m_A h_{oA} = -0.500 h_{oA}.$$

Lens B takes this first image as object and changes its height by a factor

$$m_B = -\frac{d_{iB}}{d_{oB}} = -\frac{50.0\,\text{cm}}{50.0\,\text{cm}} = -1.000.$$

The second lens reinverts the image (the minus sign) but doesn't change its size. The final image height is (remember h_{oB} is the same as h_{iA})

$$h_{iB} = m_B h_{oB} = m_B h_{iA} = m_B m_A h_{oA} = (m_{\text{total}}) h_{oA}.$$

The total magnification is the product of m_A and m_B, which here equals $m_{\text{total}} = m_B m_A = (-1.000)(-0.500) = +0.500$, or half the original height, and the final image is upright.

PROBLEM SOLVING
Total magnification is $m_{\text{total}} = m_A m_B$

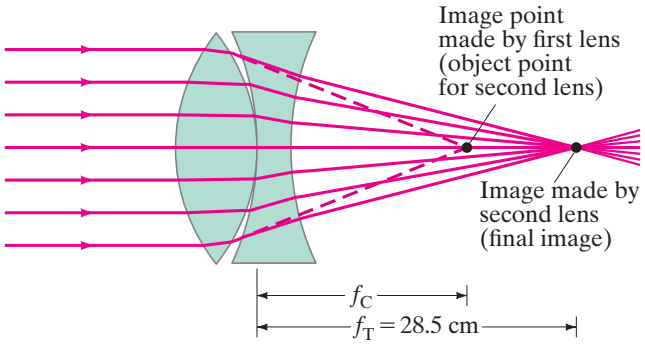


FIGURE 33–14 Determining the focal length of a diverging lens. Example 33–6.

EXAMPLE 33–6 **Measuring *f* for a diverging lens.** To measure the focal length of a diverging lens, a converging lens is placed in contact with it, as shown in Fig. 33–14. The Sun's rays are focused by this combination at a point 28.5 cm behind the lenses as shown. If the converging lens has a focal length f_C of 16.0 cm, what is the focal length f_D of the diverging lens? Assume both lenses are thin and the space between them is negligible.

APPROACH The image distance for the first lens equals its focal length (16.0 cm) since the object distance is infinity (∞). The position of this image, even though it is never actually formed, acts as the object for the second (diverging) lens. We apply the thin lens equation to the diverging lens to find its focal length, given that the final image is at $d_i = 28.5$ cm.

SOLUTION Rays from the Sun are focused 28.5 cm behind the combination, so the focal length of the total combination is $f_T = 28.5$ cm. If the diverging lens was absent, the converging lens would form the image at its focal point—that is, at a distance $f_C = 16.0$ cm behind it (dashed lines in Fig. 33–14). When the diverging lens is placed next to the converging lens, we treat the image formed by the first lens as the *object* for the second lens. Since this object lies to the right of the diverging lens, this is a situation where d_o is negative (see the sign conventions, page 963). Thus, for the diverging lens, the object is virtual and $d_o = -16.0$ cm. The diverging lens forms the image of this virtual object at a distance $d_i = 28.5$ cm (given). Thus,

CAUTION
$d_o < 0$

$$\frac{1}{f_D} = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{-16.0\,\text{cm}} + \frac{1}{28.5\,\text{cm}} = -0.0274\,\text{cm}^{-1}.$$

We take the reciprocal to find $f_D = -1/(0.0274\,\text{cm}^{-1}) = -36.5$ cm.

NOTE If this technique is to work, the converging lens must be "stronger" than the diverging lens—that is, it must have a focal length whose magnitude is less than that of the diverging lens.

33–4 Lensmaker's Equation

A useful equation, called the **lensmaker's equation**, relates the focal length of a lens (in air) to the radii of curvature R_1 and R_2 of its two surfaces and its index of refraction n:

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \qquad \textbf{(33–4)}$$

CAUTION
Sign conventions

If both surfaces are convex, R_1 and R_2 are considered positive.† For a concave surface, the radius must be considered *negative*.

Notice that Eq. 33–4 is *symmetrical* in R_1 and R_2. Thus, if a lens is turned around so that light impinges on the other surface, the focal length is the same even if the two lens surfaces are different. This confirms what we said earlier: a lens' focal length is the same on both sides of the lens.

EXAMPLE 33–7 **Calculating f for a converging lens.** A convex meniscus lens (Figs. 33–2a and 33–15) is made from glass with $n = 1.50$. The radius of curvature of the convex surface is 22.4 cm and that of the concave surface is 46.2 cm. (*a*) What is the focal length? (*b*) Where will the image be for an object 2.00 m away?

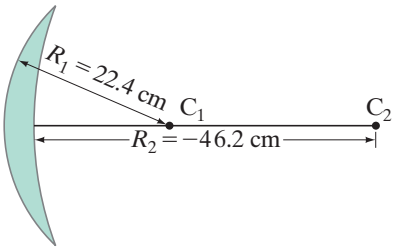


FIGURE 33–15 Example 33–7.

APPROACH We use Eq. 33–4, noting that R_2 is negative because it refers to the concave surface.

SOLUTION (*a*) $R_1 = 22.4\,\text{cm}$ and $R_2 = -46.2\,\text{cm}$.
Then

$$\frac{1}{f} = (1.50 - 1.00)\left(\frac{1}{22.4\,\text{cm}} - \frac{1}{46.2\,\text{cm}}\right) = 0.0115\,\text{cm}^{-1}.$$

So

$$f = \frac{1}{0.0115\,\text{cm}^{-1}} = 87.0\,\text{cm}$$

and the lens is converging. Notice that if we turn the lens around so that $R_1 = -46.2\,\text{cm}$ and $R_2 = +22.4\,\text{cm}$, we get the same result.

(*b*) From the lens equation, with $f = 0.870\,\text{m}$ and $d_o = 2.00\,\text{m}$, we have

$$\frac{1}{d_i} = \frac{1}{f} - \frac{1}{d_o} = \frac{1}{0.870\,\text{m}} - \frac{1}{2.00\,\text{m}} = 0.649\,\text{m}^{-1},$$

so $d_i = 1/0.649\,\text{m}^{-1} = 1.54\,\text{m}$.

EXERCISE C A Lucite planoconcave lens (see Fig. 33–2b) has one flat surface and the other has $R = -18.4\,\text{cm}$. What is the focal length? Is the lens converging or diverging?

*Derivation of the Lensmaker's Equation

Now we show that parallel rays are brought to a focus at a single point for a thin lens, and also derive the lensmaker's equation, Eq. 33–4.

In Fig. 33–16, a ray parallel to the axis of a lens (in air) is refracted at the front surface of the lens at point A_1 and is refracted at the back surface at point A_2. This ray then passes through point F, which we call the focal point for this ray. Point A_1 is a height h_1 above the axis, and point A_2 is height h_2 above the axis. C_1 and C_2 are the centers of curvature of the two lens surfaces; so the length $C_1A_1 = R_1$, the radius of curvature of the front surface, and $C_2A_2 = R_2$ is the radius of the second surface. We consider a double convex lens and by convention choose the radii of both lens surfaces as positive. The thickness of the lens has been grossly exaggerated so that the various angles would be clear. But we will assume that the lens is actually very thin and that angles between the rays and the axis are small. In this approximation, the sines and tangents of all the angles will be equal to the angles themselves in radians. For example, $\sin\theta_1 \approx \tan\theta_1 \approx \theta_1$ (radians).

†Some books use a different convention: R_1 and R_2 may be considered positive if their centers of curvature are to the right of the lens; then a minus sign appears in their version of Eq. 33–4, in place of our + sign.

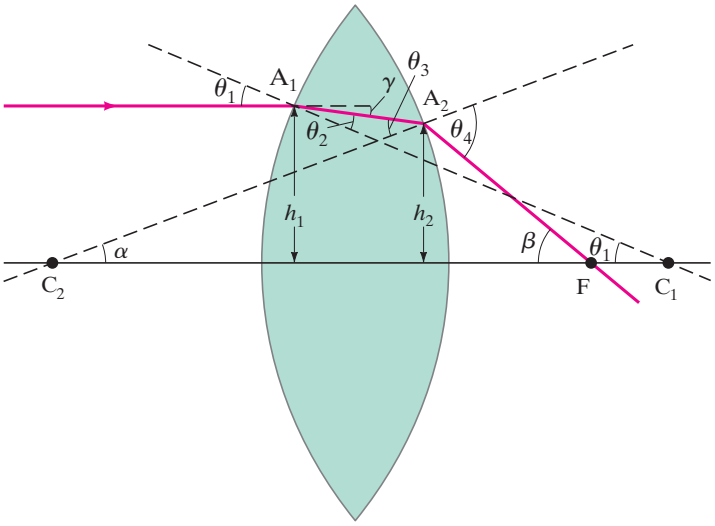


FIGURE 33–16 Diagram of a ray passing through a lens for derivation of the lensmaker's equation.

To this approximation, then, Snell's law tells us that

$$\theta_1 = n\theta_2$$
$$\theta_4 = n\theta_3$$

where n is the index of refraction of the lens, and we assume that the lens is surrounded by air $(n = 1)$. Notice also in Fig. 33–16 that

$$\theta_1 \approx \sin\theta_1 = \frac{h_1}{R_1}$$
$$\alpha \approx \frac{h_2}{R_2}$$
$$\beta \approx \frac{h_2}{f}.$$

The last follows because the distance from F to the lens (assumed very thin) is f. From the diagram, the angle γ is

$$\gamma = \theta_1 - \theta_2.$$

A careful examination of Fig. 33–16 shows also that

$$\alpha = \theta_3 - \gamma.$$

This can be seen by drawing a horizontal line to the left from point A_2, which divides the angle θ_3 into two parts. The upper part equals γ and the lower part equals α. (The opposite angles between an oblique line and two parallel lines are equal, Fig. A–1 in Appendix A.) Thus, $\theta_3 = \gamma + \alpha$. Furthermore, by drawing a horizontal line to the right from point A_2, we divide θ_4 into two parts. The upper part is α and the lower is β. Thus

$$\theta_4 = \alpha + \beta.$$

We now combine all these equations:

$$\alpha = \theta_3 - \gamma = \frac{\theta_4}{n} - (\theta_1 - \theta_2) = \frac{\alpha}{n} + \frac{\beta}{n} - \theta_1 + \theta_2,$$

or

$$\frac{h_2}{R_2} = \frac{h_2}{nR_2} + \frac{h_2}{nf} - \frac{h_1}{R_1} + \frac{h_1}{nR_1}.$$

Because the lens is thin, $h_1 \approx h_2$ and all h's can be canceled from all the numerators. We then multiply through by n and rearrange to find that

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right),$$

which is Eq. 33–4, the **lensmaker's equation**. Notice that f for a thin lens does not depend on h_1 or h_2. Thus the position of the point F does not depend on where the ray strikes the lens. Hence, all rays parallel to the axis of a thin lens will pass through the same point F, which we wished to show.

In this derivation of the lensmaker's equation, both surfaces are convex and R_1 and R_2 are considered positive. Equation 33–4 also works for lenses with one or both surfaces concave; but for a concave surface, the radius must be considered *negative* (see page 968, first paragraph).

[The lensmaker's equation, and our derivation of it, assume the lens is in air or vacuum. If the lens, of index of refraction n_2, is submerged in water or other liquid with index n_1, then our Snell's laws equations become

$$n_1\theta_1 = n_2\theta_2 \quad \text{and} \quad n_2\theta_3 = n_1\theta_4.$$

In our final result, Eq. 33–4, we would need to substitute n with (n_2/n_1). See Problem 31.]

33–5 Cameras: Film and Digital

The basic elements of a **camera** are a lens, a light-tight box, a shutter (that "clicks") to let light pass through the lens only briefly, and an electronic sensor in a digital camera or a piece of film in a traditional camera (Fig. 33–17). When the shutter is opened for a brief "exposure," light from external objects in the field of view is focused by the lens as an image on the sensor or film.

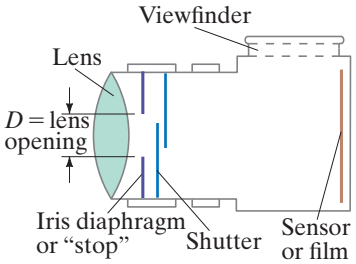


FIGURE 33–17 A simple camera.

You can see the image yourself if you are able to remove the back of a conventional camera, keeping the shutter open, and view through a piece of tissue paper (on which an image can form) placed where the film should be.

Film contains a thin **emulsion** (= a coating) with light-sensitive chemicals that change when light strikes them. The film is then developed by chemicals dissolved in water, which causes the most changed areas (brightest light) to turn opaque, so the image is recorded on the film.[†] We might call film "chemical photography" as compared to digital, which is electronic.

PHYSICS APPLIED
Cameras, film and digital

Digital Cameras, Electronic Sensors (CCD, CMOS)

In a **digital camera**, the film is replaced by a *semiconductor* sensor. Two types are common: **CCD** (*charge-coupled device*) and **CMOS** (*complementary metal oxide semiconductor*). Either sensor is made up of millions of tiny semiconductor **pixels** ("picture elements")—see Figs. 33–18 and 34–36. A 12-MP (12-megapixel) sensor might contain about 4000 pixels horizontally by 3000 pixels vertically over an area of perhaps 16×12 mm, and preferably larger such as 36×24 mm like 35-mm film. Light reaching any pixel liberates electrons within its semiconductor,[‡] which is itself attached to a tiny metal "gate" electrode (this combination is called a "photodiode"). The gate has a positive voltage applied to it and attracts the liberated electrons. The more intense the light, the more electrons are emitted and more charge accumulates during the brief exposure time. After exposure, the charge on each pixel has to be "read" (measured) and stored. In a CCD sensor, a reader-circuit at the end of each pixel row or column reads the charge on the pixel nearest it. To do so, it attracts the charge Q that was collected on "pixel-1" by being given a higher voltage itself than pixel-1. Immediately after, the charge on pixel-2 is attracted in to be measured ("read") and so on. All this information (the brightness of each pixel) goes to a central processor that stores it digitally (analog-to-digital converter) and allows re-formation of the image later onto the camera's screen, or a computer screen, or a printer. After all the pixel charge information is transferred to memory (Section 29–8), a new picture can be taken.

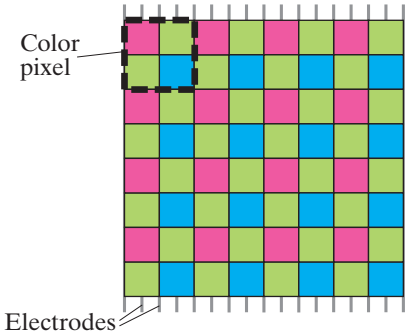


FIGURE 33–18 Portion of a typical Bayer array sensor. A square group of four pixels $\begin{matrix}\text{RG}\\\text{GB}\end{matrix}$ is sometimes called a "color pixel." (R is red, G green, B blue.)

[†]This is called a *negative*, because the black areas correspond to bright objects and vice versa. The same process occurs during printing to produce a black-and-white "positive" picture from the negative. Color film has three emulsion layers (or dyes) corresponding to the three primary colors.

[‡]Specifically, photons of light knock electrons in the valence band up to the conduction band (Chapter 40). It is sometimes said that the electrons liberated by incoming photons (a particle of light as we will see in Chapter 37) and attracted to a pixel gate are contained in a "well" of finite depth. This is *not* a physical well. It is a minimum in a *potential energy diagram* (as in Fig. 8–23, Section 8–9), and such a potential well (Section 38–8 and 38–9) can hold only a limited number of electrons because of quantum theory's exclusion principle (Section 39–4). If a "well" overflows, the charge enters neighboring pixels, referred to as "cross-talk" or "blooming", which blurs the picture.

A CMOS sensor also uses a silicon semiconductor, but incorporates transistor electronics within each pixel so each pixel charge can be read directly, somewhat like the similar MOSFET array that was shown in Fig. 29–31. CMOS costs less to manufacture, but for scientific work (such as astronomical telescopes) CCDs are usually used because they provide greater precision.

Sensor sizes are typically in the ratio 4:3 or 3:2. A larger sensor is better because it can hold more pixels, and/or each pixel can be larger and hold more charge (free electrons) to provide a wider range of brightness, better color accuracy, and better sensitivity in low-light conditions.

In the most common array of pixels, referred to as a **Bayer mosaic**, color is achieved by red, green, and blue filters over alternating pixels as shown in Fig. 33–18, similar to what a color LCD TV screen does (Sections 23–10 and 34–8). The sensor type shown in Fig. 33–18 contains twice as many green pixels as red or blue, because green seems to have a stronger influence on the human eye's sensation of sharpness.

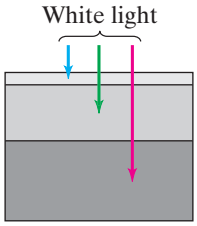
The output image from a camera provides a computer-analyzed color at many pixels, often an average with nearest-neighbor colors to reduce memory size (= compression, see Section 23–9). That is, the output image of a 12 MP camera is 12 million dots or squares, each a (computer) mixed color, not the individual colors of Fig. 33–18.

Each different color of pixel in a Bayer array is counted as a *separate* pixel. In contrast, in an LCD screen (Sections 23–10 and 34–8), a group of three subpixels is counted as one pixel, a more conservative count.

CAUTION
Camera and TV pixels are counted differently

An alternative technology, called "Foveon," uses a semiconductor layer system. Different wavelengths of light penetrate silicon to different depths, as shown in Fig. 33–19: the top layer is designed to absorb blue wavelengths, allowing green and red light to pass through. Longer wavelengths (green) are absorbed in the second layer, and the bottom layer detects the longest wavelengths (red). All three colors are detected by each "tri-pixel" site, resulting in better color resolution and fewer artifacts.

FIGURE 33–19 A layered or "Foveon" tri-pixel that includes all three colors, arranged vertically so light can pass through all three subpixels.



Digital Artifacts

Digital cameras can produce image artifacts (artificial effects in the image not present in the original) resulting from the electronic sensing of the image. One example using the Bayer mosaic of pixels (Fig. 33–18) is described in Fig. 33–20.

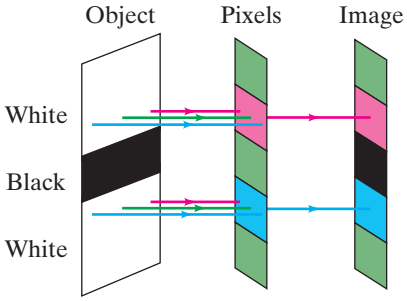


FIGURE 33–20 Suppose we take a picture that includes a black line (our object) on a white background. The *image* of this black line on a Bayer sensor has a colored halo (red above, blue below) due to the mosaic arrangement of color filter pixels, as shown here by the colors transmitted to the image plane. Computer averaging minimizes such color problems as this (the green at top and bottom of the image may average with nearby pixels to give white or nearly so) but the image is consequently "softened" or blurred.

Camera Adjustments

There are four main adjustments on good-quality cameras: shutter speed, *f*-stop, sensor or film sensitivity (ISO), and focusing. Although most cameras today make these adjustments automatically, it is valuable to understand these adjustments to use a camera effectively. For special or top-quality work, a manual camera is indispensable (Fig. 33–21).

FIGURE 33–21 On this camera, the *f*-stops and the focusing ring are on the camera lens. Shutter speeds are selected on the small wheel on top of the camera body.



Exposure time or shutter speed This refers to how long the shutter of a camera is open and the film or sensor is exposed. It could vary from a second or more ("time exposures") to $\frac{1}{1000}$ s or faster. To avoid blurring from camera movement, exposure times shorter than $\frac{1}{100}$ s are normally needed. If the object is moving, even shorter exposure times are needed to "stop" the action ($\frac{1}{1000}$ s or less). If the exposure time is not short enough, the image will be blurred by camera shake or because objects have time to move. Blurring in low light conditions is more of a problem with cheap (or most cell-phone) cameras whose inexpensive sensors need to have the shutter open longer to collect enough light. Digital still cameras or cell phones that take short videos must have a fast enough "sampling" time and fast "clearing" (of the charge) so as to take pictures at least 15 frames per second; preferable is 24 fps (like movies) or 30, 60, or 120 fps like TV refresh rates (25, 50, or 100 in areas like Europe where 50 Hz is the normal line voltage frequency).

***f*-stop and exposure** The amount of light reaching the sensor or film depends on the area of the **lens opening** as well as shutter speed, and must be carefully controlled to avoid **underexposure** (too little light so the picture is dark and only the brightest objects show up) or **overexposure** (too much light, so that bright objects have a lack of contrast and "washed-out" appearance). A good camera may control the exposure with a "stop" or iris diaphragm, whose opening is of variable diameter, placed behind the lens (Fig. 33–17). The lens opening is controlled (automatically or manually) to compensate for bright or dark lighting conditions, the sensitivity of the sensor or film (ISO), and for different shutter speeds. The size of the opening is specified by the ***f*-stop** or ***f*-number**, defined as

$$f\text{-stop} = \frac{f}{D},$$

where f is the focal length of the lens and D is the diameter of the lens opening (see Fig. 33–17). For example, when a 50-mm-focal-length lens has an opening $D = 25\,\text{mm}$, then $f/D = 50\,\text{mm}/25\,\text{mm} = 2$, so we say it is set at $f/2$. When this lens is set at $f/5.6$, the opening is only $9\,\text{mm}\ (50/9 = 5.6)$. For faster shutter speeds, or low light conditions, a wider lens opening must be used to get a proper exposure, which corresponds to a smaller f-stop number. The smaller the f-stop number, the larger the opening and the more light passes through the lens to the sensor or film. The smallest f-number of a lens (largest opening) is referred to as the *speed* of the lens.[†] The best lenses may have a speed of $f/2.0$, or even faster. The advantage of a fast lens is that it lets more light in to take pictures under poor lighting conditions. Standard f-stops are

$$1.0,\ 1.4,\ 2.0,\ 2.8,\ 4.0,\ 5.6,\ 8,\ 11,\ 16,\ 22,\ \text{and}\ 32$$

(see Fig. 33–21, which only goes to $f/16$). Each of these stops corresponds to a diameter reduction by a factor of $\sqrt{2} \approx 1.4$. Because the amount of light reaching the film is proportional to the *area* of the opening, and therefore proportional to the diameter squared, each standard f-stop corresponds to a factor of 2 change in light intensity reaching the film. If you change the shutter speed by a factor of 2, say 1/250 to 1/500 of a second, half as much light comes in so you need to open the lens by one stop (such as $f11$ to $f8$) to get the same exposure.

ISO Good quality cameras have an adjustment for the sensitivity to light of the sensor or film, called the ISO number (once called ASA). On many digital cameras, you can adjust the ISO. Film, on the other hand, is manufactured to be "fast" (high ISO) or "slow" (low ISO) and you set your camera for the film chosen. For normal bright days you might prefer ISO of 100 or 200. But at night or with low light conditions, you may want a high ISO, maybe 1600 or 3200 or even higher, in order to get a good exposure. ISO is a bit like shutter speed: if you double it (say from 200 to 400) you can use one f-stop smaller, or double the shutter speed, and get the same exposure. To get a good exposure, you or the camera needs an "exposure meter" to measure the amount of light available. And there are trade-offs. In low light, you can: (1) slow the shutter speed, but if too slow any motion is blurred; (2) use a wider lens opening, but that will reduce depth-of-field and perhaps reduce sharpness (lens aberrations); or (3) increase the ISO (inconvenient film replacement for a film camera). A high ISO has disadvantages too: pictures are grainier or "noisier" and less sharp, and have poorer color rendition. High ISO works better for very high quality sensors whose pixels have greater "depth", meaning holding more electrons for any given light level.

Focusing Focusing is the operation of placing the lens at the correct position relative to the sensor or film for the sharpest image. The image distance is smallest for objects at infinity (the symbol ∞ is used for infinity—see Fig. 33–21) and is equal to the focal length, as we saw in Section 33–1. For closer objects, the image distance is greater than the focal length, as can be seen from the lens equation, $1/f = 1/d_o + 1/d_i$ (Eq. 33–2). To focus on nearby objects, the lens must therefore be moved away from the sensor or film (by turning a ring on the lens for a manual camera).

[†]Lens apertures and film are referred to as "fast" because they allow for faster shutter speeds.

If the lens is focused on a nearby object, a sharp image of it will be formed (a point will be a point of the image), but the image of distant objects may be blurry (Fig. 33–22). The rays from a point on the distant object will be out of focus—instead of a point, they will form a circle on the sensor or film as shown (exaggerated) in Fig. 33–23. The distant object will thus produce an image consisting of overlapping circles and will be blurred. These circles are called **circles of confusion**. To have near and distant objects sharp in the same photo, you (or the camera) can try setting the lens focus at an intermediate position. For a given distance setting, there is a range of distances over which the circles of confusion will be small enough that the images will be reasonably sharp. This is called the **depth of field**. For a sensor or film width of 36 mm (including 35-mm film cameras), the depth of field is usually based on a maximum circle of confusion diameter of 0.030 mm, even 0.02 mm or 0.01 mm for critical work or very large photographs. The depth of field varies with the lens opening. If the lens opening is smaller, only rays through the central part of the lens pass through, and these form smaller circles of confusion for a given object distance. Hence, at smaller lens openings, a greater range of object distances will fit within the circle of confusion criterion, so the depth of field is greater. Smaller lens openings, however, result in reduced resolution due to diffraction (discussed in Chapter 35). Best resolution is typically found around $f/5.6$ or $f/8$.

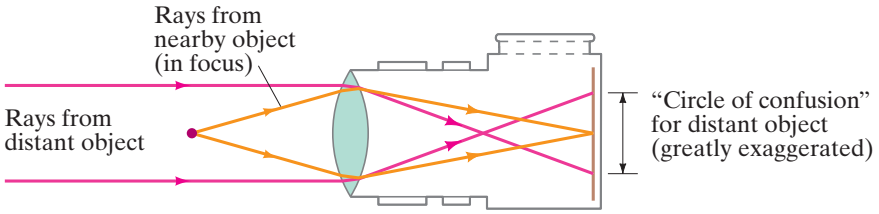


FIGURE 33–23 When the lens is positioned to focus on a nearby object, points on a distant object produce circles and are therefore blurred. (The effect is shown greatly exaggerated.)

CONCEPTUAL EXAMPLE 33–8 **Shutter speed.** To improve the depth of field, you "stop down" your camera lens by two *f*-stops from $f/4$ to $f/8$. What should you do to the shutter speed to maintain the same exposure?

RESPONSE The amount of light admitted by the lens is proportional to the area of the lens opening. Reducing the lens opening by two *f*-stops reduces the diameter by a factor of 2, and the area by a factor of $2^2 = 4$. To maintain the same exposure, the shutter must be open four times as long. If the shutter speed had been $\frac{1}{500}$ s, you would have to increase the exposure time to $\frac{1}{125}$ s.

Picture Sharpness

The sharpness of a picture depends not only on accurate focusing and short exposure times, but also on the graininess of the film, or number of pixels on a digital sensor. Fine-grained films and tiny pixels are "slower," meaning they require longer exposures for a given light level. All pixels are rarely used because digital cameras have averaging (or "compression") programs, such as JPEG, which reduce memory size by averaging over pixels where little contrast is detected. But some detail is inevitably lost. For example, a small blue lake may seem uniform, and coding 600 pixels as identical takes less memory than specifying all 600. Any slight variation of the water surface is lost. Full "RAW" data uses more memory. Film has the advantage that it always records everything (down to its grain size) with no computer averaging. So does RAW if your camera offers it. The processor also averages over pixels in low light conditions, resulting in a less sharp photo.

The quality of the lens strongly affects the image quality, and we discuss lens resolution and diffraction effects in Sections 33–10 and 35–4. The sharpness, or **resolution**, of a lens is often given as so many lines per millimeter, measured by photographing a standard set of parallel black lines on a white background (sometimes said as "line pairs/mm") on fine-grain film or high-quality sensor, or as so many dots per inch (dpi).

FIGURE 33–22 Photos taken with a camera lens (a) focused on a nearby object with distant object blurry, and (b) focused on a more distant object with nearby object blurry.

(a)



(b)



The minimum spacing of distinguishable lines or dots on sensor or film at the focal point gives the lens resolution. A lens that can give 50 lines/mm is reasonable, 100 lines/mm is very good (= 100 dots/mm ≈ 2500 dpi). Electronic sensors also have a resolution and it is sometimes given as line pairs across the full sensor width.

A "full" Bayer pixel (upper left in Fig. 33–18) is 4 regular pixels: for example, to make a white dot as part of a white line (between two black lines when determining lens resolution), all 4 Bayer pixels (RGGB) would have to be bright. For a Foveon, all three colors of one pixel need to be bright to produce a white dot.[†]

EXAMPLE 33–9 **Pixels and resolution.** A 12-MP digital camera offers a maximum resolution of 4000×3000 pixels on a 32-mm × 24-mm sensor. How sharp should the lens be to make use of this resolution in RAW?

APPROACH We find the number of pixels per millimeter and require the lens to be at least that good.

SOLUTION We can take either the image height (3000 pixels in 24 mm) or the width (4000 pixels in 32 mm):

$$\frac{3000 \text{ pixels}}{24 \text{ mm}} = 125 \text{ pixels/mm}.$$

We would like the lens to match this resolution of 125 lines or dots per mm, which would be quite a good lens. If the lens is not this good, fewer pixels and less memory could be used.

NOTE Increasing lens resolution is a tougher problem today than is squeezing more pixels on a CCD or CMOS sensor. The sensor for high-quality cameras must also be physically larger for better image accuracy (more brightness levels, more colors, etc.) and greater light sensitivity in low light conditions.

PHYSICS APPLIED
When is a photo sharp?

EXAMPLE 33–10 **Blown-up photograph.** A photograph looks sharp at normal viewing distances if the dots or lines are resolved to perhaps 10 dots/mm (or 250 dots/inch since 1 inch = 25 mm). Would an 8 × 10-inch enlargement of a photo taken by the camera in Example 33–9 seem sharp?

APPROACH We assume the image is 4000×3000 pixels on a 32×24-mm sensor as in Example 33–9, or 125 pixels/mm. We make an enlarged photo 8×10 in. = 20 cm × 25 cm.

SOLUTION The short side of the sensor is 24 mm = 2.4 cm long, and that side of the 8×10 inch photograph is 8 inches or 20 cm. Thus the size is increased by a factor of 20 cm/2.4 cm ≈ 8× (or 25 cm/3.2 cm ≈ 8×). To fill the 8 × 10-in. paper, we assume the enlargement is 8×. The pixels are thus enlarged 8×. So the given sensor resolution of 125 dots/mm becomes 125/8 = 15 dots/mm on the print. Hence an 8 × 10-inch print would be a sharp photograph. We could go 50% larger, 11×14 inch or even 12×18.

Alternate Solution When the 24×32-mm sensor image is enlarged, the 3000 pixel side fills the short 8-inch (= 200 mm) side of the print. The resolution is 3000 dots/200 mm = 15 dots/ mm (or 375 dpi). Similar for the long side, 4000 dots/10 in. = 400 dpi.

In order to make very large photographic prints, large-format cameras are used such as 6 cm × 6 cm ($2\frac{1}{4}$ inch square)—either film or sensor—and even 4 × 5 inch and 8 × 10 inch (using sheet film or glass plates).

EXERCISE D The criterion of 0.030 mm as the diameter of a circle of confusion as acceptable sharpness is how many dots per mm on the camera sensor?

[†]Consider a 4000×3000-pixel array. For a Foveon, each "full pixel" (Fig. 33–19) has all 3 colors, each of which can be counted as a pixel, so it may be considered as $4000 \times 3000 \times 3 = 36$ MP. For a Bayer sensor, Fig. 33–18, 4000×3000 is 12 MP (6 MP of green, 3 MP each of red and blue). There are more green pixels because they are most important in our eyes' ability to note resolution. So the distance between green pixels is a rough guide to the sharpness of a Bayer. To match a 4000×3000 Foveon (36 MP, or 12 MP of tri-pixel sites), a Bayer would need to have about 24 MP (because it would then have 12 MP of green). This "equivalence" is only a rough approximation.

Telephotos and Wide-Angle Lenses

Camera lenses are categorized into normal, telephoto, and wide angle, according to focal length and film size. A **normal lens** covers the sensor or film with a field of view that corresponds approximately to that of normal vision. A "normal" lens for 35-mm film has a focal length of 50 mm. The best digital cameras aim for a sensor of the same size[†] (24 mm × 36 mm). (If the sensor is smaller, digital cameras sometimes specify focal lengths to correspond with classic 35-mm cameras.) **Telephoto lenses** act like telescopes to magnify images. They have longer focal lengths than a normal lens: as we saw in Section 33–2 (Eq. 33–3), the height of the image for a given object distance is proportional to the image distance, and the image distance will be greater for a lens with longer focal length. For distant objects, the image height is very nearly proportional to the focal length. Thus a 200-mm telephoto lens for use with a 35-mm camera gives a 4× magnification over the normal 50-mm lens. A **wide-angle lens** has a shorter focal length than normal: a wider field of view is observed, and objects appear smaller. A **zoom lens** is one whose focal length can be changed (by changing the distance between the thin lenses that make up the compound lens, Section 33–10) so that you seem to zoom up to, or away from, the subject as you change the focal length.

Digital cameras may have an **optical zoom** meaning the lens can change focal length and maintain resolution. But an "electronic" or **digital zoom** just enlarges the dots (pixels) with loss of sharpness.

Different types of viewing systems are used in cameras. In some cameras, you view through a small window just above the lens as in Fig. 33–17. In a **single-lens reflex** camera (SLR or DSLR were D means digital), you actually view through the lens with the use of prisms and mirrors (Fig. 33–24). A mirror hangs at a 45° angle behind the lens and flips up out of the way just before the shutter opens. SLRs have the advantage that you can see almost exactly what you will get. Digital cameras use an electronic LCD display on the back. Some mirrorless DSLRs have a tiny electronic LCD display in the viewfinder, and it too can show what you will get on the photo if it is carefully designed.

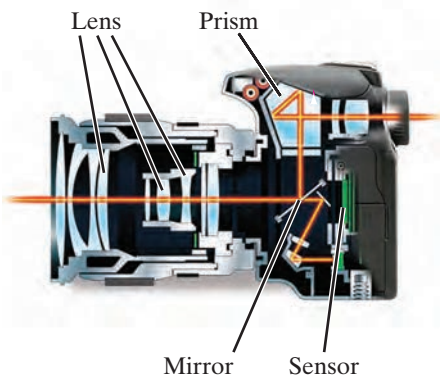


FIGURE 33–24 Single-lens reflex (SLR) camera, showing how the image is viewed through the lens with the help of a movable mirror and prism. The ray entering the camera is split, part going through the prism to be viewed directly, part to a sensor to display on a screen on the camera back.

33–6 The Human Eye; Corrective Lenses

The human eye resembles a camera in its basic structure (Fig. 33–25), but is far more sophisticated. The interior of the eye is filled with a transparent gel-like substance called the *vitreous humor* with index of refraction $n = 1.337$. Light enters this enclosed volume through the cornea and lens. Between the cornea and lens is a watery fluid, the aqueous humor (*aqua* is "water" in Latin) with $n = 1.336$. A diaphragm, called the **iris** (the colored part of your eye), adjusts automatically to control the amount of light entering the eye, similar to a camera. The hole in the iris through which light passes (the **pupil**) is black because no light is reflected from it (it's a hole), and very little light is reflected back out from the interior of the eye. The **retina**, which plays the role of the film or sensor in a camera, is on the curved rear surface of the eye. The retina consists of a complex array of nerves and receptors known as *rods* and *cones* which act to change light energy into electrical signals that travel along the nerves. The reconstruction of the image from all these tiny receptors is done mainly in the brain, although some analysis may also be done in the complex interconnected nerve network at the retina itself. At the center of the retina is a small area called the **fovea**, about 0.25 mm in diameter, where the cones are very closely packed and the sharpest image and best color discrimination are found.

PHYSICS APPLIED
The eye

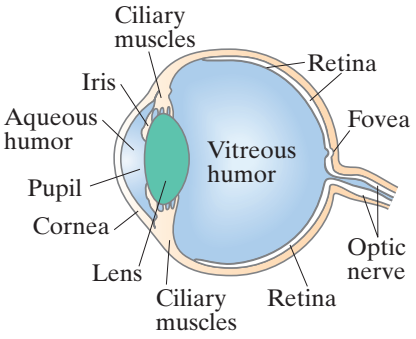


FIGURE 33–25 Diagram of a human eye.

[†]A "35-mm camera" uses film that is physically 35 mm wide; that 35 mm is not to be confused with a focal length. 35-mm film has sprocket holes (for making and showing movies), so only 24 mm of its height is used for the photo; the width is usually 36 mm for stills. Thus one frame is 36 mm × 24 mm. Movie frames on 35-mm film are 24 mm × 18 mm.

Unlike a camera, the eye contains no shutter. The equivalent operation is carried out by the nervous system, which analyzes the signals to form images at the rate of about 30 per second. This can be compared to motion picture or television cameras, which operate by taking a series of still pictures at a rate of 24 (movies) and 60 or 30 (U.S. television) per second. Their rapid projection on the screen gives the appearance of motion.

The lens of the eye ($n = 1.386$ to 1.406) does little of the bending of the light rays. Most of the refraction is done at the front surface of the **cornea** ($n = 1.376$) at its interface with air ($n = 1.0$). The lens acts as a fine adjustment for focusing at different distances. This is accomplished by the ciliary muscles (Fig. 33–25), which change the curvature of the lens so that its focal length is changed. To focus on a distant object, the ciliary muscles of the eye are relaxed and the lens is thin, as shown in Fig. 33–26a, and parallel rays focus at the focal point (on the retina). The retina–lens distance in the eye is fixed. So, unlike a camera, the image distance in the eye is fixed. To focus on a nearby object, the ciliary muscles contract, causing the center of the lens to thicken, Fig. 33–26b, thus shortening the focal length so that images of nearby objects can be focused on the retina, behind the new focal point. This focusing adjustment is called **accommodation**.

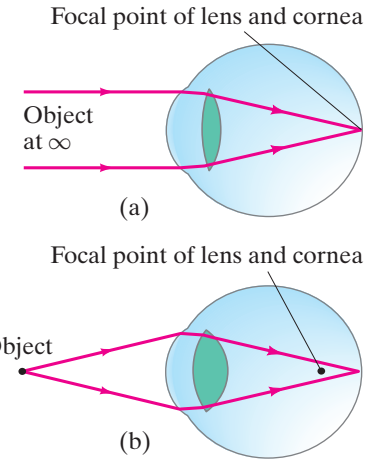


FIGURE 33–26 Accommodation by a normal eye: (a) lens relaxed, focused at infinity; (b) lens thickened, focused on a nearby object.

The closest distance at which the eye can focus clearly is called the **near point** of the eye. For young adults it is typically 25 cm, although younger children can often focus on objects as close as 10 cm. As people grow older, the ability to accommodate is reduced and the near point increases. A given person's **far point** is the farthest distance at which an object can be seen clearly. For some purposes it is useful to speak of a **normal eye** (a sort of average over the population), defined as an eye having a near point of 25 cm and a far point of infinity. To check your own near point, place this book close to your eye and slowly move it away until the letters are sharp. If you cannot focus on this page unless it is, let's say, 55 cm away, then your near point is 55 cm; but at that distance, the letters may appear too small to read, so you need reading glasses (see below).

PROBLEM SOLVING
Eye near point = 25 cm.
Far point = ∞. (Average)

The "normal" eye is sort of an ideal. Many people have eyes that do not accommodate within the "normal" range of 25 cm to infinity, or have some other defect. Two common defects are nearsightedness and farsightedness. Both can be corrected to a large extent with lenses—either eyeglasses or contact lenses.

In **nearsightedness**, or **myopia**, the human eye can focus only on nearby objects. The far point is not infinity but some shorter distance, so distant objects are not seen clearly. Nearsightedness is usually caused by an eyeball that is too long, although sometimes it is the curvature of the cornea that is too great. In either case, images of distant objects are focused in front of the retina. A diverging lens, because it causes parallel rays to diverge, allows the rays to be focused at the retina (Figs. 33–27a and b) and thus can correct nearsightedness.

PHYSICS APPLIED
Corrective lenses

In **farsightedness**, or *hyperopia*, the eye cannot focus on nearby objects. Although distant objects are usually seen clearly, the near point is somewhat

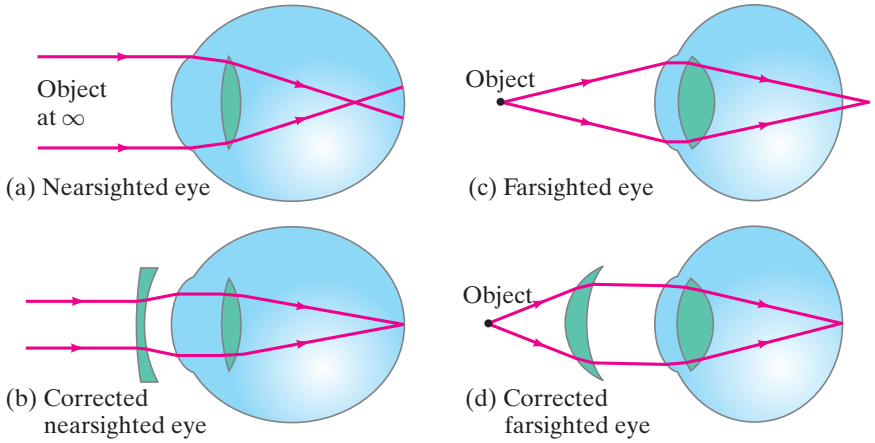
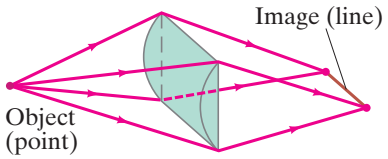


FIGURE 33–27 Correcting eye defects with lenses. (a) A nearsighted eye, which cannot focus clearly on distant objects (focal point is in front of retina), can be corrected (b) by use of a diverging lens. (c) A farsighted eye, which cannot focus clearly on nearby objects (image point behind retina), can be corrected (d) by use of a converging lens.

greater than the "normal" 25 cm, which makes reading difficult. This defect is caused by an eyeball that is too short or (less often) by a cornea that is not sufficiently curved. It is corrected by a converging lens, Figs. 33–27c and d. Similar to hyperopia is *presbyopia*, which is the lessening ability of the eye to accommodate as a person ages, and the near point moves out. Converging lenses also compensate for this, and serve as "reading glasses" for many people.

Astigmatism is usually caused by an out-of-round cornea or lens so that point objects are focused as short lines, which blurs the image. It is as if the cornea was partly spherical with a cylindrical section superimposed. As shown in Fig. 33–28, a cylindrical lens focuses a point into a line parallel to its axis. An astigmatic eye may focus rays in one plane, such as the vertical plane, at a shorter distance than it does for rays in a horizontal plane. Astigmatism is corrected with the use of a compensating cylindrical lens. Lenses for eyes that are nearsighted or farsighted as well as astigmatic are ground with superimposed spherical and cylindrical surfaces, so that the radius of curvature of the correcting lens is different in different planes.

FIGURE 33–28 A cylindrical lens forms a line image of a point object because it is converging in one plane only.



EXAMPLE 33–11 **Farsighted eye.** A farsighted eye has a near point of 100 cm. Reading glasses must have what lens power so that a newspaper can be read at a distance of 25 cm? Assume the lens is very close to the eye.

APPROACH When the object is placed 25 cm from the lens ($= d_o$), we want the image to be 100 cm away on the *same* side of the lens (so the eye can focus on it), and so the image is virtual, as shown in Fig. 33–29, and $d_i = -100\,\text{cm}$ will be negative. We use the thin lens equation (Eq. 33–2) to determine the needed focal length. Optometrists' prescriptions specify the power ($P = 1/f$, Eq. 33–1) given in diopters ($1\,\text{D} = 1\,\text{m}^{-1}$).

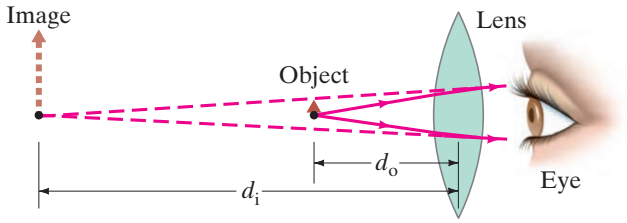


FIGURE 33–29 Lens of reading glasses (Example 33–11).

SOLUTION Given that $d_o = 25\,\text{cm}$ and $d_i = -100\,\text{cm}$, the thin lens equation gives

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{25\,\text{cm}} + \frac{1}{-100\,\text{cm}} = \frac{4-1}{100\,\text{cm}} = \frac{1}{33\,\text{cm}}.$$

So $f = 33\,\text{cm} = 0.33\,\text{m}$. The power P of the lens is $P = 1/f = +3.0\,\text{D}$. The plus sign indicates that it is a converging lens.

NOTE We chose the image position to be where the eye can actually focus. The lens needs to put the image there, given the desired placement of the object (newspaper).

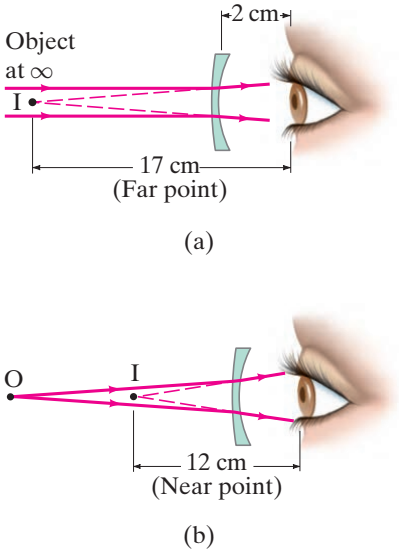


FIGURE 33–30 Example 33–12.

EXAMPLE 33–12 **Nearsighted eye.** A nearsighted eye has near and far points of 12 cm and 17 cm, respectively. (*a*) What lens power is needed for this person to see distant objects clearly, and (*b*) what then will be the near point? Assume that the eyeglass lens is 2.0 cm from the eye (which is typical).

APPROACH For a distant object $(d_o = \infty)$, the lens must put the image at the far point of the eye as shown in Fig. 33–30a, 17 cm in front of the eye. We can use the thin lens equation to find the focal length of the lens, and from this its lens power. The new near point (as shown in Fig. 33–30b) can be calculated for the lens by again using the thin lens equation.

SOLUTION (*a*) For an object at infinity $(d_o = \infty)$, the image must be in front of the lens 17 cm from the eye or $(17\text{ cm} - 2\text{ cm}) = 15\text{ cm}$ from the lens; hence $d_i = -15\text{ cm}$. We use the thin lens equation to solve for the focal length of the needed lens:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{\infty} + \frac{1}{-15\text{ cm}} = -\frac{1}{15\text{ cm}}.$$

So $f = -15\text{ cm} = -0.15\text{ m}$ or $P = 1/f = -6.7\text{ D}$. The minus sign indicates that it must be a diverging lens for the myopic eye.

(*b*) The near point when glasses are worn is where an object is placed (d_o) so that the lens forms an image at the "near point of the naked eye," namely 12 cm from the eye. That image point is $(12\text{ cm} - 2\text{ cm}) = 10\text{ cm}$ in front of the lens, so $d_i = -0.10\text{ m}$ and the thin lens equation gives

$$\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_i} = -\frac{1}{0.15\text{ m}} + \frac{1}{0.10\text{ m}} = \frac{-2 + 3}{0.30\text{ m}} = \frac{1}{0.30\text{ m}}.$$

So $d_o = 30\text{ cm}$, which means the near point when the person is wearing glasses is 30 cm in front of the lens, or 32 cm from the eye.

Contact Lenses

PHYSICS APPLIED
Contact lenses—different f and P

Suppose contact lenses are used to correct the eye in Example 33–12. Since contacts are placed directly on the cornea, we would not subtract out the 2.0 cm for the image distances. That is, for distant objects $d_i = f = -17\text{ cm}$, so $P = 1/f = -5.9\text{ D}$. The new near point would be 41 cm $(1/d_o = -1/0.17\text{ m} + 1/0.12\text{ m} = 2.45\text{ m}^{-1})$. Thus we see that a contact lens and an eyeglass lens will require slightly different powers, or focal lengths, for the same eye because of their different placements relative to the eye. We also see that glasses in this case give a better near point than contacts.

PHYSICS APPLIED
Glasses give closer near point than contact lenses

EXERCISE E What power of contact lens is needed for an eye to see distant objects if its far point is 25 cm?

Underwater Vision

PHYSICS APPLIED
Underwater vision

When your eyes are under water, distant underwater objects look blurry because at the water–cornea interface, the difference in indices of refraction is very small: $n = 1.33$ for water, 1.376 for the cornea. Hence light rays are bent very little and are focused far behind the retina, Fig. 33–31a. If you wear goggles or a face mask, you restore an air–cornea interface ($n = 1.0$ and 1.376, respectively) and the rays can be focused, Fig. 33–31b.

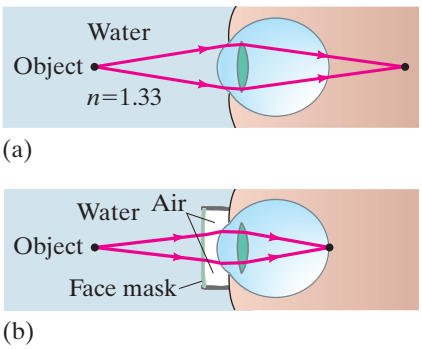


FIGURE 33–31 (a) Under water, we see a blurry image because light rays are bent much less than in air. (b) If we wear goggles, we again have an air–cornea interface and can see clearly.

33–7 Magnifying Glass

Much of the remainder of this Chapter will deal with optical devices that are used to produce magnified images of objects. We first discuss the **simple magnifier**, or **magnifying glass**, which is simply a converging lens (see Chapter-Opening Photo, page 958).

How large an object appears, and how much detail we can see on it, depends on the size of the image it makes on the retina. This, in turn, depends on the angle subtended by the object at the eye. For example, a penny held 30 cm from the eye looks twice as tall as one held 60 cm away because the angle it subtends is twice as great (Fig. 33–32). When we want to examine detail on an object, we bring it up close to our eyes so that it subtends a greater angle. However, our eyes can accommodate only up to a point (the near point), and we will assume a standard distance of $N = 25\text{ cm}$ as the near point in what follows.

A magnifying glass allows us to place the object closer to our eye so that it subtends a greater angle. As shown in Fig. 33–33a, the object is placed at the focal point or just within it. Then the converging lens produces a virtual image, which must be at least 25 cm from the eye if the eye is to focus on it. If the eye is relaxed, the image will be at infinity, and in this case the object is exactly at the focal point. (You make this slight adjustment yourself when you "focus" on the object by moving the magnifying glass.)

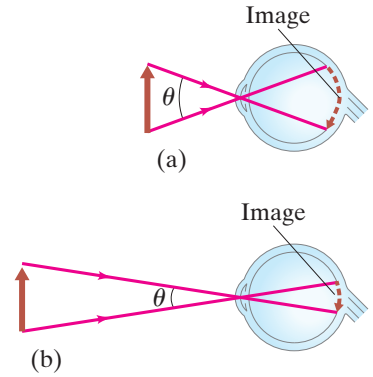


FIGURE 33–32 When the same object is viewed at a shorter distance, the image on the retina is greater, so the object appears larger and more detail can be seen. The angle θ that the object subtends in (a) is greater than in (b). *Note*: This is not a normal ray diagram because we are showing only one ray from each point.

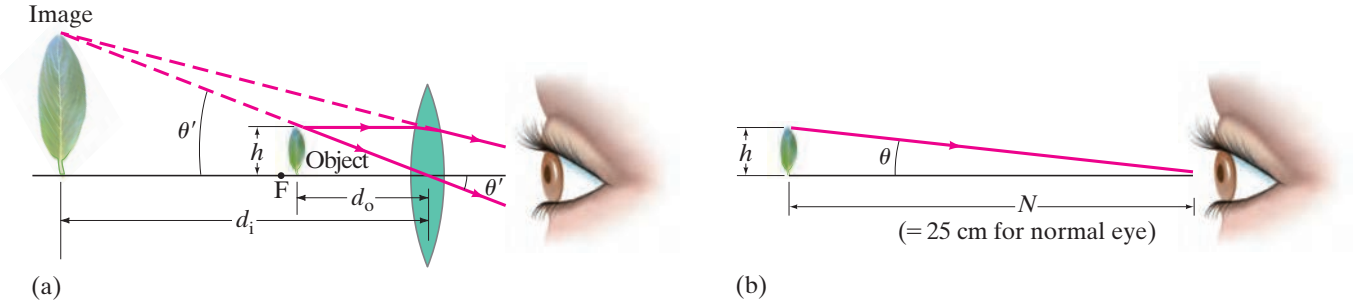


FIGURE 33–33 Leaf viewed (a) through a magnifying glass, and (b) with the unaided eye. The eye is focused at its near point in both cases.

A comparison of part (a) of Fig. 33–33 with part (b), in which the same object is viewed at the near point with the unaided eye, reveals that the angle the object subtends at the eye is much larger when the magnifier is used. The **angular magnification** or **magnifying power**, M, of a lens is defined as the ratio of: the angle θ' subtended by an object when using the lens, to the angle θ subtended without the lens with the object at the near point N of the eye ($N = 25\text{ cm}$ for a normal eye):

$$M = \frac{\theta'}{\theta}. \qquad \textbf{(33–5)}$$

The angles θ and θ' are shown in Fig. 33–33. We can write M in terms of the focal length by noting that $\theta = h/N$ (Fig. 33–33b) and $\theta' = h/d_o$ (Fig. 33–33a), where h is the height of the object and we assume the angles are small so θ and θ' (in radians) equal their sines and tangents. If the eye is relaxed (for least eye strain), the image will be at infinity and the object will be precisely at the focal point; see Fig. 33–34. Then $d_o = f$ and $\theta' = h/f$, whereas $\theta = h/N$ as before (Fig. 33–33b). Thus

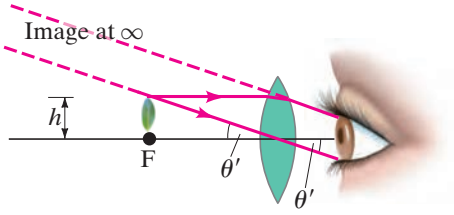
$$M = \frac{\theta'}{\theta} = \frac{h/f}{h/N} = \frac{N}{f}. \quad \left[\begin{array}{c}\text{eye focused at } \infty;\\ N = 25\text{ cm for normal eye}\end{array}\right] \qquad \textbf{(33–6a)}$$

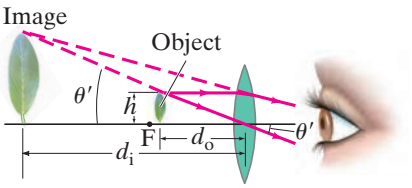
We see that the shorter the focal length of the lens, the greater the magnification.†

We use capital M here for angular magnification, which is essentially the same M we used in Chapter 32, Fig. 32–18. We need to distinguish it from lateral magnification m (Eq. 33–3), which relates the image height as made by a lens or mirror to the original object height. Here, with M of a magnifying glass, the human eye is involved, as it was in Sections 32–4 and 5.

†Note: Simple single-lens magnifiers are usually only 3 or 4× because of distortion due to spherical and other aberrations (Section 33–10).

FIGURE 33–34 With the eye relaxed, the object is placed at the focal point, and the image is at infinity. Compare to Fig. 33–33a where the image is at the eye's near point.





(a)

FIGURE 33–33a (Repeated.)

The magnification of a given lens can be increased a bit by moving the lens and adjusting your eye so it focuses on the image at the eye's near point. In this case, $d_i = -N$ (as in Fig. 33–33a) if your eye is very near the magnifier. Then the object distance d_o is given by

$$\frac{1}{d_o} = \frac{1}{f} - \frac{1}{d_i} = \frac{1}{f} + \frac{1}{N}.$$

We see from this equation that $d_o = fN/(f + N) < f$, as shown in Fig. 33–33a, since $N/(f + N)$ must be less than 1. With $\theta' = h/d_o$ the magnification is

$$M = \frac{\theta'}{\theta} = \frac{h/d_o}{h/N} = N\left(\frac{1}{d_o}\right) = N\left(\frac{1}{f} + \frac{1}{N}\right)$$

or

$$M = \frac{N}{f} + 1. \qquad \left[\begin{array}{c}\text{eye focused at near point, } N;\\ N = 25\text{ cm for normal eye;}\\ \text{magnifier near eye}\end{array}\right] \qquad \textbf{(33–6b)}$$

We see that the magnification is greater when the eye is focused at its near point, as compared to when it is relaxed. Equation 33–6b is valid only if the magnifier is near the eye. M decreases if the magnifier is moved away from your eye. But Eq. 33–6a is valid even if the magnifier is not close to the eye (relaxed eye—see Fig. 33–34) as long as the object is at the focal point of the magnifier.

EXAMPLE 33–13 ESTIMATE **A jeweler's "loupe."** An 8-cm-focal-length converging lens is used as a "jeweler's loupe," which is a magnifying glass. Estimate (*a*) the magnification when the eye is relaxed, and (*b*) the magnification if the eye is focused at its near point $N = 25$ cm.

APPROACH The magnification when the eye is relaxed is given by Eq. 33–6a. When the eye is focused at its near point, we use Eq. 33–6b and we assume the lens is near the eye.

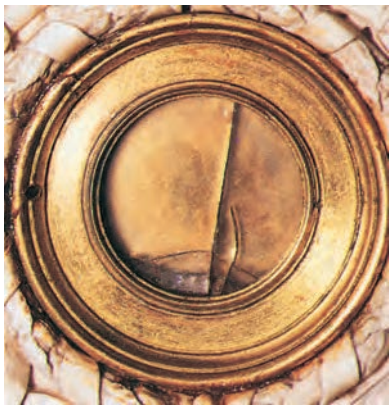
SOLUTION (*a*) With the relaxed eye focused at infinity,

$$M = \frac{N}{f} = \frac{25\text{ cm}}{8\text{ cm}} \approx 3\times.$$

(*b*) The magnification when the eye is focused at its near point ($N = 25$ cm), and the lens is near the eye, is

$$M = 1 + \frac{N}{f} = 1 + \frac{25}{8} \approx 4\times.$$

FIGURE 33–35 (a) Objective lens (mounted now in an ivory frame) from the telescope with which Galileo made his world-shaking discoveries, including the moons of Jupiter. (b) Telescopes made by Galileo (1609).



(a)



(b)

33–8 Telescopes

A telescope is used to magnify objects that are very far away. In most cases, the object can be considered to be at infinity.

Galileo, although he did not invent it,[†] developed the telescope into a usable and important instrument. He was the first to examine the heavens with the telescope (Fig. 33–35), and he made world-shaking discoveries, including the moons of Jupiter, the phases of Venus, sunspots, the structure of the Moon's surface, and that the Milky Way is made up of a huge number of individual stars.

Several types of **astronomical telescope** exist. The common **refracting** type, sometimes called **Keplerian**, contains two converging lenses located at opposite

[†]Galileo built his first telescope in 1609 after having heard of such an instrument existing in Holland. The first telescopes magnified only three to four times, but Galileo soon made a 30-power instrument. The first Dutch telescopes date from about 1604 and probably were copies of an Italian telescope built around 1590. Kepler (see Chapter 6) gave a ray description (in 1611) of the Keplerian telescope, which is named for him because he first described it, although he did not build it.

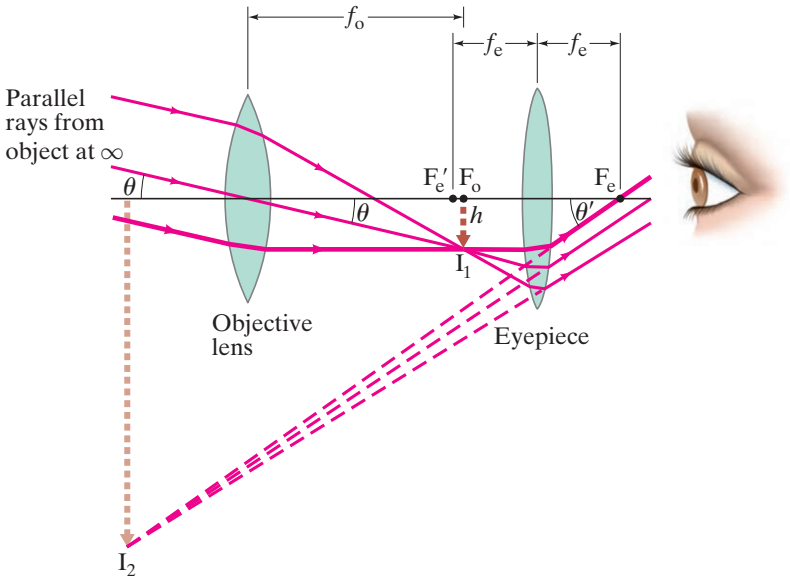


FIGURE 33–36 Astronomical telescope (refracting). Parallel light from one point on a distant object $(d_o = \infty)$ is brought to a focus by the objective lens in its focal plane. This image (I_1) is magnified by the eyepiece to form the final image I_2, which is inverted (and not much use for viewing objects on Earth). Only two of the rays shown entering the objective lens are standard rays (2 and 3) as described in Fig. 33–6.

ends of a long tube, as illustrated in Fig. 33–36. The lens closest to the object is called the **objective lens** (focal length f_o) and forms a real image I_1 of the distant object in the plane of its focal point F_o (or near it if the object is not at infinity). The second lens, called the **eyepiece** (focal length f_e), acts as a magnifier. That is, the eyepiece magnifies the image I_1 formed by the objective lens to produce a second, greatly magnified image, I_2, which is virtual and inverted. (Because the image is upside-down, this type of telescope is useful only for astronomical objects.) If the viewing eye is relaxed, the eyepiece is adjusted so the image I_2 is at infinity. Then the real image I_1 is at the focal point F_e' of the eyepiece, and the distance between the lenses is $f_o + f_e$ for an object at infinity.

PROBLEM SOLVING
Distance between lenses = $f_o + f_e$ for relaxed eye

To find the total angular magnification of this telescope, we note that the angle an object subtends as viewed by the unaided eye is just the angle θ subtended at the telescope objective. From Fig. 33–36 we can see that $\theta \approx h/f_o$, where h is the height of the image I_1 and we assume θ is small so that $\tan\theta \approx \theta$. Note, too, that the thickest of the three rays drawn in Fig. 33–36 is parallel to the axis before it strikes the eyepiece and therefore is refracted through the eyepiece focal point F_e on the far side of the eyepiece. Thus, $\theta' \approx h/f_e$ and the **total magnifying power** (that is, angular magnification, which is what is always quoted) of this telescope is

$$M = \frac{\theta'}{\theta} = \frac{(h/f_e)}{(h/f_o)} = -\frac{f_o}{f_e}, \qquad \left[\begin{matrix}\text{telescope}\\ \text{magnification}\end{matrix}\right] \quad \textbf{(33–7)}$$

where we used Eq. 33–5 and we inserted a minus sign to indicate that the image is inverted. To achieve a large magnification, the objective lens should have a long focal length and the eyepiece a short focal length.

In high-quality telescopes the eyepiece and objective are compound lenses, each consisting of several simple lenses to reduce aberrations (Section 33–10).

FIGURE 33–37 This large refracting telescope was built in 1897 and is housed at Yerkes Observatory in Wisconsin. The objective lens is 102 cm (40 inches) in diameter, and the telescope tube is about 19 m long. Example 33–14.



EXAMPLE 33–14 **Telescope magnification.** The largest optical refracting telescope in the world is located at the Yerkes Observatory in Wisconsin, Fig. 33–37. It is referred to as a "40-inch" telescope, meaning that the diameter of the objective is 40 in., or 102 cm. The objective lens has a focal length of 19 m, and the eyepiece has a focal length of 10 cm. (*a*) Calculate the total magnifying power of this telescope. (*b*) Estimate the length of the telescope.

APPROACH Equation 33–7 gives the magnification. The length of the telescope is the distance between the two lenses.

SOLUTION (*a*) From Eq. 33–7 we find

$$M = -\frac{f_o}{f_e} = -\frac{19\text{ m}}{0.10\text{ m}} = -190\times.$$

(*b*) For a relaxed eye, the image I_1 is at the focal point of both the eyepiece and the objective lenses. The distance between the two lenses is thus $f_o + f_e \approx 19$ m, which is essentially the length of the telescope.

EXERCISE F A 40× telescope has a 1.2-cm focal length eyepiece. What is the focal length of the objective lens?

For an astronomical telescope to produce bright images of faint stars, the objective lens must be large to allow in as much light as possible. Indeed, the diameter of the objective lens (and hence its "light-gathering power") is an important parameter for an astronomical telescope, which is why the largest ones are specified by giving the objective diameter (such as the 10-meter Keck telescope in Hawaii). The construction and grinding of large lenses is very difficult. And they would be very heavy. Therefore, the largest telescopes are **reflecting telescopes** which use a curved mirror as the objective, Fig. 33–38. A mirror has only one surface to be ground and can be supported along its entire surface† (a large lens, supported at its edges, would sag under its own weight). Often, the eyepiece lens or mirror (see Fig. 33–38) is removed so that the real image formed by the objective mirror can be recorded directly on film or on an electronic sensor (CCD or CMOS, Section 33–5).

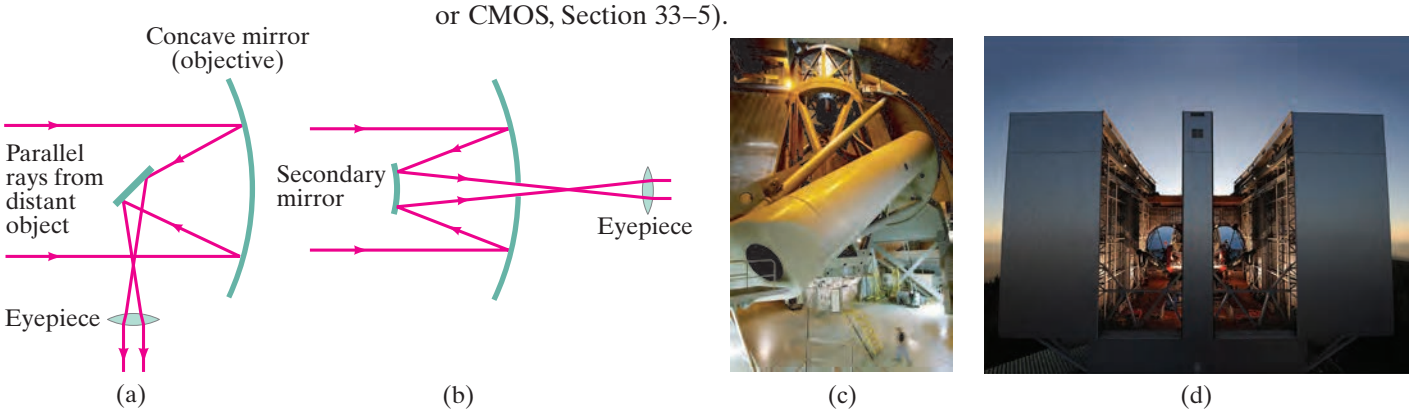
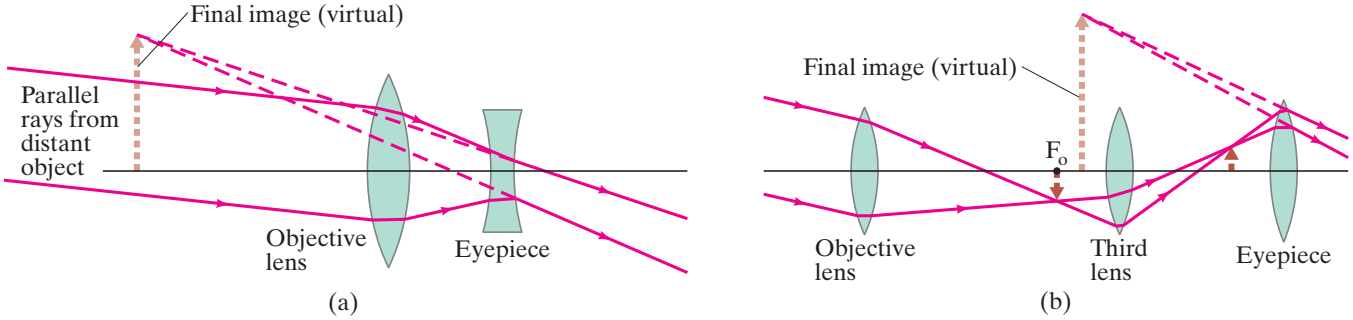


FIGURE 33–38 A concave mirror can be used as the objective of an astronomical telescope. Arrangement (a) is called the Newtonian focus, and (b) the Cassegrainian focus. Other arrangements are also possible. (c) The 200-inch (mirror diameter) Hale telescope on Palomar Mountain in California. (d) The Large Binocular Telescope in Arizona uses two 8.4-m-diameter objective reflectors; it was constructed in Italy.

A **terrestrial telescope**, for viewing objects on Earth, must provide an upright image—seeing normal objects upside down would be difficult (much less important for viewing stars). Two designs are shown in Fig. 33–39. The **Galilean** type, which Galileo used for his great astronomical discoveries, has a diverging lens as eyepiece which intercepts the converging rays from the objective lens before they reach a focus, and acts to form a virtual upright image, Fig. 33–39a. This design is still used in opera glasses. The tube is reasonably short, but the field of view is small. The second type, shown in Fig. 33–39b, is often called a **spyglass** and makes use of a third convex lens that acts to make the image upright as shown. A spyglass must be quite long. The most practical design today is the **prism binocular** which was shown in Fig. 32–34. The objective and eyepiece are converging lenses. The prisms reflect the rays by total internal reflection and shorten the physical size of the device, and they also act to produce an upright image. One prism reinverts the image in the vertical plane, the other in the horizontal plane.

FIGURE 33–39 Terrestrial telescopes that produce an upright image: (a) Galilean; (b) spyglass, or erector type.



†Another advantage of mirrors is that they exhibit no chromatic aberration because the light doesn't refract through them; and mirrors can be ground into a parabolic shape to correct for spherical aberration (Section 33–10). The reflecting telescope was first proposed by Newton.

33–9 Compound Microscope

The compound **microscope**, like the telescope, has both objective and eyepiece (or ocular) lenses, Fig. 33–40. The design is different from that for a telescope because a microscope is used to view objects that are very close, so the object distance is very small. The object is placed just beyond the objective's focal point as shown in Fig. 33–40a. The image I_1 formed by the objective lens is real, quite far from the objective lens, and much enlarged. The eyepiece is positioned so that this image I_1 (which serves as object for the eyepiece) is near the eyepiece focal point F_e. The image I_1 is magnified by the eyepiece into a very large virtual image, I_2, which is seen by the eye and is inverted. Modern microscopes use a third "tube" lens behind the objective, but we will analyze the simpler arrangement shown in Fig. 33–40a.

PHYSICS APPLIED
Microscopes

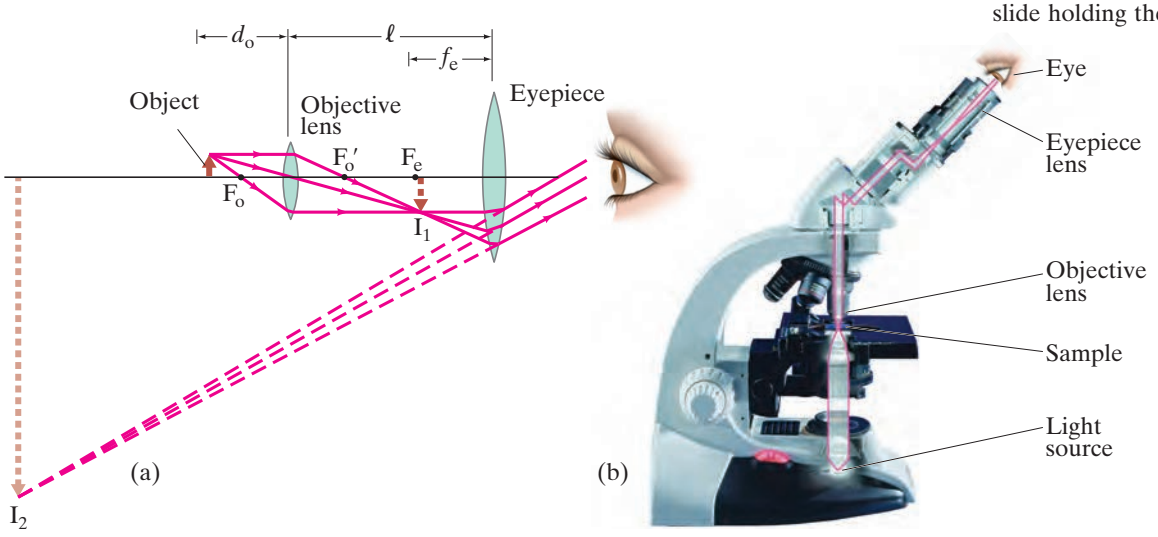


FIGURE 33–40 Compound microscope: (a) ray diagram, (b) photograph (illumination comes from below, outlined in red, then up through the slide holding the sample or object).

The overall magnification of a microscope is the product of the magnifications produced by the two lenses. The image I_1 formed by the objective lens is a factor m_o greater than the object itself, so we use the lateral magnification, Eq. 33–3. From Fig. 33–40a, the lateral magnification of the objective lens is

$$m_o = \frac{h_i}{h_o} = \frac{d_i}{d_o} = \frac{\ell - f_e}{d_o}, \qquad \textbf{(33–8)}$$

where d_o and d_i are the object and image distances for the objective lens, and ℓ is the distance between the lenses (equal to the length of the barrel); we ignored the minus sign in Eq. 33–3 which only tells us that the image is inverted. We wrote $d_i = \ell - f_e$, which is exact only if the eye is relaxed, so that the image I_1 is at the eyepiece focal point F_e. The eyepiece acts like a simple magnifier, so we use the angular magnification M_e (Eq. 33–6a) assuming the eye is relaxed:

$$M_e = \frac{N}{f_e}, \qquad \textbf{(33–9)}$$

where the near point $N = 25\,\text{cm}$ for the normal eye. Since the eyepiece enlarges the image formed by the objective, the overall angular magnification M is the product of the magnification of the objective lens, m_o, times the angular magnification, M_e, of the eyepiece lens (Eqs. 33–8 and 33–9):

$$M = M_e m_o = \left(\frac{N}{f_e}\right)\left(\frac{\ell - f_e}{d_o}\right) \qquad \left[\begin{matrix}\text{microscope}\\ \text{magnification}\end{matrix}\right] \qquad \textbf{(33–10a)}$$

$$\approx \frac{N\ell}{f_e f_o}. \qquad [f_o \text{ and } f_e \ll \ell] \qquad \textbf{(33–10b)}$$

The approximation, Eq. 33–10b, is accurate when f_e and f_o are small compared to ℓ, so $\ell - f_e \approx \ell$, and the object is near F_o so $d_o \approx f_o$ (Fig. 33–40a). This is a good approximation for large magnifications, which are obtained when f_o and f_e are very small (they are in the denominator of Eq. 33–10b). These microscope lenses of very short focal length are compound lenses involving several elements to avoid serious aberrations, as discussed in the next Section.

EXAMPLE 33–15 Microscope. A compound microscope consists of a 10× eyepiece and a 50× objective 17.0 cm apart. Determine (*a*) the overall magnification, (*b*) the focal length of each lens, and (*c*) the position of the object when the final image is in focus with the eye relaxed. Assume a normal eye, so $N = 25\,\text{cm}$.

APPROACH The overall magnification is the product of the eyepiece magnification and the objective magnification. The focal length of the eyepiece is found from Eq. 33–6a or 33–9 for the magnification of a simple magnifier. For the objective lens, it is easier to next find d_o (part *c*) using Eq. 33–8 before we find f_o.

SOLUTION (*a*) The overall magnification is $(10\times)(50\times) = 500\times$.

(*b*) The eyepiece focal length is (Eq. 33–9) $f_e = N/M_e = 25\,\text{cm}/10 = 2.5\,\text{cm}$. Next we solve Eq. 33–8 for d_o, and find

$$d_o = \frac{\ell - f_e}{m_o} = \frac{(17.0\,\text{cm} - 2.5\,\text{cm})}{50} = 0.29\,\text{cm}.$$

Then, from the thin lens equation for the objective with $d_i = \ell - f_e = 14.5\,\text{cm}$ (see Fig. 33–40a),

$$\frac{1}{f_o} = \frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{0.29\,\text{cm}} + \frac{1}{14.5\,\text{cm}} = 3.52\,\text{cm}^{-1};$$

so $f_o = 1/(3.52\,\text{cm}^{-1}) = 0.28\,\text{cm}$.

(*c*) We just calculated $d_o = 0.29\,\text{cm}$, which is very close to f_o.

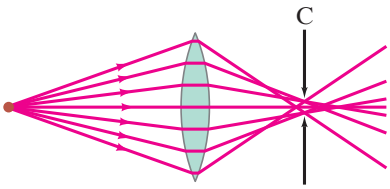
33–10 Aberrations of Lenses and Mirrors

PHYSICS APPLIED
Lens aberrations

Earlier in this Chapter we developed a theory of image formation by a thin lens. We found, for example, that all rays from each point on an object are brought to a single point as the image point. This result, and others, were based on approximations for a thin lens, mainly that all rays make small angles with the axis and that we can use $\sin\theta \approx \theta$. Because of these approximations, we expect deviations from the simple theory, which are referred to as **lens aberrations**. There are several types of aberration; we will briefly discuss each of them separately, but all may be present at one time.

Consider an object at any point (even at infinity) on the axis of a lens with spherical surfaces. Rays from this point that pass through the outer regions of the lens are brought to a focus at a different point than those that pass through the center of the lens. This is called **spherical aberration**, and is shown exaggerated in Fig. 33–41. Consequently, any object point will have its image on a screen or film not as a point but as a tiny circular patch of light. If the sensor or film is placed at point C, as indicated, the circle will have its smallest diameter, which is referred to as the **circle of least confusion**. Spherical aberration is present whenever spherical surfaces are used. It can be reduced by using nonspherical (= aspherical) lens surfaces, but grinding such lenses is difficult and expensive. Spherical aberration can be reduced by the use of several lenses in combination (a **compound lens**), and by using primarily the central part of lenses.

FIGURE 33–41 Spherical aberration (exaggerated). Circle of least confusion is at C.



For object points off the lens axis, additional aberrations occur. Rays passing through the different parts of the lens cause spreading of the image that is noncircular. There are two effects: **coma** (because the image of a point is comet-shaped rather than a tiny circle) and **off-axis astigmatism**.† Furthermore, the image points for objects off the axis but at the same distance from the lens do not fall on a flat plane but on a curved surface—that is, the focal plane is not flat. (We expect this because the points on a flat plane, such as the film in a camera, are not equidistant

†Although the effect is the same as for astigmatism in the eye (Section 33–6), the cause is different. Off-axis astigmatism is no problem in the eye because objects are clearly seen only at the fovea, on the lens axis.

from the lens.) This aberration is known as **curvature of field** and is a problem in cameras and other devices where the sensor or film is a flat plane. In the eye, however, the retina is curved, which compensates for this effect.

Another aberration, **distortion**, is a result of variation of magnification at different distances from the lens axis. Thus a straight-line object some distance from the axis may form a curved image. A square grid of lines may be distorted to produce *barrel distortion*, or *pincushion distortion*, Fig. 33–42. The former is common in extreme wide-angle lenses.

All the above aberrations occur for monochromatic light and hence are referred to as *monochromatic aberrations*. Normal light is not monochromatic, and there will also be **chromatic aberration**. This aberration arises because of dispersion—the variation of index of refraction of transparent materials with wavelength (Section 32–6). For example, blue light is bent more than red light by glass. So if white light is incident on a lens, the different colors are focused at different points, Fig. 33–43, and have slightly different magnifications resulting in colored fringes in the image. Chromatic aberration can be eliminated for any two colors (and reduced greatly for all others) by the use of two lenses made of different materials with different indices of refraction and dispersion. Normally one lens is converging and the other diverging, and they are often cemented together (Fig. 33–44). Such a lens combination is called an **achromatic doublet** (or "color-corrected" lens).

To reduce aberrations, high-quality lenses are **compound lenses** consisting of many simple lenses, referred to as **elements**. A typical high-quality camera lens may contain six to eight (or more) elements. For simplicity we will usually indicate lenses in diagrams as if they were simple lenses.

The human eye is also subject to aberrations, but they are minimal. Spherical aberration, for example, is minimized because (1) the cornea is less curved at the edges than at the center, and (2) the lens is less dense at the edges than at the center. Both effects cause rays at the outer edges to be bent less strongly, and thus help to reduce spherical aberration. Chromatic aberration is partially compensated for because the lens absorbs some of the shorter wavelengths and the retina is less sensitive to the blue and violet wavelengths. This is just the region of the spectrum where dispersion—and thus chromatic aberration—is greatest (Fig. 32–28).

Spherical mirrors (Section 32–3) also suffer aberrations including spherical aberration (see Fig. 32–13). Mirrors can be ground in a parabolic shape to correct for aberrations, but they are much harder to make and therefore very expensive. Spherical mirrors do not, however, exhibit chromatic aberration because the light does not pass through them (no refraction, no dispersion).

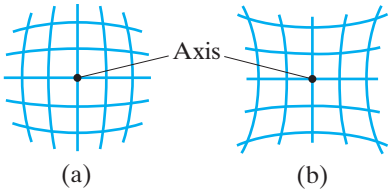
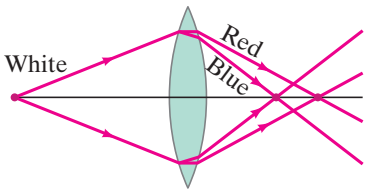


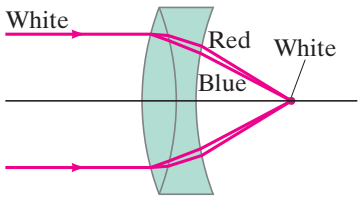
FIGURE 33–42 Distortion: lenses may image a square grid of perpendicular lines to produce (a) barrel distortion or (b) pincushion distortion. These distortions can be seen in the photograph of Fig. 33–2d for a simple lens.

FIGURE 33–43 Chromatic aberration. Different colors are focused at different points.



PHYSICS APPLIED
Human eye

FIGURE 33–44 Achromatic doublet.



Summary

A lens uses refraction to produce a real or virtual image. Parallel rays of light are focused to a point, the **focal point**, by a **converging** lens. The distance of the focal point from a thin lens is the **focal length** f of the lens. It is the same on both sides of the lens.

After parallel rays pass through a **diverging** lens, they appear to diverge from a point in front of the lens, which is its focal point; and the corresponding focal length is considered negative.

The **power** P of a lens, which is

$$P = \frac{1}{f} \qquad \textbf{(33–1)}$$

is given in diopters, which are units of inverse meters (m^{-1}).

For a given object, the position and size of the image formed by a lens can be found approximately by ray tracing. Algebraically, the relation between image and object distances, d_i and d_o, and the focal length f of a thin lens, is given by the **thin lens equation**:

$$\frac{1}{d_\mathrm{o}} + \frac{1}{d_\mathrm{i}} = \frac{1}{f}. \qquad \textbf{(33–2)}$$

The ratio of image height to object height, which equals the **lateral magnification** m for a lens, is

$$m = \frac{h_\mathrm{i}}{h_\mathrm{o}} = -\frac{d_\mathrm{i}}{d_\mathrm{o}}. \qquad \textbf{(33–3)}$$

When using the equations of geometric optics, you must remember the **sign conventions** for all quantities involved: carefully review them (page 963) when doing Problems.

When two (or more) thin lenses are used in combination to produce an image, the thin lens equation can be used for each lens in sequence. The image produced by the first lens acts as the object for the second lens.

The **lensmaker's equation** relates the radii of curvature of the lens surfaces and the lens' index of refraction to the focal length of the lens.

A **camera** lens forms an image on film, or on an electronic sensor (CCD or CMOS) in a digital camera. Light is allowed in briefly through a shutter. The image is focused by moving the lens relative to the film or sensor, and the ***f*-stop** (or lens opening) must be adjusted for the brightness of the scene and the chosen shutter speed. The *f*-stop is defined as the ratio of the focal length to the diameter of the lens opening.

The human **eye** also adjusts for the available light—by opening and closing the iris. It focuses not by moving the lens, but by using muscles to adjust the shape of the lens to vary its focal length. The image is formed on the retina, which contains an array of receptors known as rods and cones.

Diverging eyeglass or contact lenses are used to correct the defect of a nearsighted eye, which cannot focus well on distant objects. Converging lenses are used to correct for defects in which the eye cannot focus on close objects.

A **simple magnifier**, or **magnifying glass**, is a converging lens that forms a virtual image of an object placed at (or within) the focal point.

The **angular magnification**, when viewed by a relaxed eye, is

$$M = \frac{N}{f}, \qquad \textbf{(33–6a)}$$

where f is the focal length of the lens and N is the near point of the eye (25 cm for a "normal" eye).

An **astronomical telescope** consists of an **objective** lens or mirror, and an **eyepiece** that magnifies the real image formed by the objective. The **magnification** is equal to the ratio of the objective and eyepiece focal lengths, and the image is inverted:

$$M = -\frac{f_o}{f_e}. \qquad \textbf{(33–7)}$$

A **terrestrial telescope**, for which there are several designs, produces an upright image.

A compound **microscope** also uses objective and eyepiece lenses, and the final image is inverted. The total magnification is the product of the magnifications of the two lenses and is approximately

$$M \approx \frac{N\ell}{f_e f_o}, \qquad \textbf{(33–10b)}$$

where ℓ is the distance between the lenses, N is the near point of the eye, and f_o and f_e are the focal lengths of objective and eyepiece, respectively.

Microscopes, telescopes, and other optical instruments are limited in the formation of sharp images by **lens aberrations**. These include **spherical aberration**, in which rays passing through the edge of a lens are not focused at the same point as rays that pass near the center; and **chromatic aberration**, in which different colors are focused at different points. Compound lenses, consisting of several elements, can largely correct for aberrations.

Questions

1. Where must the film be placed if a camera lens is to make a sharp image of an object far away? Explain.
2. A photographer moves closer to his subject and then refocuses. Does the camera lens move farther away from, or closer to, the camera film or sensor? Explain.
3. Use ray diagrams to show that a real image formed by a thin lens is always inverted, whereas a virtual image is always upright if the object is real.
4. Light rays are said to be "reversible." Is this consistent with the thin lens equation? Explain.
5. A thin converging lens is moved closer to a nearby object. Does the real image formed change (*a*) in position, (*b*) in size? If yes, describe how.
6. Explain how you could have a virtual object.
7. A dog with its tail in the air stands facing a converging lens. If the nose and the tail are each focused on a screen in turn, which will have the greater magnification?
8. A cat with its tail in the air stands facing a converging lens. Under what circumstances (if any) would the image of the nose be virtual and the image of the tail be real? Where would the image of the rest of the cat be?
9. The thicker a double convex lens is in the center as compared to its edges, the shorter its focal length for a given lens diameter. Explain.
10. Consider two converging lenses separated by some distance. An object is placed so that the image from the first lens lies exactly at the focal point of the second lens. Will this combination produce an image? If so, where? If not, why not?
11. Example 33–6 shows how to use a converging lens to measure the focal length of a diverging lens. (*a*) It is said that for this to work, the converging lens must be stronger than the diverging lens. What is meant by "stronger," and why is this statement true? (*b*) Why can't you measure the focal length of a diverging lens directly?
12. A non-symmetrical lens (say, planoconvex) forms an image of a nearby object. Use the lensmaker's equation to explain if the image point changes when the lens is turned around.
13. If a glass converging lens is placed in water, its focal length in water will be (*a*) longer, (*b*) shorter, or (*c*) the same as in air. Explain.
14. (*a*) Does the focal length of a lens depend on the fluid in which it is immersed? (*b*) What about the focal length of a spherical mirror? Explain.
15. A lens is made of a material with an index of refraction $n = 1.25$. In air, it is a converging lens. Will it still be a converging lens if placed in water? Explain, using a ray diagram.
16. An underwater lens consists of a carefully shaped thin-walled plastic container filled with air. What shape should it have in order to be (*a*) converging (*b*) diverging? Use ray diagrams to support your answer.
17. Will a nearsighted person who wears corrective lenses in her glasses be able to see clearly underwater when wearing those glasses? Use a diagram to show why or why not.

18. Why can a nearsighted person, wearing distance glasses, read a book more easily by moving his glasses down a bit on his nose? [*Hint*: see paragraph on contact lenses, page 978.]

19. You can tell whether people are nearsighted or farsighted by looking at the width of their face through their glasses. If a person's face appears narrower through the glasses (Fig. 33–45), is the person farsighted or nearsighted? Try to explain, but also check experimentally with friends who wear glasses.



FIGURE 33–45
Question 19.

20. In attempting to discern distant details, people will sometimes squint. Why does this help?

21. Is the image formed on the retina of the human eye upright or inverted? Discuss the implications of this for our perception of objects.

22. Reading glasses use converging lenses. A simple magnifier is also a converging lens. Are reading glasses therefore magnifiers? Discuss the similarities and differences between converging lenses as used for these two different purposes.

23. Nearsighted people often look over (or under) their glasses when they want to see something small up close, like a cell phone screen. Why?

24. Why are bifocals needed mainly by older persons and not generally by younger people?

25. Why is the depth of field greater, and the image sharper, when a camera lens is "stopped down" to a larger *f*-number? (Ignore diffraction.)

26. (*a*) Describe the difference between a real image and a virtual image. (*b*) Can your eyes tell the difference? (*c*) How can you tell the difference on a ray diagram? (*d*) How could you tell the difference between a virtual image and a real image experimentally? (*e*) If you were to take a photograph of a virtual image, would you see the image in the photograph? (*f*) If you were to put a CCD digital sensor at the location of a virtual image, would the image be captured on the sensor? (*g*) Explain any differences in your answers to parts (*e*) and (*f*).

27. Digital cameras may offer an optical zoom or a digital zoom. An optical zoom uses a variable focal-length lens, so only the central part of the field of view fills the entire sensor; a digital zoom electronically includes only the central pixels of the sensor, so objects are larger in the final picture. Discuss which is better, and why.

28. Spherical aberration in a thin lens is minimized if rays are bent equally by the two surfaces. If a planoconvex lens is used to form a real image of an object at infinity, which surface should face the object? Use ray diagrams to show why.

29. For both converging and diverging lenses, discuss how the focal length for red light differs from that for violet light.

30. Describe at least four advantages of using mirrors rather than lenses for an astronomical telescope.

31. Which aberrations present in a simple lens are not present (or are greatly reduced) in the human eye?

MisConceptual Questions

1. The focal length of a converging lens is
   (*a*) the diameter of the lens.
   (*b*) the thickness of the lens.
   (*c*) the distance from the lens at which incoming parallel rays bend to intersect at a point.
   (*d*) the distance from the lens at which all real images are formed.

2. Virtual images can be formed by
   (*a*) only mirrors.
   (*b*) only lenses.
   (*c*) only plane mirrors.
   (*d*) only curved mirrors or lenses.
   (*e*) plane and curved mirrors, and lenses.

3. Virtual images cannot be
   (*a*) seen.
   (*b*) photographed.
   (*c*) projected on a screen.
   (*d*) All of the above.
   (*e*) None of the above.

4. Thin lenses have two focal points. These are
   (*a*) always the image and object locations.
   (*b*) usually at different distances from the lens.
   (*c*) sometimes on the same side of the lens.
   (*d*) the same distance from the lens but on opposite sides.
   (*e*) one is the real image location and the other is the virtual image location.

5. The only means to create a real image with a single lens would be to place
   (*a*) the object inside the focal length of a converging lens.
   (*b*) the object inside the focal length of a diverging lens.
   (*c*) the object outside the focal length of a converging lens.
   (*d*) the object outside the focal length of a diverging lens.
   (*e*) Any of the above, given the correct distance from the focal point.

6. You cover half of a lens that is forming an image on a screen. Compare what happens when you cover the top half of the lens versus the bottom half.
   (*a*) When you cover the top half of the lens, the top half of the image disappears; when you cover the bottom half of the lens, the bottom half of the image disappears.
   (*b*) When you cover the top half of the lens, the bottom half of the image disappears; when you cover the bottom half of the lens, the top half of the image disappears.
   (*c*) The image becomes half as bright in both cases.
   (*d*) Nothing changes in either case.
   (*e*) The image disappears in both cases.

7. A lens can be characterized by its *power*, which
   (*a*) is the same as the magnification.
   (*b*) tells how much light the lens can focus.
   (*c*) depends on where the object is located.
   (*d*) is the reciprocal of the focal length.

8. Which of the following can form an image?
(*a*) A plane mirror.
(*b*) A curved mirror.
(*c*) A lens curved on both sides.
(*d*) A lens curved on only one side.
(*e*) All of the above.

9. The image of a nearby object formed by a camera lens is
(*a*) at the lens' focal point.
(*b*) always blurred.
(*c*) at the same location as the image of an object at infinity.
(*d*) farther from the lens than the lens' focal point.

10. While you are photographing a dog, it begins to move away. What must you do to keep it in focus?
(*a*) Increase the *f*-stop value.
(*b*) Decrease the *f*-stop value.
(*c*) Move the lens away from the sensor or film.
(*d*) Move the lens closer to the sensor or film.
(*e*) None of the above.

11. A converging lens, such as a typical magnifying glass,
(*a*) always produces a magnified image (image taller than the object).
(*b*) always produces an image smaller than the object.
(*c*) always produces an upright image.
(*d*) always produces an inverted image (upside down).
(*e*) None of these statements are true.

12. If the distance from your eye's lens to the retina is shorter than for a normal eye, you will struggle to see objects that are
(*a*) nearby.
(*b*) far away.
(*c*) colorful.
(*d*) moving fast.

13. When a nearsighted person looks at a distant object through her glasses, the image produced by the glasses should be
(*a*) about 25 cm from her eye.
(*b*) at her eye's far point.
(*c*) at her eye's near point.
(*d*) at the far point for a normal eye.

14. How do eyeglasses help a nearsighted person see more clearly?
(*a*) Diverging lenses bend light entering the eye, so the image focuses farther from the front of the eye.
(*b*) Diverging lenses bend light entering the eye, so the image focuses closer to the front of the eye.
(*c*) Converging lenses bend light entering the eye, so the image focuses farther from the front of the eye.
(*d*) Converging lenses bend light entering the eye, so the image focuses closer to the front of the eye.
(*e*) Lenses adjust the distance from the cornea to the back of the eye.

15. The image produced on the retina of the eye is _____ compared to the object being viewed.
(*a*) inverted
(*b*) upright
(*c*) sideways
(*d*) enlarged

16. Which of the following statements is true? (See Section 33–6.)
(*a*) Contact lenses and eyeglasses for the same person would have the same power.
(*b*) Farsighted people can see far clearly but not near.
(*c*) Nearsighted people cannot see near or far clearly.
(*d*) Astigmatism in vision is corrected by using different spherical lenses for each eye.

Problems

33–1 and 33–2 Thin Lenses

1. (I) A sharp image is located 373 mm behind a 235-mm-focal-length converging lens. Find the object distance (*a*) using a ray diagram, (*b*) by calculation.

2. (I) Sunlight is observed to focus at a point 16.5 cm behind a lens. (*a*) What kind of lens is it? (*b*) What is its power in diopters?

3. (I) (*a*) What is the power of a 26.5-cm-focal-length lens? (*b*) What is the focal length of a −6.25-D lens? Are these lenses converging or diverging?

4. (II) A certain lens focuses an object 2.15 m away as an image 48.3 cm on the other side of the lens. What type of lens is it and what is its focal length? Is the image real or virtual?

5. (II) A 105-mm-focal-length lens is used to focus an image on the sensor of a camera. The maximum distance allowed between the lens and the sensor plane is 132 mm. (*a*) How far in front of the sensor should the lens (assumed thin) be positioned if the object to be photographed is 10.0 m away? (*b*) 3.0 m away? (*c*) 1.0 m away? (*d*) What is the closest object this lens could photograph sharply?

6. (II) A stamp collector uses a converging lens with focal length 28 cm to view a stamp 15 cm in front of the lens. (*a*) Where is the image located? (*b*) What is the magnification?

7. (II) It is desired to magnify reading material by a factor of 3.0× when a book is placed 9.0 cm behind a lens. (*a*) Draw a ray diagram and describe the type of image this would be. (*b*) What type of lens is needed? (*c*) What is the power of the lens in diopters?

8. (II) How far from a converging lens with a focal length of 25 cm should an object be placed to produce a real image which is the same size as the object?

9. (II) An object is located 1.35 m from an 8.0-D lens. By how much does the image move if the object is moved (*a*) 0.90 m closer to the lens, and (*b*) 0.90 m farther from the lens?

10. (II) (*a*) How far from a 75.0-mm-focal-length lens must an object be placed if its image is to be magnified 2.50× and be real? (*b*) What if the image is to be virtual and magnified 2.50×?

11. (II) Repeat Problem 10 for a −75.0-mm-focal-length lens. [*Hint*: Consider objects real or virtual (formed by some other piece of optics).]

12. (II) A 3.10-cm-high insect is 1.30 m from a 135-mm-focal-length lens. (*a*) Where is the image, how high is it, and what type is it? (*b*) What if $f = -135$ mm?

13. (II) A bright object and a viewing screen are separated by a distance of 86.0 cm. At what location(s) between the object and the screen should a lens of focal length 16.0 cm be placed in order to produce a sharp image on the screen? [*Hint*: First draw a diagram.]

14. (II) How far apart are an object and an image formed by a 78-cm-focal-length converging lens if the image is 2.95× larger than the object and is real?

15. (II) Show algebraically that the image formed by a converging lens (*a*) is real and inverted if the object is beyond the focal point $(d_o > f)$, and (*b*) is virtual and upright if the object is within the focal point $(d_o < f)$. (*c*) Next, describe the image if the object is itself an image (formed by another lens), and its position is on the opposite side of the lens from the incoming light, for $-d_o > f$, and for $0 < -d_o < f$.

16. (II) In a film projector, the film acts as the object whose image is projected on a screen (Fig. 33–46). If a 105-mm-focal-length lens is to project an image on a screen 22.5 m away, how far from the lens should the film be? If the film is 24 mm wide, how wide will the picture be on the screen?



FIGURE 33–46 Film projector, Problem 16.

17. (III) (*a*) Show that the thin lens equation can be written in the *Newtonian form*:

$$xx' = f^2,$$

where x is the distance of the object from the focal point on the front side of the lens, and x' is the distance of the image to the focal point on the other side of the lens. (*b*) Calculate the location of an image if the object is placed 58.0 cm in front of a convex lens with a focal length of 36.0 cm using the standard form of the thin lens formula, and (*c*) again by using the Newtonian form, derived above.

18. (III) A bright object is placed on one side of a converging lens of focal length f, and a white screen for viewing the image is on the opposite side. The distance $d_T = d_i + d_o$ between the object and the screen is kept fixed, but the lens can be moved. (*a*) Show that if $d_T > 4f$, there will be two positions where the lens can be placed and a sharp image will be produced on the screen. (*b*) If $d_T < 4f$, show that there will be no lens position where a sharp image is formed. (*c*) Determine a formula for the distance between the two lens positions in part (*a*), and the ratio of the image sizes.

33–3 Lens Combinations

19. (II) Two 28.0-cm-focal-length converging lenses are placed 16.5 cm apart. An object is placed 35.0 cm in front of one lens. (*a*) Where will the final image formed by the second lens be located? (*b*) What is the total magnification?

20. (II) A diverging lens with $f = -36.5$ cm is placed 14.0 cm behind a converging lens with $f = 20.0$ cm. Where will an object at infinity be focused?

21. (II) A 38.0-cm-focal-length converging lens is 24.0 cm behind a diverging lens. Parallel light strikes the diverging lens. After passing through the converging lens, the light is again parallel. What is the focal length of the diverging lens? [*Hint*: First draw a ray diagram.]

22. (II) A diverging lens with a focal length of −14 cm is placed 12 cm to the right of a converging lens with a focal length of 18 cm. An object is placed 29 cm to the left of the converging lens. (*a*) Where will the final image be located? (*b*) Where will the image be if the diverging lens is 38 cm from the converging lens?

23. (II) Two lenses, one converging with focal length 20.0 cm and one diverging with focal length −10.0 cm, are placed 25.0 cm apart. An object is placed 60.0 cm in front of the converging lens. Determine (*a*) the position and (*b*) the magnification of the final image formed. (*c*) Sketch a ray diagram for this system.

24. (II) A diverging lens is placed next to a converging lens of focal length f_C, as in Fig. 33–14. If f_T represents the focal length of the combination, show that the focal length of the diverging lens, f_D, is given by

$$\frac{1}{f_D} = \frac{1}{f_T} - \frac{1}{f_C}.$$

25. (II) A lighted candle is placed 36 cm in front of a converging lens of focal length $f_1 = 13$ cm, which in turn is 56 cm in front of another converging lens of focal length $f_2 = 16$ cm (see Fig. 33–47). (*a*) Draw a ray diagram and estimate the location and the relative size of the final image. (*b*) Calculate the position and relative size of the final image.

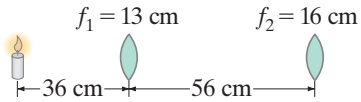


FIGURE 33–47 Problem 25.

33–4 Lensmaker's Equation

26. (I) A double concave lens has surface radii of 33.4 cm and 28.8 cm. What is the focal length if $n = 1.63$?

27. (I) Both surfaces of a double convex lens have radii of 31.4 cm. If the focal length is 26.3 cm, what is the index of refraction of the lens material?

28. (II) An object is placed 96.0 cm from a glass lens $(n = 1.52)$ with one concave surface of radius 22.0 cm and one convex surface of radius 18.5 cm. (*a*) Where is the final image? (*b*) What is the magnification?

29. (II) A prescription for an eyeglass lens calls for +3.50 diopters. The lensmaker grinds the lens from a "blank" with $n = 1.56$ and convex front surface with radius of curvature of 30.0 cm. What should be the radius of curvature of the other surface?

30. (II) A symmetric double convex lens with a focal length of 20.5 cm is to be made from glass with an index of refraction of 1.52. What should be the radius of curvature for each surface?

*31. (III) Derive the lensmaker's equation (like Eq. 33–4) for a lens immersed in a liquid of index of refraction n_1. The lens has index of refraction n_2 and radii R_1 and R_2.

33–5 Cameras

32. (I) A 65-mm-focal-length lens has f-stops ranging from $f/1.4$ to $f/22$. What is the corresponding range of lens diaphragm diameters?

33. (I) A light meter reports that a camera setting of $\frac{1}{500}$ s at $f/5.6$ will give a correct exposure. But the photographer wishes to use $f/11$ to increase the depth of field. What should the shutter speed be?

34. (I) A television camera lens has a 17-cm focal length and a lens diameter of 6.0 cm. What is its f-number?

35. (II) A "pinhole" camera uses a tiny pinhole instead of a lens. Show, using ray diagrams, how reasonably sharp images can be formed using such a pinhole camera. In particular, consider two point objects 2.0 cm apart that are 1.0 m from a 1.0-mm-diameter pinhole. Show that on a piece of film 7.0 cm behind the pinhole the two objects produce two separate circles that do not overlap.

36. (II) Suppose that a correct exposure is $\frac{1}{250}$ s at $f/11$. Under the same conditions, what exposure time would be needed for a pinhole camera (Problem 35) if the pinhole diameter is 1.0 mm and the film is 7.0 cm from the hole?

37. (II) Human vision normally covers an angle of roughly 40° horizontally. A "normal" camera lens then is defined as follows: When focused on a distant horizontal object which subtends an angle of 40°, the lens produces an image that extends across the full horizontal extent of the camera's light-recording medium (film or electronic sensor). Determine the focal length f of the "normal" lens for the following types of cameras: (*a*) a 35-mm camera that records images on film 36 mm wide; (*b*) a digital camera that records images on a digital electronic sensor 1.00 cm wide.

38. (II) A nature photographer wishes to shoot a 34-m-tall tree from a distance of 55 m. What focal-length lens should be used if the image is to fill the 24-mm height of the sensor?

39. (II) How large is the image of the Sun on a camera sensor when using (*a*) a 28-mm-focal-length lens, (*b*) a 50-mm-focal-length lens, and (*c*) a 135-mm-focal-length lens? (*d*) If the 50-mm lens is considered normal for this camera, what relative magnification does each of the other two lenses provide? The Sun has diameter 1.4×10^6 km, and it is 1.5×10^8 km away.

40. (II) If a 155-mm telephoto lens is designed to cover object distances from 1.30 m to ∞, over what distance must the lens move relative to the plane of the sensor or film?

41. (II) A 200-mm-focal-length lens can be adjusted so that it is 200.0 mm to 208.2 mm from the film. For what range of object distances can it be adjusted?

42. (III) Show that for objects very far away (assume infinity), the magnification of any camera lens is proportional to its focal length.

33–6 Eye and Corrective Lenses

43. (I) A human eyeball is about 2.0 cm long and the pupil has a maximum diameter of about 8.0 mm. What is the "speed" of this lens?

44. (II) One lens of a nearsighted person's eyeglasses has a focal length of −23.0 cm and the lens is 1.8 cm from the eye. If the person switches to contact lenses placed directly on the eye, what should be the focal length of the corresponding contact lens?

45. (II) Reading glasses of what power are needed for a person whose near point is 105 cm, so that he can read a computer screen at 55 cm? Assume a lens–eye distance of 1.8 cm.

46. (II) A person struggles to read by holding a book at arm's length, a distance of 55 cm away, so the print seems too small to read. What power of reading glasses should be prescribed for her, assuming they will be placed 2.0 cm from the eye and she wants to read at the "normal" near point of 25 cm?

47. (II) An eye is corrected by a −3.50-D lens, 2.0 cm from the eye. (*a*) Is this eye near- or farsighted? (*b*) What is this eye's far point without glasses?

48. (II) A person's right eye can see objects clearly only if they are between 25 cm and 88 cm away. (*a*) What power of contact lens is required so that objects far away are sharp? (*b*) What will be the near point with the lens in place?

49. (II) A person has a far point of 13 cm. What power glasses would correct this vision if the lenses are 2.0 cm from the eye? What power contact lenses, placed on the eye, would the person need?

50. (II) What is the focal length of the eye–lens system when viewing an object (*a*) at infinity, and (*b*) 45 cm from the eye? Assume that the cornea–retina distance is 2.0 cm.

51. (II) A nearsighted person has near and far points of 10.6 and 18.0 cm respectively. If she puts on contact lenses with power $P = -4.00$ D, what are her new near and far points?

52. (II) The closely packed cones in the fovea of the eye each have a diameter of about 2 μm. For the eye to discern two images on the fovea as distinct, assume that the images must be separated by at least one cone that is not excited. If these images are of two point-like objects at the eye's 25-cm near point, how far apart are these barely resolvable objects? Assume the diameter of the eye (cornea-to-fovea distance) is 2.0 cm.

53. (II) About how much longer is the nearsighted eye in Example 33–12 than the 2.0 cm of a normal eye?

33–7 Magnifying Glass

54. (I) A magnifier is rated at 2.5× for a normal eye focusing on an image at the near point. (*a*) What is its focal length? (*b*) What is its focal length if the 2.5× refers to a relaxed eye?

55. (I) What is the magnification of a lens used with a relaxed eye if its focal length is 16 cm?

56. (II) Sherlock Holmes is using an 8.80-cm-focal-length lens as his magnifying glass. To obtain maximum magnification, where must the object be placed (assume a normal eye), and what will be the magnification?

57. (II) A 3.40-mm-wide bolt is viewed with a 9.20-cm-focal-length lens placed near the eye. A normal eye views the image at its near point. Calculate (*a*) the angular magnification, (*b*) the width of the image, and (*c*) the object distance from the lens.

58. (II) A magnifying glass with a focal length of 9.8 cm is used to read print placed at a distance of 8.3 cm. Calculate (*a*) the position of the image; (*b*) the angular magnification.

59. (II) A magnifying glass is rated at 3.0× for a normal eye that is relaxed. What would be the magnification for a relaxed eye whose near point is (*a*) 57 cm, and (*b*) 17 cm? Explain the differences.

60. (II) A converging lens of focal length $f = 12\,\text{cm}$ is being used by a writer as a magnifying glass to read some fine print on his book contract. Initially, the writer holds the lens above the fine print so that its image is at infinity. To get a better look, he then moves the lens so that the image is at his 25-cm near point. How far, and in what direction (toward or away from the fine print) did the writer move the lens? Assume his eye is adjusted to remain always very near the magnifying glass.

33–8 Telescopes

61. (I) What is the magnification of an astronomical telescope whose objective lens has a focal length of 86 cm, and whose eyepiece has a focal length of 2.8 cm? What is the overall length of the telescope when adjusted for a relaxed eye?

62. (I) The overall magnification of an astronomical telescope is desired to be $35\times$. If an objective of 78 cm focal length is used, what must be the focal length of the eyepiece? What is the overall length of the telescope when adjusted for use by the relaxed eye?

63. (II) An $8.0\times$ binocular has 3.0-cm-focal-length eyepieces. What is the focal length of the objective lenses?

64. (II) A Galilean telescope adjusted for a relaxed eye is 33.8 cm long. If the objective lens has a focal length of 36.0 cm, what is the magnification?

65. (II) An astronomical telescope has its two lenses spaced 78.0 cm apart. If the objective lens has a focal length of 75.5 cm, what is the magnification of this telescope? Assume a relaxed eye.

66. (II) The Moon's image appears to be magnified $140\times$ by a reflecting astronomical telescope with an eyepiece having a focal length of 3.1 cm. What are the focal length and radius of curvature of the main (objective) mirror?

67. (II) What is the magnifying power of an astronomical telescope using a reflecting mirror whose radius of curvature is 6.1 m and an eyepiece whose focal length is 2.8 cm?

68. (II) A $120\times$ astronomical telescope is adjusted for a relaxed eye when the two lenses are 1.15 m apart. What is the focal length of each lens?

69. (II) An astronomical telescope longer than about 50 cm is not easy to hold by hand. Estimate the maximum angular magnification achievable for a telescope designed to be handheld. Assume its eyepiece lens, if used as a magnifying glass, provides a magnification of $5\times$ for a relaxed eye with near point $N = 25\,\text{cm}$.

70. (III) A reflecting telescope (Fig. 33–38b) has a radius of curvature of 3.25 m for its objective mirror and a radius of curvature of $-1.75\,\text{m}$ for its secondary mirror. If the distance between the two mirrors is 0.90 m, how far in front of the secondary mirror should you place the electronic sensor to record the image of a star?

71. (III) A $6.5\times$ pair of binoculars has an objective focal length of 26 cm. If the binoculars are focused on an object 4.0 m away (from the objective), what is the magnification? (The $6.5\times$ refers to objects at infinity; Eq. 33–7 holds only for objects at infinity and not for nearby ones.)

33–9 Microscopes

72. (I) A microscope uses an eyepiece with a focal length of 1.50 cm. Using a normal eye with a final image at infinity, the barrel length is 19.5 cm and the focal length of the objective lens is 0.65 cm. What is the magnification of the microscope?

73. (I) A $630\times$ microscope uses a 0.40-cm-focal-length objective lens. If the barrel length is 17.5 cm, what is the focal length of the eyepiece? Assume a normal eye and that the final image is at infinity.

74. (II) A microscope has a $13.0\times$ eyepiece and a $62.0\times$ objective lens 20.0 cm apart. Calculate (*a*) the total magnification, (*b*) the focal length of each lens, and (*c*) where the object must be for a normal relaxed eye to see it in focus.

75. (II) Repeat Problem 74 assuming that the final image is located 25 cm from the eyepiece (near point of a normal eye).

76. (II) A microscope has a 1.8-cm-focal-length eyepiece and a 0.80-cm objective. Assuming a relaxed normal eye, calculate (*a*) the position of the object if the distance between the lenses is 16.8 cm, and (*b*) the total magnification.

77. (II) The eyepiece of a compound microscope has a focal length of 2.80 cm and the objective lens has $f = 0.740\,\text{cm}$. If an object is placed 0.790 cm from the objective lens, calculate (*a*) the distance between the lenses when the microscope is adjusted for a relaxed eye, and (*b*) the total magnification.

78. (II) An inexpensive instructional lab microscope allows the user to select its objective lens to have a focal length of 32 mm, 15 mm, or 3.9 mm. It also has two possible eyepieces with magnifications $5\times$ and $15\times$. Each objective forms a real image 160 mm beyond its focal point. What are the largest and smallest overall magnifications obtainable with this instrument?

79. (III) Given two 12-cm-focal-length lenses, you attempt to make a crude microscope using them. While holding these lenses a distance 55 cm apart, you position your microscope so that its objective lens is distance d_o from a small object. Assume your eye's near point $N = 25\,\text{cm}$. (*a*) For your microscope to function properly, what should d_o be? (*b*) Assuming your eye is relaxed when using it, what magnification M does your microscope achieve? (*c*) Since the length of your microscope is not much greater than the focal lengths of its lenses, the approximation $M \approx N\ell/f_e f_o$ is not valid. If you apply this approximation to your microscope, what % error do you make in your microscope's true magnification?

33–10 Lens Aberrations

80. (II) A planoconvex lens (Fig. 33–2a) has one flat surface and the other has $R = 15.3\,\text{cm}$. This lens is used to view a red and yellow object which is 62.0 cm away from the lens. The index of refraction of the glass is 1.5106 for red light and 1.5226 for yellow light. What are the locations of the red and yellow images formed by the lens?

81. (II) An achromatic lens is made of two very thin lenses, placed in contact, that have focal lengths $f_1 = -28\,\text{cm}$ and $f_2 = +25\,\text{cm}$. (*a*) Is the combination converging or diverging? (*b*) What is the net focal length?

General Problems

82. A small object is 25.0 cm from a diverging lens as shown in Fig. 33–48. A converging lens with a focal length of 12.0 cm is 30.0 cm to the right of the diverging lens. The two-lens system forms a real inverted image 17.0 cm to the right of the converging lens. What is the focal length of the diverging lens?

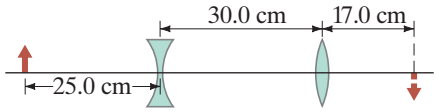


FIGURE 33–48 Problem 82.

83. A converging lens with focal length of 13.0 cm is placed in contact with a diverging lens with a focal length of −18.0 cm. What is the focal length of the combination, and is the combination converging or diverging?

84. Figure 33–49 is a photograph of an eyeball with the image of a boy in a doorway. (*a*) Is the eye here acting as a lens or as a mirror? (*b*) Is the eye being viewed right side up or is the camera taking this photo upside down? (*c*) Explain, based on all possible images made by a convex mirror or lens.



FIGURE 33–49 Problem 84.

85. Which of the two lenses shown in Fig. 33–50 is converging, and which is diverging? Explain using ray diagrams and show how each image is formed.



FIGURE 33–50 Problem 85.

86. (*a*) An object 37.5 cm in front of a certain lens is imaged 8.60 cm in front of that lens (on the same side as the object). What type of lens is this, and what is its focal length? Is the image real or virtual? (*b*) If the image were located, instead, 44.5 cm in front of the lens, what type of lens would it be and what focal length would it have?

87. An object is placed 30.0 cm from a +5.0-D lens. A spherical mirror with focal length +25 cm is placed 65 cm behind the lens. Where is the final image? (Note that the mirror reflects light back through the lens.) Be sure to draw a diagram.

88. (*a*) Show that if two thin lenses of focal lengths f_1 and f_2 are placed in contact with each other, the focal length of the combination is given by $f_T = f_1 f_2/(f_1 + f_2)$. (*b*) Show that the power P of the combination of two lenses is the sum of their separate powers, $P = P_1 + P_2$.

89. Two converging lenses are placed 30.0 cm apart. The focal length of the lens on the right is 20.0 cm, and the focal length of the lens on the left is 15.0 cm. An object is placed to the left of the 15.0-cm-focal-length lens. A final image from both lenses is inverted and located halfway between the two lenses. How far to the left of the 15.0-cm-focal-length lens is the original object?

90. Figure 33–51 was taken from the NIST Laboratory (National Institute of Standards and Technology) in Boulder, CO, 2.0 km from the hiker in the photo. The Sun's image was 15 mm across on the film. Estimate the focal length of the camera lens (actually a telescope). The Sun has diameter 1.4×10^6 km, and it is 1.5×10^8 km away.



FIGURE 33–51 Problem 90.

91. A movie star catches a reporter shooting pictures of her at home. She claims the reporter was trespassing. To prove her point, she gives as evidence the film she seized. Her 1.75-m height is 8.55 mm high on the film, and the focal length of the camera lens was 220 mm. How far away from the subject was the reporter standing?

92. As early morning passed toward midday, and the sunlight got more intense, a photographer noted that, if she kept her shutter speed constant, she had to change the f-number from $f/5.6$ to $f/16$. By what factor had the sunlight intensity increased during that time?

93. A person with normal vision adjusts a microscope for a good image when her eye is relaxed. She then places a camera where her eye was. For what object distance should the camera be set? Explain.

94. A child has a near point of 15 cm. What is the maximum magnification the child can obtain using an 8.2-cm-focal-length magnifier? What magnification can a normal eye obtain with the same lens? Which person sees more detail?

95. A woman can see clearly with her right eye only when objects are between 45 cm and 175 cm away. Prescription bifocals should have what powers so that she can see distant objects clearly (upper part) and be able to read a book 25 cm away (lower part) with her right eye? Assume that the glasses will be 2.0 cm from the eye.

***96.** In a science-fiction novel, an intelligent ocean-dwelling creature's eye functions underwater with a near point of 25 cm. This creature would like to create an underwater magnifier out of a thin plastic container filled with air. What shape should the air-filled plastic container have (i.e., determine radii of curvature of its surfaces) in order for it to be used by the creature as a $3.5\times$ magnifier? Assume the eye is focused at its near point. [*Note*: see Problem 31.]

97. A physicist lost in the mountains tries to make a telescope using the lenses from his reading glasses. They have powers of $+2.0$ D and $+4.5$ D, respectively. (*a*) What maximum magnification telescope is possible? (*b*) Which lens should be used as the eyepiece?

98. A 50-year-old man uses $+2.5$-D lenses to read a newspaper 25 cm away. Ten years later, he must hold the paper 32 cm away to see clearly with the same lenses. What power lenses does he need now in order to hold the paper 25 cm away? (Distances are measured from the lens.)

99. An object is moving toward a converging lens of focal length f with constant speed v_o, but its distance d_o from the lens is always greater than f. (*a*) Determine the velocity v_i of the image as a function of d_o. (*b*) Which direction (toward or away from the lens) does the image move? (*c*) For what d_o does the image's speed equal the object's speed?

100. Two converging lenses, one with $f = 4.0$ cm and the other with $f = 41$ cm, are made into a telescope. (*a*) Determine the telescope's length and its magnification. Which lens should be the eyepiece? (*b*) Assume these lenses are now combined to make a microscope; if the magnification needs to be $125\times$, how long would the microscope be? Explain why the result in (*b*) is so impractical.

101. Both a converging lens and a concave mirror can produce virtual images that are larger than the object. Concave mirrors can be used as magnifying makeup or shaving mirrors, but converging lenses cannot be. (*a*) Draw ray diagrams to explain why not. (*b*) If a concave mirror has the same focal length as a converging lens, and an object is placed first at a distance of $\frac{1}{2}f$ from the lens and then at a distance of $\frac{1}{2}f$ from the mirror, how will the magnification of the object compare in the two cases?

102. What is the magnifying power of a $+4.0$-D lens used as a magnifier? Assume a relaxed normal eye.

103. Sam purchases $+3.70$-D eyeglasses which correct his faulty vision to put his near point at 25 cm. (Assume he wears the lenses 2.0 cm from his eyes.) (*a*) Calculate the focal length of Sam's glasses. (*b*) Calculate Sam's near point without glasses. (*c*) Pam, who has normal eyes with near point at 25 cm, puts on Sam's glasses. Calculate Pam's near point with Sam's glasses on.

104. The proper functioning of certain optical devices (e.g., optical fibers and spectrometers) requires that the input light be a collection of diverging rays within a cone of half-angle θ (Fig. 33–52). If the light initially exists as a collimated beam (i.e., parallel rays), show that a single lens of focal length f and diameter D can be used to create the required input light if $D/f = 2\tan\theta$. If $\theta = 3.5°$ for a certain device, what focal length lens should be used if the lens diameter is 5.0 cm?

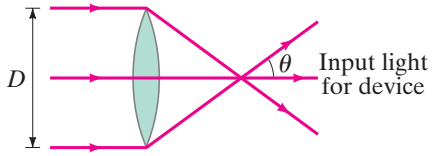


FIGURE 33–52 Problem 104.

105. The power of one lens in a pair of eyeglasses is -3.5 D. The radius of curvature of the outside surface is 14.0 cm. What is the radius of curvature of the inside surface? The lens is made of plastic with $n = 1.62$.

106. Draw ray diagrams for the following two cases: (*a*) A 10-mm-high object 100 mm from a 40-mm-focal-length converging lens. (*b*) A 10-mm-high object 100 mm from a 40-mm-focal-length convex mirror.

107. Figure 33–53 shows a converging lens held above three equal-sized letters A. In (a) the lens is 5 cm from the paper, and in (b) the lens is 15 cm from the paper. Estimate the focal length of the lens. What is the image position for each case?

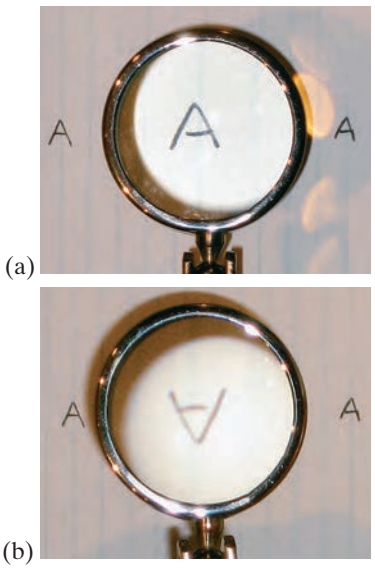


FIGURE 33–53 Problem 107

108. An astronomical telescope, Fig. 33–36, produces an inverted image. One way to make a telescope that produces an upright image is to insert a third lens between the objective and the eyepiece, Fig. 33–39b. To have the same magnification, the non-inverting telescope will be longer. Suppose lenses of focal length 150 cm, 1.5 cm, and 10 cm are available. Where should these three lenses be placed to make a non-inverting telescope with magnification 100×?

109. An object is placed so it is touching one surface of a thin lens. Where is the image in this case (i.e., when d_o approaches 0)?

110. Students in a physics lab are assigned to find the location where a bright object may be placed in order that a converging lens with $f = 12\,\text{cm}$ will produce an image three times the size of the object. Two students complete the assignment at different times using identical equipment, but when they compare notes later, they discover that their answers for the object distance are not the same. Explain why they do not necessarily need to repeat the lab, and justify your response with a calculation.

111. Redo Examples 33–9 and 33–10 assuming the sensor has only 6 MP. Explain the different results and their impact on finished photographs.

112. The focal length f of a converging lens can be found by placing an object of known size at various locations in front of the lens and measuring the resulting real-image distances d_i and their associated magnifications m (minus sign indicates that image is inverted). The data taken in such an experiment are given here:

d_i (cm)	20	25	30	35	40
m	−0.43	−0.79	−1.14	−1.50	−1.89

(*a*) Show algebraically that a graph of m vs. d_i should produce a straight line. What are the theoretically expected values for the slope and the y-intercept of this line? [*Hint*: d_o is not constant.] (*b*) Using the data above, graph m vs. d_i and show that a straight line does indeed result. Use the slope of this line to determine the focal length of the lens. Does the y-intercept of your plot have the expected value?

113. A telephoto lens system obtains a large magnification in a compact package. A simple such system can be constructed out of two lenses, one converging and one diverging, of focal lengths f_1 and $f_2 = -\frac{1}{2}f_1$, respectively, separated by a distance $\ell = \frac{3}{4}f_1$ as shown in Fig. 33–54. (*a*) For a distant object located at distance d_o from the first lens, show that the first lens forms an image with lateral magnification $m_1 \approx -f_1/d_o$ located very close to its focal point. Go on to show that the total magnification for the two-lens system is $m \approx -2f_1/d_o$. (*b*) For an object located at infinity, show that the two-lens system forms an image that is a distance $\frac{5}{4}f_1$ behind the first lens. (*c*) A single 250-mm-focal-length thin lens would have to be mounted about 250 mm from a camera's film in order to produce an image of a distant object at d_o with magnification $-(250\,\text{mm})/d_o$. To produce an image of this object with the same magnification using the two-lens system, what value of f_1 should be used and how far in front of the film should the first lens be placed? How much smaller is the "focusing length" (i.e., first lens-to-final image distance) of this two-lens system in comparison with the 250-mm "focusing length" of the equivalent single lens?

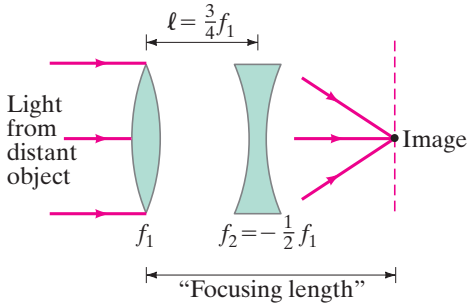


FIGURE 33–54 Problem 113.

ANSWERS TO EXERCISES

A: Closer to it.
B: (*b*) and (*d*) are true.
C: Use Table 32–1: $f = 1/[(0.50)(0 - 1/18.4\,\text{cm})] = -37\,\text{cm}$; diverging.
D: 33 dots/mm.
E: $P = -4.0\,\text{D}$.
F: 48 cm.



The beautiful colors from the surface of this soap bubble can be nicely explained by the wave theory of light. A soap bubble is a very thin spherical film filled with air. Light reflected from the outer and inner surfaces of this thin film of soapy water interferes constructively to produce the bright colors. Which color we see at any point depends on the thickness of the soapy water film at that point and also on the viewing angle. Near the top of the bubble, we see a small black area surrounded by a silver or white area. The bubble's thickness is smallest at that black spot, perhaps only about 30 nm thick, and is fully transparent (we see the black background).

We cover fundamental aspects of the wave nature of light, including two-slit interference, interference in thin films, polarization and its applications, and lumens.

CHAPTER 34

The Wave Nature of Light: Interference and Polarization

CHAPTER-OPENING QUESTION—Guess now!

When a thin layer of oil lies on top of water or wet pavement, you can often see swirls of color. We also see swirls of color on the soap bubble shown above. What causes these colors?

(a) Additives in the oil or soap reflect various colors.
(b) Chemicals in the oil or soap absorb various colors.
(c) Dispersion due to differences in index of refraction in the oil or soap.
(d) The interactions of the light with a thin boundary layer where the oil (or soap) and the water have mixed irregularly.
(e) Light waves reflected from the top and bottom surfaces of the thin oil or soap film can add up constructively for particular wavelengths.

Light carries energy. You can feel it on your skin on a hot day. Also you may have noted that focusing the Sun's rays with a magnifying glass on a piece of paper can burn a hole in it. But how does light travel, and in what form is this energy carried? In our discussion of waves in Chapter 15, we noted that energy can be carried from place to place in basically two ways: by particles or by waves. In the first case, material objects or particles can carry energy, such as an avalanche of rocks or rushing water. In the second case, water waves and sound waves, for example, can carry energy over long distances even though the oscillating particles of the medium do not travel these distances.

CONTENTS

34 – 1 Waves vs. Particles; Huygens' Principle and Diffraction
34 – 2 Huygens' Principle and the Law of Refraction; Mirages
34 – 3 Interference—Young's Double-Slit Experiment
*34 – 4 Intensity in the Double-Slit Interference Pattern
34 – 5 Interference in Thin Films
34 – 6 Michelson Interferometer
34 – 7 Polarization
*34 – 8 Liquid Crystal Displays (LCD)
*34 – 9 Scattering of Light by the Atmosphere
34–10 Brightness: Lumens and Luminous Intensity
*34–11 Efficiency of Lightbulbs

What can we say about the nature of light: does light travel as a stream of particles away from its source, or does light travel in the form of waves that spread outward from the source?

Historically, this question has turned out to be a difficult one. For one thing, light does not reveal itself in any obvious way as being made up of tiny particles; nor do we see tiny light waves passing by as we do water waves. The evidence seemed to favor first one side and then the other until about 1830, when most physicists had accepted the wave theory. By the end of the nineteenth century, light was considered to be an *electromagnetic wave* (Chapter 31). In the early twentieth century, light was shown to have a particle nature as well, as we shall discuss in Chapter 37. We now speak of the wave–particle duality of light. The wave theory of light remains valid and has proved very successful. In this Chapter we investigate the evidence for the wave theory and how it has been used to explain a wide range of phenomena.

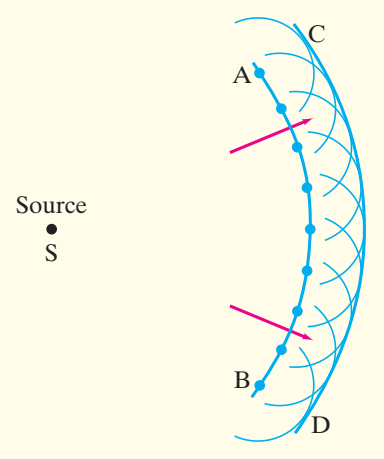
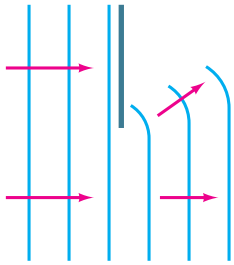
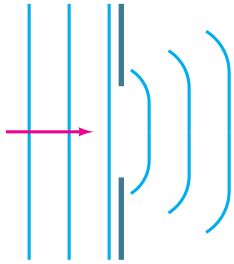


FIGURE 34–1 Huygens' principle, used to determine wave front CD when wave front AB is given.

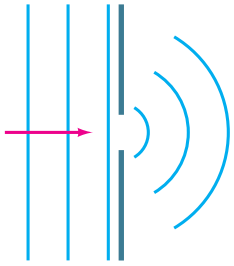
FIGURE 34–2 Huygens' principle is consistent with diffraction (a) around the edge of an obstacle, (b) through a large hole, (c) through a small hole whose size is on the order of the wavelength of the wave.



(a)



(b)



(c)

34–1 Waves vs. Particles; Huygens' Principle and Diffraction

The Dutch scientist Christian Huygens (1629–1695), a contemporary of Newton, proposed a wave theory of light that had much merit. Still useful today is a technique Huygens developed for predicting the future position of a wave front when an earlier position is known. By a **wave front**, we mean all the points along a two- or three-dimensional wave that form a wave *crest*—what we often just call a "wave" as seen on the ocean. Wave fronts are perpendicular to rays as discussed in Chapter 15 (Fig. 15–20). **Huygens' principle** can be stated as follows: *Every point on a wave front can be considered as a source of tiny wavelets that spread out in the forward direction at the speed of the wave itself. The new wave front is the envelope of all the wavelets*—that is, *the tangent to all of them*.

As an example of the use of Huygens' principle, consider the wave front AB in Fig. 34–1, which is traveling away from a source S. We assume the medium is *isotropic*—that is, the speed v of the waves is the same in all directions. To find the wave front a short time t after it is at AB, tiny circles are drawn at points along AB with radius $r = vt$. The centers of these tiny circles are shown as blue dots on the original wave front AB, and the circles represent Huygens' (imaginary) wavelets. The tangent to all these wavelets, the curved line CD, is the new position of the wave front after a time t.

Huygens' principle is particularly useful for analyzing what happens when waves run into an obstacle and the wave fronts are partially interrupted. Huygens' principle predicts that waves bend in behind an obstacle, as shown in Fig. 34–2. This is just what water waves do, as we saw in Chapter 15 (Figs. 15–31 and 15–32). The bending of waves behind obstacles into the "shadow region" is known as **diffraction**. Since diffraction occurs for waves, but not for particles, it can serve as one means for distinguishing the nature of light.

Note, as shown in Fig. 34–2, that diffraction is most prominent when the size of the opening is on the order of the wavelength of the wave. If the opening is much larger than the wavelength, diffraction may go unnoticed.

Does light exhibit diffraction? In the mid-seventeenth century, the Jesuit priest Francesco Grimaldi (1618–1663) had observed that when sunlight entered a darkened room through a tiny hole in a screen, the spot on the opposite wall was larger than would be expected from geometric rays. He also observed that the border of the image was not clear but was surrounded by colored fringes. Grimaldi attributed this to the diffraction of light.

The wave model of light nicely accounts for diffraction, and we discuss diffraction in detail in the next Chapter. But the ray model (Chapter 32) cannot account for diffraction, and it is important to be aware of such limitations to the ray model. Geometric optics using rays is successful in a wide range of situations only because normal openings and obstacles are much larger than the wavelength of the light, and so relatively little diffraction or bending occurs.

34–2 Huygens' Principle and the Law of Refraction; Mirages

The laws of reflection and refraction were well known in Newton's time. The law of reflection could not distinguish between the two theories we just discussed: waves versus particles. For when waves reflect from an obstacle, the angle of incidence equals the angle of reflection (Fig. 15–21). The same is true of particles—think of a tennis ball without spin striking a flat surface.

The law of refraction is another matter. Consider a ray of light entering a medium where it is bent toward the normal, as when traveling from air into water. As shown in Fig. 34–3, this bending can be constructed using Huygens' principle if we assume the speed of light is less in the second medium $(v_2 < v_1)$. In time t, point B on wave front AB (perpendicular to the incoming ray) travels a distance $v_1 t$ to reach point D. Point A on the wave front, traveling in the second medium, goes a distance $v_2 t$ to reach point C, and $v_2 t < v_1 t$. Huygens' principle is applied to points A and B to obtain the curved wavelets shown at C and D. The wave front is tangent to these two wavelets, so the new wave front is the line CD. Hence the rays, which are perpendicular to the wave fronts, bend toward the normal if $v_2 < v_1$, as drawn.

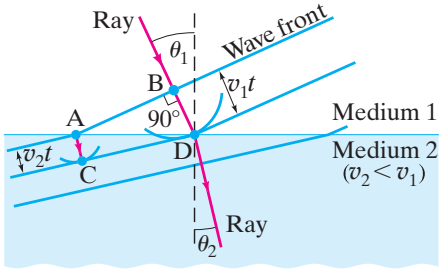


FIGURE 34–3 Refraction explained, using Huygens' principle. Wave fronts are perpendicular to the rays.

Newton favored a particle theory of light which predicted the opposite result, that the speed of light would be greater in the second medium $(v_2 > v_1$ if $\theta_2 < \theta_1)$. Thus the wave theory predicts that the speed of light in water, for example, is less than in air; and Newton's particle theory predicts the reverse. An experiment to actually measure the speed of light in water was performed in 1850 by the French physicist Jean Foucault, and it confirmed the wave-theory prediction. By then, however, the wave theory was already fully accepted because of interference phenomena, as we shall see in the next Section.

Snell's law of refraction follows directly from Huygens' principle, given that the speed of light v in any medium is related to the speed in a vacuum, c, and the index of refraction, n, by Eq. 32–4: that is, $v = c/n$. From the Huygens' construction of Fig. 34–3, angle ADC is equal[†] to θ_2 and angle BAD is equal to θ_1. Then for the two triangles that have the common side AD, we have

$$\sin\theta_1 = \frac{v_1 t}{\text{AD}}, \qquad \sin\theta_2 = \frac{v_2 t}{\text{AD}}.$$

We divide these two equations and obtain

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2}.$$

Then, by Eq. 32–4, $v_1 = c/n_1$ and $v_2 = c/n_2$, so we have

$$n_1 \sin\theta_1 = n_2 \sin\theta_2,$$

which is Snell's law of refraction, Eq. 32–5. (The law of reflection can be derived from Huygens' principle in a similar way: Problem 1 at the end of this Chapter.)

When a light wave travels from one medium to another, its frequency is not affected, but its wavelength does change. This can be seen from Fig. 34–3, where each of the blue lines representing a wave front corresponds to a crest (peak) of the wave. Then

$$\frac{\lambda_2}{\lambda_1} = \frac{v_2 t}{v_1 t} = \frac{v_2}{v_1} = \frac{n_1}{n_2},$$

where, in the last step, we used Eq. 32–4, $v = c/n$. If medium 1 is a vacuum (or air), we have $n_1 = 1$ and $v_1 = c$, and we call λ_1 simply λ, then the wavelength in another medium of index of refraction $n\ (= n_2)$ will be

$$\lambda_n = \frac{\lambda}{n}. \qquad \textbf{(34–1)}$$

This result is consistent with the frequency f being unchanged no matter what medium the wave is traveling in, since $c = f\lambda$.

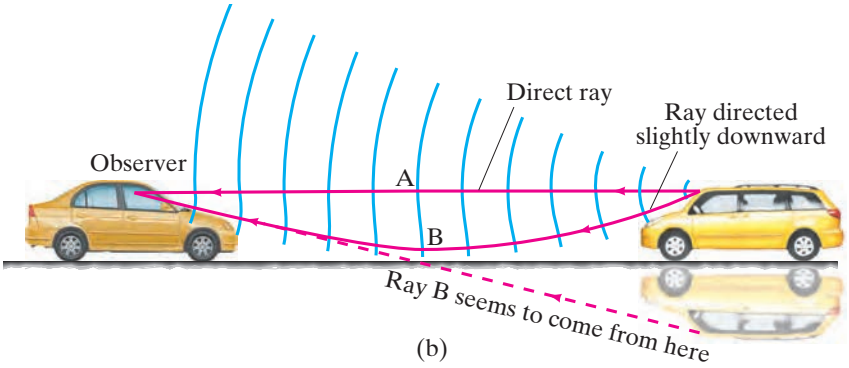
CAUTION
Frequency is fixed, wavelength can change

EXERCISE A A light beam in air, with wavelength = 500 nm, frequency = 6.0×10^{14} Hz, and speed = 3.0×10^8 m/s, goes into glass which has an index of refraction = 1.5. What are the wavelength, frequency, and speed of the light in the glass?

[†]Right sides of the 2 angles are perpendicular, the left sides are perpendicular: see Appendix page A–2, Fig. A–2.



(a)



(b)

FIGURE 34–4 (a) A highway mirage (there is not really water on the road). (b) Drawing (greatly exaggerated) showing wave fronts and rays to explain highway mirages. Note how sections of the wave fronts near the ground are slightly farther apart, which means they are moving faster.

PHYSICS APPLIED
Highway mirages

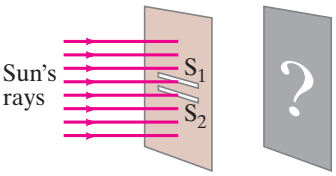
Wave fronts can be used to explain how **mirages** are produced by refraction of light. For example, on a hot day motorists sometimes see a mirage of water on the highway ahead of them, with distant vehicles seemingly reflected in it (Fig. 34–4a). On a hot day, there can be a layer of very hot air next to the roadway (made hot by the Sun beating down on the road). Hot air is less dense than cooler air, so the index of refraction is slightly lower in the hot air. In Fig. 34–4b, we see a diagram of light coming from one point on a distant car (on the right) heading left toward the observer. Wave fronts and two rays (perpendicular to the wave fronts) are shown. Ray A heads directly at the observer and follows a straight-line path, and represents the normal view of the distant car. Ray B is a ray initially directed slightly downward but, instead of hitting the road, it refracts slightly as it moves through layers of air of different index of refraction. The wave fronts, shown in blue in Fig. 34–4b, move slightly faster in the layers of (less dense) air nearer the ground. Thus ray B is bent as shown, and seems to the observer to be coming from below (dashed line) as if reflected off the road. Hence the mirage.

34–3 Interference—Young's Double-Slit Experiment

In 1801, the Englishman Thomas Young (1773–1829) obtained convincing evidence for the wave nature of light and was even able to measure wavelengths for visible light. Figure 34–5a shows a schematic diagram of Young's famous double-slit experiment. To have light from a single source, Young used the sunlight passing through a very narrow slit in a window covering. This beam of parallel rays falls on a screen containing two closely spaced slits, S_1 and S_2. (The slits and their separation are very narrow, not many times greater than the wavelength of the light.) If light consists of tiny particles, we would expect to see two bright lines on a screen placed behind the slits as in (b). But instead, a series of bright lines are seen as in (c). Young was able to explain this result as a **wave-interference** phenomenon.

To understand why, consider **plane waves**† of light of a single wavelength—called **monochromatic**, meaning "one color"—falling on the two slits as shown in Fig. 34–6. Because of diffraction, the waves leaving the two small slits spread out as shown. This is equivalent to the interference pattern produced when two rocks are thrown into a lake (Fig. 15–23), or when sound from two loudspeakers interferes (Fig. 16–15). Recall Section 15–8 on wave interference.

FIGURE 34–5 (a) Young's double-slit experiment. The slits and their separation need to be very thin. (b) If light consists of particles, we would expect to see two bright lines on the screen behind the slits. (c) In fact, many lines are observed.



(a) Viewing screen



(b) Viewing screen (particle theory prediction)



(c) Viewing screen (actual)

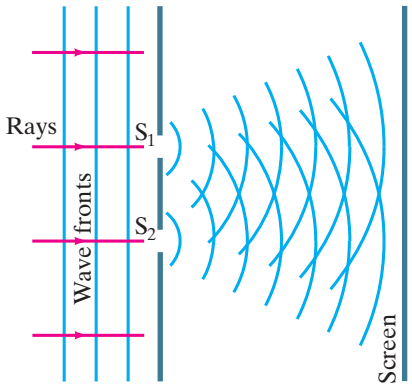


FIGURE 34–6 Plane waves (parallel flat wave fronts) fall on two slits. If light is a wave, light passing through one of two slits should interfere with light passing through the other slit.

†See Figs. 15–20 and 31–8, pages 443 and 904.

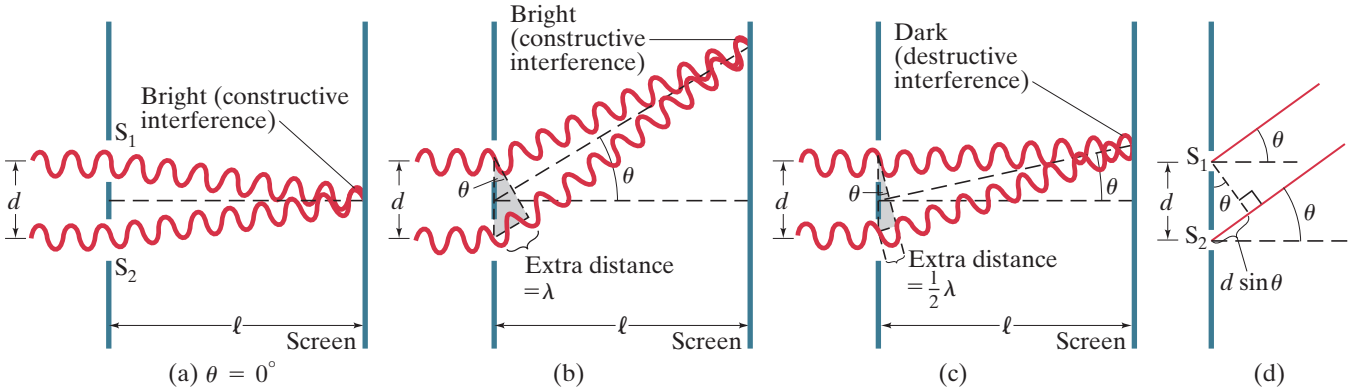


FIGURE 34–7 How the wave theory explains the pattern of lines seen on a viewing screen due to a double-slit experiment. (a) At the center of the screen, the waves from each slit travel the same distance and are in phase. [Assume $\ell \gg d$.] (b) At this angle θ, the lower wave travels an extra distance of one whole wavelength, and the waves are in phase; note from the shaded triangle that the path difference equals $d \sin\theta$. (c) For this angle θ, the lower wave travels an extra distance equal to one-half wavelength, so the two waves arrive at the screen fully out of phase. (d) A more detailed diagram showing the geometry for parts (b) and (c).

To see how an interference pattern is produced on the screen, we make use of Fig. 34–7. Waves of wavelength λ are shown entering the slits S_1 and S_2, which are a distance d apart. The waves continually spread out in all directions after passing through the slits (Fig. 34–6), but they are shown in Fig. 34–7a, b, and c only for three different angles θ. In Fig. 34–7a, the waves reaching the center of the screen are shown ($\theta = 0°$). Waves from the two slits travel the same distance, so they are always **in phase**: crests of one wave arrive at the same time as crests of the other wave. Hence the amplitudes of the two waves add to form a larger amplitude as shown in Fig. 34–8a. This is **constructive interference**, and there is a bright line at the center of the screen. Constructive interference also occurs when the paths of the two rays differ by one wavelength (or any whole number of wavelengths), as shown in Fig. 34–7b; also here there will be a bright line on the screen. But if one ray travels an extra distance of one-half wavelength (or $\frac{3}{2}\lambda$, $\frac{5}{2}\lambda$, and so on), the two waves are always exactly **out of phase** (Section 15–8) when they reach the screen: the crests of one wave arrive at the same time as the troughs of the other wave, and so they add to produce zero amplitude (Fig. 34–8b). This is **destructive interference**, and the screen is dark, Fig. 34–7c. Thus, there will be a series of bright and dark lines (or **fringes**) on the viewing screen.

To determine exactly where the bright lines fall, first note that Fig. 34–7 is somewhat exaggerated; in real situations, the distance d between the slits is very small compared to the distance ℓ to the screen. The rays from each slit for each case will therefore be essentially parallel, and θ is the angle they make with the horizontal as shown in Fig. 34–7d. From the shaded right triangles shown in Figs. 34–7b and c, we can see that the extra distance traveled by the lower ray is $d \sin\theta$ (seen more clearly in Fig. 34–7d). Constructive interference will occur, and a bright fringe will appear on the screen, when the *path difference*, $d \sin\theta$, equals a whole number of wavelengths:

$$d \sin\theta = m\lambda, \qquad m = 0, 1, 2, \cdots. \qquad \begin{bmatrix}\text{constructive}\\ \text{interference}\\ \text{(bright)}\end{bmatrix} \qquad \textbf{(34–2a)}$$

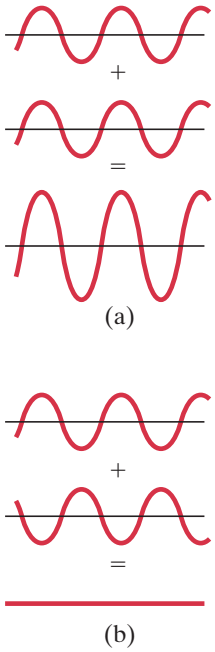
The value of m is called the **order** of the interference fringe. The first order ($m = 1$), for example, is the first fringe on each side of the central fringe (which is at $\theta = 0$, $m = 0$). Destructive interference occurs when the path difference $d \sin\theta$ is $\frac{1}{2}\lambda$, $\frac{3}{2}\lambda$, and so on:

$$d \sin\theta = \left(m + \tfrac{1}{2}\right)\lambda, \qquad m = 0, 1, 2, \cdots. \qquad \begin{bmatrix}\text{destructive}\\ \text{interference}\\ \text{(dark)}\end{bmatrix} \qquad \textbf{(34–2b)}$$

The bright fringes are peaks or maxima of light intensity, the dark fringes are minima.[†]

[†]The plural of minimun and maximum are minima and maxima, from Latin, and not minimums or maximums.

FIGURE 34–8 Two traveling waves are shown undergoing (a) constructive interference, (b) destructive interference. (See also Section 15–8.)



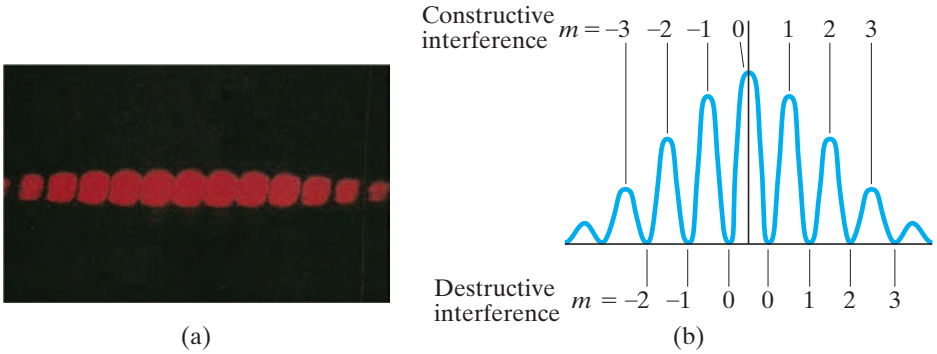


FIGURE 34–9 (a) Interference fringes produced by a double-slit experiment and detected by photographic film placed on the viewing screen. (b) Graph of the intensity of light in the interference pattern. Also shown are values of m for Eq. 34–2a (constructive interference) and Eq. 34–2b (destructive interference).

The intensity of the bright fringes in the double-slit pattern is greatest for the central fringe $(m = 0)$ and decreases for higher orders as shown in Fig. 34–9. How much the intensity decreases with increasing order depends on the width of the two slits.

CONCEPTUAL EXAMPLE 34–1 **Interference pattern lines.** (a) Will there be an infinite number of points on the viewing screen where constructive and destructive interference occur, or only a finite number of points? (b) Are neighboring points of constructive interference uniformly spaced, or is the spacing between neighboring points of constructive interference not uniform?

RESPONSE (a) When you look at Eqs. 34–2a and b you might be tempted to say, given the statement $m = 0, \pm 1, 2, \cdots$. beside the equations, that there are an infinite number of points of constructive and destructive interference. However, recall that $\sin\theta$ cannot exceed 1. Thus, there is an upper limit to the values of m that can be used in these equations. For Eq. 34–2a, the maximum value of m is the integer closest in value but smaller than d/λ. So there are a *finite* number of points of constructive and destructive interference no matter how large the screen. (b) The spacing between neighboring points of constructive or destructive interference is not uniform: The spacing gets larger as θ gets larger, because as θ increases, a doubling of $\sin\theta$ more than doubles θ (see trig Table, page A–5 in Appendix A). For small values of θ the spacing is nearly uniform as we will see in Example 34–2.

EXAMPLE 34–2 **Line spacing for double-slit interference.** A screen containing two slits 0.100 mm apart is 1.20 m from the viewing screen. Light of wavelength $\lambda = 500$ nm falls on the slits from a distant source. Approximately how far apart will adjacent bright interference fringes be on the screen?

CAUTION
Use the approximation $\theta \approx \tan\theta$ or $\theta \approx \sin\theta$ only if θ is small and in radians

APPROACH The angular position of bright (constructive interference) fringes is found using Eq. 34–2a. The distance between the first two fringes (say) can be found using right triangles as shown in Fig. 34–10.

SOLUTION Given $d = 0.100\,\text{mm} = 1.00 \times 10^{-4}\,\text{m}$, $\lambda = 500 \times 10^{-9}\,\text{m}$, and $\ell = 1.20\,\text{m}$, the first-order fringe $(m = 1)$ occurs at an angle θ given by Eq. 34–2a:

$$\sin\theta_1 = \frac{m\lambda}{d} = \frac{(1)(500 \times 10^{-9}\,\text{m})}{1.00 \times 10^{-4}\,\text{m}} = 5.00 \times 10^{-3}.$$

This is a very small angle, so we can take $\sin\theta \approx \theta$, with θ in radians. The first-order fringe will occur a distance x_1 above the center of the screen (see Fig. 34–10), given by $x_1/\ell = \tan\theta_1 \approx \theta_1$, so

$$x_1 \approx \ell\theta_1 = (1.20\,\text{m})(5.00 \times 10^{-3}) = 6.00\,\text{mm}.$$

The second-order fringe $(m = 2)$ will occur at

$$x_2 \approx \ell\theta_2 = \ell\frac{2\lambda}{d} = 12.0\,\text{mm}$$

above the center, and so on. Thus the lower-order fringes are 6.00 mm apart.

NOTE The spacing between fringes is essentially uniform until the approximation $\sin\theta \approx \theta$ is no longer valid.

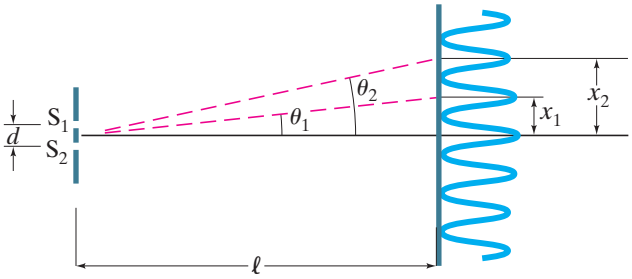


FIGURE 34–10 Examples 34–2 and 34–3. For small angles θ (give θ in radians), the interference fringes occur at distance $x = \theta\ell$ above the center fringe $(m = 0)$; θ_1 and x_1 are for the first-order fringe $(m = 1)$; θ_2 and x_2 are for $m = 2$.

CONCEPTUAL EXAMPLE 34–3 **Changing the wavelength.** (*a*) What happens to the interference pattern shown in Fig. 34–10, Example 34–2, if the incident light (500 nm) is replaced by light of wavelength 700 nm? (*b*) What happens instead if the wavelength stays at 500 nm but the slits are moved farther apart?

RESPONSE (*a*) When λ increases in Eq. 34–2a but d stays the same, the angle θ for bright fringes increases and the interference pattern spreads out. (*b*) Increasing the slit spacing d reduces θ for each order, so the lines are closer together.

From Eqs. 34–2 we can see that, except for the zeroth-order fringe at the center, the position of the fringes depends on wavelength. When white light falls on the two slits, as Young found in his experiments, the central fringe is white, but the first (and higher) order fringes contain a spectrum of colors like a rainbow. Using Eq. 34–2a, we can see that θ is smallest for violet light and largest for red (Fig. 34–11). By measuring the position of these fringes, Young was the first to determine the wavelengths of visible light. In doing so, he showed that what distinguishes different colors physically is their wavelength (or frequency), an idea put forward earlier by Grimaldi in 1665.

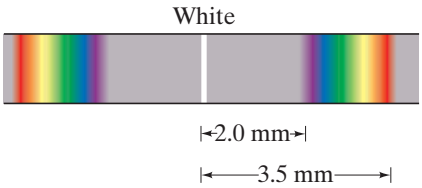


FIGURE 34–11 First-order fringes for a double slit are a full spectrum, like a rainbow. Also Example 34–4.

EXAMPLE 34–4 **Wavelengths from double-slit interference.** White light passes through two slits 0.50 mm apart, and an interference pattern is observed on a screen 2.5 m away. The first-order fringe resembles a rainbow with violet and red light at opposite ends. The violet light is about 2.0 mm and the red 3.5 mm from the center of the central white fringe (Fig. 34–11). Estimate the wavelengths for the violet and red light.

APPROACH We find the angles for violet and red light from the distances given and the diagram of Fig. 34–10. Then we use Eq. 34–2a to obtain the wavelengths. Because 3.5 mm is much less than 2.5 m, we can use the small-angle approximation.

SOLUTION We use Eq. 34–2a $(d \sin\theta = m\lambda)$ with $m = 1$, $d = 5.0 \times 10^{-4}$ m, and $\sin\theta \approx \tan\theta \approx \theta$. Also $\theta \approx x/\ell$ (Fig. 34–10), so for violet light, $x = 2.0$ mm, and

$$\lambda = \frac{d \sin\theta}{m} \approx \frac{d\theta}{m} \approx \frac{d}{m}\frac{x}{\ell} = \left(\frac{5.0 \times 10^{-4}\,\text{m}}{1}\right)\left(\frac{2.0 \times 10^{-3}\,\text{m}}{2.5\,\text{m}}\right) = 4.0 \times 10^{-7}\,\text{m},$$

or 400 nm. For red light, $x = 3.5$ mm, so

$$\lambda \approx \frac{d}{m}\frac{x}{\ell} = \left(\frac{5.0 \times 10^{-4}\,\text{m}}{1}\right)\left(\frac{3.5 \times 10^{-3}\,\text{m}}{2.5\,\text{m}}\right) = 7.0 \times 10^{-7}\,\text{m} = 700\,\text{nm}.$$

These values correspond to the visible spectrum as discussed in Section 32–8.

EXERCISE B For the setup in Example 34–4, how far from the central white fringe is the first-order fringe for green light $\lambda = 500$ nm?

Coherence

The two slits in Figs. 34–6 and 34–7 act as if they were two sources of radiation. They are called **coherent sources** because the waves leaving them have the same wavelength and frequency, and bear the same phase relationship to each other at all times. This happens because the waves come from a single source to the left of the two slits. An interference pattern is observed only when the sources are coherent.

If two tiny lightbulbs replaced the two slits, an interference pattern would not be seen. The light emitted by one lightbulb would have a random phase with respect to the second bulb, and the screen would be more or less uniformly illuminated. Two such sources, whose output waves have phases that bear no fixed relationship to each other over time, are called **incoherent sources**.

*34–4 Intensity in the Double-Slit Interference Pattern

We saw in Section 34–3 that the interference pattern produced by the coherent light from two slits, S_1 and S_2 (Figs. 34–7 and 34–9), produces a series of bright and dark fringes. If the two monochromatic waves of wavelength λ are in phase at the slits, the maxima (brightest points) occur at angles θ given by (Eqs. 34–2)

$$d \sin\theta = m\lambda,$$

and the minima (darkest points) when

$$d \sin\theta = \left(m + \tfrac{1}{2}\right)\lambda,$$

where m is an integer $(m = 0, \pm 1, \pm 2, \cdots)$.

We now determine the intensity of the light at all points in the pattern, assuming that if either slit were covered, the light passing through the other would diffract sufficiently to illuminate a large portion of the screen uniformly. The intensity I of the light at any point is proportional to the square of its wave amplitude (Section 15–3). Treating light as an electromagnetic wave, I is proportional to the square of the electric field E (or to the magnetic field B, Section 31–8): $I \propto E^2$. The electric field $\vec{\mathbf{E}}$ at any point P (see Fig. 34–12) will be the sum of the electric field vectors of the waves coming from each of the two slits, $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$. Since $\vec{\mathbf{E}}_1$ and $\vec{\mathbf{E}}_2$ are essentially parallel (on a screen far away compared to the slit separation), the magnitude of the electric field at angle θ (that is, at point P) will be

$$E_\theta = E_1 + E_2.$$

Both E_1 and E_2 vary sinusoidally with frequency $f = c/\lambda$, but they differ in phase, depending on their different travel distances from the slits. The electric field at P can then be written for the light from each of the two slits, using $\omega = 2\pi f$, as

$$\begin{aligned} E_1 &= E_{10} \sin\omega t \\ E_2 &= E_{20} \sin(\omega t + \delta) \end{aligned} \qquad \textbf{(34–3)}$$

where E_{10} and E_{20} are their respective amplitudes and δ is the phase difference. The value of δ depends on the angle θ, so let us now determine δ as a function of θ.

At the center of the screen (point 0), $\delta = 0$. If the difference in path length from P to S_1 and S_2 is $d \sin\theta = \lambda/2$, the two waves are exactly out of phase so $\delta = \pi$ (or 180°). If $d \sin\theta = \lambda$, the two waves differ in phase by $\delta = 2\pi$. In general, then, δ is related to θ by

$$\frac{\delta}{2\pi} = \frac{d \sin\theta}{\lambda}$$

or (note that if $d \sin\theta = \lambda$, then $\delta = 2\pi$ as we just observed)

$$\delta = \frac{2\pi}{\lambda} d \sin\theta. \qquad \textbf{(34–4)}$$

To determine $E_\theta = E_1 + E_2$, we add the two scalars E_1 and E_2 which are sine functions differing by the phase δ. One way to determine the sum of E_1 and E_2 is to use a **phasor diagram**. (We used this technique before, in Chapter 30.) As shown in Fig. 34–13, we draw an arrow of length E_{10} to represent the amplitude of E_1 (Eq. 34–3); and an arrow of length E_{20} for the amplitude of E_2, which we draw to make a fixed angle δ with E_{10}. When the diagram rotates at angular frequency ω about the origin, the projections of E_{10} and E_{20} on the vertical axis represent E_1 and E_2 as a function of time (Eqs. 34–3). We let $E_{\theta 0}$ be the "vector" sum† of E_{10} and E_{20}; $E_{\theta 0}$ is the amplitude of the sum $E_\theta = E_1 + E_2$, and the projection of $E_{\theta 0}$ on the vertical axis is E_θ, Fig. 34–13.

†We are not adding the actual electric field vectors; instead we are using the "phasor" technique to add the amplitudes, taking into account the phase difference of the two waves.

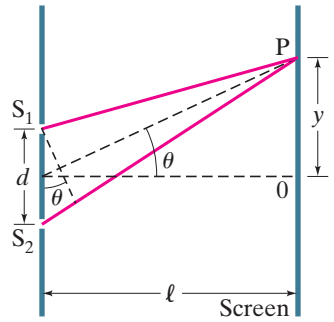


FIGURE 34–12 Determining the intensity in a double-slit interference pattern. Not to scale: in fact $\ell \gg d$, and the two rays become essentially parallel.

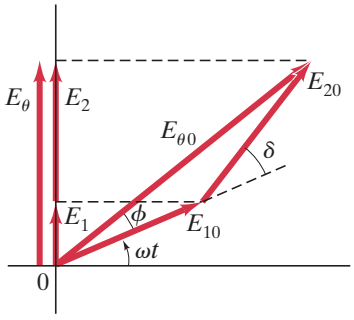


FIGURE 34–13 Phasor diagram for double-slit interference pattern.

If the two slits provide equal illumination, so that $E_{10} = E_{20} = E_0$, then from *symmetry* in Fig. 34–13, the angle $\phi = \delta/2$, and we can write

$$E_\theta = E_{\theta 0} \sin\left(\omega t + \frac{\delta}{2}\right). \qquad \textbf{(34–5a)}$$

From Fig. 34–13 we can also see that

$$E_{\theta 0} = 2E_0 \cos\phi = 2E_0 \cos\frac{\delta}{2}. \qquad \textbf{(34–5b)}$$

Combining Eqs. 34–5a and b, we obtain

$$E_\theta = 2E_0 \cos\frac{\delta}{2} \sin\left(\omega t + \frac{\delta}{2}\right), \qquad \textbf{(34–5c)}$$

where δ is given by Eq. 34–4.

We are not really interested in E_θ as a function of time, since for visible light the frequency (10^{14} to 10^{15} Hz) is much too high to be noticeable. We are interested in the *average* intensity, which is proportional to the amplitude squared, $E_{\theta 0}^2$. We drop the word "average," and let $I_\theta (I_\theta \propto E_\theta^2)$ be the intensity at any point P at an angle θ to the horizontal (Fig. 34–12). We let I_0 be the intensity at point O, the center of the screen, where $\theta = \delta = 0$, so $I_0 \propto (E_{10} + E_{20})^2 = (2E_0)^2$. Then the ratio I_θ/I_0 is equal to the ratio of the squares of the electric-field amplitudes at points P and O (Fig. 34–12), so

$$\frac{I_\theta}{I_0} = \frac{E_{\theta 0}^2}{(2E_0)^2} = \cos^2\frac{\delta}{2}$$

where we used Eq. 34–5b. Thus the intensity I_θ at any point is related to the maximum intensity at the center of the screen by

$$I_\theta = I_0 \cos^2\frac{\delta}{2} = I_0 \cos^2\left(\frac{\pi d \sin\theta}{\lambda}\right) \qquad \textbf{(34–6)}$$

PROBLEM SOLVING
Light intensity in double-slit interference pattern

where δ was given by Eq. 34–4. This is the relation we sought.

From Eq. 34–6 we see that maxima occur where $\cos(\delta/2) = \pm 1$, which corresponds to $\delta = 0, \pm 2\pi, \pm 4\pi, \cdots$. From Eq. 34–4, δ has these values when

$$d \sin\theta = m\lambda, \qquad m = 0, \pm 1, \pm 2, \cdots.$$

Minima occur where $\cos(\delta/2) = 0$, so $\delta = \pm\pi, \pm 3\pi, \pm 5\pi, \cdots$, which corresponds to

$$d \sin\theta = \left(m + \tfrac{1}{2}\right)\lambda, \qquad m = 0, \pm 1, \pm 2, \cdots.$$

These are the same results we obtained in Section 34–3. But now we know not only the position of maxima and minima, but from Eq. 34–6 we can determine the intensity at all points on the viewing screen.

In the usual situation the distance ℓ to the screen from the slits is large compared to the slit separation d. Then, with $\ell \gg d$, if we consider only points P (Fig. 34–12) whose distance y from the center (point O) is small compared to ℓ ($y \ll \ell$ in Fig. 34–12) then

$$\sin\theta \approx \frac{y}{\ell}.$$

From this it follows (see Eq. 34–4) that

$$\delta = \frac{2\pi}{\lambda}\frac{d}{\ell}y.$$

Equation 34–6 then becomes approximately

$$I_\theta = I_0\left[\cos\left(\frac{\pi d}{\lambda \ell}y\right)\right]^2. \qquad [y \ll \ell,\ d \ll \ell] \qquad \textbf{(34–7)}$$

EXERCISE C What are the values for the intensity I_θ when (*a*) $y = 0$, (*b*) $y = \lambda\ell/4d$, and (*c*) $y = \lambda\ell/2d$?

The intensity I_θ as a function of the phase difference δ is plotted in Fig. 34–14. In the approximation of Eq. 34–7, the horizontal axis could as well be y, the position on the screen.

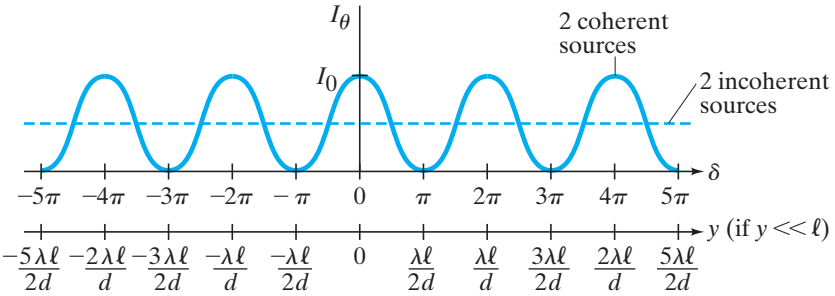


FIGURE 34–14 Intensity I as a function of phase difference δ and position on screen y (assuming $y \ll \ell$).

The intensity pattern expressed in Eqs. 34–6 and 34–7, and plotted in Fig. 34–14, shows a series of maxima of equal height, and is based on the assumption that each slit (alone) would illuminate the screen uniformly. This is never quite true, as we shall see when we discuss diffraction in the next Chapter. We will see that the center maximum is strongest and each succeeding maximum to each side is a little less strong. See also Fig. 34–9.

EXAMPLE 34–5 **Antenna intensity.** Two radio antennas are located close to each other as shown in Fig. 34–15, separated by a distance d. The antennas radiate in phase with each other, emitting waves of intensity I_0 at wavelength λ. (*a*) Calculate the net intensity as a function of θ for points very far from the antennas. (*b*) For $d = \lambda$, determine I and find in which directions I is a maximum and a minimum. (*c*) Repeat part (*b*) when $d = \lambda/2$.

APPROACH This setup is similar to Young's double-slit experiment.

SOLUTION (*a*) Points of constructive and destructive interference are still given by Eqs. 34–2a and b, and the net intensity as a function of θ is given by Eq. 34–6.

(*b*) We let $d = \lambda$ in Eq. 34–6, and find for the intensity,

$$I = I_0 \cos^2(\pi \sin\theta).$$

I is a maximum, equal to I_0, when $\sin\theta = 0,\ 1,$ or -1, meaning $\theta = 0, 90°, 180°,$ and $270°$. I is zero when $\sin\theta = \frac{1}{2}$ and $-\frac{1}{2}$, for which $\theta = 30°, 150°, 210°,$ and $330°$.

(*c*) For $d = \lambda/2$, $I = I_0 \cos^2(\pi \sin\theta/2)$ and is maximized for $\theta = 0$ and $180°$, and minimized for $90°$ and $270°$.

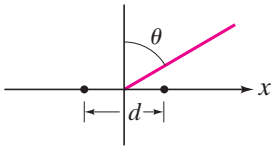


FIGURE 34–15 Example 34–5. The two dots represent the antennas.

34–5 Interference in Thin Films

Interference of light gives rise to many everyday phenomena such as the bright colors reflected from soap bubbles and from thin oil or gasoline films on water, Fig. 34–16. In these and other cases, the colors are a result of constructive interference between light reflected from the two surfaces of the thin film.



FIGURE 34–16 Thin-film interference patterns seen in (a) a soap bubble, (b) a thin film of soapy water, and (c) a thin layer of oil on wet pavement.

The effect is observed only if the thickness of the film is on the order of a few wavelengths of the light. If the film thickness is much greater, the effect gets washed out.

To see how this **thin-film interference** happens, consider a smooth surface of water on top of which is a thin uniform layer of another substance, say an oil whose index of refraction is less than that of water (we'll see why we assume this shortly); see Fig. 34–17. Assume for now that the incident light is of a single wavelength, and strikes the surface close to perpendicular. Part of the incident light is reflected at A on the top surface, and part of the light transmitted is reflected at B on the lower surface. The part reflected at the lower surface must travel the extra distance ABC. If this *path difference* ABC equals one or a whole number of wavelengths in the film (λ_n), the two waves will reach the eye in phase and interfere constructively. Hence the region AC on the surface film will appear bright. But if ABC equals $\frac{1}{2}\lambda_n, \frac{3}{2}\lambda_n$, and so on, the two waves will be exactly out of phase and destructive interference occurs: the area AC on the film will show no reflection—it will be dark (transparent to the dark material below). The wavelength λ_n is *the wavelength in the film*: $\lambda_n = \lambda/n$, where n is the index of refraction in the film and λ is the wavelength in vacuum. See Eq. 34–1.

When white light falls on such a film, the path difference ABC will equal λ_n (or $m\lambda_n$, with m = an integer) for only one wavelength at a given viewing angle. The color corresponding to λ (λ in air) will be seen as very bright. For light viewed at a slightly different angle, the path difference ABC will be longer or shorter and a different color will undergo constructive interference. Thus, for an extended (nonpoint) source emitting white light, a series of bright colors will be seen next to one another (Fig. 34–16). Variations in thickness of the film will also alter the path difference ABC and therefore affect the color of light that is most strongly reflected.

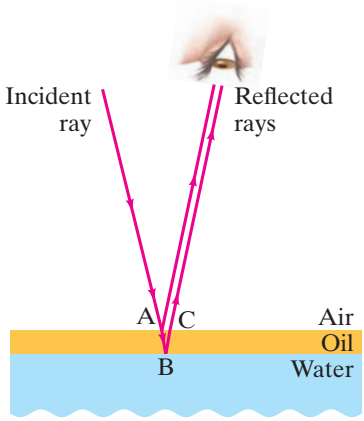


FIGURE 34–17 Light reflected from the upper and lower surfaces of a thin film of oil lying on water. We assume the light rays are (nearly) perpendicular to the oil surface, but are drawn at an angle so we can distinguish them.

EXERCISE D Return to the Chapter-Opening Question, page 995, and answer it again now. Try to explain why you may have answered differently the first time.

When a curved glass surface is placed in contact with a flat glass surface, Fig. 34–18, a series of concentric rings is seen when illuminated from above by either white light (as shown) or by monochromatic light. These are called **Newton's rings**[†] and they are due to interference between waves reflected by the top and bottom surfaces of the very thin *air gap* between the two pieces of glass. Because this gap (which is equivalent to a thin film) increases in width from the central contact point out to the edges, the extra path length for the lower ray (equal to BCD) varies. Where it equals $0, \frac{1}{2}\lambda, \lambda, \frac{3}{2}\lambda, 2\lambda$, and so on, it corresponds to constructive or destructive interference; and this gives rise to the series of bright colored circles seen in Fig. 34–18b. The color you see at a given radius corresponds to constructive interference; at that radius, other colors partially or fully destructively interfere. (If monochromatic light is used, the rings are alternately bright and dark.)

The point of contact of the two glass surfaces (A in Fig. 34–18a) is not bright in Fig. 34–18b. Since the path difference is zero here, our previous analysis would suggest that the waves reflected from each surface are in phase—so this central area ought to be bright. But it is dark, which tells us the two waves must be completely out of phase.

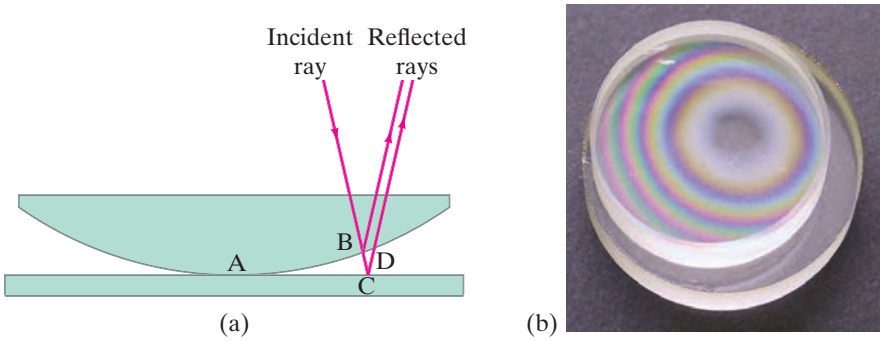


FIGURE 34–18 Newton's rings. (a) Light rays reflected from upper and lower surfaces of the thin air gap can interfere. (b) Photograph of interference patterns using white light.

[†]Although Newton gave an elaborate description of them, they had been first observed and described by his contemporary, Robert Hooke.

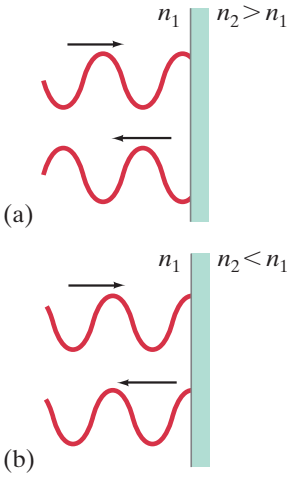
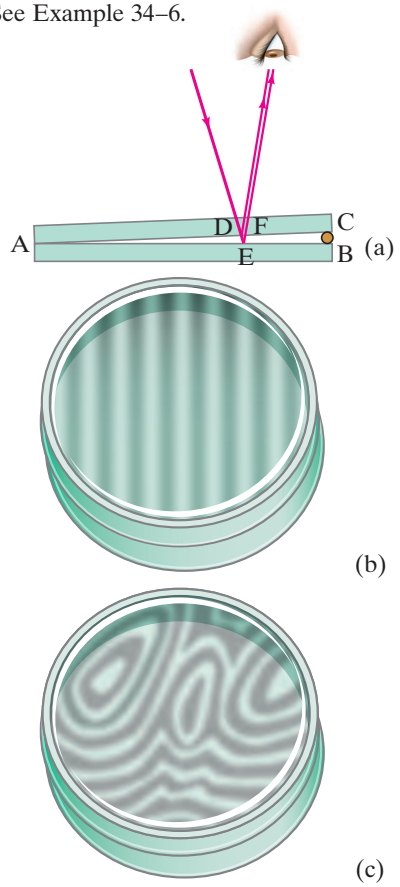


FIGURE 34–19 (a) Reflected ray changes phase by 180° or $\frac{1}{2}$ cycle if $n_2 > n_1$, but (b) does not if $n_2 < n_1$.

FIGURE 34–20 (a) Light rays reflected from the upper and lower surfaces of a thin wedge of air (between two glass plates) interfere to produce bright and dark bands. (b) Pattern observed when glass plates are optically flat; (c) pattern when plates are not so flat. See Example 34–6.



PHYSICS APPLIED
Testing glass for flatness

Let us see why the pattern is dark at the center of Fig. 34–18b (repeated here on facing page). This can happen only if one of the waves, upon reflection, flips over—a crest becomes a trough—see Fig. 34–19. We say that the reflected wave has undergone a **phase shift** of 180°, or of half a wave cycle ($\frac{1}{2}\lambda$). Indeed, this and other experiments reveal that, at normal incidence,

a beam of light, reflected by a material with index of refraction greater than that of the material in which it is traveling, changes phase by 180° or $\frac{1}{2}$ cycle;

see Fig. 34–19. This phase shift acts just like a path difference of $\frac{1}{2}\lambda$. If the index of refraction of the reflecting material is less than that of the material in which the light is traveling, *no phase change occurs.*† Thus the wave reflected at the curved surface above the air gap in Fig. 34–18a undergoes no change in phase. But the wave reflected at the lower surface, where the beam in air strikes the glass, undergoes a $\frac{1}{2}$-cycle phase shift, equivalent to a $\frac{1}{2}\lambda$ path difference. Thus the two waves reflected near the point of contact A of the two glass surfaces (where the air gap approaches zero thickness) will be a half cycle (or 180°) out of phase, and a dark spot occurs, Fig. 34–18b.

Bright colored rings will occur when the path difference is $\frac{1}{2}\lambda, \frac{3}{2}\lambda$, and so on, because the phase shift at one surface effectively adds a path difference of $\frac{1}{2}\lambda$ $(= \frac{1}{2}\text{ cycle})$. (If monochromatic light is used, the bright Newton's rings will be separated by dark bands which occur when the path difference BCD in Fig. 34–18a is equal to an integral number of wavelengths.)

Returning for a moment to Fig. 34–17, the light reflecting at both interfaces, air–oil and oil–water, *each* underwent a phase change of 180° equivalent to a path difference of $\frac{1}{2}\lambda$, since we assumed $n_{\text{water}} > n_{\text{oil}} > n_{\text{air}}$. Because the two phase shifts were equal, they didn't affect our analysis.

EXAMPLE 34–6 **Thin film of air, wedge-shaped.** A very fine wire 7.35×10^{-3} mm in diameter is placed between two flat glass plates as in Fig. 34–20a. The two plates touch at the other end (left). Light whose wavelength in air is 600 nm falls (and is viewed) perpendicular to the plates and a series of bright and dark bands is seen, Fig. 34–20b. How many light and dark bands will there be in this case? Will the area next to the wire be bright or dark?

APPROACH We need to consider two effects: (1) path differences for rays reflecting from the two close surfaces (thin wedge of air between the two glass plates), and (2) the $\frac{1}{2}$-cycle phase shift at the lower surface (point E in Fig. 34–20a), where rays in air can enter glass (or be reflected). Because there is a phase shift only at the lower surface, there will be a dark band (no reflection) when the path difference is 0, λ, 2λ, 3λ, and so on. Assuming the light rays are perpendicular to the plates, the extra path length (DEF) equals $2t$, where t is the thickness of the air gap at any point.

SOLUTION Dark bands will occur where

$$2t = m\lambda, \qquad m = 0, 1, 2, \cdots.$$

Bright bands occur when $2t = (m + \frac{1}{2})\lambda$, where m is an integer. At the position of the wire, $t = 7.35 \times 10^{-6}$ m. At this point there will be $2t/\lambda = (2)(7.35 \times 10^{-6}\text{ m})/(6.00 \times 10^{-7}\text{ m}) = 24.5$ wavelengths. This is a "half integer," so the area next to the wire will be bright. There will be a total of 25 dark lines along the plates, corresponding to path lengths DEF of 0λ, 1λ, 2λ, $3\lambda, \cdots$, 24λ, including the one at the point of contact A $(m = 0)$. Between them, there will be 24 bright lines plus the one at the end, or 25.

NOTE The bright and dark bands will be straight only if the glass plates are extremely flat. If they are not, the pattern is uneven, as in Fig. 34–20c. Thus we see a very precise way of testing a glass surface for flatness. Spherical lens surfaces can be tested for precision by placing the lens on a flat glass surface and observing Newton's rings (Fig. 34–18b) for perfect circularity.

†This result corresponds to the reflection of a wave traveling along a cord when it reaches the end. As we saw in Fig. 15–18, if the end is tied down, the wave changes phase and the pulse flips over, but if the end is free, no phase shift occurs.

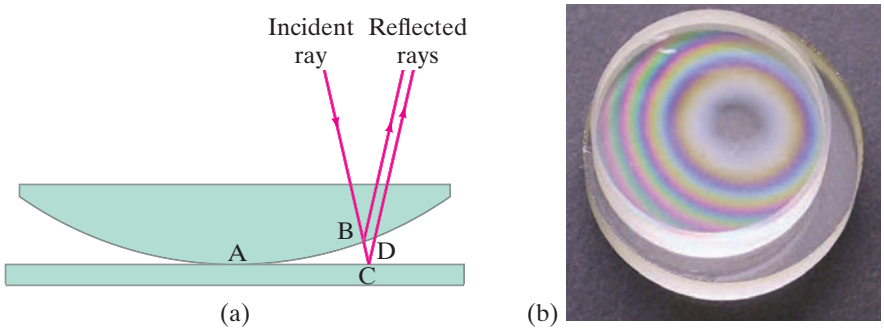


FIGURE 34–18 (Repeated.) Newton's rings. (a) Light rays reflected from upper and lower surfaces of the thin air gap can interfere. (b) Photograph of interference patterns using white light.

If the wedge between the two glass plates of Example 34–6 is filled with some transparent substance other than air—say, water—the pattern shifts because the wavelength of the light changes. In a material where the index of refraction is n, the wavelength is $\lambda_n = \lambda/n$, where λ is the wavelength in vacuum (Eq. 34–1). For instance, if the thin wedge of Example 34–6 were filled with water, then $\lambda_n = 600\,\text{nm}/1.33 = 450\,\text{nm}$; instead of 25 dark lines, there would be 33.

When white light (rather than monochromatic light) is incident on the thin wedge of air in Figs. 34–18a or 34–20a, a colorful series of fringes is seen because constructive interference occurs for different wavelengths in the reflected light at different thicknesses along the wedge.

A soap bubble (Fig. 34–16a and Chapter-Opening Photo) is a thin spherical shell (or film) with air inside. The variations in thickness of a soap bubble film give rise to bright colors reflected from the soap bubble. (There is air on both sides of the bubble film.) Similar variations in film thickness produce the bright colors seen reflecting from a thin layer of oil or gasoline on a puddle or lake (Fig. 34–16c). Which wavelengths appear brightest also depends on the viewing angle.

EXAMPLE 34–7 **Thickness of soap bubble skin.** A soap bubble appears green ($\lambda = 540\,\text{nm}$) at the point on its front surface nearest the viewer. What is the smallest thickness the soap bubble film could have? Assume $n = 1.35$.

APPROACH Assume the light is reflected perpendicularly from the point on a spherical surface nearest the viewer, Fig. 34–21. The light rays also reflect from the inner surface of the soap bubble film as shown. The path difference of these two reflected rays is $2t$, where t is the thickness of the soap film. Light reflected from the first (outer) surface undergoes a 180° phase change (index of refraction of soapy water is greater than that of air), whereas reflection at the second (inner) surface does not. To determine the thickness t for an interference maximum, we must use the wavelength of light in the soap ($n = 1.35$).

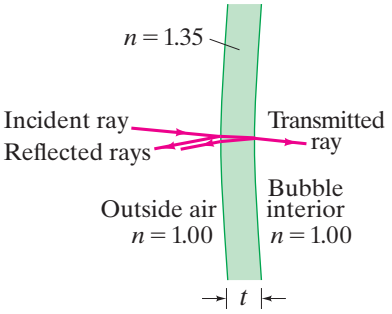
SOLUTION The 180° phase change at only one surface is equivalent to a $\frac{1}{2}\lambda$ path difference. Therefore, green light is bright when the minimum path difference equals $\frac{1}{2}\lambda_n$. Thus, $2t = \lambda_n/2$, so

$$t = \frac{\lambda_n}{4} = \frac{\lambda}{4n} = \frac{(540\,\text{nm})}{(4)(1.35)} = 100\,\text{nm}.$$

This is the smallest thickness.

NOTE At this small thickness, blue (450 nm) and red (600 nm) also would reflect fairly constructively, so the bubble would appear almost white. The green color is more likely to be seen at the *next* thickness that gives constructive interference, $2t = 3\lambda/2n$, because other colors would be more fully cancelled by destructive interference. Then t would be $t = 3\lambda/4n = 300\,\text{nm}$. Note that green is seen in air, so $\lambda = 540\,\text{nm}$ (not λ/n).

FIGURE 34–21 Soap bubble. Example 34–7. The incident and reflected rays are assumed to be perpendicular to the bubble's surface. They are shown at a slight angle so we can distinguish them.



CAUTION

A formula is not enough: you must also check for phase changes at surfaces

Colors in a Thin Soap Film

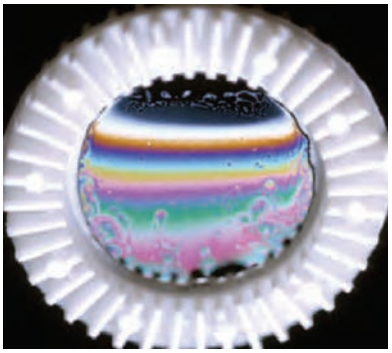


FIGURE 34–16b (Repeated.)

The thin film of soapy water (in a plastic loop) shown in Fig. 34–16b (repeated here) has stood vertically for a long time. Gravity has pulled the soapy water downward, so the film increases in thickness going toward the bottom. The top section of the soapy film in Fig. 34–16b is so thin (perhaps 30 nm thick $\ll \lambda$) that light reflected from the front and back surfaces have almost zero path difference. Thus the 180° phase change at the front surface assures that the two reflected waves are 180° out of phase for all wavelengths of visible light. The white light incident on this thin film does not reflect at the top part of the film, so the top is transparent and we see the background which is black.

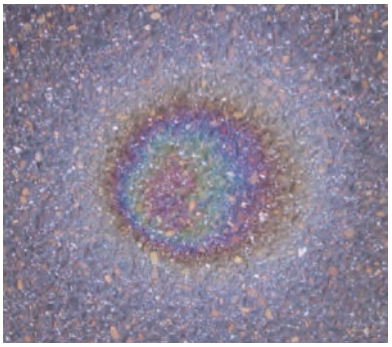
Below the black area at the top of Fig. 34–16b, there is a thin blue line, and then a white band. The film has thickened to perhaps 75 to 100 nm, so the shortest wavelength (blue) light begins to partially interfere constructively. But just below, where the thickness is slightly greater (100 nm), the path difference is reasonably close to $\lambda/2$ for much of the spectrum and we see white or silver.[†]

Immediately below the white band in this Figure we see a brown band, where $t \approx 200$ nm, and many wavelengths (not all) are close to λ—and those colors destructively interfere, leaving only a few colors to partially interfere constructively, giving us murky brown.

Farther down in Fig. 34–16b, with increasing thickness t, a path length $2t = 510$ nm corresponds nicely to $\frac{3}{2}\lambda$ for blue, but not for other colors, so we see blue ($\frac{3}{2}\lambda$ path difference plus $\frac{1}{2}\lambda$ phase change = constructive interference). Other colors experience constructive interference (at $\frac{3}{2}\lambda$ and then at $\frac{5}{2}\lambda$) at still greater thicknesses, so going down we see a series of separated colors something like a rainbow.

In the soap bubble of our Chapter-Opening Photo (page 995), similar things happen: at the top (where the film is thinnest) we see black and then silver-white, just as in the soap film shown in Fig. 34–16b.

FIGURE 34–16c (Repeated.)



Next consider the oil film on wet pavement shown in Fig. 34–16c (repeated here). An oil film is thickest at the center and thins out toward the edges. Notice the whitish outer ring where most colors constructively interfere, which would suggest a thickness on the order of 100 nm as discussed above for the white band in the soap film. Beyond the outer white band of the oil film, Fig. 34–16c, there is still some oil, but the film is so thin that reflected light from upper and lower surfaces destructively interfere and you can see right through this very thin oil film.

Lens Coatings

PHYSICS APPLIED
Lens coatings

An important application of thin-film interference is in the coating of glass to make it "nonreflecting," particularly for lenses. A glass surface reflects about 4% of the light incident upon it. Good-quality cameras, microscopes, and other optical devices may contain six to ten thin lenses. Reflection from all these surfaces can reduce the light transmitted considerably, and multiple reflections produce a background haze that reduces the quality of the image. By reducing reflection, transmission is increased.

A very thin coating on the lens surfaces can reduce reflections considerably. The thickness of the film is chosen so that light (at least for one wavelength) reflecting from the front and rear surfaces of the film destructively interferes, so all the light is transmitted. The amount of reflection at a boundary depends on the difference in index of refraction between the two materials. Ideally, the coating material should have an index of refraction which is the geometric mean $(= \sqrt{n_1 n_2})$ of those for air and glass, so that the amount of reflection at each surface is about equal. Then destructive interference can occur nearly completely for one particular wavelength depending on the thickness of the coating. Nearby wavelengths will at least partially destructively interfere, but a single coating cannot eliminate reflections for all wavelengths. Nonetheless, a single coating can reduce total reflection from 4% to 1% of the incident light. Often the coating is designed to eliminate the center of the reflected spectrum (around 550 nm).

[†]Why? Recall that red starts at 600 nm in air; so most colors in the spectrum lie between 450 nm and 600 nm in air; but in water the wavelengths are $n = 1.33$ times smaller, 340 nm to 450 nm, so a 100-nm thickness is a 200-nm path difference, not far from $\lambda/2$ for most colors.

The extremes of the spectrum—red and violet—will not be reduced as much by the lens coating. Since a mixture of red and violet produces purple, the light seen reflected from such coated lenses is purple (Fig. 34–22). Lenses containing two or three separate coatings can more effectively reduce a wider range of reflecting wavelengths, and increase transmission.



FIGURE 34–22 A coated lens. Note color of light reflected from the front lens surface.

PROBLEM SOLVING

Interference

1. **Interference effects** depend on the simultaneous arrival of two or more waves at the same point in space.
2. **Constructive interference** occurs when waves of a single wavelength continually arrive in phase with each other: crests of one wave arrive at the same time as crests of the other wave. The amplitudes of the waves then add to form a larger amplitude. Constructive interference also occurs when the path difference is exactly one full wavelength or any integer multiple of a full wavelength: $1\lambda, 2\lambda, 3\lambda, \cdots$.
3. **Destructive interference** occurs when crests of one wave continually arrive at the same time as troughs of the other wave. The amplitudes add, but they are of opposite sign, so the total amplitude is reduced to zero if the two amplitudes are equal. Destructive interference occurs whenever the phase difference is half a wave cycle, or the path difference is a half-integral number of wavelengths. Thus, the total amplitude will be zero if two identical waves arrive one-half wavelength out of phase, or $(m + \frac{1}{2})\lambda$ out of phase, where m is an integer.
4. For thin-film interference, an extra half-wavelength **phase shift** occurs when light **reflects** from an optically more dense medium (going from a material of lesser toward greater index of refraction).

EXAMPLE 34–8 **Nonreflective coating.** What is the thickness t of an optical coating of MgF_2 whose index of refraction is $n = 1.38$ and which is designed to eliminate reflected light at wavelengths (in air) around 550 nm when incident normally on glass for which $n = 1.50$?

APPROACH We explicitly follow the procedure outlined in the Problem Solving Strategy above.

SOLUTION

1. **Interference effects**. Consider two rays reflected from the front and rear surfaces of the coating on the lens as shown in Fig. 34–23. The rays are drawn not quite perpendicular to the lens so we can see each of them. These two reflected rays will interfere with each other.
2. **Constructive interference**. We want to eliminate reflected light, so we do not consider constructive interference.
3. **Destructive interference**. To eliminate reflection, we want reflected rays 1 and 2 to be $\frac{1}{2}$ cycle out of phase with each other so that they destructively interfere. The phase difference is due to the path difference $2t$ traveled by ray 2, as well as any phase change in either ray due to reflection.
4. **Reflection phase shift**. Rays 1 and 2 *both* undergo a change of phase by $\frac{1}{2}$ cycle when they reflect from the coating's front and rear surfaces, respectively (at both surfaces the index of refraction increases). Thus there is no net change in phase due to the reflections. The net phase difference will be due to the extra path $2t$ taken by ray 2 in the coating, where $n = 1.38$. We want $2t$ to equal $\frac{1}{2}\lambda_n$ so that destructive interference occurs, where $\lambda_n = \lambda/n$ is the wavelength in the coating. With $2t = \lambda_n/2 = \lambda/2n$, then

$$t = \frac{\lambda_n}{4} = \frac{\lambda}{4n} = \frac{(550\,\text{nm})}{(4)(1.38)} = 99.6\,\text{nm}.$$

NOTE We could have set $2t = (m + \frac{1}{2})\lambda_n$, where m is an integer. The smallest thickness $(m = 0)$ is usually chosen because destructive interference will occur over the widest angle.

NOTE Complete destructive interference occurs only for the given wavelength of visible light. Longer and shorter wavelengths will have only partial cancellation.

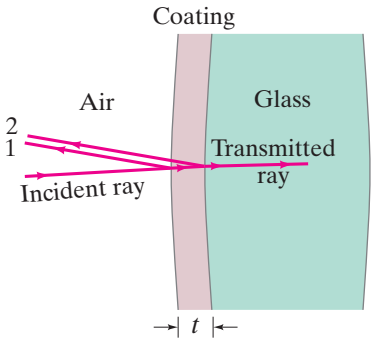


FIGURE 34–23 Example 34–8. Incident ray of light is partially reflected at the front surface of a lens coating (ray 1) and again partially reflected at the rear surface of the coating (ray 2), with most of the energy passing as the transmitted ray into the glass.

34–6 Michelson Interferometer

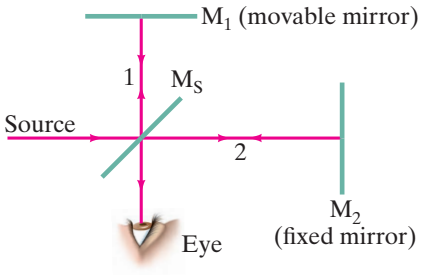


FIGURE 34–24 Michelson interferometer.

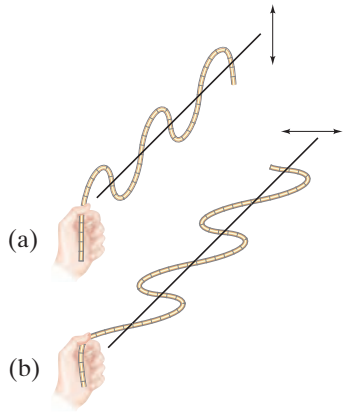
A useful instrument involving wave interference is the **Michelson interferometer** (Fig. 34–24),† invented by the American Albert A. Michelson (Section 31–7). Monochromatic light from a single point on an extended source is shown striking a half-silvered mirror M_S. This **beam splitter** mirror M_S has a thin layer of silver that reflects only half the light that hits it, so that half of the beam passes through to a fixed mirror M_2, where it is reflected back. The other half is reflected by M_S to a mirror M_1 that is movable (by a fine-thread screw), where it is also reflected back. Upon its return, part of beam 1 passes through M_S and reaches a sensor or the eye; and part of beam 2, on its return, is reflected by M_S into the eye. If the two path lengths are identical, the two coherent beams entering the eye constructively interfere and brightness will be seen. If the movable mirror is moved a distance $\lambda/4$, one beam will travel an extra distance equal to $\lambda/2$ (because it travels back and forth over the distance $\lambda/4$). In this case, the two beams will destructively interfere and darkness will be seen. As M_1 is moved farther, brightness will recur (when the path difference is λ), then darkness, and so on.

Very precise length measurements can be made with an interferometer. The motion of mirror M_1 by only $\frac{1}{4}\lambda$ produces a clear difference between brightness and darkness. For $\lambda = 400\,\text{nm}$, this means a precision of 100 nm or 10^{-4} mm! If mirror M_1 is tilted very slightly, the bright or dark spots are seen instead as a series of bright and dark lines or "fringes" that move as M_1 moves. By counting the number of fringes (or fractions thereof) that pass a reference line, extremely precise length measurements can be made.

34–7 Polarization

An important and useful property of light is that it can be *polarized*. To see what this means, let us examine waves traveling on a rope. A rope can oscillate in a vertical plane, Fig. 34–25a, or in a horizontal plane, Fig. 34–25b. In either case, the wave can be said to be **linearly polarized** or **plane-polarized**—*the oscillations are in a plane*.

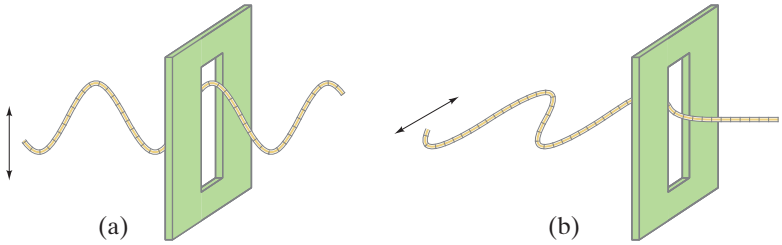
FIGURE 34–25 Transverse waves on a rope polarized (a) in a vertical plane and (b) in a horizontal plane.



If we now place an obstacle containing a vertical slit in the path of the wave, Fig. 34–26, a vertically polarized wave passes through the vertical slit, but a horizontally polarized wave will not. If a horizontal slit were used, the vertically polarized wave would be stopped. If both types of slit were used, both types of wave would be stopped by one slit or the other. Note that polarization can exist *only* for *transverse waves*, and not for longitudinal waves such as sound. The latter oscillate only along the direction of motion, and neither orientation of slit would stop them.

Maxwell's theory of light as electromagnetic (EM) waves predicted that light can be polarized since an EM wave is a transverse wave. The direction of polarization in a plane-polarized EM wave is taken as the direction of the electric field vector $\vec{\mathbf{E}}$.

FIGURE 34–26 (a) A vertically polarized wave passes through a vertical slit, but (b) a horizontally polarized wave will not.



†There are other types of interferometer, but Michelson's is the best known.

Light is not necessarily polarized. It can also be **unpolarized**, which means that the source has oscillations in many planes at once, as shown in Fig. 34–27a. Ordinary lightbulbs emit unpolarized light, as does the Sun.

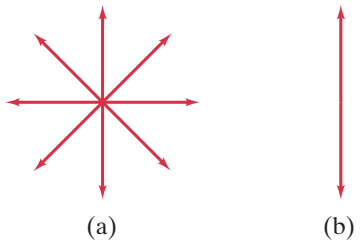


FIGURE 34–27 (a) Oscillation of the electric field vectors in unpolarized light. The light is traveling into or out of the page. (b) Electric field in linear polarized light.

Polaroids (Polarization by Absorption)

Plane-polarized light can be obtained from unpolarized light using certain crystals such as tourmaline. Or, more commonly, we use a **Polaroid sheet**. (Polaroid materials were invented in 1929 by Edwin Land.) A Polaroid sheet consists of long complex molecules arranged parallel to one another. Such a Polaroid acts like a series of parallel slits to allow one orientation of polarization to pass through nearly undiminished. This direction is called the *transmission axis* of the Polaroid. Polarization perpendicular to this direction is absorbed almost completely by the Polaroid.

Absorption by a Polaroid can be explained at the molecular level. An electric field $\vec{\mathbf{E}}$ that oscillates parallel to the long molecules can set electrons into motion along the molecules, thus doing work on them and transferring energy. Hence, if $\vec{\mathbf{E}}$ is parallel to the molecules, it gets absorbed. An electric field $\vec{\mathbf{E}}$ perpendicular to the long molecules does not have this possibility of doing work and transferring its energy, and so passes through freely. When we speak of the *transmission axis* of a Polaroid, we mean the direction for which $\vec{\mathbf{E}}$ is passed, so a Polaroid axis is *perpendicular* to the long molecules. [If we want to think of there being slits between the parallel molecules in the sense of Fig. 34–26, then Fig. 34–26 would apply for the $\vec{\mathbf{B}}$ field in the EM wave which is perpendicular to the $\vec{\mathbf{E}}$ field.]

If a beam of plane-polarized light strikes a Polaroid whose transmission axis is at an angle θ to the incident polarization direction, the beam will emerge plane-polarized parallel to the Polaroid transmission axis, and the amplitude of E will be reduced to $E\cos\theta$, Fig. 34–28. Thus, a Polaroid passes only that component of polarization (the electric field vector, $\vec{\mathbf{E}}$) that is parallel to its transmission axis. Because the intensity of a light beam is proportional to the square of the amplitude (Sections 15–3 and 31–8), the intensity of a plane-polarized beam transmitted by a polarizer is proportional to $(E_0\cos\theta)^2$, a relation called Malus' law,

$$I = I_0\cos^2\theta, \qquad \left[\begin{array}{c}\text{intensity of plane-polarized}\\ \text{light passed by polarizer}\end{array}\right] \qquad \textbf{(34–8)}$$

CAUTION
Intensity of previously polarized light

where I_0 is the incoming intensity and θ is the angle between the polarizer transmission axis and the plane of polarization of the incoming wave.

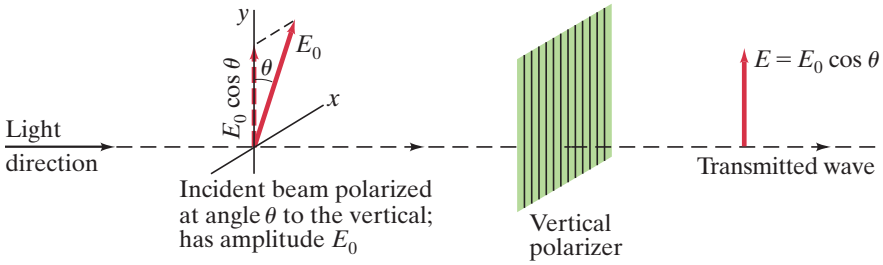


FIGURE 34–28 Vertical Polaroid transmits only the vertical component of a wave (electric field) incident upon it.

A Polaroid can be used as a **polarizer** to *produce* plane-polarized light from unpolarized light, since only the component of light parallel to the axis is transmitted. A Polaroid can also be used as an **analyzer** to determine (1) if light is polarized and (2) the plane of polarization. A Polaroid acting as an analyzer will pass the same amount of light independent of the orientation of its axis if the light is unpolarized; try rotating one lens of a pair of Polaroid sunglasses while looking through it at a lightbulb. If the light is polarized, however, when you rotate the Polaroid the transmitted light will be a maximum when the plane of polarization is parallel to the Polaroid's transmission axis, and a minimum when perpendicular to it. If you do this while looking at the sky, preferably at right angles to the Sun's direction, you will see that skylight is polarized. (Direct sunlight is unpolarized, but don't look directly at the Sun, even through a polarizer, for damage to the eye may occur.) If the light transmitted by an analyzer Polaroid falls to zero at one orientation, then the light is 100% plane-polarized. If it merely reaches a minimum, the light is *partially polarized*.

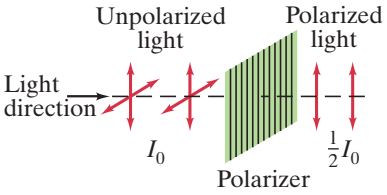


FIGURE 34–29 Unpolarized light has equal intensity vertical and horizontal components. After passing through a polarizer, one of these components is eliminated. The intensity of the light is reduced to half.

Unpolarized light consists of light with random directions of polarization. Each of these polarization directions can be resolved into components along two mutually perpendicular directions. On average, an unpolarized beam can be thought of as two plane-polarized beams of equal magnitude perpendicular to one another. When unpolarized light passes through a polarizer, one component is eliminated. So the unpolarized light passing through a polarizer not only becomes polarized, but also its intensity is reduced by half because half the light is eliminated (Fig. 34–29):

$$I = \tfrac{1}{2}I_0. \qquad \left[\begin{array}{c}\text{intensity of initially}\\ \text{unpolarized light after}\\ \text{passing through polarizer}\end{array}\right]$$

(Note that Eq. 34–8 refers to already-polarized light.)

When two Polaroids are *crossed*—that is, their polarizing axes are perpendicular to one another—unpolarized light can be entirely stopped. As shown in Fig. 34–30, unpolarized light is made plane-polarized by the first Polaroid (the polarizer). The second Polaroid, the analyzer, then eliminates this component since its transmission axis is perpendicular to the first.

FIGURE 34–30 Crossed Polaroids completely eliminate light.

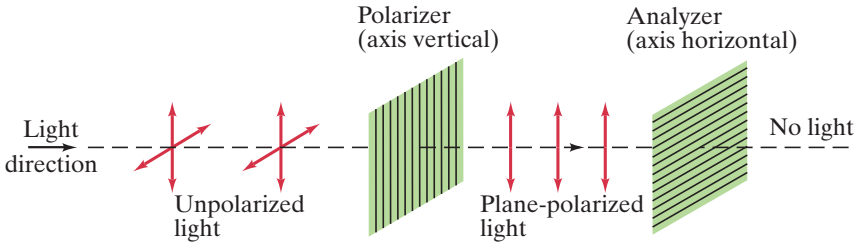


FIGURE 34–31 Crossed Polaroids. When the two polarized sunglass lenses overlap, with axes perpendicular, almost no light passes through.



You can try this with Polaroid sunglasses (Fig. 34–31). Note that Polaroid sunglasses eliminate 50% of unpolarized light because of their polarizing property; they absorb even more because they are colored. Plane-polarized light in any direction is also stopped by crossed Polaroids.

EXAMPLE 34–9 **Two Polaroids at 60°.** Unpolarized light passes through two Polaroids; the axis of the first is vertical and that of the second is at 60° to the vertical. Describe the orientation and intensity of the transmitted light.

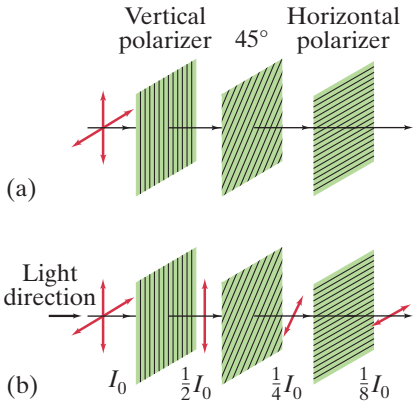
APPROACH Half of the unpolarized light is absorbed by the first Polaroid, and the remaining light emerges plane-polarized vertically. When that light passes through the second Polaroid, the intensity is further reduced according to Eq. 34–8, and the plane of polarization is then along the axis of the second Polaroid.

SOLUTION The first Polaroid eliminates half the light, so the intensity is reduced by half: $I_1 = \frac{1}{2}I_0$. The light reaching the second polarizer is vertically polarized and so is reduced in intensity (Eq. 34–8) to

$$I_2 = I_1(\cos 60°)^2 = \tfrac{1}{4}I_1.$$

Thus, $I_2 = \frac{1}{8}I_0$. The transmitted light has an intensity one-eighth that of the original and is plane-polarized at a 60° angle to the vertical. If the Polaroids are colored, even more light is lost.

FIGURE 34–32 Example 34–10.



CONCEPTUAL EXAMPLE 34–10 **Three Polaroids.** We saw in Fig. 34–30 that when unpolarized light falls on two crossed Polaroids (axes at 90°), no light passes through. What happens if a third Polaroid, with axis at 45° to each of the other two, is placed between them (Fig. 34–32a)?

RESPONSE We start just as in Example 34–9 and recall again that light emerging from each Polaroid is polarized parallel to that Polaroid's axis. Thus the angle in Eq. 34–8 is that between the transmission axes of each pair of Polaroids taken in turn. The first Polaroid changes the unpolarized light to plane-polarized and reduces the intensity from I_0 to $I_1 = \frac{1}{2}I_0$. The second polarizer further reduces the intensity by $(\cos 45°)^2$, Eq. 34–8:

$$I_2 = I_1(\cos 45°)^2 = \tfrac{1}{2}I_1 = \tfrac{1}{4}I_0.$$

The light leaving the second polarizer is plane-polarized at 45° (Fig. 34–32b) relative to the third polarizer, so the third one reduces the intensity to

$$I_3 = I_2(\cos 45°)^2 = \tfrac{1}{2}I_2,$$

or $I_3 = \frac{1}{8}I_0$. Thus $\frac{1}{8}$ of the original intensity gets transmitted.

NOTE If we don't insert the 45° Polaroid, zero intensity results (Fig. 34–30).

EXERCISE E How much light would pass through if the 45° polarizer in Example 34–10 was placed not between the other two polarizers but (*a*) before the vertical (first) polarizer, or (*b*) after the horizontal polarizer?

Polarization by Reflection

Another means of producing polarized light from unpolarized light is by reflection. When light strikes a nonmetallic surface at any angle other than perpendicular, the reflected beam is found to be polarized preferentially in the plane parallel to the surface, Fig. 34–33. In other words, the component with polarization in the plane perpendicular to the surface is preferentially transmitted or absorbed. You can check this by rotating Polaroid sunglasses while looking through them at a flat surface of a lake or road. Since most outdoor surfaces are horizontal, Polaroid sunglasses are made with their axes vertical to eliminate the more strongly reflected horizontal component, and thus reduce glare.

People who go fishing wear Polaroids to eliminate reflected glare from the surface of a lake or stream and thus see beneath the water more clearly (Fig. 34–34).

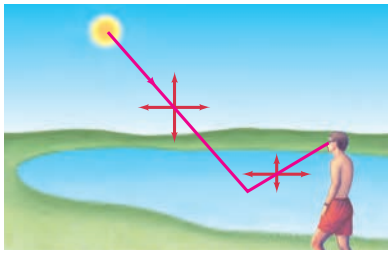


FIGURE 34–33 Light reflected from a nonmetallic surface, such as the smooth surface of water in a lake, is partially polarized parallel to the surface.

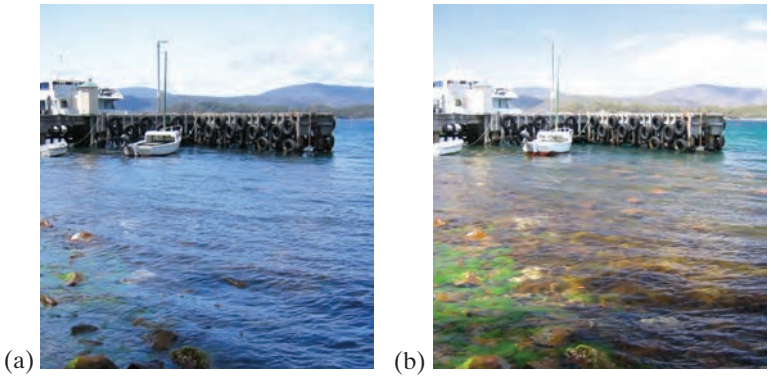


FIGURE 34–34 Photographs of a lake, (a) allowing all light into the camera lens, and (b) using a polarizer. The polarizer is adjusted to absorb most of the (polarized) light reflected from the water's surface, allowing the dimmer light from the bottom of the lake, and any fish lying there, to be seen more readily.

The amount of polarization in the reflected beam depends on the angle, varying from no polarization at normal incidence to 100% polarization at an angle known as the **polarizing angle** θ_p.[†] This angle is related to the index of refraction of the two materials on either side of the boundary by the equation

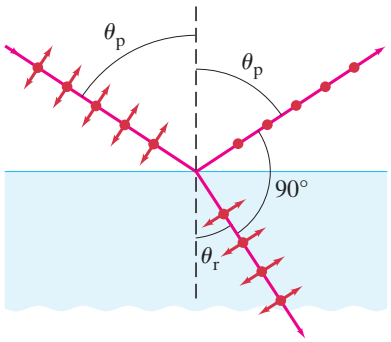
$$\tan\theta_p = \frac{n_2}{n_1}, \qquad \textbf{(34–9a)}$$

where n_1 is the index of refraction of the material in which the incident beam is traveling, and n_2 is that of the medium beyond the reflecting boundary. If the beam is traveling in air, $n_1 = 1$, and Eq. 34–9a becomes

$$\tan\theta_p = n. \qquad \textbf{(34–9b)}$$

The polarizing angle θ_p is also called **Brewster's angle**, and Eqs. 34–9 *Brewster's law*, after the Scottish physicist David Brewster (1781–1868), who worked it out experimentally in 1812. Equations 34–9 can be derived from the electromagnetic wave theory of light. It is interesting that at Brewster's angle, the reflected ray and the transmitted (refracted) ray make a 90° angle to each other; that is, $\theta_p + \theta_r = 90°$, where θ_r is the refraction angle (Fig. 34–35). This can be seen by substituting Eq. 34–9a, $n_2 = n_1\tan\theta_p = n_1\sin\theta_p/\cos\theta_p$, into Snell's law, $n_1\sin\theta_p = n_2\sin\theta_r$, which gives $\cos\theta_p = \sin\theta_r$ which can only hold if $\theta_p = 90° - \theta_r$ (see Trigonometric identities inside back cover or Appendix A).

FIGURE 34–35 At θ_p the reflected light is plane-polarized parallel to the surface, and $\theta_p + \theta_r = 90°$, where θ_r is the refraction angle. (The large dots represent electric field vibrations perpendicular to the page.)



[†]Only a fraction of the incident light is reflected at the surface of a transparent medium. Although this reflected light is 100% polarized (if $\theta = \theta_p$), the remainder of the light, which is transmitted into the new medium, is only partially polarized.

EXAMPLE 34–11 **Polarizing angle.** (*a*) At what incident angle is sunlight reflected from a lake perfectly plane-polarized? (*b*) What is the refraction angle?

APPROACH The polarizing angle at the surface is Brewster's angle, Eq. 34–9b. We find the angle of refraction from Snell's law.

SOLUTION (*a*) We use Eq. 34–9b with $n = 1.33$, so $\tan\theta_p = 1.33$ giving $\theta_p = 53.1°$.
(*b*) From Snell's law, $\sin\theta_r = \sin\theta_p/n = \sin 53.1°/1.33 = 0.601$ giving $\theta_r = 36.9°$.

NOTE $\theta_p + \theta_r = 53.1° + 36.9° = 90.0°$, as expected.

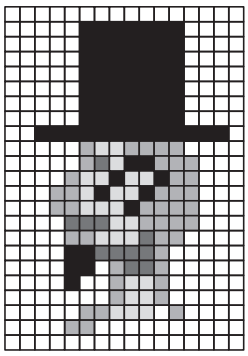


FIGURE 34–36 Example of an image made up of many small picture elements or *pixels*.

*34–8 Liquid Crystal Displays (LCD)

A wonderful use of polarization is in a **liquid crystal display** (LCD). LCDs are used as the display in cell phones, other hand-held electronic devices, and flat-panel computer and television screens.

A liquid crystal display is made up of many tiny rectangles called **pixels**, or "picture elements." The picture you see depends on which pixels are dark or light and of what color, as suggested in Fig. 34–36 for a simple black and white picture. (Hopefully you can recognize Abe Lincoln.)

Liquid crystals are organic materials that at room temperature exist in a phase that is neither fully solid nor fully liquid. They are sort of gooey, and their molecules display a randomness of position characteristic of liquids, as discussed in Section 17–1 and Fig. 17–2b. They also show some of the orderliness of a solid crystal (Fig. 17–2a), but only in one dimension.

FIGURE 34–37 Liquid crystal molecules tend to align in one dimension (parallel to each other) but have random positions (left-right, up-down).



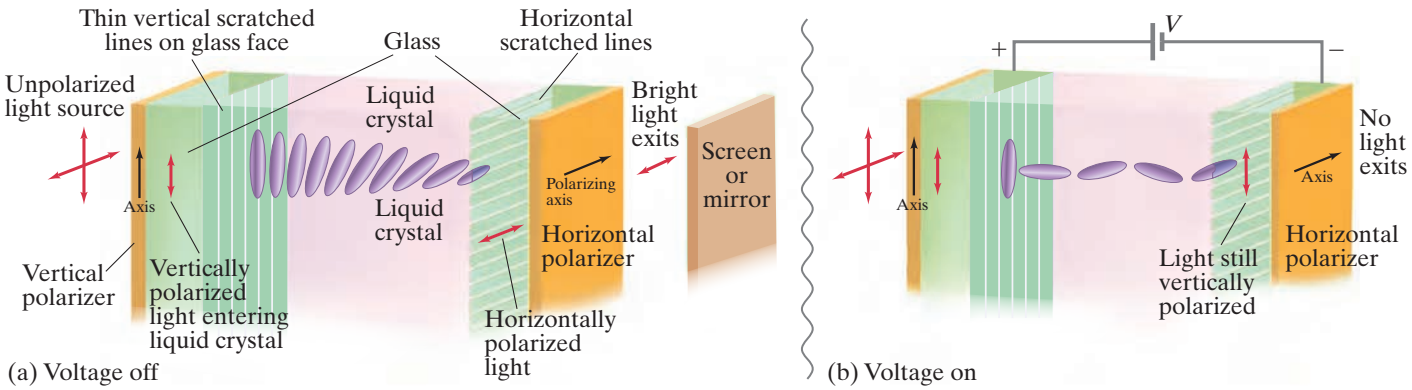
The liquid crystals we find useful are made up of relatively rigid rod-like molecules that interact weakly with each other and tend to align parallel to each other, as shown in Fig. 34–37.

In a simple LCD, each pixel (picture element) contains liquid crystal material sandwiched between two glass plates (green in Fig. 34–38) whose inner surfaces have nanometer-wide parallel scratched lines. The rod-like liquid crystal molecules (violet) in contact with the scratched lines tend to line up along those lines. The front and back plates usually have their scratched lines at 90° to each other, and the weak electric forces between the rod-like molecules tend to keep them nearly aligned with their nearest neighbors (and lying in a plane parallel to the glass plates), resulting in a twisted pattern, molecules being vertical on the left, rotating gradually until horizontal on the right, Fig. 34–38a.

The outer surfaces of the two glass plates each have a thin film polarizer (orange), which are also oriented at 90° to each other. Unpolarized light incident from the left immediately becomes plane-polarized, and the liquid crystal molecules (purple) keep this polarization aligned with their rod-like shape. That is, the plane of polarization of the light rotates with the molecules as the light passes through the liquid crystal (pink). The light emerges with its plane of polarization rotated by 90°, and readily passes through the second polarizer, Fig. 34–38a. A tiny LCD pixel in this situation will appear bright.

Now suppose a voltage is applied to transparent electrodes on each glass plate of the pixel. The rod-like molecules are polar (or can acquire an internal separation of charge due to the applied electric field). The applied voltage tends to align

FIGURE 34–38 (a) "Twisted" form of liquid crystal. Light polarization plane is rotated 90°, and so is transmitted by the horizontal polarizer. Only one series of molecules is shown. (b) Molecules disoriented by electric field. The plane of polarization does not get changed, so light does not pass through the horizontal polarizer. (Transparent electrodes not shown.)



the molecules end-to-end, and they no longer follow the careful twisted pattern shown in Fig. 34–38a. Instead the applied electric field tends to align the molecules end-to-end, left to right perpendicular to the glass plates, Fig. 34–38b, and so they don't affect the light polarization significantly. The entering plane-polarized light no longer has its plane of polarization rotated as it passes through the liquid crystal, and no light exits through the second (horizontal) polarizer (Fig. 34–38b). With the voltage on, the pixel appears dark.†

Simple display screens, such as for watches and calculators, use ambient light as the unpolarized source on the left in Fig. 34–38 (you can't see the display in the dark), with a mirror behind the LCD to reflect the light back. There are only a few pixels, corresponding to the straight segments needed to form the numbers from 0 to 9 (and letters in some displays), as seen in Fig. 34–39. Any pixels to which a voltage is applied appear dark and form part of a number. With no voltage, pixels pass light through the polarizers to the mirror and back out, which forms a bright background. Displays with white numbers on a dark background have the voltages reversed.

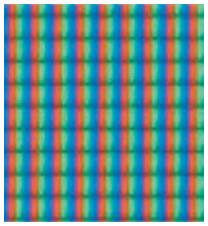
Television, cell phone, and computer LCDs are more sophisticated. A color pixel consists of three cells, or subpixels, each covered with a red, green, or blue filter (Fig. 34–40). Varying brightnesses of these three primary colors can yield almost any natural color. A good-quality screen consists of millions of pixels. Behind this array of pixels is a light source, sometimes thin fluorescent tubes the diameter of a straw, or light-emitting diodes (LEDs). The light passes through the liquid crystal subpixels, or not, depending on the voltage applied to each, as we discussed in detail in Section 23–10. See especially Figs. 23–31 and 23–32.

To obtain a range of gray scale or range of color brightness, each subpixel cannot simply go on or off as in Fig. 34–38. Several techniques can be used depending on the construction of the LCD. (1) If the voltage applied in Fig. 34–38b is small enough, the disorientation of the molecules may be small, allowing some rotation of the polarization vector and thus some light can pass through, the actual amount depending on the voltage. (2) Alternatively, each subpixel can be pulsed—the length of time it is *on* affects the perceived brightness. (3) The effect of stronger or weaker brightness can instead be provided by the number of nearby subpixels of the same color that are turned on or off; this third system lets the eye "average" over many pixels, but reduces the sharpness or resolution of the picture.



FIGURE 34–39 Watch-face LCD display with altimeter. The black segments or pixels have a voltage applied to them. Note that the 8 uses all seven segments (pixels); other numbers use fewer.

FIGURE 34–40 Arrangement of subpixels on a TV or computer display (enlarged).

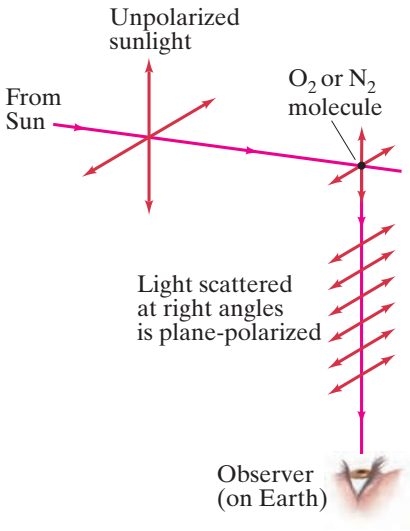


*34–9 Scattering of Light by the Atmosphere

Sunsets are red, the sky is blue, and skylight is polarized (at least partially). These phenomena can be explained on the basis of the *scattering* of light by the molecules of the atmosphere. In Fig. 34–41 we see unpolarized light from the Sun impinging on a molecule of the Earth's atmosphere. The electric field of the EM wave sets the electric charges within the molecule into oscillation, and the N_2 or O_2 molecule absorbs some of the incident radiation. But the molecule quickly reemits this light since the charges are oscillating. As discussed in Section 31–4 oscillating electric charges produce EM waves. The intensity is strongest along the direction perpendicular to the oscillation, and drops to zero along the line of oscillation (Section 31–4). In Fig. 34–41 the motion of the charges is resolved into two components. An observer at right angles to the direction of the sunlight, as shown, will see plane-polarized light because no light is emitted along the line of the other component of the oscillation. (When viewing along the line of an oscillation, you don't see that oscillation, and hence see no waves made by it.) At other viewing angles, both components will be present; one will be stronger, however, so the light appears partially polarized. Thus, this process of scattering explains the polarization of skylight.

Scattering of light by the Earth's atmosphere depends on wavelength λ. For particles much smaller than the wavelength of light (such as molecules of air), the particles will be less of an obstruction to long wavelengths than to short ones.

FIGURE 34–41 Unpolarized sunlight scattered by molecules of the air. An observer at right angles sees plane-polarized light, since the component of oscillation along the line of sight emits no light along that line.



†Some displays use an opposite system: the polarizers are parallel to each other (the scratches remain at 90° to maintain the twist). Then voltage *off* results in *black* (no light), and voltage *on* results in bright light.

PHYSICS APPLIED

Why the sky is blue
Why sunsets are red
Why clouds are white

The scattering decreases, in fact, as $1/\lambda^4$. Blue and violet light are thus scattered much more than red and orange, which is why the sky looks blue. At sunset, the Sun's rays pass through a maximum length of atmosphere. Much of the blue has been taken out by scattering. The light that reaches us at this low angle where the Sun is near the horizon, and reflects off clouds and haze, is thus lacking in blue. That is why sunsets appear reddish.

The dependence of scattering on $1/\lambda^4$ is valid only if the scattering objects are much smaller than the wavelength of the light. This is valid for oxygen and nitrogen molecules whose diameters are about 0.2 nm. Clouds, however, contain water droplets or crystals that are much larger than λ. They scatter all frequencies of light nearly uniformly. Hence clouds appear white (or gray, if shadowed).

34–10 Brightness: Lumens and Luminous Intensity

The *intensity* of light, as for any electromagnetic wave, is measured by the Poynting vector in W/m^2, and the total power output of a source can be measured in watts (the *radiant flux*). But for measuring the visual sensation we call *brightness*, we must consider *only* the visible spectrum as well as the eye's sensitivity to different wavelengths—the eye is most sensitive in the central, 550-nm (green), portion of the spectrum.

These factors are taken into account by the quantity **luminous flux**, F_ℓ, whose unit is the **lumen** (lm). One lumen is equivalent to $\frac{1}{683}$ watts of 555-nm light.

Since the luminous flux from a source may not be uniform over all directions, we define the **luminous intensity** I_ℓ as the luminous flux per unit solid angle[†] (steradian). Its unit is the **candela** (cd) where $1\text{ cd} = 1\text{ lm/sr}$.

The **illuminance**, E_ℓ, is the luminous flux incident on a surface per unit area of the surface: $E_\ell = F_\ell/A$. Its unit is the lumen per square meter (lm/m^2) and is a measure of the illumination falling on a surface.[‡]

EXAMPLE 34–12 **Lightbulb illuminance.** The brightness of a particular type of 100-W lightbulb is rated at 1700 lm. Determine (*a*) the luminous intensity and (*b*) the illuminance at a distance of 2.0 m.

APPROACH Assume the light output is uniform in all directions.

SOLUTION (*a*) A full sphere corresponds to 4π sr.[†] Hence, $I_\ell = 1700\text{ lm}/4\pi\text{ sr} = 135\text{ cd}$. It does not depend on distance.

(*b*) At $d = 2.0\text{ m}$ from the source, the luminous flux per unit area is

$$E_\ell = \frac{F_\ell}{A} = \frac{F_\ell}{4\pi d^2} = \frac{1700\text{ lm}}{(4\pi)(2.0\text{ m})^2} = 34\text{ lm/m}^2.$$

The illuminance decreases as the square of the distance.

*34–11 Efficiency of Lightbulbs

We can compare the efficiencies of different types of lightbulbs using the lumen. A 100-W traditional incandescent bulb (no halogen capsule) emits EM waves of a wide variety of wavelengths (Section 37–1), including about 1500 lumens of visible light, depending on the filament (a finer filament gets hotter and gives a more natural spectrum and more light output, but has a shorter life); output also depends on any coating on the inside of the glass bulb. "Soft-white" bulbs are coated and are considered by many as ideal for reading. As an average, let us estimate that a normal 100 W (input) incandescent bulb produces 1500 lm, which means an efficiency of $1500\text{ lm}/100\text{ W} = 15\text{ lm/W}$. Lifetimes are around 1000 hours.

[†]A solid angle is a sort of two-dimensional angle and is measured in steradians (sr). Think of a solid angle starting at a point and intercepting an area ΔA on a sphere of radius r surrounding that point. The solid angle has magnitude $\Delta A/r^2$ steradians. A solid angle including all of space intercepts the full surface area of the sphere, $4\pi r^2$, and so has magnitude $4\pi r^2/r^2 = 4\pi$ steradians. (Compare to a normal angle, for which a full circle subtends 2π radians.)

[‡]The British unit is the foot-candle, or lumen per square foot.

A halogen bulb (incandescent having the filament in a capsule filled with halogen gas) does better, about 20 lm/W, plus a life that can reach about 2000 h. Compact fluorescent bulbs (CFL) for use in ordinary sockets are spiral-shaped (or have sharp bends) and can produce about 70 lm/W and may last up to 10,000 h if well made (often not the case). CFLs are less wasteful of energy and can last longer; but the light they produce can be displeasing to many people, they cost more, and contain toxic material. LED bulbs are even more efficient, as much as 160 or 180 lm/W for a good white color (a 300 lm/W bulb has been claimed) and a life of perhaps 20,000 h (again a reputable brand). How CFLs and LEDs (light emitting diodes) work is discussed in Chapter 39 and 40 respectively.

The spectrum of light emitted by a filament depends on its temperature (see Section 37–1), whereas CFL and LED may specify an "effective" temperature (technically, "correlated color temperature").†

We can look at all this in another way which connects with our definition in Section 34–10 that 683 lumens is equivalent to 1 watt of 555-nm light. For white light, the sensitivity of the eye peaks at about 555 nm and is less sensitive for longer or shorter wavelengths, dropping to zero around 400 nm and 700 nm. So let us estimate over the full visible spectrum ("white") that 1 watt of light entering the eye would correspond to about half of what it is at 555 nm (683 lm) or roughly 300 or 400 lumens. Then a 100-W incandescent bulb puts out about $(1500\,\text{lm})/(300\,\text{lm/W}) \approx 5\,\text{W}$ of visible light (100 W input to the bulb, 5 W output of visible light).

†LED bulbs may specify a temperature, but that is *not* its own temperature. A 3000 K LED means the mix of different color LEDs gives a warm yellowish light like a 3000 K incandescent bulb filament, and that would be a comfortable bulb to read by. If labeled 5000 K, which is close to the Sun's temperature of 6000 K, it is more white like outdoor light, and may be too harsh to read by. Temperature and color are connected for a filament (discussed in Section 37–1) but not for LEDs (Chapter 40).

Summary

The wave theory of light is strongly supported by the observations that light exhibits **interference** and **diffraction**. Wave theory also explains the refraction of light and the fact that light travels more slowly in transparent solids and liquids than it does in air.

An aid to predicting wave behavior is **Huygens' principle**, which states that every point on a wave front can be considered as a source of tiny wavelets that spread out in the forward direction at the speed of the wave itself. The new wave front is the envelope (the common tangent) of all the wavelets.

The wavelength of light in a medium with index of refraction n is

$$\lambda_n = \frac{\lambda}{n}, \qquad \textbf{(34–1)}$$

where λ is the wavelength in vacuum; the frequency is not changed.

Young's double-slit experiment clearly demonstrated the interference of light. The observed bright lines of the interference pattern are explained as constructive interference between the beams coming through the two slits, where the beams differ in path length by zero or an integral number of wavelengths. The dark areas in between are due to destructive interference when the path lengths differ by $\frac{1}{2}\lambda$, $\frac{3}{2}\lambda$, and so on. The angles θ at which **constructive interference** occurs are given by

$$\sin\theta = m\frac{\lambda}{d}, \qquad \textbf{(34–2a)}$$

where λ is the wavelength of the light, d is the separation of the slits, and m is an integer $(0, 1, 2, \cdots)$. **Destructive interference** occurs at angles θ given by

$$\sin\theta = \left(m + \tfrac{1}{2}\right)\frac{\lambda}{d}, \qquad \textbf{(34–2b)}$$

where m is an integer $(0, 1, 2, \cdots)$.

[*The light intensity I_θ at any point in a double-slit interference pattern can be calculated using a phasor diagram, which predicts that

$$I_\theta = I_0\cos^2\frac{\delta}{2} \qquad \textbf{(34–6)}$$

where I_0 is the intensity at $\theta = 0$ and the phase angle δ is

$$\delta = \frac{2\pi d}{\lambda}\sin\theta. \qquad \textbf{(34–4)}]$$

Two sources of light are perfectly **coherent** if the waves leaving them are of the same single frequency and maintain the same phase relationship at all times. If the light waves from the two sources have a random phase with respect to each other over time (as for two lightbulbs), the two sources are **incoherent**.

Light reflected from the front and rear surfaces of a thin film of transparent material can interfere constructively or destructively, depending on the path difference. A phase change of 180° or $\frac{1}{2}\lambda$ occurs when the light reflects at a surface where the index of refraction increases. Such **thin-film interference** has many practical applications, such as lens coatings and using Newton's rings to check the uniformity of glass surfaces.

In **unpolarized light**, the electric field vectors oscillate in all transverse directions. If the electric vector oscillates only in one plane, the light is said to be **plane-polarized**. Light can also be partially polarized.

When an unpolarized light beam passes through a **Polaroid** sheet, the emerging beam is plane-polarized. When a light beam is polarized and passes through a Polaroid, the intensity varies as the Polaroid is rotated. Thus a Polaroid can act as a **polarizer** or as an **analyzer**.

The intensity I of a plane-polarized light beam incident on a Polaroid is reduced by the factor

$$I = I_0\cos^2\theta \qquad \textbf{(34–8)}$$

where θ is the angle between the axis of the Polaroid and the initial plane of polarization.

Light can also be partially or fully **polarized by reflection**. If light traveling in air is reflected from a medium of index of refraction n, the reflected beam will be *completely* plane-polarized if the incident angle θ_p is given by

$$\tan\theta_p = n. \qquad \textbf{(34–9b)}$$

The fact that light can be polarized shows that it must be a transverse wave.

The brightness of a lightbulb may be specified in *lumens*, which takes into account the sensitivity of the eye.

Questions

1. Does Huygens' principle apply to sound waves? To water waves? Explain how Huygens' principle makes sense for water waves, where each point vibrates up and down.
2. What is the evidence that light is energy?
3. Why is light sometimes described as rays and sometimes as waves?
4. We can hear sounds around corners but we cannot see around corners; yet both sound and light are waves. Explain the difference.
5. Can the wavelength of light be determined from reflection or refraction measurements?
6. Two rays of light from the same source destructively interfere if their path lengths differ by how much?
7. Monochromatic red light is incident on a double slit and the interference pattern is viewed on a screen some distance away. Explain how the fringe pattern would change if the red light source is replaced by a blue light source.
8. If Young's double-slit experiment were submerged in water, how would the fringe pattern be changed?
9. Compare Figs. 34–5, 34–6, and 34–7, which are different representations of the double-slit experiment. For each Figure state the direction the light is traveling. Where are the wave crests in terms of this direction? How are they represented in each Figure? Give one advantage of each Figure in helping you understand the double-slit experiment and interference.
10. Compare a double-slit experiment for sound waves to that for light waves. Discuss the similarities and differences.
11. Suppose white light falls on the two slits of Fig. 34–7, but one slit is covered by a red filter (700 nm) and the other by a blue filter (450 nm). Describe the pattern on the screen.
12. Why doesn't the light from the two headlights of a distant car produce an interference pattern?
13. Why are interference fringes noticeable only for a *thin* film like a soap bubble and not for a thick piece of glass, say?
14. Why are Newton's rings (Fig. 34–18) closer together farther from the center?
15. Some coated lenses appear greenish yellow when seen by reflected light. What reflected wavelengths do you suppose the coating is designed to eliminate completely?
16. A drop of oil on a pond appears bright at its edges where its thickness is much less than the wavelengths of visible light. What can you say about the index of refraction of the oil compared to that of water?
17. What does polarization tell us about the nature of light?
18. Explain the advantage of polarized sunglasses over plain tinted sunglasses.
19. How can you tell if a pair of sunglasses is polarizing or not?
*20. What would be the color of the sky if the Earth had no atmosphere?
*21. If the Earth's atmosphere were 50 times denser than it is, would sunlight still be white, or would it be some other color?

MisConceptual Questions

1. When light passes through a narrow slit, it spreads out rather than traveling in a straight line. The reason for the spreading is
   (*a*) reflections from the edges of the slit.
   (*b*) refraction by fluctuations in the air density within the slit.
   (*c*) because light is a wave, and this is a basic wave property.
2. Light passing through a double-slit arrangement is viewed on a distant screen. The interference pattern observed on the screen would have the widest spaced fringes for the case of
   (*a*) red light and a small slit spacing.
   (*b*) blue light and a small slit spacing.
   (*c*) red light and a large slit spacing.
   (*d*) blue light and a large slit spacing.
3. In a double-slit experiment (Figs. 34–9 and 10), when the wavelength of the light is increased, the interference pattern
   (*a*) spreads out.
   (*b*) stays the same.
   (*c*) shrinks together.
   (*d*) disappears.
4. If the slits of the previous Question are moved farther apart (without changing the wavelength), the interference pattern
   (*a*) spreads out.
   (*b*) stays the same.
   (*c*) shrinks together.
   (*d*) disappears.
5. Destructive interference occurs when you look at monochromatic light reflecting from a very thin sheet of glass $(n = 1.5)$ surrounded by air. If the air on one side of the glass is replaced with water, then
   (*a*) constructive interference will occur.
   (*b*) destructive interference will still occur.
   (*c*) either constructive or destructive interference could occur, depending on the wavelength of the light.
6. Reflected light from a thin film of oil on glass shows destructive interference for a certain wavelength of visible light. If the thickness of the oil $(n_{oil} < n_{glass})$ is increased by 50%, that wavelength of light will experience
   (*a*) constructive interference.
   (*b*) destructive interference.
   (*c*) about 50% destructive (and therefore 50% constructive) interference.
   (*d*) Not enough information; we need to know the wavelength of light and the indices of refraction.
7. What is the difference in the interference pattern you would see with light reflecting from a thin soap film $(n = 1.42)$ in air versus the same film on water $(n = 1.33)$?
   (*a*) There would be no difference.
   (*b*) There would be a shift due to the phase change upon reflection at the soap-water interface.
   (*c*) There would be a shift due to the absence of a phase change upon reflection at the soap-water interface.
   (*d*) There would be a shift due to the phase change upon reflection at the air-soap interface.

8. Light strikes perpendicular to a thin film in air. Considering the wavelength of light inside the film, then if the film has a thickness that is
(*a*) $\frac{1}{4}$ of a wavelength, constructive interference will always occur.
(*b*) $\frac{1}{4}$ of a wavelength, destructive interference will always occur.
(*c*) $\frac{1}{2}$ of a wavelength, constructive interference will always occur.
(*d*) $\frac{1}{2}$ of a wavelength, destructive interference will always occur.
(*e*) None of the above is always true.

9. Suppose you have a detector between two identical, coherent red lasers (λ) on either side pointing at each other. The detector midway between detects total destructive interference. To detect total constructive interference, you need to move the detector
(*a*) a distance λ toward one of the lasers.
(*b*) a distance $\frac{1}{2}\lambda$ toward one of the lasers.
(*c*) a distance $\frac{1}{4}\lambda$ toward one of the lasers.
(*d*) None of the above.

10. The two waves shown in Fig. 34–42 are
(*a*) out of phase by 180°.
(*b*) out of phase by 90°.
(*c*) out of phase by 45°.
(*d*) out of phase by 360°.
(*e*) in phase.

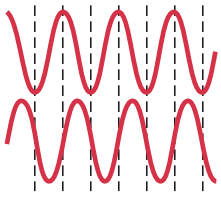


FIGURE 34–42
MisConceptual Question 10.

11. Reflection from a thin film (index of refraction n) in air shows destructive interference. This means the thickness of the film could be
(*a*) $\frac{\lambda}{4n}$.
(*b*) $\frac{\lambda}{2n}$.
(*c*) $\frac{\lambda}{n}$.
(*d*) Not enough information.

12. If unpolarized light is incident from the left on three polarizers as shown in Fig. 34–43, in which case will some light get through?
(*a*) Case 1 only.
(*b*) Case 2 only.
(*c*) Case 3 only.
(*d*) Case 1 and 3.
(*e*) All three cases.

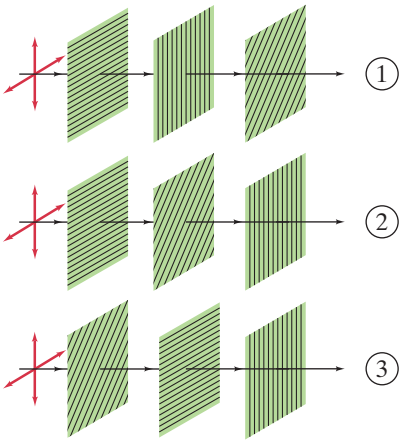


FIGURE 34–43
MisConceptual Question 12.

Problems

34–2 Huygens' Principle

1. (II) Derive the law of reflection—namely, that the angle of incidence equals the angle of reflection from a flat surface—using Huygens' principle for waves.

34–3 Double-Slit Interference

2. (I) Monochromatic light falling on two slits 0.018 mm apart produces the fifth-order bright fringe at a 12° angle. What is the wavelength of the light used?

3. (I) The third-order bright fringe of 610-nm light is observed at an angle of 25° when the light falls on two narrow slits. How far apart are the slits?

4. (II) Monochromatic light falls on two very narrow slits 0.048 mm apart. Successive fringes on a screen 6.50 m away are 8.5 cm apart near the center of the pattern. Determine the wavelength and frequency of the light.

5. (II) Assume that light of a single color, rather than white light, passes through the two-slit setup described in Example 34–4. If the distance from the central fringe to a first-order fringe is measured to be 2.9 mm on the screen, determine the light's wavelength (in nm) and color (see Fig. 34–11).

6. (II) A red laser from the physics lab is marked as producing 632.8-nm light. When light from this laser falls on two closely spaced slits, an interference pattern formed on a wall several meters away has bright fringes spaced 6.00 mm apart near the center of the pattern. When the laser is replaced by a small laser pointer, the fringes are 5.04 mm apart. What is the wavelength of light produced by the pointer?

7. (II) If 720-nm and 660-nm light passes through two slits 0.61 mm apart, how far apart are the second-order fringes for these two wavelengths on a screen 1.0 m away?

8. (II) Water waves having parallel crests 4.5 cm apart pass through two openings 7.5 cm apart in a board. At a point 3.0 m beyond the board, at what angle relative to the "straight-through" direction would there be little or no wave action?

9. (II) Light of wavelength 680 nm falls on two slits and produces an interference pattern in which the third-order bright red fringe is 38 mm from the central fringe on a screen 2.2 m away. What is the separation of the two slits?

10. (II) A parallel beam of light from a He–Ne laser, with a wavelength 633 nm, falls on two very narrow slits 0.068 mm apart. How far apart are the fringes in the center of the pattern on a screen 4.2 m away?

11. (II) A physics professor wants to perform a lecture demonstration of Young's double-slit experiment for her class using the 633-nm light from a He–Ne laser. Because the lecture hall is very large, the interference pattern will be projected on a wall that is 4.6 m from the slits. For easy viewing by all students in the class, the professor wants the distance between the $m = 0$ and $m = 1$ maxima to be 25 cm. What slit separation is required in order to produce the desired interference pattern?

12. (II) Light of wavelength λ passes through a pair of slits separated by 0.17 mm, forming a double-slit interference pattern on a screen located a distance 35 cm away. Suppose that the image in Fig. 34–9a is an actual-size reproduction of this interference pattern. Use a ruler to measure a pertinent distance on this image; then utilize this measured value to determine λ (nm).

13. (II) Suppose a thin piece of glass is placed in front of the lower slit in Fig. 34–7 so that the two waves enter the slits 180° out of phase (Fig. 34–44). Describe in detail the interference pattern on the screen.

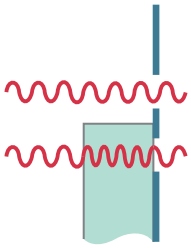


FIGURE 34–44
Problem 13.

14. (II) A very thin sheet of plastic $(n = 1.60)$ covers one slit of a double-slit apparatus illuminated by 640-nm light. The center point on the screen, instead of being a maximum, is dark. What is the (minimum) thickness of the plastic?

15. (II) Two narrow slits separated by 1.4 mm are illuminated by 544-nm light. Find the distance between adjacent bright fringes on a screen 5.0 m from the slits.

16. (II) In a double-slit experiment, the third-order maximum for light of wavelength 500 nm is located 12 mm from the central bright spot on a screen 1.3 m from the slits. Light of wavelength 650 nm is then projected through the same slits. How far from the central bright spot will the second-order maximum of this light be located?

17. (II) Light of wavelength 474 nm in air shines on two slits 6.00×10^{-2} mm apart. The slits are immersed in water, as is a viewing screen 60.0 cm away. How far apart are the fringes on the screen?

18. (II) In a double-slit experiment it is found that blue light of wavelength 480 nm gives a second-order maximum at a certain location on the screen. What wavelength of visible light would have a minimum at the same location?

*34–4 Intensity in Two-Slit Interference

*19. (I) If one slit in Fig. 34–12 is covered, by what factor does the intensity at the center of the screen change?

*20. (II) In a two-slit interference experiment, the path length to a certain point P on the screen differs for one slit in comparison with the other by 1.25λ. (*a*) What is the phase difference between the two waves arriving at point P? (*b*) Determine the intensity at P, expressed as a fraction of the maximum intensity I_0 on the screen.

*21. (II) Show that the angular full width at half maximum of the central peak in a double-slit interference pattern is given by $\Delta\theta = \lambda/2d$ if $\lambda \ll d$.

*22. (II) Derive an expression similar to Eq. 34–2 which gives the angles for which the double-slit intensity is one-half its maximum value, $I_\theta = \frac{1}{2} I_0$.

*23. (III) Suppose that one slit of a double-slit apparatus is wider than the other so that the intensity of light passing through it is twice as great. Determine the intensity I as a function of position (θ) on the screen for coherent light.

*24. (III) (*a*) Consider three equally spaced and equal-intensity coherent sources of light (such as adding a third slit to the two slits of Fig. 34–12). Use the phasor method to obtain the intensity as a function of the phase difference δ (Eq. 34–4). (*b*) Determine the positions of maxima and minima.

34–5 Thin-Film Interference

25. (I) If a soap bubble is 120 nm thick, what wavelength is most strongly reflected at the center of the outer surface when illuminated normally by white light? Assume that $n = 1.35$.

26. (I) How far apart are the dark fringes in Example 34–6 if the glass plates are each 26.3 cm long?

27. (II) (*a*) What is the smallest thickness of a soap film $(n = 1.33)$ that would appear black if illuminated with 480-nm light? Assume there is air on both sides of the soap film. (*b*) What are two other possible thicknesses for the film to appear black? (*c*) If the thickness t was much less than λ, why would the film also appear black?

28. (II) A thin oil slick $(n_o = 1.50)$ floats on water $(n_w = 1.33)$. When a beam of white light strikes this film at normal incidence from air, the only enhanced reflected colors are red (653 nm) and violet (392 nm). From this information, deduce the (minimum) thickness t of the oil slick.

29. (II) A thin film of oil $(n_o = 1.50)$ with varying thickness floats on water $(n_w = 1.33)$. When it is illuminated from above by white light, the reflected colors are as shown in Fig. 34–45. In air, the wavelength of yellow light is 580 nm. (*a*) Why are there no reflected colors at point A? (*b*) What is the oil's thickness t at point B?

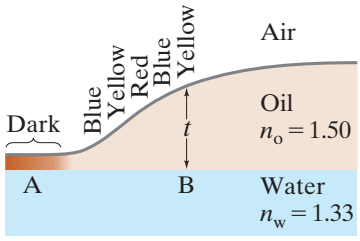


FIGURE 34–45
Problem 29.

30. (II) A lens appears greenish yellow ($\lambda = 570$ nm is strongest) when white light reflects from it. What minimum thickness of coating $(n = 1.25)$ do you think is used on such a glass lens $(n = 1.56)$, and why?

31. (II) A total of 35 bright and 35 dark Newton's rings (not counting the dark spot at the center) are observed when 560-nm light falls normally on a planoconvex lens resting on a flat glass surface (Fig. 34–18). How much thicker is the center than the edges?

32. (II) A fine metal foil separates one end of two pieces of optically flat glass, as in Fig. 34–20. When light of wavelength 670 nm is incident normally, 24 dark bands are observed (with one at each end). How thick is the foil?

33. (II) (*a*) How thick (minimum) should the air layer be between two flat glass surfaces if the reflected light is to appear bright when 480-nm light is incident normally? (*b*) What if the glass is to appear dark?

34. (II) A uniform thin film of alcohol $(n = 1.36)$ lies on a flat glass plate $(n = 1.56)$. When monochromatic light, whose wavelength can be changed, is incident normally, the reflected light is a minimum for $\lambda = 492$ nm and a maximum for $\lambda = 615$ nm. What is the minimum thickness of the film?

35. (II) If the wedge between the glass plates of Example 34–6 is filled with some transparent substance other than air—say, water—the pattern shifts because the wavelength of the light changes. In a material where the index of refraction is n, the wavelength is $\lambda_n = \lambda/n$, where λ is the wavelength in vacuum (Eq. 34–1). How many dark bands would there be if the wedge of Example 34–6 is filled with water?

36. (II) Let's explore why only "thin" layers exhibit thin-film interference. Assume a layer of water, sitting atop a flat glass surface, is illuminated from the air above by white light (all wavelengths from 400 nm to 700 nm). Further, assume that the water layer's thickness t is much greater than a micron ($= 1000$ nm); in particular, let $t = 200\ \mu$m. Take the index of refraction for water to be $n = 1.33$ for all visible wavelengths and the index for glass to be $n_g = 1.50$. (*a*) Show that a visible color will be reflected from the water layer if its wavelength is $\lambda = 2nt/m$, where m is an integer. (*b*) Show that the two extremes in wavelengths (400 nm and 700 nm) of the incident light are both reflected from the water layer and determine the m-value associated with each. (*c*) How many other visible wavelengths, besides $\lambda = 400$ nm and 700 nm, are reflected from the "thick" layer of water? (*d*) How does this explain why such a thick layer does not reflect colorfully, but is white or grey?

37. (II) Show that the radius r of the m^{th} dark Newton's ring, as viewed from directly above (Fig. 34–18), is given by $r = \sqrt{m\lambda R}$ where R is the radius of curvature of the curved glass surface and λ is the wavelength of light used. Assume that the thickness of the air gap is much less than R at all points and that $r \ll R$. [*Hint*: Use the binomial expansion.]

38. (II) Use the result of Problem 37 to show that the distance between adjacent dark Newton's rings is

$$\Delta r \approx \sqrt{\frac{\lambda R}{4m}}$$

for the m^{th} ring, assuming $m \gg 1$.

39. (II) When a Newton's ring apparatus (Fig. 34–18) is immersed in a liquid, the diameter of the tenth dark ring decreases from 2.92 cm to 2.54 cm. What is the refractive index of the liquid? [*Hint*: See Problem 37.]

40. (II) A planoconvex Lucite lens 3.4 cm in diameter is placed on a flat piece of glass as in Fig. 34–18. When 580-nm light is incident normally, 41 bright rings are observed, the last one right at the edge. What is the radius of curvature of the lens surface, and the focal length of the lens? [*Hint*: See Problem 37.]

41. (II) How many uncoated thin lenses in an optical instrument would reduce the amount of light passing through the instrument to 50% or less? (Assume the same transmission percent at each of the two surfaces—see page 1008.)

42. (III) A single optical coating reduces reflection to zero for $\lambda = 550$ nm. By what factor is the intensity reduced by the coating for $\lambda = 430$ nm and $\lambda = 670$ nm as compared to no coating? Assume normal incidence.

34–6 Michelson Interferometer

43. (II) How far must the mirror M_1 in a Michelson interferometer be moved if 730 fringes of 589-nm light are to pass by a reference line?

44. (II) What is the wavelength of the light entering an interferometer if 368 bright fringes are counted when the movable mirror moves 0.125 mm?

45. (II) A micrometer is connected to the movable mirror of an interferometer. When the micrometer is tightened down on a thin metal foil, the net number of bright fringes that move, compared to the empty micrometer, is 272. What is the thickness of the foil? The wavelength of light used is 611 nm.

46. (III) One of the beams of an interferometer (Fig. 34–46) passes through a small evacuated glass container 1.155 cm deep. When a gas is allowed to slowly fill the container, a total of 163 dark fringes are counted to move past a reference line. The light used has a wavelength of 632.8 nm. Calculate the index of refraction of the gas at its final density, assuming that the interferometer is in vacuum.

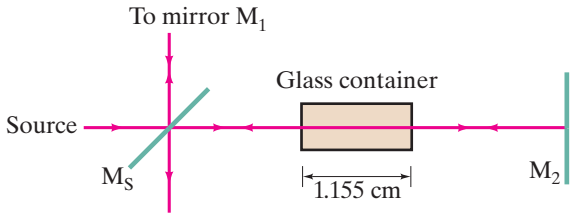


FIGURE 34–46 Problem 46.

47. (III) The yellow sodium D lines have wavelengths of 589.0 and 589.6 nm. When they are used to illuminate a Michelson interferometer, it is noted that the interference fringes disappear and reappear periodically as the mirror M_1 is moved. Why does this happen? How far must the mirror move between one disappearance and the next?

34–7 Polarization

48. (I) Two polarizers are oriented at 75° to one another. Unpolarized light falls on them. What fraction of the light intensity is transmitted?

49. (I) What is Brewster's angle for an air-glass ($n = 1.56$) surface? Specify two answers.

50. (I) What is Brewster's angle for a diamond submerged in water? Specify two answers. Will the diamond sparkle as nicely as in air? (See Section 32–9.)

51. (II) Two polarizers are oriented at 44.0° to one another. Light polarized at a 22.0° angle to each polarizer passes through both. What is the transmitted intensity (%)?

52. (II) At what angle should the axes of two Polaroids be placed so as to reduce the intensity of the incident unpolarized light to (*a*) $\frac{1}{3}$, (*b*) $\frac{1}{10}$?

53. (II) The critical angle for total internal reflection at a boundary between two materials is 48°. What is Brewster's angle at this boundary? Give two answers, one for each material.

54. (II) What would Brewster's angle be for reflections off the surface of water for light coming from beneath the surface? Compare to the angle for total internal reflection, and to Brewster's angle from above the surface.

55. (II) Two Polaroids are aligned so that the initially unpolarized light passing through them is a maximum. At what angle should one of them be placed so the transmitted intensity is subsequently reduced by half?

56. (II) A piece of material, suspected of being a stolen diamond ($n = 2.42$), is submerged in oil of refractive index 1.43 and illuminated by unpolarized light. It is found that the reflected light is completely polarized at an angle of 56°. Is it diamond? Explain.

57. (II) Three perfectly polarizing sheets are spaced 2 cm apart and in parallel planes. The transmission axis of the second sheet is 30° relative to the first one. The transmission axis of the third sheet is 90° relative to the *first* one. Unpolarized light impinges on the first polarizing sheet. What percent of this light is transmitted out through the third polarizer?

58. (II) Unpolarized light of intensity I_0 passes through six successive Polaroid sheets each of whose axis makes a 38° angle with the previous one. What is the intensity of the transmitted beam?

59. (II) Two polarizers A and B are aligned so that their transmission axes are vertical and horizontal, respectively. A third polarizer is placed between these two with its axis aligned at angle θ with respect to the vertical. Assuming vertically polarized light of intensity I_0 is incident upon polarizer A, find an expression for the light intensity I transmitted through this three-polarizer sequence. Calculate the derivative $dI/d\theta$; then use it to find the angle θ that maximizes I.

60. (III) The percent polarization P of a partially polarized beam of light is defined as

$$P = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \times 100$$

where $I_{\max}$ and $I_{\min}$ are the maximum and minimum intensities that are obtained when the light passes through a polarizer that is slowly rotated. Such light can be considered as the sum of two unequal plane-polarized beams of intensities $I_{\max}$ and $I_{\min}$ perpendicular to each other. Show that the light transmitted by a polarizer, whose axis makes an angle ϕ to the direction in which $I_{\max}$ is obtained, has intensity

$$\frac{1 + p\cos 2\phi}{1 + p} I_{\max}$$

where $p = P/100$ is the "fractional polarization."

34–10 Brightness: Lumens and Luminous Intensity

61. (II) The illuminance of direct sunlight on Earth is about $10^5\,\text{lm/m}^2$. Estimate the luminous flux and luminous intensity of the Sun.

62. (II) The *luminous efficiency* of a lightbulb is the ratio of luminous flux to electric power input. (*a*) What is the luminous efficiency of a 100-W, 1600-lm bulb? (*b*) How many 40-W, 60-lm/W fluorescent lamps would be needed to provide an illuminance of $250\,\text{lm/m}^2$ on a factory floor of area $22\,\text{m} \times 30\,\text{m}$? Assume the lights are 10 m above the floor and that half their flux reaches the floor.

General Problems

63. A series of polarizers are each rotated 10° from the previous polarizer. Unpolarized light is incident on this series of polarizers. How many polarizers does the light have to go through before it is $\frac{1}{6}$ of its original intensity?

64. Television and radio waves reflecting from mountains or airplanes can interfere with the direct signal from the station. (*a*) What kind of interference will occur when 75-MHz television signals arrive at a receiver directly from a distant station, and are reflected from a nearby airplane 122 m directly above the receiver? Assume $\frac{1}{2}\lambda$ change in phase of the signal upon reflection. (*b*) What kind of interference will occur if the plane is 22 m closer to the receiver?

65. *Stealth aircraft* are designed to not reflect radar, whose wavelength is typically 2 cm, by using an antireflecting coating. Ignoring any change in wavelength in the coating, estimate its thickness.

66. Calculate the minimum thickness needed for an antireflective coating ($n = 1.42$) applied to a glass lens in order to eliminate (*a*) blue (450 nm), or (*b*) red (720 nm) reflections for light at normal incidence.

67. Light of wavelength $5.0 \times 10^{-7}\,\text{m}$ passes through two parallel slits and falls on a screen 5.0 m away. Adjacent bright bands of the interference pattern are 2.0 cm apart. (*a*) Find the distance between the slits. (*b*) The same two slits are next illuminated by light of a different wavelength, and the fifth minimum for this light occurs at the same point on the screen as the fourth minimum for the previous light. What is the wavelength of the second source of light?

68. A radio station operating at 92.5 MHz broadcasts from two identical antennas at the same elevation but separated by a 9.0-m horizontal distance d, Fig. 34–47. A maximum signal is found along the midline, perpendicular to d at its midpoint and extending horizontally in both directions. If the midline is taken as 0°, at what other angle(s) θ is a maximum signal detected? A minimum signal? Assume all measurements are made much farther than 9.0 m from the antenna towers.

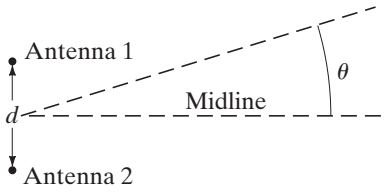


FIGURE 34–47
Problem 68.

69. Light of wavelength 690 nm passes through two narrow slits 0.66 mm apart. The screen is 1.75 m away. A second source of unknown wavelength produces its second-order fringe 1.23 mm closer to the central maximum than the 690-nm light. What is the wavelength of the unknown light?

70. Monochromatic light of variable wavelength is incident normally on a thin sheet of plastic film in air. The reflected light is a maximum only for $\lambda = 491.4$ nm and $\lambda = 688.0$ nm in the visible spectrum. What is the thickness of the film ($n = 1.62$)? [*Hint*: Assume successive values of m.]

71. A "60-watt equivalent" LED lightbulb is rated at 7.5 W and 800 lumens. How much power goes into light? (The rest goes into heat.)

72. At what angle above the horizon is the Sun when light reflecting off a smooth lake is polarized most strongly?

73. Light from a green laser of wavelength 530 nm passes through two slits that are 35 microns apart. The resulting pattern formed on a screen in front of the slits is shown in Fig. 34–48. If point A is the same distance from both slits, how much closer is point B to one slit than to the other?

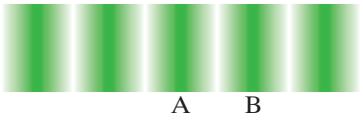


FIGURE 34–48 Problem 73.

74. Suppose the mirrors in a Michelson interferometer are perfectly aligned and the path lengths to mirrors M_1 and M_2 are identical. With these initial conditions, an observer sees a bright maximum at the center of the viewing area. Now one of the mirrors is moved a distance x. Determine a formula for the intensity at the center of the viewing area as a function of x, the distance the movable mirror is moved from the initial position.

75. A highly reflective mirror can be made for a particular wavelength at normal incidence by using two thin layers of transparent materials of indices of refraction n_1 and n_2 $(1 < n_1 < n_2)$ on the surface of the glass $(n > n_2)$. What should be the minimum thicknesses d_1 and d_2 in Fig. 34–49 in terms of the incident wavelength λ, to maximize reflection?

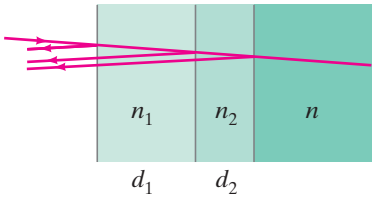


FIGURE 34–49 Problem 75.

76. What is the index of refraction of a clear material if a minimum thickness of 125 nm, when laid on glass, is needed to reduce reflection to nearly zero when light of 655 nm is incident normally upon it? Do you have a choice for an answer?

77. Light of wavelength λ strikes a screen containing two slits a distance d apart at an angle θ_i to the normal. Determine the angle θ_m at which the m^{th}-order maximum occurs.

78. Consider two antennas radiating 6.8-MHz radio waves in phase with each other. They are located at points S_1 and S_2, separated by a distance $d = 175$ m, Fig. 34–50. Determine the points on the positive y axis where the signals from the two sources will be out of phase (crests of one meet troughs of the other).

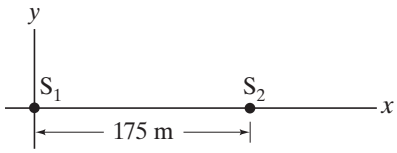


FIGURE 34–50 Problem 78.

79. What is the minimum (non-zero) thickness for the air layer between two flat glass surfaces if the glass is to appear dark when 650-nm light is incident normally? What if the glass is to appear bright?

80. Suppose you viewed the light *transmitted* through a thin coating layered on a flat piece of glass. Draw a diagram, similar to Fig. 34–17 or 34–23, and describe the conditions required for maxima and minima. Consider all possible values of index of refraction. Discuss the relative intensity of the minima compared to the maxima and to zero.

81. Consider the antennas of Example 34–5, Fig. 34–15. Let $d = \lambda/2$, and suppose that the two antennas are now 180° out of phase with each other. Find the directions for constructive and destructive interference, and compare with the case when the sources are in phase. (These results illustrate the basis for directional antennas.)

82. *Lloyd's mirror* provides one way of obtaining a double-slit interference pattern from a single source so the light is coherent. As shown in Fig. 34–51, the light that reflects from the plane mirror appears to come from the virtual image of the slit. Describe in detail the interference pattern on the screen.

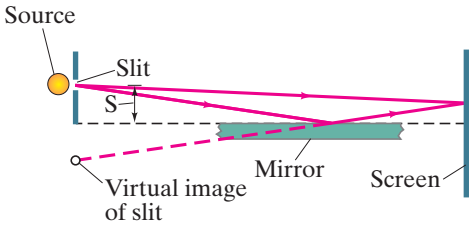


FIGURE 34–51 Problem 82.

83. A thin film of soap $(n = 1.34)$ coats a piece of flat glass $(n = 1.52)$. What is the minimum film thickness if it reflects 643-nm red light most strongly when illuminated normally by white light?

84. Two narrow slits 0.070 mm apart are illuminated by a very bright 488-nm light source forming an interference pattern on a screen 4.0 m away. Calculate (*a*) the distance between the $m = 0$ and $m = 1$ lines in the pattern and (*b*) the distance between the $m = 100$ and $m = 101$ lines.

85. Unpolarized light falls on two polarizer sheets whose axes are at right angles. (*a*) What fraction of the incident light intensity is transmitted? (*b*) What fraction is transmitted if a third polarizer is placed between the first two so that its axis makes a 58° angle with the axis of the first polarizer? (*c*) What if the third polarizer is in front of the other two?

86. Four polarizers are placed in succession with their axes vertical, at 30.0° to the vertical, at 60.0° to the vertical, and at 90.0° to the vertical. (*a*) Calculate what fraction of the incident unpolarized light is transmitted by the four polarizers. (*b*) Can the transmitted light be *decreased* by removing one of the polarizers? If so, which one? (*c*) Can the transmitted light intensity be extinguished by removing polarizers? If so, which one(s)?

87. A two-slit interference set-up with slit separation $d = 0.10$ mm produces interference fringes at a particular set of angles θ_m (where $m = 0, 1, 2, \ldots$) for red light of frequency $f = 4.6 \times 10^{14}$ Hz. If one wishes to construct an analogous two-slit interference set-up that produces interference fringes at the same set of angles θ_m for room-temperature sound of middle-C frequency $f_S = 262$ Hz, what should the slit separation d_S be for this analogous set-up?

88. Two polarizers are oriented at 55° to each other and plane-polarized light is incident on them. If only 25% of the light gets through both of them, what was the initial polarization direction of the incident light?

89. At what angle should the axes of two Polaroids be placed so as to reduce the intensity of the incident unpolarized light by an additional factor (after the first Polaroid cuts it in half) of (*a*) 4, (*b*) 10, (*c*) 100?

90. Two identical sources S_1 and S_2, separated by distance d, coherently emit light of wavelength λ uniformly in all directions. Defining the x axis with its origin at S_1 as shown in Fig. 34–52, find the locations (expressed as multiples of λ) where the signals from the two sources are out of phase along this axis for $x > 0$, if $d = 3\lambda$.



FIGURE 34–52
Problem 90.

91. In a compact disc (CD), digital information is stored as a sequence of raised surfaces called "pits" and recessed surfaces called "lands." Both pits and lands are highly reflective and are embedded in a thick plastic material with index of refraction $n = 1.55$ (Fig. 34–53). As a 780-nm wavelength (in air) laser scans across the pit–land sequence, the transition between a neighboring pit and land is sensed by monitoring the intensity of reflected laser light from the CD. At the moment when half the width of the laser beam is reflected from the pit and the other half from the land, we want the two reflected halves of the beam to be 180° out of phase with each other. What should be the (minimum) height difference t between a pit and land? [When this light enters a detector, cancellation of the two out-of-phase halves of the beam produces a minimum detector output.]

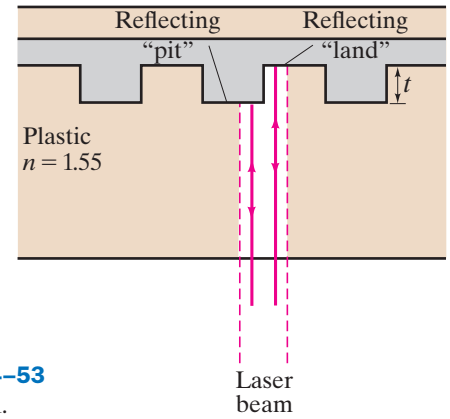


FIGURE 34–53
Problem 91.

92. Describe how to rotate the plane of polarization of a plane-polarized beam of light by 90° and produce only a 10% loss in intensity, using polarizers. Let N be the number of polarizers and θ be the (same) angle between successive polarizers.

93. What percent of visible light is reflected from plain glass? Assume your answer refers to transmission through each surface, front and back. How does the presence of multiple lenses in a good camera degrade the image? What is suggested in Section 34–5 to reduce this reflection? Explain in words, and sketch how this solution works. For a 6-element glass lens in air, about how much improvement does this solution provide?

94. A radio telescope, whose two antennas are separated by 55 m, is designed to receive 3.0-MHz radio waves produced by astronomical objects. The received radio waves create 3.0-MHz electronic signals in the telescope's left and right antennas. These signals then travel by equal-length cables to a centrally located amplifier, where they are added together. The telescope can be "pointed" to a certain region of the sky by adding the instantaneous signal from the right antenna to a "time-delayed" signal received by the left antenna a time Δt ago. (This time delay of the left signal can be easily accomplished with the proper electronic circuit.) If a radio astronomer wishes to "view" radio signals arriving from an object oriented at a 12° angle to the vertical as in Fig. 34–54, what time delay Δt is necessary?

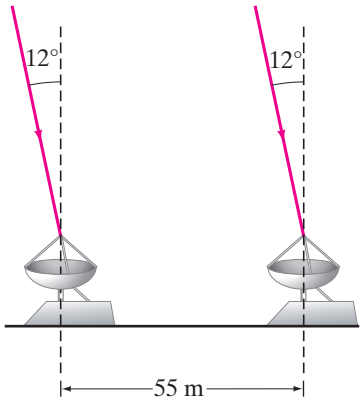
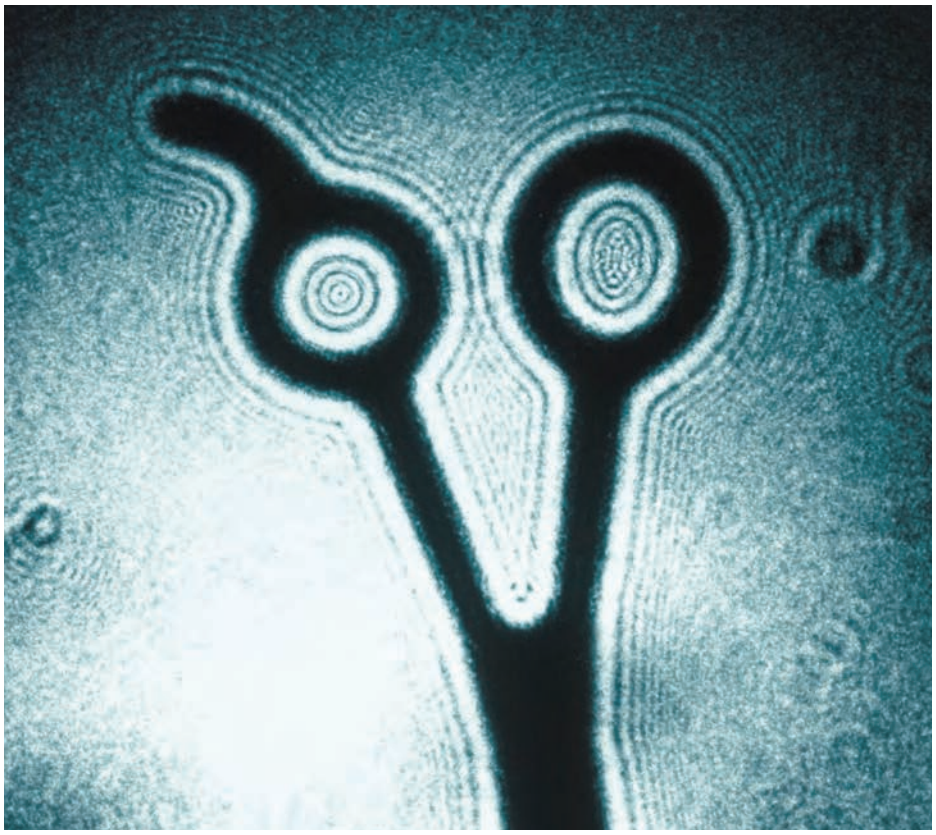


FIGURE 34–54
Problem 94.

ANSWERS TO EXERCISES

A: 333 nm; 6.0×10^{14} Hz; 2.0×10^8 m/s.
B: 2.5 mm.
C: (*a*) I_0; (*b*) $0.50I_0$; (*c*) 0.
D: (*e*).
E: Zero for both (*a*) and (*b*), because the two successive polarizers at 90° cancel all light. The 45° Polaroid must be inserted *between* the other two if any transmission is to occur.



Parallel coherent light from a laser, which acts as nearly a point source, illuminates these shears. Instead of a clean shadow, there is a dramatic diffraction pattern, which is a strong confirmation of the wave theory of light. Diffraction patterns are washed out when typical extended sources of light are used, and hence are not seen, although a careful examination of shadows will reveal fuzziness. We will examine diffraction by a single slit, and how it affects the double-slit pattern. We also discuss diffraction gratings and diffraction of X-rays by crystals. We will see how diffraction affects the resolution of optical instruments, and that the ultimate resolution can never be greater than the wavelength of the radiation used.

CHAPTER 35

Diffraction

CHAPTER-OPENING QUESTION—Guess now!

A microscope using visible light has a useful magnification of about

(a) 50×;
(b) 100×;
(c) 500×;
(d) 2000×;
(e) 5000×;

and the smallest objects it can resolve have a size of about

(a) 10 nm;
(b) 100 nm;
(c) 500 nm;
(d) 2500 nm;
(e) 5500 nm.

Young's double-slit experiment put the wave theory of light on a firm footing. But full acceptance came only with studies on diffraction more than a decade later, in the 1810s and 1820s.

We have already discussed diffraction briefly with regard to water waves (Section 15–11) as well as for light (Section 34–1). We have seen that it refers to the spreading or bending of waves around edges. Now we look at diffraction in more detail, including its important practical effects of limiting the amount of detail, or *resolution*, that can be obtained with any optical instrument such as telescopes, cameras, and the eye.

CONTENTS

35 – 1 Diffraction by a Single Slit or Disk
*35 – 2 Intensity in Single-Slit Diffraction Pattern
*35 – 3 Diffraction in the Double-Slit Experiment
35 – 4 Interference vs. Diffraction
35 – 5 Limits of Resolution; Circular Apertures
35 – 6 Resolution of Telescopes and Microscopes; the λ Limit
35 – 7 Resolution of the Human Eye and Useful Magnification
35 – 8 Diffraction Grating
35 – 9 The Spectrometer and Spectroscopy
*35–10 Peak Widths and Resolving Power for a Diffraction Grating
35–11 X-Rays and X-Ray Diffraction
*35–12 X-Ray Imaging and Computed Tomography (CT Scan)
*35–13 Specialty Microscopes and Contrast

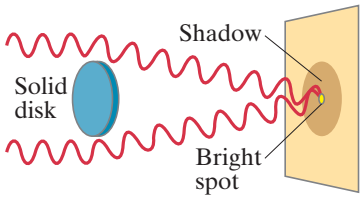
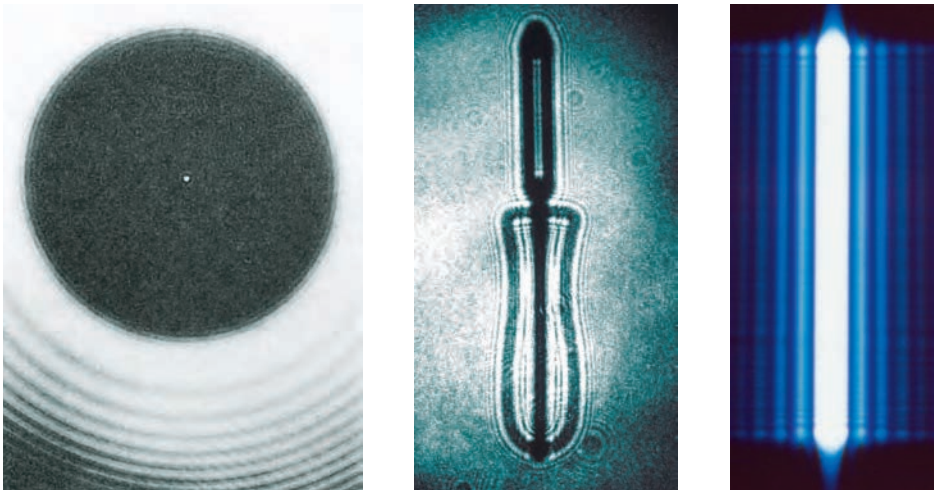


FIGURE 35–1 If light is a wave, a bright spot will appear at the center of the shadow of a solid disk illuminated by a point source of monochromatic light.

In 1819, Augustin Fresnel (1788–1827) presented to the French Academy a wave theory of light that predicted and explained interference and diffraction effects. Almost immediately Siméon Poisson (1781–1840) pointed out a counter-intuitive inference: according to Fresnel's wave theory, if light from a point source were to fall on a solid disk, part of the incident light would be diffracted around the edges and would constructively interfere at the center of the shadow (Fig. 35–1), producing a bright spot. That prediction seemed very unlikely. But when the experiment was actually carried out by François Arago, the bright spot was seen at the very center of the shadow (Fig. 35–2a). This was strong evidence for the wave theory.

Figure 35–2a is a photograph of the shadow cast by a coin using a coherent point source of light, a laser in this case. The bright spot is clearly present at the center. Note also the bright and dark fringes beyond the shadow. These resemble the interference fringes of a double slit. Indeed, they are due to interference of waves diffracted around the outer edge of the disk, and the group of fringes is referred to as a **diffraction pattern**. A diffraction pattern exists around any sharp-edged object illuminated by a point source, as shown in Fig. 35–2b and c. We are not always aware of diffraction because most sources of light in everyday life are not points (or coherent), so light from different parts of the source washes out the pattern.



(a) (b) (c)

FIGURE 35–2 Diffraction pattern of (a) a circular disk (a coin), (b) a potato peeler, (c) a single slit, each illuminated by a nearly point source of coherent monochromatic light, such as a laser.

35–1 Diffraction by a Single Slit or Disk

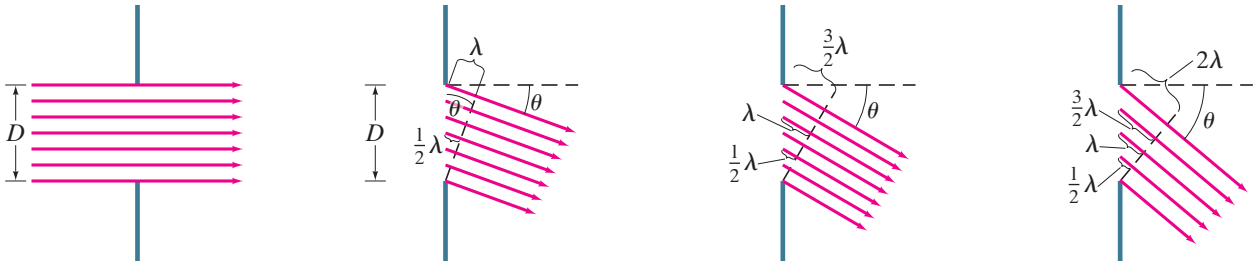
To see how a diffraction pattern arises, we will analyze the important case of monochromatic light passing through a narrow slit, as in Fig. 35–2c. We assume that parallel rays (plane waves) of light pass through a slit whose narrow width is D, and continue to a viewing screen very far away. If the viewing screen is not far away, lenses can be used to make the rays parallel.[†] As we know from studying water waves and from Huygens' principle, waves passing through the slit spread out in all directions, Fig. 35–3. We will now examine how the waves passing through different parts of the slit interfere with each other.

FIGURE 35–3 Water waves impinging on a small opening of a water break, pass through and spread out to either side. See also Section 15–11.



Parallel rays of monochromatic light pass through the narrow slit as shown in Fig. 35–4a. The slit width D is assumed to be on the order of the wavelength λ of the light, but the slit's length (into and out of page) can be large compared to λ. The light falls on a screen as in the double-slit experiment, page 998; the screen is assumed to be very far away, so the rays heading toward any point are very nearly parallel before they meet at the screen.

[†]Such a diffraction pattern, involving parallel rays, is called *Fraunhofer diffraction*. If the screen is close and no lenses are used, it is called *Fresnel diffraction*. The analysis in the latter case is rather involved, so we consider only the limiting case of Fraunhofer diffraction.



(a) $\theta = 0$
Bright

(b) $\sin\theta = \frac{\lambda}{D}$
Dark

(c) $\sin\theta = \frac{3\lambda}{2D}$
Bright

(d) $\sin\theta = \frac{2\lambda}{D}$
Dark

FIGURE 35–4 Analysis of diffraction pattern formed by light passing through a narrow slit of width D.

First we consider rays that pass straight through as in Fig. 35–4a. They are all in phase, so there will be a central bright rectangle or line on the screen (see Fig. 35–2c). In Fig. 35–4b, we consider rays reaching the screen that move at an angle θ such that the ray from the top of the slit travels exactly one wavelength farther than the ray from the bottom of the slit. The ray passing through the very center of the slit will travel one-half wavelength $(\frac{1}{2}\lambda)$ farther than the ray at the bottom. These two rays will be exactly out of phase with one another and so will destructively interfere when they overlap at the screen. Similarly, a ray slightly above the bottom one will cancel a ray that is the same distance above the central one. Indeed, each ray passing through the lower half of the slit will cancel with a corresponding ray passing through the upper half, a distance $D/2$ above. Thus, all the rays destructively interfere in pairs, and so the light intensity will be zero on the viewing screen at this angle. The angle θ at which this takes place can be seen from Fig. 35–4b to occur when $\lambda = D\sin\theta$, so

$$\sin\theta = \frac{\lambda}{D}. \qquad \text{[first minimum]} \quad \textbf{(35–1)}$$

The light intensity is a maximum at $\theta = 0°$ and decreases to a minimum (intensity = zero) at the angle θ given by Eq. 35–1.

Now consider a larger angle θ such that the top ray travels $\frac{3}{2}\lambda$ farther than the bottom ray, as in Fig. 35–4c. In this case, the rays from the bottom third of the slit will cancel in pairs with those in the middle third because they will be $\lambda/2$ out of phase. However, light from the top third of the slit will still reach the screen, so there will be a bright spot or fringe centered near $\sin\theta \approx 3\lambda/2D$, but it will not be nearly as bright as the central spot at $\theta = 0°$. For an even larger angle θ such that the top ray travels 2λ farther than the bottom ray, Fig. 35–4d, rays from the bottom quarter of the slit will cancel with those in the quarter just above it because the path lengths differ by $\lambda/2$. And the rays through the quarter of the slit just above center will cancel with those through the top quarter. At this angle there will again be a minimum of zero intensity in the diffraction pattern. A plot of the intensity as a function of angle is shown in Fig. 35–5. This corresponds well with the photo of Fig. 35–2c. Notice that minima (zero intensity) occur on both sides of center at

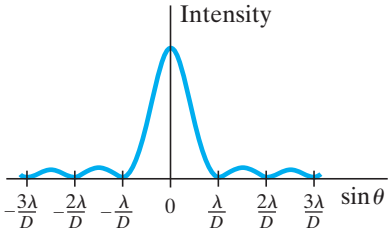
$$D\sin\theta = m\lambda, \qquad m = \pm 1, \pm 2, \pm 3, \cdots, \qquad \text{[minima]} \quad \textbf{(35–2a)}$$

but *not* at $m = 0$ where there is the *strongest maximum* ($\theta = 0$). Between the minima, smaller intensity maxima occur at approximately (not exactly) $m \approx \frac{3}{2}, \frac{5}{2}, \cdots$:

$$D\sin\theta = m\lambda, \qquad m \approx \pm\frac{3}{2}, \pm\frac{5}{2}, \cdots. \qquad \text{[weak maxima]} \quad \textbf{(35–2b)}$$

Note that the *minima* for a diffraction pattern, Eq. 35–2a, satisfy a criterion that looks very similar to that for the *maxima* (bright spots or fringes) for double-slit interference, Eq. 34–2a. Also note that D is a single slit width, whereas d in Eq. 34–2 is the distance between two slits.

FIGURE 35–5 Intensity in the diffraction pattern of a single slit as a function of $\sin\theta$. Note that the central maximum is not only much higher than the maxima to each side, but it is also twice as wide ($2\lambda/D$ wide) as compared to the others (only λ/D wide each).



⚠ CAUTION

Don't confuse Eqs. 35–2 for diffraction with Eqs. 34–2 for 2-slit interference: note the differences

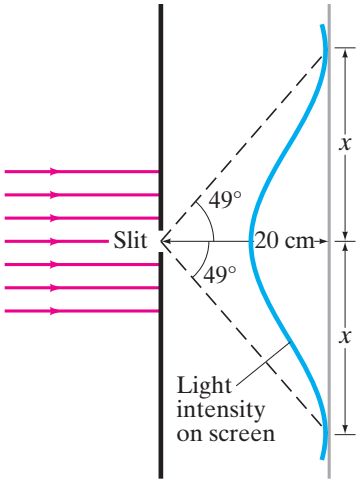


FIGURE 35–6 Example 35–1.

EXAMPLE 35–1 **Single-slit diffraction maximum.** Light of wavelength 750 nm passes through a slit 1.00×10^{-3} mm wide. How wide is the central maximum (*a*) in degrees, and (*b*) in centimeters on a screen 20 cm away?

APPROACH The width of the central maximum goes from the first minimum on one side to the first minimum on the other side. We use Eq. 35–1 to find the angular position of the first single-slit diffraction minimum.

SOLUTION (*a*) The first minimum occurs at

$$\sin\theta = \frac{\lambda}{D} = \frac{7.50 \times 10^{-7}\,\text{m}}{1.00 \times 10^{-6}\,\text{m}} = 0.75.$$

So $\theta = 49°$. This is the angle between the center and the first minimum, Fig. 35–6. The angle subtended by the whole central maximum, between the minima above and below the center, is twice this, or 98°.

(*b*) The width of the central maximum is $2x$, where $\tan\theta = x/20\,\text{cm}$. So $2x = 2(20\,\text{cm})(\tan 49°) = 46\,\text{cm}$.

NOTE A large width of the screen will be illuminated, but it will not normally be very bright since the amount of light that passes through such a small slit will be small and it is spread over a large area. Note also that we *cannot* use the small-angle approximation here ($\theta \approx \sin\theta \approx \tan\theta$) because θ is large.

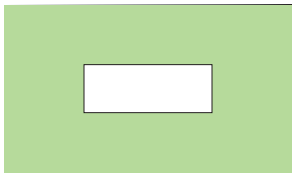
EXERCISE A In Example 35–1, red light ($\lambda = 750\,\text{nm}$) was used. If instead yellow light ($\lambda = 550\,\text{nm}$) had been used, would the central maximum be wider or narrower?

CONCEPTUAL EXAMPLE 35–2 **Diffraction spreads.** Light shines through a small rectangular slit that is narrower in the vertical direction than the horizontal, Fig. 35–7. (*a*) Would you expect the diffraction pattern to be more spread out in the vertical direction or in the horizontal direction? (*b*) Should a rectangular loudspeaker horn at a stadium be tall and narrow, or wide and flat?

RESPONSE (*a*) From Eq. 35–1 we can see that if we make the slit width D smaller, the pattern spreads out more (θ will be larger in Eq. 35–1). This is consistent with our study of waves in Chapter 15. The diffraction through the rectangular slit will be wider vertically, since the opening is smaller in that direction.

(*b*) For a stadium loudspeaker, the sound pattern desired is one spread out horizontally, so the horn should be tall and narrow (rotate Fig. 35–7 by 90°).

FIGURE 35–7 Example 35–2.



*35–2 Intensity in Single-Slit Diffraction Pattern

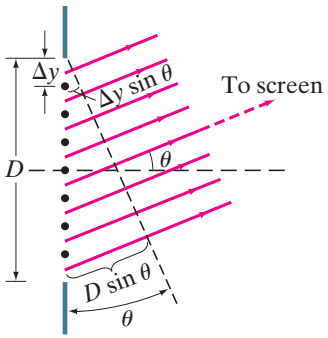
We have determined the positions of the minima in the diffraction pattern produced by light passing through a single slit, Eq. 35–2a. We now discuss a method for predicting the amplitude and intensity at any point in the pattern using the phasor technique already discussed in Section 34–4.

Let us consider the slit divided into N very thin strips of width Δy as indicated in Fig. 35–8. Each strip sends light in all directions toward a screen on the right. Again we take the rays heading for any particular point on the distant screen to be parallel, all making an angle θ with the horizontal as shown. We choose the strip width Δy to be much smaller than the wavelength λ of the monochromatic light falling on the slit, so all the light from a given strip is in phase. The strips are of equal size, and if the whole slit is uniformly illuminated, we can take the electric field wave amplitudes ΔE_0 from each thin strip to be equal as long as θ is not too large. However, the separate amplitudes from the different strips will differ in phase. The phase difference in the light coming from adjacent strips will be (see Section 34–4, Eq. 34–4)

$$\Delta\beta = \frac{2\pi}{\lambda}\Delta y \sin\theta \qquad \textbf{(35–3)}$$

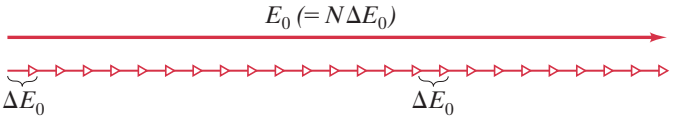
since the difference in path length is $\Delta y \sin\theta$.

FIGURE 35–8 Slit of width D divided into N strips of width Δy.

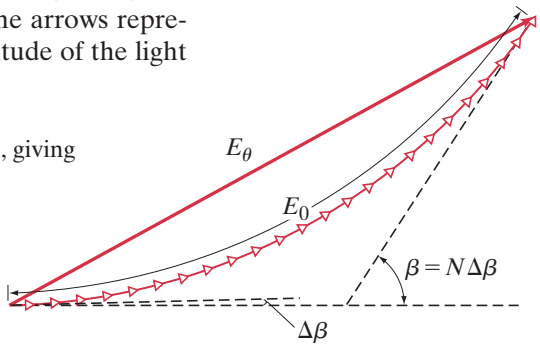


The total amplitude on the screen at any angle θ will be the sum of the separate wave amplitudes due to each strip. These wavelets have the same amplitude ΔE_0 but differ in phase. To obtain the total amplitude, we can use a phasor diagram as we did in Section 34–4 (Fig. 34–13). The phasor diagrams for four different angles θ are shown in Fig. 35–9. At the center of the screen, $\theta = 0$, the waves from each strip are all in phase ($\Delta\beta = 0$, Eq. 35–3), so the arrows representing each ΔE_0 line up as shown in Fig. 35–9a. The total amplitude of the light arriving at the center of the screen is then $E_0 = N\,\Delta E_0$.

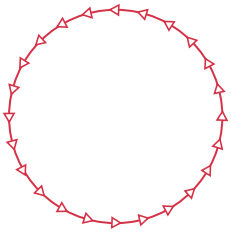
FIGURE 35–9 Phasor diagram for single-slit diffraction, giving the total amplitude E_θ at four different angles θ.



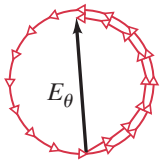
(a) At center, $\theta = 0$.



(b) Between center and first minimum (length of arc of tiny arrows = E_0).



(c) First minimum, $E_\theta = 0$ ($\beta = 2\pi = 360°$).



(d) Near secondary maximum.

At a small angle θ, for a point on the distant screen not far from the center, Fig. 35–9b shows how the wavelets of amplitude ΔE_0 add up to give E_θ, the total amplitude on the screen at this angle θ. Note that each wavelet differs in phase from the adjacent one by $\Delta\beta$. The phase difference between the wavelets from the top and bottom edges of the slit is

$$\beta = N\,\Delta\beta = \frac{2\pi}{\lambda} N\,\Delta y \sin\theta = \frac{2\pi}{\lambda} D \sin\theta \qquad \textbf{(35–4)}$$

where $D = N\,\Delta y$ is the total width of the slit. The "arc" of tiny arrows in Fig. 35–9b has length $N\Delta E_0$, and so equals E_0. But the amplitude of the total wave E_θ at angle θ is the *vector* sum of each wavelet amplitude and so is equal to the length of the straight arrow E_θ. The arrow E_θ is shorter than the arc, so $E_\theta < E_0$.

For greater θ, we eventually come to the case, illustrated in Fig. 35–9c, where the chain of arrows closes on itself. In this case the vector sum is zero, so $E_\theta = 0$ for this angle θ. This corresponds to the first minimum. Since $\beta = N\,\Delta\beta$ is 360° or 2π in this case, we have from Eq. 35–3,

$$2\pi = N\,\Delta\beta = N\left(\frac{2\pi}{\lambda}\,\Delta y \sin\theta\right)$$

or, since the slit width $D = N\,\Delta y$,

$$\sin\theta = \frac{\lambda}{D}.$$

Thus the first minimum $(E_\theta = 0)$ occurs where $\sin\theta = \lambda/D$, which is the same result we obtained in the previous Section, Eq. 35–1.

For even greater values of θ, the chain of arrows spirals beyond 360°. Figure 35–9d shows the case near the secondary maximum next to the first minimum. Here $\beta = N\,\Delta\beta \approx 360° + 180° = 540°$ or 3π. When greater angles θ are considered, new maxima and minima occur. But since the total length of the coil remains constant, equal to $N\,\Delta E_0\,(= E_0)$, each succeeding maximum is smaller and smaller as the coil winds in on itself.

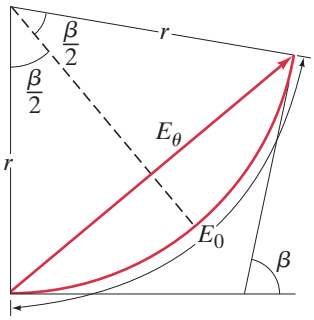


FIGURE 35–10 Determining amplitude E_θ as a function of θ for single-slit diffraction.

To obtain a quantitative expression for the amplitude (and intensity) for any point on the screen (that is, for any angle θ), we now consider the limit $N \to \infty$ so Δy becomes the infinitesimal width dy. In this case, the diagrams of Fig. 35–9 become smooth curves, one of which is shown in Fig. 35–10. For any angle θ, the wave amplitude on the screen is E_θ, equal to the length E_θ of the straight arrow in Fig. 35–10. The length of the arc is E_0, as before. If r is the radius of curvature of the arc, then

$$\frac{E_\theta}{2} = r \sin\frac{\beta}{2}.$$

Using radian measure for $\beta/2$, we also have

$$\frac{E_0}{2} = r\frac{\beta}{2}.$$

We combine these to obtain

$$E_\theta = E_0 \frac{\sin\beta/2}{\beta/2}. \qquad \textbf{(35–5)}$$

The angle β is the phase difference between the waves from the top and bottom edges of the slit. The path difference for these two rays is $D\sin\theta$ (Fig. 35–8 and Eq. 35–4), so

$$\beta = \frac{2\pi}{\lambda} D \sin\theta. \qquad \textbf{(35–6)}$$

Intensity is proportional to the square of the wave amplitude, so the intensity I_θ at any angle θ is, from Eq. 35–5,

$$I_\theta = I_0\left(\frac{\sin\beta/2}{\beta/2}\right)^2 \qquad \textbf{(35–7)}$$

where $I_0\ (\propto E_0^2)$ is the intensity at $\theta = 0$ (the central maximum). We can combine Eqs. 35–7 and 35–6 (although it is often simpler to leave them as separate equations) to obtain

$$I_\theta = I_0\left[\frac{\sin\left(\frac{\pi D\sin\theta}{\lambda}\right)}{\left(\frac{\pi D\sin\theta}{\lambda}\right)}\right]^2. \qquad \textbf{(35–8)}$$

According to Eq. 35–8, minima $(I_\theta = 0)$ occur where $\sin(\pi D\sin\theta/\lambda) = 0$, which means $\pi D\sin\theta/\lambda$ must be $\pm\pi$, $\pm 2\pi$, $\pm 3\pi$, and so on, or

$$D\sin\theta = m\lambda, \qquad m = \pm 1, \pm 2, \pm 3, \cdots \qquad \text{[minima]}$$

which is what we have obtained previously, Eq. 35–2a. Notice that m cannot be zero: when $\beta/2 = \pi D\sin\theta/\lambda = 0$, the denominator as well as the numerator in Eqs. 35–7 or 35–8 equals zero. We can evaluate the intensity in this case by taking the limit as $\theta \to 0$ (or $\beta \to 0$); for very small angles, $\sin\beta/2 \approx \beta/2$, so $(\sin\beta/2)/(\beta/2) \to 1$ and $I_\theta = I_0$, the *maximum* at the center of the pattern.

The intensity I_θ as a function of θ, as given by Eq. 35–8, corresponds to the intensity pattern shown in Fig. 35–5.

EXAMPLE 35–3 **ESTIMATE** **Intensity at secondary maxima.** Estimate the intensities of the first two secondary maxima to either side of the central maximum.

APPROACH The secondary maxima occur close to halfway between the minima, at about

$$\frac{\beta}{2} = \frac{\pi D\sin\theta}{\lambda} \approx \left(m + \tfrac{1}{2}\right)\pi. \qquad m = \pm 1, \pm 2, \pm 3, \cdots$$

The actual maxima are not quite at these points—their positions can be determined by differentiating Eq. 35–7 (see Problem 14)—but we are only seeking an estimate.

SOLUTION Using these values for β in Eq. 35–7 or 35–8, with $\sin\left[\left(m+\frac{1}{2}\right)\pi\right]=1$, gives

$$I_\theta = \frac{I_0}{\left(m+\frac{1}{2}\right)^2\pi^2}. \qquad m = 1, 2, 3, \cdots$$

For $m = 1$ and 2, we get

$$I_\theta = \frac{I_0}{22.2} = 0.045I_0 \qquad [m = 1]$$

$$I_\theta = \frac{I_0}{61.7} = 0.016I_0. \qquad [m = 2]$$

The first maximum to the side of the central peak has only 1/22, or 4.5%, the intensity of the central peak, and succeeding ones are smaller still, just as we can see in Fig. 35–5 and the photo of Fig. 35–2c.

Diffraction by a circular opening produces a similar pattern (though circular rather than rectangular), and is of great practical importance because lenses are essentially circular apertures through which light passes. We will discuss this in Section 35–4 and see how diffraction limits the resolution (or sharpness) of images.

*35–3 Diffraction in the Double-Slit Experiment

When we analyzed Young's double-slit experiment in Section 34–4, we assumed that the central portion of the screen was uniformly illuminated. This is equivalent to assuming the slits are infinitesimally narrow, so that the central diffraction peak is spread out over the whole screen. This can never be the case for real slits; diffraction reduces the intensity of the bright interference fringes to the side of center so they are not all of the same height as they were shown in Fig. 34–14. (They were shown more correctly in Fig. 34–9b.)

To calculate the intensity in a double-slit interference pattern including diffraction, let us assume the slits have equal widths D and their centers are separated by a distance d. Since the distance to the screen is large compared to the slit separation d, the wave amplitude due to each slit is essentially the same at each point on the screen. Then the total wave amplitude at any angle θ will no longer be

$$E_{\theta 0} = 2E_0\cos\frac{\delta}{2},$$

as was given by Eq. 34–5b (where $\delta = 2\pi d\sin\theta/\lambda$). Rather, this equation must be modified using Eq. 35–5 because each of the two slits exhibits its own diffraction, so that

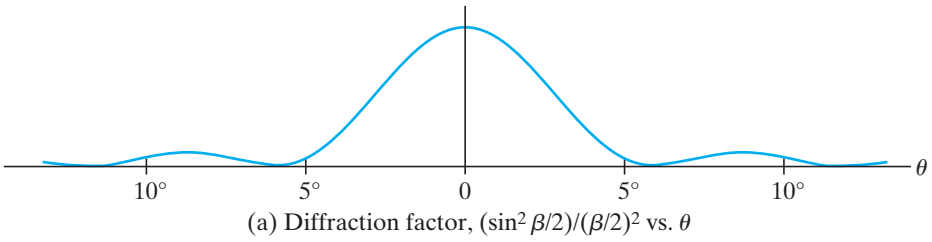
$$E_{\theta 0} = 2E_0\left(\frac{\sin\beta/2}{\beta/2}\right)\cos\frac{\delta}{2}.$$

Thus the intensity will be given by $(I \propto E^2)$

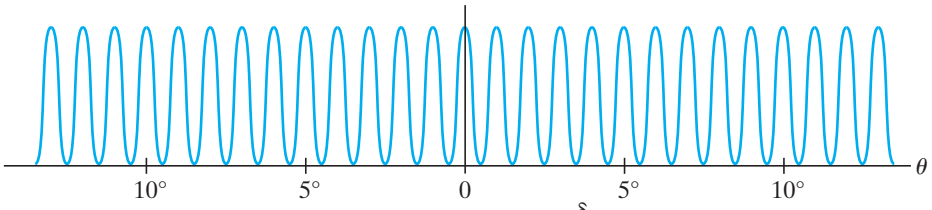
$$I_\theta = I_0\left(\frac{\sin\beta/2}{\beta/2}\right)^2\left(\cos\frac{\delta}{2}\right)^2 \qquad \textbf{(35–9)}$$

where $I_0 = 4E_0^2$, and from Eqs. 35–6 and 34–4 we have

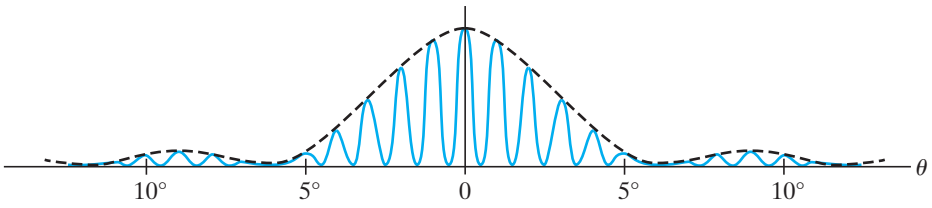
$$\frac{\beta}{2} = \frac{\pi}{\lambda}D\sin\theta \qquad \text{and} \qquad \frac{\delta}{2} = \frac{\pi}{\lambda}d\sin\theta.$$



(a) Diffraction factor, $(\sin^2 \beta/2)/(\beta/2)^2$ vs. θ



(b) Interference factor, $\cos^2 \frac{\delta}{2}$ vs. θ



(c) Intensity, I_θ vs. θ

FIGURE 35–11
(a) Diffraction factor,
(b) interference factor, and
(c) the resultant intensity I_θ, plotted as a function of θ for $d = 6D = 60\lambda$.

Equation 35–9 for the intensity in a double-slit pattern, as we just saw, is

$$I_\theta = I_0\left(\frac{\sin\beta/2}{\beta/2}\right)^2\left(\cos\frac{\delta}{2}\right)^2. \qquad \textbf{(35–9)}$$

The first term in parentheses is sometimes called the "diffraction factor" and the second one the "interference factor." These two factors are plotted in Fig. 35–11a and b for the case when $d = 6D$, and $D = 10\lambda$. (Figure 35–11b is essentially the same as Fig. 34–14.) Figure 35–11c shows the product of these two curves (times I_0) which is the actual intensity as a function of θ (or as a function of position on the screen for θ not too large) as given by Eq. 35–9. As indicated by the dashed lines in Fig. 35–11c, the diffraction factor acts as a sort of envelope that limits the interference peaks.

EXAMPLE 35–4 **Diffraction plus interference.** Show why the central diffraction peak in Fig. 35–11c contains 11 interference fringes.

APPROACH The first minimum in the diffraction pattern occurs where

$$\sin\theta = \frac{\lambda}{D}.$$

Since $d = 6D$,

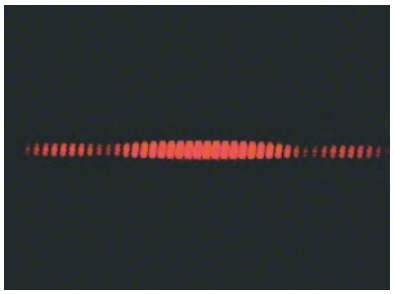
$$d\sin\theta = 6D\left(\frac{\lambda}{D}\right) = 6\lambda.$$

SOLUTION From Eq. 34–2a, interference peaks (maxima) occur for $d\sin\theta = m\lambda$ where m can be $0, 1, \cdots$ or any integer. Thus the diffraction minimum $(d\sin\theta = 6\lambda)$ coincides with $m = 6$ in the interference pattern, so the $m = 6$ peak won't appear. Hence the central diffraction peak encloses the central interference peak $(m = 0)$ and five peaks $(m = 1$ to $5)$ on each side for a total of 11. Since the sixth order doesn't appear, it is said to be a "missing order."

Notice from Example 35–4 that the number of interference fringes in the central diffraction peak depends only on the ratio d/D. It does not depend on wavelength λ. The actual spacing (in angle, or in position on the screen) does depend on λ. For the case illustrated, $D = 10\lambda$, and so the first diffraction minimum occurs at $\sin\theta = \lambda/D = 0.10$ or about 6°.

The decrease in intensity of the interference fringes away from the center, as graphed in Fig. 35–11, is shown in Fig. 35–12.

FIGURE 35–12 Photographs of a double-slit interference pattern using a laser beam showing effects of diffraction. In both cases $d = 0.50$ mm, whereas $D = 0.040$ mm in (a) and 0.080 mm in (b).


(a)


(b)

35–4 Interference vs. Diffraction

The patterns due to interference and diffraction arise from the same phenomenon—the superposition of coherent waves of different phase. The distinction between them is thus not so much physical as for convenience of description, as in Section 35–3 where we analyzed the two-slit pattern in terms of interference and diffraction separately. In general, we use the word "diffraction" when referring to an analysis by superposition of many infinitesimal and usually contiguous sources, such as when we subdivide a source into infinitesimal parts. We use the term "interference" when we superpose the wave from a finite (and usually small) number of coherent sources.

⚠ CAUTION
Diffraction and interference

35–5 Limits of Resolution; Circular Apertures

No lens can image a point object as a perfect point image. The image is always spread out a bit, as a blob, because of *aberrations* (Section 33–10) and *diffraction*.

The ability of a lens to produce distinct images of two point objects very close together is called the **resolution** of the lens. The closer the two images can be and still be seen as distinct (rather than overlapping blobs), the higher the resolution. The resolution of a camera lens, for example, is often specified as so many dots or lines per millimeter, as mentioned in Section 33–5.

In Section 35–1, we saw that because light travels as a wave, light from a point source passing through a slit is spread out into a diffraction pattern (Figs. 35–2 and 35–5). A lens, because it has edges, acts like a round slit. When a lens forms the image of a point object, the image is actually a tiny diffraction pattern. Thus *an image would be blurred even if lens aberrations were absent*.

In the analysis that follows, we assume that the lens is free of aberrations, so we can concentrate on diffraction effects and how much they limit the resolution of a lens. In Fig. 35–5 we saw that the diffraction pattern produced by light passing through a rectangular slit has a central maximum in which most of the light falls. This central peak falls to a minimum on either side of its center at an angle θ given by $\sin\theta = \lambda/D$ (this is Eq. 35–1), where D is the slit width and λ the wavelength of light used. θ is the *angular half-width* of the central maximum. For small angles (in radians), $\theta \approx \sin\theta$, so we can write

$$\theta \approx \frac{\lambda}{D}. \qquad \text{[slit]}$$

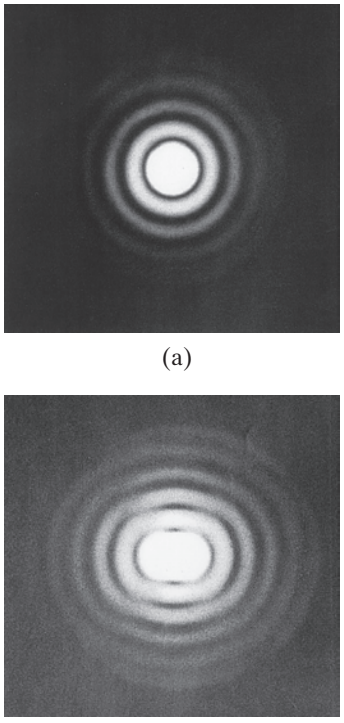
There are also low-intensity fringes beyond.

For a lens, or any circular hole, the image of a point object will consist of a *circular* central peak (called the *diffraction spot* or *Airy disk*) surrounded by faint circular fringes, as shown in Fig. 35–13a. The central maximum has an angular half width θ (in radians) given by

$$\theta \approx \frac{1.22\lambda}{D}, \qquad \text{[circular opening]}$$

where D is the diameter of the circular opening. This is a *theoretical* result for a perfect circle or lens, obtained by integrating over the circular opening. The formula above differs from that for a slit (Eq. 35–1) by the factor 1.22. This factor appears because the width of a circular hole is not uniform (like a rectangular slit) but varies from its diameter D at the center to zero at the edges. A mathematical analysis shows that the "average" width is $D/1.22$. Hence we get the equation above rather than Eq. 35–1. The intensity of light in the diffraction pattern from a point source of light passing through a circular opening is shown in Fig. 35–14. The image for a non-point source is a superposition of such patterns. Usually, we consider only the central spot, since the concentric rings are so much dimmer.

If two point objects are very close, the diffraction patterns of their images will overlap as shown in Fig. 35–13b. As the objects are moved closer, a separation is reached where you can't tell if there are two overlapping images or a single image.

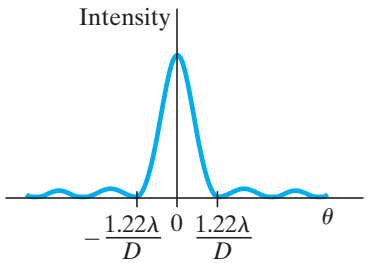


(a)

(b)

FIGURE 35–13 Photographs of images (greatly magnified) formed by a lens, showing the diffraction pattern of an image for: (a) a single point object; (b) two point objects whose images are barely resolved.

FIGURE 35–14 Intensity of light across the diffraction pattern of a circular hole.



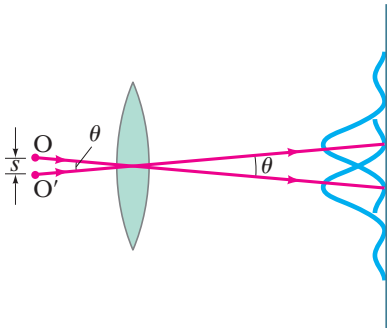


FIGURE 35–15 The *Rayleigh criterion*. Two images are just resolvable when the center of the diffraction peak of one is directly over the first minimum in the diffraction pattern of the other. The two point objects O and O′ subtend an angle θ at the lens; only one ray (it passes through the center of the lens) is drawn for each object, to indicate the center of the diffraction pattern of its image.

The separation at which this happens may be judged differently by different observers. However, a generally accepted criterion is that proposed by Lord Rayleigh (1842–1919). This **Rayleigh criterion** states that *two images are just resolvable when the center of the diffraction disk of one image is directly over the first minimum in the diffraction pattern of the other*. This is shown in Fig. 35–15. Since the first minimum is at an angle $\theta = 1.22\lambda/D$ from the central maximum, Fig. 35–15 shows that two objects can be considered *just resolvable* if they are separated by at least an angle θ given by

$$\theta = \frac{1.22\lambda}{D}. \qquad \left[\begin{array}{l}\text{2 points just resolvable;}\\ \theta \text{ in radians}\end{array}\right] \qquad \textbf{(35–10)}$$

In this equation, D is the diameter of the lens, and applies also to a mirror diameter. This is the limit on resolution set by the wave nature of light due to diffraction. A smaller angle means better resolution: you can make out closer objects. We see from Eq. 35–10 that using a shorter wavelength λ can reduce θ and thus increase resolution.

EXERCISE B Green light (550 nm) passes through a 25-mm-diameter camera lens. What is the angular half-width of the resulting diffraction pattern? (*a*) 2.7×10^{-5} degrees, (*b*) 1.5×10^{-3} degrees, (*c*) 3.2°, (*d*) 27°, (*e*) 1.5×10^{3} degrees.


(a)


(b)

FIGURE 35–16 (a) Hubble Space Telescope, with Earth in the background. The flat orange panels are solar cells that collect energy from the Sun to power the equipment. (b) Hubble image looking deep into space, showing some of the most distant visible galaxies in the universe.

EXAMPLE 35–5 **Hubble Space Telescope.** The Hubble Space Telescope (HST) is a reflecting telescope that was placed in orbit in 1990 above the Earth's atmosphere, so its resolution would not be limited by turbulence in the atmosphere (Fig. 35–16). It is still sending back remarkable images (Fig. 35–16b). Its objective diameter is 2.4 m. For visible light, say $\lambda = 550$ nm, estimate the improvement in resolution the Hubble offers over Earth-bound telescopes, which are limited in resolution by movement of the Earth's atmosphere to about half an arc second. (Each degree is divided into 60 minutes each containing 60 seconds, so $1° = 3600$ arc seconds.)

APPROACH Angular resolution for the Hubble is given (in radians) by Eq. 35–10. The resolution for Earth telescopes is given, and we first convert it to radians so we can compare.

SOLUTION Earth-bound telescopes are limited to an angular resolution of

$$\theta = \tfrac{1}{2}\left(\frac{1}{3600}\right)^{\circ}\left(\frac{2\pi \text{ rad}}{360^{\circ}}\right) = 2.4 \times 10^{-6}\,\text{rad}.$$

The Hubble, on the other hand, is limited by diffraction (Eq. 35–10) which for $\lambda = 550\,\text{nm}$ is

$$\theta = \frac{1.22\lambda}{D} = \frac{1.22(550 \times 10^{-9}\,\text{m})}{2.4\,\text{m}} = 2.8 \times 10^{-7}\,\text{rad},$$

thus giving almost ten times better resolution $(2.4 \times 10^{-6}\,\text{rad}/2.8 \times 10^{-7}\,\text{rad} \approx 9\times)$.

EXAMPLE 35–6 **ESTIMATE** **Eye resolution.** You are in an airplane at an altitude of 10,000 m. If you look down at the ground, estimate the minimum separation s between objects that you could distinguish. Could you count cars in a parking lot? Consider only diffraction, and assume your pupil is about 3.0 mm in diameter and $\lambda = 550\,\text{nm}$.

PHYSICS APPLIED
How well the eye can see

APPROACH We use the Rayleigh criterion, Eq. 35–10 and Fig. 35–15, to estimate θ. The separation s of objects is $s = \ell\theta$, where $\ell = 10^4\,\text{m}$ and θ is in radians.

SOLUTION In Eq. 35–10, we set $D = 3.0\,\text{mm}$ for the opening of the eye:

$$s = \ell\theta = \ell\frac{1.22\lambda}{D} = \frac{(10^4\,\text{m})(1.22)(550 \times 10^{-9}\,\text{m})}{3.0 \times 10^{-3}\,\text{m}} = 2.2\,\text{m}.$$

Yes, you could just resolve a car (roughly 2 m wide by 3 or 4 m long) and so, considering only diffraction, could count the number of cars in the lot.

EXERCISE C Someone claims a spy satellite camera can see 3-cm-high newspaper headlines from an altitude of 100 km. If diffraction were the only limitation ($\lambda = 550\,\text{nm}$), use Eq. 35–10 to determine what diameter lens the camera would have.

35–6 Resolution of Telescopes and Microscopes; the λ Limit

You might think that a microscope or telescope could be designed to produce any desired magnification, depending on the choice of focal lengths and quality of the lenses. But this is not possible, because of diffraction. An increase in magnification above a certain point merely results in *magnification of the diffraction patterns*. This can be highly misleading since we might think we are seeing details of an object when we are really seeing details of the diffraction pattern.

CAUTION
What may seem details under magnification may be only a diffraction pattern

Telescope Resolution

To examine this problem, we apply the Rayleigh criterion: two objects (or two nearby points on one object) are just resolvable if they are separated by an angle θ (Fig. 35–15) given by Eq. 35–10:

$$\theta = \frac{1.22\lambda}{D}.$$

Telescope Resolution

This formula is valid for either a microscope or a telescope, where D is the diameter of the objective lens or mirror. For a telescope, the resolution is specified by stating θ as given by this equation.†

†Earth-bound telescopes with large-diameter objectives are usually limited not by diffraction but by other effects such as turbulence in the atmosphere. The resolution of a high-quality microscope, on the other hand, normally *is* limited by diffraction; microscope objectives are complex compound lenses containing many elements of small diameter (since f is small), thus reducing aberrations.



FIGURE 35–17 The 300-meter radiotelescope in Arecibo, Puerto Rico, uses radio waves (Fig. 31–12) instead of visible light.

EXAMPLE 35–7 Telescope resolution (radio wave vs. visible light). What is the theoretical minimum angular separation of two stars that can just be resolved by (*a*) the 200-inch telescope on Palomar Mountain (Fig. 33–38c); and (*b*) the Arecibo radiotelescope (Fig. 35–17), whose diameter is 300 m? Assume $\lambda = 550\,\text{nm}$ for the visible-light telescope in part (*a*), and $\lambda = 4\,\text{cm}$ (the shortest wavelength at which the radiotelescope has operated) in part (*b*).

APPROACH We apply the Rayleigh criterion (Eq. 35–10) for each telescope.

SOLUTION (*a*) With $D = 200\,\text{in.} = 5.1\,\text{m}$, we have from Eq. 35–10 that

$$\theta = \frac{1.22\lambda}{D} = \frac{(1.22)(5.50 \times 10^{-7}\,\text{m})}{(5.1\,\text{m})} = 1.3 \times 10^{-7}\,\text{rad},$$

or 0.75×10^{-5} deg. (Note that this is equivalent to resolving two points less than 1 cm apart from a distance of 100 km!)

(*b*) For radio waves with $\lambda = 0.04\,\text{m}$ emitted by stars, the resolution is

$$\theta = \frac{(1.22)(0.04\,\text{m})}{(300\,\text{m})} = 1.6 \times 10^{-4}\,\text{rad}.$$

The resolution is less because the wavelength is so much larger, but the larger objective (300 m) collects more radiation and thus detects fainter objects.

NOTE In both cases, we determined the limit set by diffraction. The resolution for a visible-light Earth-bound telescope is not this good because of aberrations and, more importantly, turbulence in the atmosphere. In fact, large-diameter objectives are not justified by increased resolution, but by their greater *light-gathering ability*—they allow more light in, so fainter objects can be seen. Radiotelescopes are not hindered by atmospheric turbulence, and the resolution found in (*b*) is a good estimate.

Microscope Resolving Power

For a microscope, it is more convenient to specify the actual distance, s, between two points that are just barely resolvable: see Fig. 35–15. Since objects are normally placed near the focal point of the microscope objective, the angle θ (in radians) subtended by two objects is $\theta = s/f$, so $s = f\theta$. If we combine this with Eq. 35–10, we obtain the **resolving power** (**RP**) of a microscope

$$\text{RP} = s = f\theta = \frac{1.22\lambda f}{D}, \qquad \textbf{(35–11)}$$

where f is the objective lens' focal length (not frequency) and D its diameter. The distance s is called the resolving power of the lens because it is the minimum separation of two object points that can just be resolved—assuming the highest quality lens since this limit is imposed by the wave nature of light. A smaller RP means better resolution, better detail.

EXERCISE D What is the resolving power ($\lambda = 550\,\text{nm}$) of a microscope with a 5-mm-diameter objective which has $f = 9\,\text{mm}$? (*a*) 550 nm, (*b*) 750 nm, (*c*) 1200 nm, (*d*) 0.05 nm, (*e*) 0.005 nm.

Ultimate Resolution Is λ

Diffraction sets an ultimate limit on the detail that can be seen on any object. In Eq. 35–11 for resolving power of a microscope, the focal length of the lens cannot practically be made less than (approximately) the radius of the lens, and even that is very difficult (see the lensmaker's equation, Eq. 33–4). In this best optimistic case, Eq. 35–11 gives, with $f \approx D/2$,

$$\text{RP} \approx \frac{\lambda}{2}. \qquad \textbf{(35–12)}$$

Thus we can say, to within a factor of 2 or so, that

λ limits resolution

it is not possible to resolve detail of objects smaller than the wavelength of the radiation being used.

This is an important and useful rule of thumb.

Compound lenses in microscopes are now designed so well that the actual limit on resolution is often set by diffraction—that is, by the wavelength of the light used. To obtain greater detail, one must use radiation of shorter wavelength. The use of UV radiation can increase the resolution by a factor of perhaps 2. Far more important, however, was the discovery in the early twentieth century that electrons have wave properties (Chapter 37) and that their wavelengths can be very small. The wave nature of electrons is utilized in the electron microscope (Section 37–8), which can magnify 100 to 1000 times more than a visible-light microscope because of the much shorter wavelengths. X-rays, too, have very short wavelengths and are often used to study objects in great detail (Section 35–10).

35–7 Resolution of the Human Eye and Useful Magnification

The resolution of the human eye is limited by several factors, all of roughly the same order of magnitude. The resolution is best at the fovea (Section 33–6), where the cone spacing is smallest, about $3\,\mu\text{m}$ $(=3000\,\text{nm})$. The diameter of the pupil varies from about 1 mm to about 8 mm. So for $\lambda = 550\,\text{nm}$ (where the eye's sensitivity is greatest), the diffraction limit is about $\theta \approx 1.22\lambda/D \approx 8\times10^{-5}\,\text{rad}$ to $6\times10^{-4}\,\text{rad}$. The eye is about 2 cm long, giving a resolving power (Eq. 35–11) of $s \approx (2\times10^{-2}\,\text{m})(8\times10^{-5}\,\text{rad}) \approx 2\,\mu\text{m}$ at best, to about $10\,\mu\text{m}$ at worst (pupil small). Spherical and chromatic aberration also limit the resolution to about $10\,\mu\text{m}$. The net result is that the eye can just resolve objects whose angular separation is around

$$5\times10^{-4}\,\text{rad}. \qquad \left[\begin{array}{c}\text{best eye}\\ \text{resolution}\end{array}\right]$$

This corresponds to objects separated by 1 cm at a distance of about 20 m. (This contradicts, by a factor of 2, Example 35–6 which considered only diffraction.)

The typical near point of a human eye is about 25 cm. At this distance, the eye can just resolve objects that are $(25\,\text{cm})(5\times10^{-4}\,\text{rad}) \approx 10^{-4}\,\text{m} = \frac{1}{10}\,\text{mm}$ apart.[†] Since the best light microscopes can resolve objects no smaller than about 200 nm at best (Eq. 35–12 for violet light, $\lambda = 400\,\text{nm}$), the useful magnification [= (resolution by naked eye)/(resolution by microscope)] is limited to about

$$\frac{10^{-4}\,\text{m}}{200\times10^{-9}\,\text{m}} \approx 500\times. \qquad \left[\begin{array}{c}\text{maximum useful}\\ \text{microscope magnification}\end{array}\right]$$

In practice, magnifications of about $1000\times$ are often used to minimize eyestrain. Any greater magnification would simply make visible the diffraction pattern produced by the microscope objective lens.

EXERCISE E Return to the Chapter-Opening Questions, page 1025, and answer them again now. Try to explain why you may have answered differently the first time.

35–8 Diffraction Grating

A large number of equally spaced parallel slits is called a **diffraction grating**, although the term "interference grating" might be as appropriate. Gratings can be made by precision machining of very fine parallel lines on a glass plate. The untouched spaces between the lines serve as the slits. Photographic transparencies of an original grating serve as inexpensive gratings. Gratings containing 10,000 lines or slits per centimeter are common, and are very useful for precise measurements of wavelengths. A diffraction grating containing slits is called a **transmission grating**. Another type of diffraction grating is the **reflection grating**, made by ruling fine lines on a metallic or glass surface from which light is reflected and analyzed. The analysis is basically the same as for a transmission grating, which we now discuss.

[†]A nearsighted eye that needs −8 or −10 D lenses can have a near point of 8 or 10 cm, and a higher resolution up close (without glasses) of a factor of $2\frac{1}{2}$ or 3, or $\approx \frac{1}{25}\,\text{mm} \approx 40\,\mu\text{m}$.

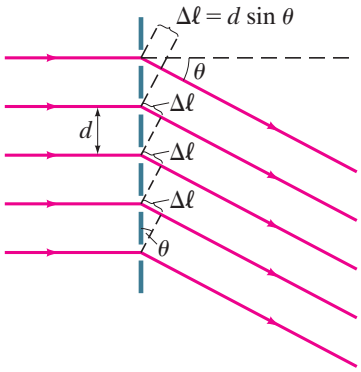
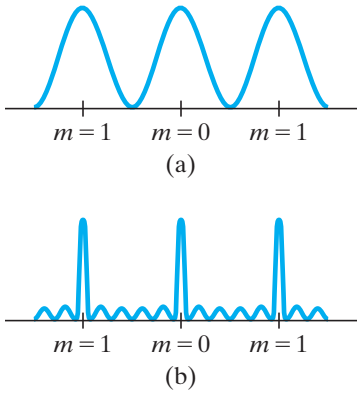


FIGURE 35–18 Diffraction grating.

⚠ CAUTION

Diffraction grating is analyzed using interference formulas, not diffraction formulas

FIGURE 35–19 Intensity as a function of viewing angle θ (or position on the screen) for (a) two slits, (b) six slits. For a diffraction grating, the number of slits is very large $(\approx 10^4)$ and the peaks are narrower still.



The analysis of a diffraction grating is much like that of Young's double-slit experiment. We assume parallel rays of light are incident on the grating as shown in Fig. 35–18. We also assume that the slits are narrow enough so that diffraction by each of them spreads light over a very wide angle on a distant screen beyond the grating, and interference can occur with light from all the other slits. Light rays that pass through each slit without deviation $(\theta = 0°)$ interfere constructively to produce a bright maximum at the center of the screen. Constructive interference also occurs at an angle θ such that rays from adjacent slits travel an extra distance of $\Delta\ell = m\lambda$, where m is an integer. If d is the distance *between* slits, then we see from Fig. 35–18 that $\Delta\ell = d \sin\theta$, and

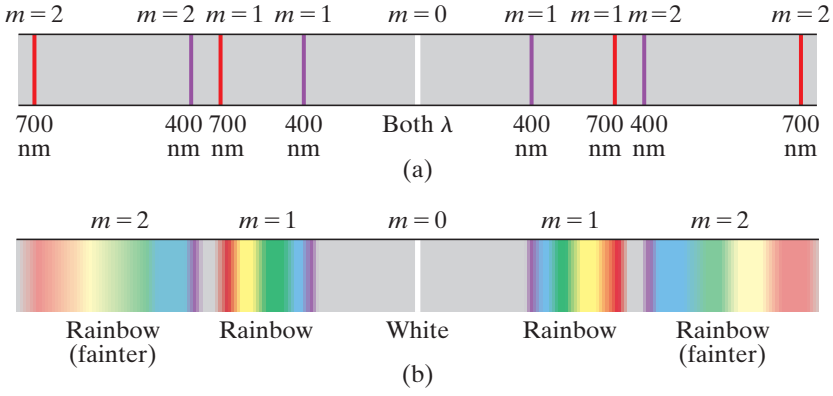
$$\sin\theta = \frac{m\lambda}{d}, \qquad m = 0, 1, 2, \cdots \qquad \left[\begin{array}{c}\text{diffraction grating,}\\ \text{principal maxima}\end{array}\right] \qquad \textbf{(35–13)}$$

is the criterion to have a brightness maximum. This is the same equation as for the double-slit situation, and again m is called the **order** of the pattern.

There is an important difference between a double-slit and a multiple-slit pattern. The bright maxima are much *sharper* and *narrower* for a grating. Why? Suppose the angle θ in Fig. 35–18 is increased just slightly beyond θ required for a maximum. For only two slits, the two waves will be only slightly out of phase, so nearly full constructive interference occurs. This means the maxima are wide (see Fig. 34–10). For a grating, the waves from two adjacent slits will also not be significantly out of phase. But waves from one slit and those from a second one a few hundred slits away may be exactly out of phase; all or nearly all the light can cancel in pairs in this way. For example, suppose the angle θ is very slightly different from its first-order maximum, so that the extra path length for a pair of adjacent slits is not exactly λ but rather 1.0010λ. The wave through one slit and another one 500 slits below will have a path difference of $1\lambda + (500)(0.0010\lambda) = 1.5000\lambda$, or $1\frac{1}{2}$ wavelengths, so the two will be out of phase and cancel. A pair of slits, one below each of these, will also cancel each other. That is, the light from slit 1 cancels with light from slit 501; light from slit 2 cancels with light from slit 502, and so on. Thus even for a tiny angle[†] corresponding to an extra path length of $\frac{1}{1000}\lambda$, there is much destructive interference, and so the maxima of a diffraction grating are very narrow. The more slits there are in a grating, the sharper will be the peaks (see Fig. 35–19). Because a grating produces much sharper lines than two slits alone, and also much brighter maxima because there are many more slits, a grating is a far more precise device for measuring wavelengths.

Suppose the light striking a diffraction grating is not monochromatic, but consists of two or more distinct wavelengths. Then for all orders other than $m = 0$, each wavelength will produce a maximum at a different angle (Eq. 35–13), forming a line on the screen as shown in Fig. 35–20a. If white light strikes a grating, the central $(m = 0)$ maximum will be a sharp white peak.

FIGURE 35–20 Spectra produced by a grating: (a) two wavelengths, 400 nm and 700 nm; (b) white light. The second order will normally be dimmer than the first order. (Higher orders are not shown.) If the grating spacing is small enough, the second and higher orders will be missing.



[†]Depending on the total number of slits, there may or may not be complete cancellation for such an angle, so there will be very tiny peaks between the main maxima (see Fig. 35–19b), but they are usually much too small to be seen.

But for all other orders, there will be a distinct spectrum of colors spread out over a certain angular width, Fig. 35–20b. Because a diffraction grating spreads out light into its component wavelengths, the resulting pattern is called a **spectrum**.

EXAMPLE 35–8 **Diffraction grating: line positions.** Determine the angular positions of the first- and second-order lines (maxima) for light of wavelength 400 nm and 700 nm incident on a grating containing 10,000 slits per centimeter. (Sometimes we say "lines per cm".)

APPROACH First we find the distance d between grating slits: if the grating has N slits in 1 m, then the distance between slits is $d = 1/N$ meters. Then we use Eq. 35–13, $\sin\theta = m\lambda/d$, to get the angles for the two wavelengths for $m = 1$ and 2.

SOLUTION The grating contains 1.00×10^4 slits/cm $= 1.00 \times 10^6$ slits/m, which means the distance between slits is $d = (1/1.00 \times 10^6)\,\text{m} = 1.00 \times 10^{-6}\,\text{m} = 1.00\,\mu\text{m}$. In first order $(m = 1)$, the angles are

$$\sin\theta_{400} = \frac{m\lambda}{d} = \frac{(1)(400 \times 10^{-9}\,\text{m})}{1.00 \times 10^{-6}\,\text{m}} = 0.400$$

$$\sin\theta_{700} = \frac{(1)(700 \times 10^{-9}\,\text{m})}{1.00 \times 10^{-6}\,\text{m}} = 0.700$$

so $\theta_{400} = 23.6°$ and $\theta_{700} = 44.4°$. In second order,

$$\sin\theta_{400} = \frac{2\lambda}{d} = \frac{(2)(400 \times 10^{-9}\,\text{m})}{1.00 \times 10^{-6}\,\text{m}} = 0.800$$

$$\sin\theta_{700} = \frac{(2)(700 \times 10^{-9}\,\text{m})}{1.00 \times 10^{-6}\,\text{m}} = 1.40$$

so $\theta_{400} = 53.1°$. But the second order does not exist for $\lambda = 700\,\text{nm}$ because $\sin\theta$ cannot exceed 1. No higher orders will appear.

NOTE The grating of Fig. 35–20 must have $d > 1.00\,\mu\text{m}$ since red is visible in both spectra for $m = 2$. $\text{Sin}\,\theta \le 1$ only if $d \ge 1.40 \times 10^{-6}\,\text{m}$ (see equation just above), so a grating must have fewer than 7140 slits/cm if the second order is to be visible.

EXAMPLE 35–9 **Spectra overlap.** White light containing wavelengths from 400 nm to 750 nm strikes a grating containing 4000 slits/cm. Show that the blue at $\lambda = 450\,\text{nm}$ of the third-order spectrum overlaps the red at 700 nm of the second order.

APPROACH We use $\sin\theta = m\lambda/d$ to calculate the angular positions of the $m = 3$ blue maximum and the $m = 2$ red one.

SOLUTION The grating spacing is $d = (1/4000)\,\text{cm} = 2.50 \times 10^{-6}\,\text{m}$. The blue of the third order occurs at an angle θ given by

$$\sin\theta = \frac{m\lambda}{d} = \frac{(3)(450 \times 10^{-9}\,\text{m})}{(2.50 \times 10^{-6}\,\text{m})} = 0.540.$$

Red in second order occurs at

$$\sin\theta = \frac{(2)(7.00 \times 10^{-7}\,\text{m})}{(2.50 \times 10^{-6}\,\text{m})} = 0.560,$$

which is a greater angle. So the second order overlaps into the beginning of the third-order spectrum.

EXERCISE F You are shown the spectra produced by red light shining through two different gratings. The maxima in spectrum A are farther apart than those in spectrum B. Which grating has more slits/cm?

35–9 The Spectrometer and Spectroscopy

A **spectrometer** or **spectroscope**, Fig. 35–21, is a device to measure wavelengths accurately using a diffraction grating (or a prism) to separate different wavelengths of light. Light from a source passes through a narrow slit S in the "collimator." The slit is at the focal point of the lens L, so parallel light falls on the grating. The movable telescope can bring the rays to a focus. Nothing will be seen in the viewing telescope unless it is positioned at an angle θ that corresponds to a diffraction peak (first order is usually used) of a wavelength emitted by the source. The angle θ can be measured to very high accuracy, so the wavelength can be determined to high accuracy using Eq. 35–13:

$$\lambda = \frac{d}{m}\sin\theta,$$

where m is an integer representing the order, and d is the distance between grating slits. The bright line you see in a spectrometer corresponding to a discrete particular wavelength is actually an image of the slit S. A narrower slit results in dimmer light, but we can measure the angular position more precisely. If the light contains a continuous range of wavelengths, then a continuous spectrum is seen in the spectroscope.

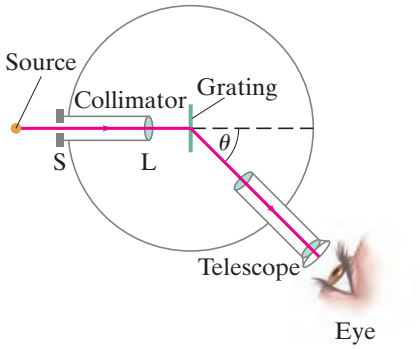


FIGURE 35–21 Spectrometer or spectroscope.

The spectrometer in Fig. 35–21 uses a transmission grating. Others may use a reflection grating, or sometimes a prism. A prism works because of dispersion (Section 32–6), bending light of different wavelengths into different angles. A prism is not a linear device and must be calibrated because λ is not proportional to $\sin\theta$; see Fig. 32–28.

An important use of a spectrometer is for the identification of atoms or molecules. When a gas is heated or an electric current is passed through it, the gas emits a characteristic **line spectrum**. That is, only certain discrete wavelengths of light are emitted, and these are different for different elements and compounds.[†] Figure 35–22 shows the line spectra for a number of elements in the gas state. Line spectra occur only for gases at high temperatures and low pressure and density. The light from heated solids, such as a lightbulb filament, and even from a dense gaseous object such as the Sun, produces a **continuous spectrum** including a wide range of wavelengths.

Figure 35–22 also shows the Sun's "continuous spectrum," which contains a number of *dark* lines (only the most prominent are shown), called **absorption lines**. Atoms and molecules can absorb light at the same wavelengths at which they emit light.

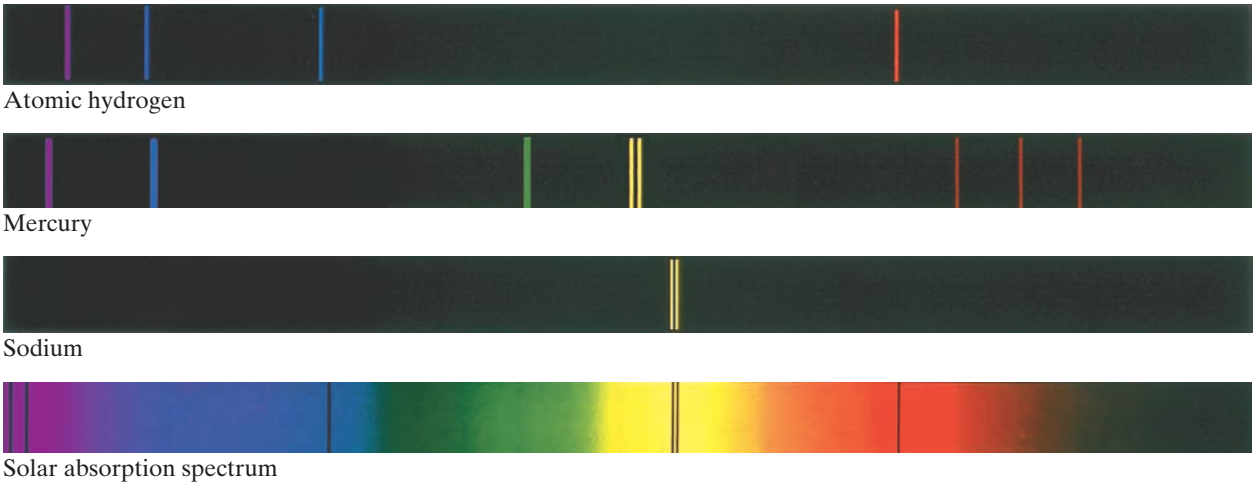


FIGURE 35–22 Line spectra for the gases indicated, and the spectrum from the Sun showing absorption lines.

[†]Why atoms and molecules emit line spectra was a great mystery for many years and played a central role in the development of modern quantum theory, as we shall see in Chapter 37.

The Sun's absorption lines are due to absorption by atoms and molecules in the cooler outer atmosphere of the Sun, as well as by atoms and molecules in the Earth's atmosphere. A careful analysis of all the Sun's thousands of absorption lines reveals that at least two-thirds of all elements are present in the Sun's atmosphere. The presence of elements in the atmosphere of nearby planets, in interstellar space, and in stars, is also determined by spectroscopy.

Spectroscopy is useful for determining the presence of certain types of molecules in laboratory specimens where chemical analysis would be difficult. For example, biological DNA and different types of protein absorb light in particular regions of the spectrum (such as in the UV). The material to be examined, which is often in solution, is placed in a monochromatic light beam whose wavelength is selected by the placement angle of a diffraction grating or prism. The amount of absorption, as compared to a standard solution without the specimen, can reveal not only the presence of a particular type of molecule, but also its concentration.

PHYSICS APPLIED
Chemical and biochemical analysis by spectroscopy

Light emission and absorption also occur outside the visible part of the spectrum, such as in the UV and IR regions. Glass absorbs light in these regions, so reflection gratings and mirrors (in place of lenses) are used. Special types of film or sensors are used for detection.

EXAMPLE 35–10 **Hydrogen spectrum.** Light emitted by hot hydrogen gas is observed with a spectroscope using a diffraction grating having 1.00×10^4 slits/cm. The spectral lines nearest to the center $(0°)$ are a violet line at 24.2°, a blue line at 25.7°, a blue-green line at 29.1°, and a red line at 41.0° from the center. What are the wavelengths of these spectral lines of hydrogen?

APPROACH We get the wavelengths from the angles by using $\lambda = (d/m)\sin\theta$ where d is the spacing between slits, and m is the order of the spectrum (Eq. 35–13).

SOLUTION Since these are the closest lines to $\theta = 0°$, this is the first-order spectrum $(m = 1)$. The slit spacing is $d = 1/(1.00 \times 10^4\,\text{cm}^{-1}) = 1.00 \times 10^{-6}\,\text{m}$. The wavelength of the violet line is

$$\lambda = \left(\frac{d}{m}\right)\sin\theta = \left(\frac{1.00 \times 10^{-6}\,\text{m}}{1}\right)\sin 24.2° = 4.10 \times 10^{-7}\,\text{m} = 410\,\text{nm}.$$

The other wavelengths are:

$$\text{blue:} \quad \lambda = (1.00 \times 10^{-6}\,\text{m})\sin 25.7° = 434\,\text{nm},$$

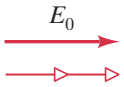
$$\text{blue-green:} \quad \lambda = (1.00 \times 10^{-6}\,\text{m})\sin 29.1° = 486\,\text{nm},$$

$$\text{red:} \quad \lambda = (1.00 \times 10^{-6}\,\text{m})\sin 41.0° = 656\,\text{nm}.$$

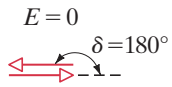
NOTE In an unknown mixture of gases, these four spectral lines need to be seen to identify that the mixture contains hydrogen.

*35–10 Peak Widths and Resolving Power for a Diffraction Grating

We now look at the pattern of maxima produced by a multiple-slit grating using phasor diagrams. We can determine a formula for the width of each peak, and we will see why there are tiny maxima between the principal maxima, as indicated in Fig. 35–19b. First of all, it should be noted that the two-slit and six-slit patterns shown in Fig. 35–19 were drawn assuming very narrow slits so that diffraction does not limit the height of the peaks. For real diffraction gratings, this is not normally the case: the slit width D is often not much smaller than the slit separation d, and diffraction thus limits the intensity of the peaks so the central peak $(m = 0)$ is brighter than the side peaks. We won't worry about this effect on intensity except to note that if a diffraction minimum coincides with a particular order of the interference pattern, that order will not appear. (For example, if $d = 2D$, all the even orders, $m = 2, 4, \cdots$, will be missing. Can you see why? *Hint*: See Example 35–4.)



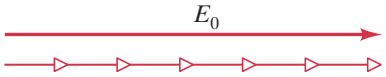
(a) Central maximum: $\theta = 0, \delta = 0$



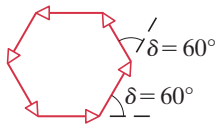
(b) Minimum: $\delta = 180°$

FIGURE 35–23 Phasor diagram for two slits (a) at the central maximum, (b) at the nearest minimum.

FIGURE 35–24 Phasor diagram for six slits (a) at the central maximum, (b) at the nearest minimum.



(a) Central maximum: $\theta = 0, \delta = 0$



(b) Minimum: $\delta = 60°, E_\theta = 0$

FIGURE 35–25 Phasor diagram for the secondary peak.



Figures 35–23 and 35–24 show phasor diagrams for a two-slit and a six-slit grating, respectively. Each short arrow represents the amplitude of a wave from a single slit, and their vector sum (as phasors) represents the total amplitude for a given viewing angle θ. Part (a) of each Figure shows the phasor diagram at $\theta = 0°$, at the center of the pattern, which is the central maximum $(m = 0)$. Part (b) of each Figure shows the condition for the adjacent minimum: where the arrows first close on themselves (add to zero) so the amplitude E_θ is zero. For two slits, this occurs when the two separate amplitudes are 180° out of phase. For six slits, it occurs when each amplitude makes a 60° angle with its neighbor. For two slits, the minimum occurs when the phase between slits is $\delta = 2\pi/2$ (in radians); for six slits it occurs when the phase δ is $2\pi/6$; and in the general case of N slits, the minimum occurs for a phase difference between adjacent slits of

$$\delta = \frac{2\pi}{N}. \qquad \textbf{(35–14)}$$

What does this correspond to in θ? First note that δ is related to θ by (Eq. 34–4)

$$\frac{\delta}{2\pi} = \frac{d \sin\theta}{\lambda} \quad \text{or} \quad \delta = \frac{2\pi}{\lambda} d \sin\theta. \qquad \textbf{(35–15)}$$

Let us call $\Delta\theta_0$ the angular position of the minimum next to the peak at $\theta = 0$. Then

$$\frac{\delta}{2\pi} = \frac{d \sin\Delta\theta_0}{\lambda}.$$

We insert Eq. 35–14 for δ and find

$$\sin\Delta\theta_0 = \frac{\lambda}{Nd}. \qquad \textbf{(35–16a)}$$

Since $\Delta\theta_0$ is usually small (N is usually very large for a grating), $\sin\Delta\theta_0 \approx \Delta\theta_0$, so in the small angle limit we can write

$$\Delta\theta_0 = \frac{\lambda}{Nd}. \qquad \textbf{(35–16b)}$$

We can see from either of the last two relations that the larger N is, the narrower will be the central peak. (For $N = 2$, $\sin\Delta\theta_0 = \lambda/2d$, which is what we obtained earlier for the double slit, Eq. 34–2b with $m = 0$.)

Either of Eqs. 35–16 shows why the peaks become narrower for larger N. The origin of the small secondary maxima between the principal peaks (see Fig. 35–19b) can be deduced from the diagram of Fig. 35–25. This is just a continuation of Fig. 35–24b (where $\delta = 60°$); but now the phase δ has been increased to almost 90°, where E_θ is a relative maximum. Note that E_θ is much less than E_0 (Fig. 35–24a), so the intensity in this secondary maximum is much smaller than in a principal peak. As δ (and θ) is increased further, E_θ again decreases to zero (a "double circle"), then reaches another tiny maximum, and so on. Eventually the diagram unfolds again and when $\delta = 360°$, all the amplitudes again lie in a straight line (as in Fig. 35–24a) corresponding to the next principal maximum ($m = 1$ in Eq. 35–13).

Equation 35–16b gives the half-width of the central $(m = 0)$ peak. To determine the half-width of higher order peaks, $\Delta\theta_m$ for order m, we differentiate Eq. 35–15 so as to relate the change $\Delta\delta$ in δ, to the change $\Delta\theta$ in the angle θ:

$$\Delta\delta \approx \frac{d\delta}{d\theta}\Delta\theta = \frac{2\pi d}{\lambda}\cos\theta\,\Delta\theta.$$

If $\Delta\theta_m$ represents the half-width of a peak of order m $(m = 1, 2, \cdots)$—that is, the angle between the peak maximum and the minimum to either side—then $\Delta\delta = 2\pi/N$ as given by Eq. 35–14. We insert this into the above relation and find

$$\Delta\theta_m = \frac{\lambda}{Nd\cos\theta_m}, \qquad \textbf{(35–17)}$$

where θ_m is the angular position of the m^{th} peak as given by Eq. 35–13.

This derivation is valid only for small $\Delta\delta\ (=2\pi/N)$ which is indeed the case for real gratings since N is on the order of 10^4 or more.

An important property of any diffraction grating used in a spectrometer is its ability to resolve two very closely spaced wavelengths (wavelength difference $= \Delta\lambda$). The **resolving power** R of a grating is defined as

$$\mathrm{R} = \frac{\lambda}{\Delta\lambda}. \qquad \textbf{(35–18)}$$

With a little work, using Eq. 35–17, we can show that $\Delta\lambda = \lambda/Nm$ where N is the total number of grating slits and m is the order. Then we have

$$\mathrm{R} = \frac{\lambda}{\Delta\lambda} = Nm. \qquad \textbf{(35–19)}$$

The larger the value of R, the closer two wavelengths can be resolvable. If R is given, the minimum separation $\Delta\lambda$ between two wavelengths near λ, is (by Eq. 35–18)

$$\Delta\lambda = \frac{\lambda}{\mathrm{R}}.$$

EXAMPLE 35–11 **Resolving two close lines.** Yellow sodium light, which consists of two wavelengths, $\lambda_1 = 589.00\,\mathrm{nm}$ and $\lambda_2 = 589.59\,\mathrm{nm}$, falls on a 7500-slit/cm diffraction grating. Determine (*a*) the maximum order m that will be present for these sodium lines, (*b*) the width of grating necessary to resolve the two sodium lines.

APPROACH We first find $d = 1\,\mathrm{cm}/7500 = 1.33 \times 10^{-6}\,\mathrm{m}$, and then use Eq. 35–13 to find m. For (*b*) we use Eqs. 35–18 and 35–19.

SOLUTION (*a*) The maximum value of m at $\lambda = 589\,\mathrm{nm}$, using Eq. 35–13 with $\sin\theta \leq 1$, is

$$m = \frac{d}{\lambda}\sin\theta \leq \frac{d}{\lambda} = \frac{1.33 \times 10^{-6}\,\mathrm{m}}{5.89 \times 10^{-7}\,\mathrm{m}} = 2.26,$$

so $m = 2$ is the maximum order present.

(*b*) The resolving power needed is

$$\mathrm{R} = \frac{\lambda}{\Delta\lambda} = \frac{589\,\mathrm{nm}}{0.59\,\mathrm{nm}} = 1000.$$

From Eq. 35–19, the total number N of slits needed for the $m = 2$ order is $N = \mathrm{R}/m = 1000/2 = 500$, so the grating need only be $500/7500\,\mathrm{cm}^{-1} = 0.0667\,\mathrm{cm}$ wide. A typical grating is a few centimeters wide, and so will easily resolve the two lines.

35–11 X-Rays and X-Ray Diffraction

In 1895, W. C. Roentgen (1845–1923) discovered that when electrons were accelerated by a high voltage in a vacuum tube and allowed to strike a glass or metal surface inside the tube, fluorescent minerals some distance away would glow, and photographic film would become exposed. Roentgen attributed these effects to a new type of radiation (different from cathode rays). They were given the name **X-rays** after the algebraic symbol x, meaning an unknown quantity. He soon found that X-rays penetrated through some materials better than others, and within a few weeks he presented the first X-ray photograph (of his wife's hand). The production of X-rays today is usually done in a tube (Fig. 35–26) similar to Roentgen's, using voltages of typically 30 kV to 150 kV.

Investigations into the nature of X-rays indicated they were not charged particles (such as electrons) since they could not be deflected by electric or magnetic fields. It was suggested that they might be a form of invisible light. However, they showed no diffraction or interference effects using ordinary gratings.

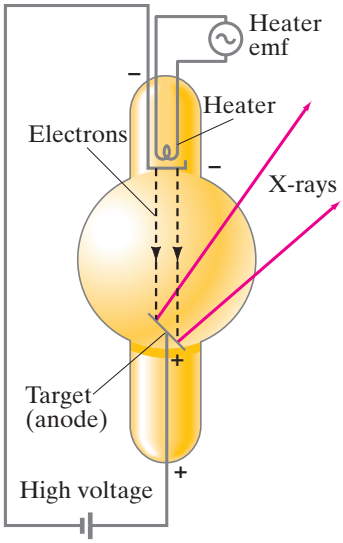


FIGURE 35–26 X-ray tube. Electrons emitted by a heated filament in a vacuum tube are accelerated by a high voltage. When they strike the surface of the anode, the "target," X-rays are emitted.

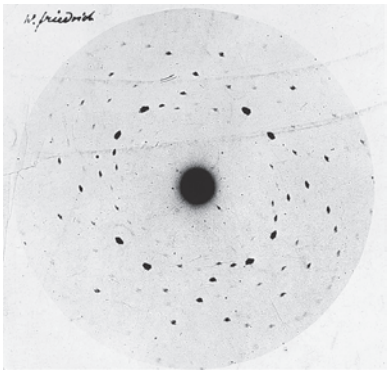
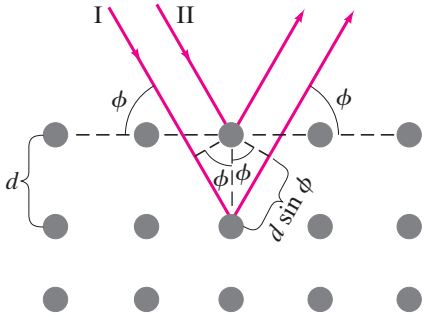


FIGURE 35–27 This X-ray diffraction pattern is one of the first observed by Max von Laue in 1912 when he aimed a beam of X-rays at a zinc sulfide crystal. The diffraction pattern was detected directly on a photographic plate.

FIGURE 35–28 X-ray diffraction by a crystal.



Indeed, if their wavelengths were much smaller than the typical grating spacing of $10^{-6}\,\text{m}$ $(=10^{3}\,\text{nm})$, no effects would be expected. Around 1912, Max von Laue (1879–1960) suggested that if the atoms in a crystal were arranged in a regular array (see Fig. 17–2a), such a crystal might serve as a diffraction grating for very short wavelengths on the order of the spacing between atoms, estimated to be about $10^{-10}\,\text{m}$ $(=10^{-1}\,\text{nm})$. Experiments soon showed that X-rays scattered from a crystal did indeed show the peaks and valleys of a diffraction pattern (Fig. 35–27). Thus it was shown, in a single blow, that X-rays have a wave nature and that atoms are arranged in a regular way in crystals. Today, X-rays are recognized as electromagnetic radiation with wavelengths in the range of about $10^{-2}\,\text{nm}$ to 10 nm, the range readily produced in an X-ray tube.

X-Ray Diffraction

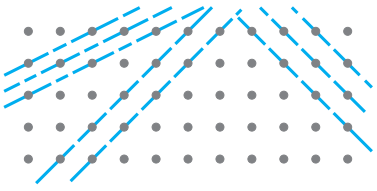
We saw in Section 35–5 that light of shorter wavelength provides greater resolution when we are examining an object microscopically. Since X-rays have much shorter wavelengths than visible light, they should in principle offer much greater resolution. However, there seems to be no effective material to use as lenses for the very short wavelengths of X-rays. Instead, the clever but complicated technique of **X-ray diffraction** (or **crystallography**) has proved very effective for examining the microscopic world of atoms and molecules. In a simple crystal such as NaCl, the atoms are arranged in an orderly cubical fashion, Fig. 35–28, with atoms spaced a distance d apart. Suppose that a beam of X-rays is incident on the crystal at an angle ϕ to the surface, and that the two rays shown are reflected from two subsequent planes of atoms as shown. The two rays will constructively interfere if the extra distance ray I travels is a whole number of wavelengths farther than the distance ray II travels. This extra distance is $2d\,\sin\phi$. Therefore, constructive interference will occur when

$$m\lambda = 2d\sin\phi, \qquad m = 1, 2, 3, \cdots, \tag{35–20}$$

where m can be any integer. (Notice that ϕ is *not* the angle with respect to the normal to the surface.) This is called the **Bragg equation** after W. L. Bragg (1890–1971), who derived it and who, together with his father W. H. Bragg (1862–1942), developed the theory and technique of X-ray diffraction by crystals in 1912–1913. If the X-ray wavelength is known and the angle ϕ is measured, the distance d between atoms can be obtained. This is the basis for X-ray crystallography.

EXERCISE G When X-rays of wavelength $0.10\times10^{-9}\,\text{m}$ are scattered from a sodium chloride crystal, a second-order diffraction peak is observed at 21°. What is the spacing between the planes of atoms for this scattering?

FIGURE 35–29 X-rays can be diffracted from many possible planes within a crystal.



Actual X-ray diffraction patterns are quite complicated. First of all, a crystal is a three-dimensional object, and X-rays can be diffracted from different planes at different angles within the crystal, as shown in Fig. 35–29. Although the analysis is complex, a great deal can be learned from X-ray diffraction about any substance that can be put in crystalline form.

X-ray diffraction has also been very useful in determining the structure of biologically important molecules, such as the double helix structure of DNA, worked out by James Watson and Francis Crick in 1953. See Fig. 35–30, and for models of the double helix, Figs. 21–48a and 21–49. Around 1960, the first detailed structure of a protein molecule, myoglobin, was elucidated with the aid of X-ray diffraction.

Soon the structure of an important constituent of blood, hemoglobin, was worked out, and since then the structures of a great many molecules have been determined with the help of X-rays.

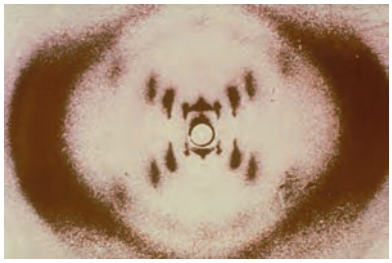


FIGURE 35–30 X-ray diffraction photo of DNA molecules taken by Rosalind Franklin in the early 1950s. The cross of spots suggested that DNA is a helix.

*35–12 X-Ray Imaging and Computed Tomography (CT Scan)

*Normal X-Ray Image

For a conventional medical or dental X-ray photograph, the X-rays emerging from the tube (Fig. 35–26) pass through the body and are detected on photographic film, a digital sensor, or a fluorescent screen, Fig. 35–31. The rays travel in very nearly straight lines through the body with minimal deviation since at X-ray wavelengths there is little diffraction or refraction. There is absorption (and scattering), however; and the difference in absorption by different structures in the body is what gives rise to the image produced by the transmitted rays. The less the absorption, the greater the transmission and the darker the film. The image is, in a sense, a "shadow" of what the rays have passed through.

An X-ray image is *not* produced by focusing rays with lenses as for the optical instruments discussed earlier in Chapter 33.

PHYSICS APPLIED
Normal X-ray image is a sort of shadow (no lenses are involved)

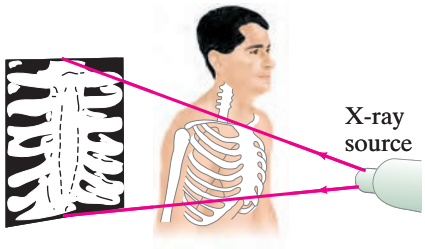


FIGURE 35–31 Conventional X-ray imaging, which is essentially shadowing.

*Tomography Images (CT)

In conventional X-ray images, a body's thickness is projected onto film or a sensor; structures overlap and in many cases are difficult to distinguish. In the 1970s, a revolutionary X-ray technique was developed called **computed tomography** (CT), which produces an image of a *slice* through the body. (The word **tomography** comes from the Greek: *tomos* = slice, *graph* = picture.) Structures and lesions previously impossible to visualize can now be seen with remarkable clarity. The principle behind CT is shown in Fig. 35–32: a thin collimated (parallel) beam of X-rays passes through the body to a detector that measures the transmitted intensity. Measurements are made at a large number of points as the source and detector are moved past the body together. The apparatus is then rotated slightly about the body axis and again scanned; this is repeated at (perhaps) 1° intervals for 180°. The intensity of the transmitted beam for the many points of each scan, and for each angle, is sent to a computer that reconstructs the image of the slice. Note that the imaged slice is perpendicular to the long axis of the body. For this reason, CT is sometimes called **computerized axial tomography** (CAT), although the abbreviation CAT, as in CAT scan, can also be read as **computer-assisted tomography**.

PHYSICS APPLIED
Computed tomography images (CT or CAT scans)

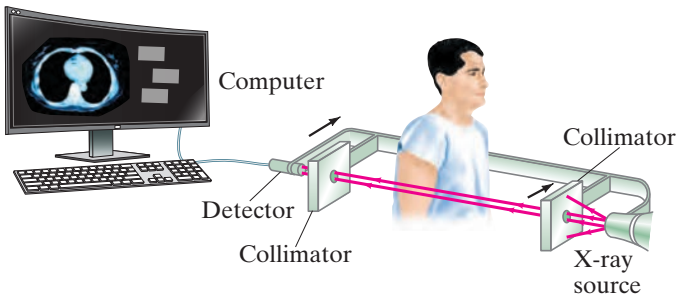
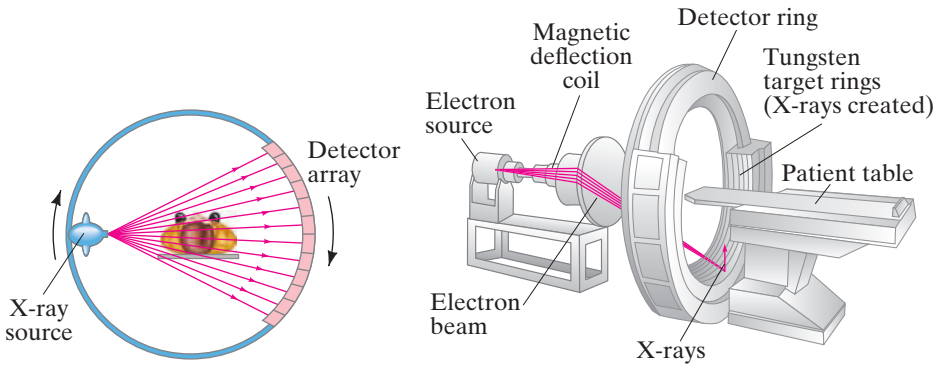
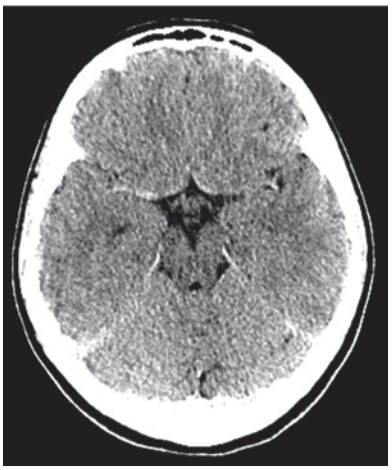


FIGURE 35–32 Tomographic imaging: the X-ray source and detector move together across the body, the transmitted intensity being measured at a large number of points. Then the "source-detector" assembly is rotated slightly (say, 1°) around a vertical axis, and another scan is made. This process is repeated for perhaps 180°. The computer reconstructs the image of the slice and it is presented on a computer screen. Normally the patient is not standing but lying down.

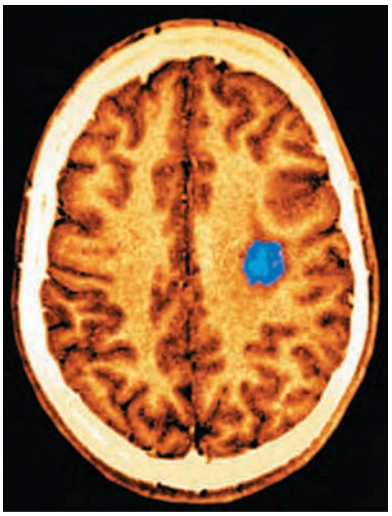


(a) (b)

FIGURE 35–33 (a) Fan-beam scanner. Rays transmitted through the entire body are measured simultaneously at each angle. The source and detector rotate to take measurements at different angles. In another type of fan-beam scanner, there are detectors around the entire 360° of the circle which remain fixed as the source moves. (b) In still another type, a beam of electrons from a source is directed by magnetic fields at tungsten targets surrounding the patient.



(a)



(b)

FIGURE 35–34 Two CT images, with different resolutions, each showing a cross section of a brain. Photo (a) is of low resolution; photo (b), of higher resolution, shows a brain tumor, and uses false color to highlight it.

The use of a single detector as in Fig. 35–32 would require a few minutes for the many scans needed to form a complete image. Much faster scanners use a fan beam, Fig. 35–33a, in which beams passing through the entire cross section of the body are detected simultaneously by many detectors. The source and detectors are then rotated about the patient, and an image requires only a few seconds. Even faster, and therefore useful for heart scans, are fixed source machines wherein an electron beam is directed (by magnetic fields) to tungsten targets surrounding the patient, creating the X-rays. See Fig. 35–33b.

*Image Formation

But how is the image formed? We can think of the slice to be imaged as being divided into many tiny picture elements (or **pixels**), which could be squares (as in Fig. 34–36). For CT, the width of each pixel is chosen according to the width of the detectors and/or the width of the X-ray beams, and this determines the resolution of the image, which might be $\frac{1}{2}$ mm. An X-ray detector measures the intensity of the transmitted beam. Subtracting this value from the intensity of the beam at the source yields the total absorption (called a "projection") along that beam line. Complicated mathematical techniques are used to analyze all the absorption projections for the huge number of beam scans measured (see the next Subsection), obtaining the absorption at each pixel and assigning each a "grayness value" according to how much radiation was absorbed. The image is made up of tiny spots (pixels) of varying shades of gray. Often the amount of absorption is color-coded. The colors in the resulting **false-color** image have nothing to do, however, with the actual color of the object. The actual images are monochromatic (various shades of gray, depending on the absorption). Only *visible* light has color; X-rays do not.

Figure 35–34 illustrates what actual CT images look like. It is generally agreed that CT scanning has revolutionized some areas of medicine by providing much less invasive, and/or more accurate, diagnosis.

Computed tomography can also be applied to ultrasound imaging (Section 16–9) and to emissions from radioisotopes and nuclear magnetic resonance (Sections 42–9 and 42–10).

*Tomographic Image Reconstruction

How can the "grayness" of each pixel be determined even though all we can measure is the total absorption along each beam line in the slice? It can be done only by using the many beam scans made at a great many different angles. Suppose the image is to be an array of 100×100 elements for a total of 10^4 pixels. If we have 100 detectors and measure the absorption projections at 100 different angles, then we get 10^4 pieces of information. From this information, an image can be reconstructed, but not precisely. If more angles are measured, the reconstruction of the image can be done more accurately.

To suggest how mathematical reconstruction is done, we consider a very simple case using the **iterative** technique ("to iterate" is from the Latin "to repeat"). Suppose our sample slice is divided into the simple 2×2 pixels as shown in Fig. 35–35. The number inside each pixel represents the amount of absorption by the material in that area (say, in tenths of a percent): that is, 4 represents twice as much absorption as 2. But we cannot directly measure these values—they are the unknowns we want to solve for. All we can measure are the *projections*—the total absorption along each beam line—and these are shown in Fig. 35–35 outside the yellow squares as the sum of the absorptions for the pixels along each line at four different angles. These projections (given at the tip of each arrow) are what we can measure, and we now want to work back from them to see how close we can get to the true absorption value for each pixel. We start our analysis with each pixel being assigned a zero value, Fig. 35–36a. In the iterative technique, we use the projections to estimate the absorption value in each square, and repeat for each angle. The angle 1 projections are 7 and 13 (Fig. 35–35). We divide each of these equally between their two squares: each square in the left column of Fig. 35–36a gets $3\frac{1}{2}$ (half of 7), and each square in the right column gets $6\frac{1}{2}$ (half of 13) as shown in Fig. 35–36b.

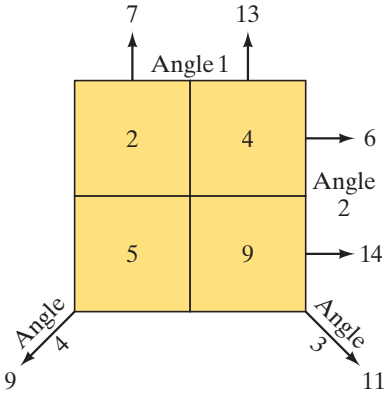


FIGURE 35–35 A simple 2×2 image showing true absorption values (inside the squares) and measured projections.

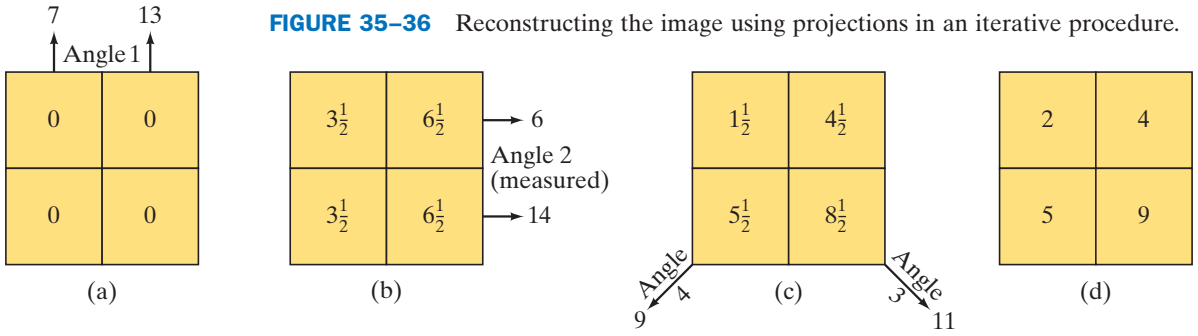


FIGURE 35–36 Reconstructing the image using projections in an iterative procedure.

Next we use the projections at angle 2. We calculate the difference between the measured projections at angle 2 (6 and 14) and the projections based on the previous estimate (top row: $3\frac{1}{2} + 6\frac{1}{2} = 10$; same for bottom row). Then we distribute this difference equally to the squares in that row. For the top row, we have

$$3\tfrac{1}{2} + \frac{6 - 10}{2} = 1\tfrac{1}{2} \quad \text{and} \quad 6\tfrac{1}{2} + \frac{6 - 10}{2} = 4\tfrac{1}{2};$$

and for the bottom row,

$$3\tfrac{1}{2} + \frac{14 - 10}{2} = 5\tfrac{1}{2} \quad \text{and} \quad 6\tfrac{1}{2} + \frac{14 - 10}{2} = 8\tfrac{1}{2}.$$

These values are inserted as shown in Fig. 35–36c. Next, the projection at angle 3 (= 11), combined with the difference as above, gives

$$\text{(upper left)} \quad 1\tfrac{1}{2} + \frac{11 - 10}{2} = 2 \quad \text{and} \quad \text{(lower right)} \quad 8\tfrac{1}{2} + \frac{11 - 10}{2} = 9;$$

and then for angle 4 we have

$$\text{(lower left)} \quad 5\tfrac{1}{2} + \frac{9 - 10}{2} = 5 \quad \text{and} \quad \text{(upper right)} \quad 4\tfrac{1}{2} + \frac{9 - 10}{2} = 4.$$

The result, shown in Fig. 35–36d, corresponds exactly to the true values. (In real situations, the true values are not known, which is why these computer techniques are required.) To obtain these numbers exactly, we used six pieces of information (two each at angles 1 and 2, one each at angles 3 and 4). For the much larger number of pixels used for actual images, exact values are generally not attained. Many iterations may be needed, and the calculation is considered sufficiently precise when the difference between calculated and measured projections is sufficiently small. The above example illustrates the "convergence" of the process: the first iteration (b to c in Fig. 35–36) changed the values by 2, the last iteration (c to d) by only $\frac{1}{2}$.

*35–13 Specialty Microscopes and Contrast

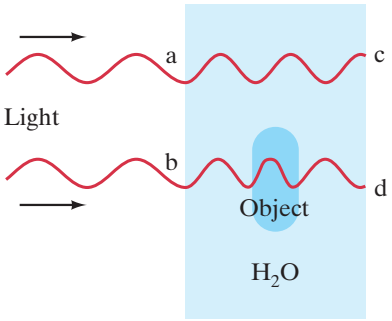


FIGURE 35–37 Object—say, a bacterium—in a water solution.

All the resolving power a microscope can attain will be useless if the object to be seen cannot be distinguished from the background. The difference in brightness between the image of an object and the image of the surroundings is called **contrast**. Achieving high contrast is an important problem in microscopy and other forms of imaging. The problem arises in biology, for example, because cells consist largely of water and are almost uniformly transparent to light. We now briefly discuss two special types of microscope that can increase contrast: the interference and phase-contrast microscopes.

PHYSICS APPLIED
Interference microscope

An **interference microscope** makes use of the wave properties of light in a direct way to increase contrast in a transparent object. Consider a transparent object—say, a bacterium in water (Fig. 35–37). Coherent light enters uniformly from the left and is in phase at all points such as a and b. If the object is as transparent as the water, the beam leaving at d will be as bright as that at c. There will be no contrast and the object will not be seen. However, if the object's refractive index is slightly different from that of the surrounding medium, the wavelength within the object will be altered as shown. Hence light waves at points c and d will differ in phase, if not in amplitude. The interference microscope changes this difference in phase into a difference of amplitude which our eyes can detect. Light that passes through the sample is superimposed onto a reference beam that does not pass through the object, so that the beams interfere. One way of doing this is shown in Fig. 35–38. Light from a source is split into two equal beams by a half-silvered mirror, MS_1. One beam passes through the object, and the second (comparison beam) passes through an identical system without the object. The two meet again and are superposed by the half-silvered mirror MS_2 before entering the eyepiece and eye. The path length (and amplitude) of the comparison beam is adjustable so that the background can be dark; that is, full destructive interference occurs. Light passing through the object (beam bd in Fig. 35–37) will also interfere with the comparison beam. But because of its different phase, the interference will not be completely destructive. Thus it will appear brighter than the background. Where the object varies in thickness, the phase difference between beams ac and bd in Fig. 35–37 will be different, thus affecting the amount of interference. Hence *variation in the thickness of the object will appear as variations in brightness in the image*.

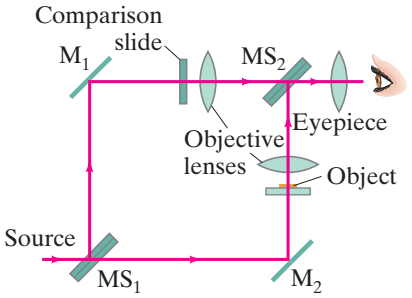


FIGURE 35–38 Diagram of an interference microscope.

PHYSICS APPLIED
Phase-contrast microscope

A **phase-contrast microscope** also makes use of interference and differences in phase to produce a high-contrast image. Contrast is achieved by a circular glass *phase plate* that has a groove (or a raised portion) in the shape of a ring, positioned so undeviated source rays pass through it, but rays deviated by the object do not pass through this ring. Because the rays deviated by the object travel through a different thickness of glass than the undeviated source rays, the two can be out of phase and can interfere destructively at the object image plane. Thus the image of the object can contrast sharply with the background. Phase-contrast microscope images tend to have "halos" around them (as a result of diffraction from the phase-plate opening), so care must be taken in the interpretation of images.

Summary

Diffraction refers to the fact that light, like other waves, bends around objects it passes, and spreads out after passing through narrow slits. This bending gives rise to a **diffraction pattern** due to interference between rays of light that travel different distances.

Light passing through a very narrow slit of width D (on the order of the wavelength λ) will produce a pattern with a bright central maximum of half-width θ given by

$$\sin\theta = \frac{\lambda}{D}, \qquad \textbf{(35–1)}$$

flanked by fainter lines to either side.

The minima in the diffraction pattern occur at

$$D\sin\theta = m\lambda \qquad \textbf{(35–2)}$$

where $m = 1, 2, 3, \cdots$, but not $m = 0$ (for which the pattern has its strongest maximum).

[*The **intensity** at any point in the single-slit diffraction pattern can be calculated using **phasor** diagrams. The same technique can be used to determine the intensity of the pattern produced by two slits.]

[*The pattern for two-slit interference can be described as a series of maxima due to interference of light from the two slits, modified by an "envelope" due to diffraction at each slit.]

The wave nature of light limits the sharpness or **resolution** of images. Because of diffraction, it is *not possible to discern details smaller than the wavelength* of the radiation being used. The useful magnification of a light microscope is limited by diffraction to about $500\times$.

A **diffraction grating** consists of many parallel slits or lines, separated from each other by a distance d. The peaks of constructive interference occur at angles θ given by

$$\sin\theta = \frac{m\lambda}{d}, \qquad \textbf{(35–13)}$$

where $m = 0, 1, 2, \cdots$. The peaks of constructive interference are much brighter and sharper for a diffraction grating than for a two-slit apparatus. [*Peak width is inversely proportional to the total number of slits in the grating.]

A diffraction grating (or a prism) is used in a **spectrometer** to separate different colors and observe **line spectra**. For a given order m, θ depends on λ. Precise determination of wavelength can be done with a spectrometer by careful measurement of θ.

X-rays are a form of electromagnetic radiation of very short wavelength. They are produced when high-speed electrons, accelerated by high voltage in an evacuated tube, strike a glass or metal target.

[***Computed tomography** (CT or CAT scan) uses many narrow X-ray beams through a section of the body to construct an image of that cross-section.]

Questions

1. Radio waves and visible light are both electromagnetic waves. Why can a radio receive a signal behind a hill when we cannot see the transmitting antenna?

2. Hold one hand close to your eye and focus on a distant light source through a narrow slit between two fingers. (Adjust your fingers to obtain the best pattern.) Describe the pattern that you see.

3. Explain why diffraction patterns are more difficult to observe with an extended light source than for a point source. Compare also a monochromatic source to white light.

4. For diffraction by a single slit, what is the effect of increasing (*a*) the slit width, (*b*) the wavelength?

5. Describe the single-slit diffraction pattern produced when white light falls on a slit having a width of (*a*) 50 nm, (*b*) 50,000 nm.

6. What happens to the diffraction pattern of a single slit if the whole apparatus is immersed in (*a*) water, (*b*) a vacuum, instead of in air?

7. In the single-slit diffraction pattern, why does the first off-center maximum not occur at exactly $\sin\theta = \frac{3}{2}\lambda/D$?

8. Discuss the similarities, and differences, of double-slit interference and single-slit diffraction.

***9.** Figure 35–11 shows a two-slit interference pattern for the case when d is larger than D. Can the reverse case occur, when d is less than D?

***10.** When both diffraction and interference are taken into account in the double-slit experiment, discuss the effect of increasing (*a*) the wavelength, (*b*) the slit separation, (*c*) the slit width.

11. Does diffraction limit the resolution of images formed by (*a*) spherical mirrors, (*b*) plane mirrors?

12. Do diffraction effects occur for virtual as well as real images?

13. Give at least two advantages for the use of large reflecting mirrors in astronomical telescopes.

14. Atoms have diameters of about 10^{-8} cm. Can visible light be used to "see" an atom? Explain.

15. Which color of visible light would give the best resolution in a microscope? Explain.

16. Could a diffraction grating just as well be called an interference grating? Discuss.

17. Suppose light consisting of wavelengths between 400 nm and 700 nm is incident normally on a diffraction grating. For what orders (if any) would there be overlap in the observed spectrum? Does your answer depend on the slit spacing?

18. What is the difference in the interference patterns formed by two slits 10^{-4} cm apart as compared to a diffraction grating containing 10^4 slits/cm?

19. By what factor can you improve resolution, other things being equal, if you use blue light ($\lambda = 450$ nm) rather than red (700 nm)?

20. The 300-meter radiotelescope in Arecibo, Puerto Rico (Fig. 35–17), is the world's largest radiotelescope, but many other radiotelescopes are also very large. Why are radiotelescopes so big? Why not make optical telescopes that are equally large? (The largest optical telescopes have diameters of about 10 meters.)

21. For a diffraction grating, what is the advantage of (*a*) many slits, (*b*) closely spaced slits?

22. White light strikes (*a*) a diffraction grating and (*b*) a prism. A rainbow appears on a wall just below the direction of the horizontal incident beam in each case. What is the color of the top of the rainbow in each case? Explain.

MisConceptual Questions

1. Blue light of wavelength λ passes through a single slit of width D and forms a diffraction pattern on a screen. If we replace the blue light by red light of wavelength 2λ, we can retain the original diffraction pattern if we change the slit width
(*a*) to $D/4$. (*c*) to $4D$. (*e*) not at all.
(*b*) to $D/2$. (*d*) to $2D$.

2. Imagine holding a circular disk in a beam of monochromatic light (Fig. 35–39). If diffraction occurs at the edge of the disk, the center of the shadow is
(*a*) darker than the rest of the shadow.
(*b*) a bright spot.
(*c*) bright or dark, depending on the wavelength.
(*d*) bright or dark, depending on the distance to the screen.

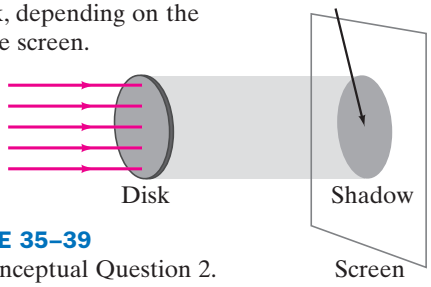


FIGURE 35–39
MisConceptual Question 2.

3. If you hold two fingers very close together and look at a bright light, you see lines between your fingers. What is happening?
(*a*) You are holding your fingers too close to your eye to be able to focus on it.
(*b*) You are seeing a diffraction pattern.
(*c*) This is a quantum-mechanical tunneling effect.
(*d*) The brightness of the light is overwhelming your eye.

4. A double-slit experiment yields an interference pattern due to the path length difference from light traveling through one slit versus the other. Why does a single slit show a diffraction pattern?
(*a*) There is a path length difference from waves originating at different parts of the slit.
(*b*) The wavelength of the light is shorter than the slit.
(*c*) The light passing through the slit interferes with light that does not pass through.
(*d*) The single slit must have something in the middle of it, causing it to act like a double slit.

5. Red light from a laser pointer shines directly on a narrow slit whose width is about 5 wavelengths of this light. A white screen is placed 0.1 m behind the slit. What is observed on the screen?
(*a*) Light that is more orange, because the wavelength shifts.
(*b*) Light that is closer to infrared, because the wavelength shifts.
(*c*) A tiny red rectangle.
(*d*) A red rectangle and dimmer red stripes on either side of it.
(*e*) No light, because the slits are too narrow.

6. A single-slit diffraction pattern is formed on a distant screen by a narrow vertical slit. If the width of the slit is reduced, what happens to the width of the central bright fringe?
(*a*) It becomes narrower.
(*b*) It becomes wider.
(*c*) It becomes brighter.
(*d*) It becomes taller.
(*e*) It doesn't change at all.

7. If someone is standing on a sidewalk around a corner from you, what is the main reason you can hear him speaking but can't see him?
(*a*) Sound travels farther in air than light does.
(*b*) Sound can travel through walls, but light cannot.
(*c*) Sound waves have long enough wavelengths to bend around a corner; light wavelengths are too short to bend much.
(*d*) Sound waves reflect off walls, but light cannot.

8. Light passes through a slit that is about 5×10^{-3} m high and 5×10^{-7} m wide. The central bright light visible on a distant screen will be
(*a*) about 5×10^{-3} m high and about 5×10^{-7} m wide.
(*b*) about 5×10^{-3} m high and wider than 5×10^{-7} m.
(*c*) about 5×10^{-3} m high and narrower than 5×10^{-7} m.
(*d*) taller than 5×10^{-3} m high and wider than 5×10^{-7} m.
(*e*) taller than 5×10^{-3} m high and about 5×10^{-7} m wide.

9. A point source of yellow light shines on a slit. If the width of the slit is increased by 20%, how will the diffraction pattern change?
(*a*) The pattern will shift to longer wavelengths.
(*b*) The pattern will shift to shorter wavelengths.
(*c*) The pattern spacing will spread out.
(*d*) The pattern spacing will be compressed.
(*e*) There will be no change.

10. Which of the following statements is true?
(*a*) A larger-diameter lens can better resolve two distant points.
(*b*) Red light can better resolve two distant points than blue light can.
(*c*) It is easier to resolve distant objects than nearer objects.
(*d*) Objects that are closer together are easier to resolve than objects that are farther apart.

11. It would be impossible to build a microscope that could use visible light to see the molecular structure of a crystal because
(*a*) lenses with enough magnification cannot be made.
(*b*) lenses cannot be ground with fine enough precision.
(*c*) lenses cannot be placed in the correct place with enough precision.
(*d*) diffraction limits the resolving power to about the size of the wavelength of the light used.
(*e*) More than one of the above is correct.

12. The size of the smallest object that can be clearly seen by the naked eye is determined by
(*a*) the diameter of the pupil.
(*b*) the light intensity.
(*c*) the eye's far point.
(*d*) the eye's near point.

13. The resolving power of a microscope is greatest when the object being observed is illuminated by
(*a*) ultraviolet light.
(*b*) infrared light.
(*c*) visible light.
(*d*) radio waves.

Problems

[Note: Assume light passing through slits is in phase, unless stated otherwise.]

35–1 Single-Slit Diffraction

1. (I) If 680-nm light falls on a slit 0.0385 mm wide, what is the angular width of the central diffraction peak?

2. (I) Monochromatic light falls on a slit that is 2.60×10^{-3} mm wide. If the angle between the first dark fringes on either side of the central maximum is 29.0° (dark fringe to dark fringe), what is the wavelength of the light used?

3. (II) When violet light of wavelength 415 nm falls on a single slit, it creates a central diffraction peak that is 9.60 cm wide on a screen that is 2.85 m away. How wide is the slit?

4. (II) Light of wavelength 580 nm falls on a slit that is 3.50×10^{-3} mm wide. Estimate how far the first brightest diffraction fringe is from the strong central maximum if the screen is 10.0 m away.

5. (II) Consider microwaves which are incident perpendicular to a metal plate which has a 1.8-cm slit in it. Determine the angles at which there are diffraction minima for wavelengths of (*a*) 0.50 cm, (*b*) 1.0 cm, and (*c*) 3.0 cm.

6. (II) Monochromatic light of wavelength 633 nm falls on a slit. If the angle between the first bright fringes on either side of the central maximum is 32°, estimate the slit width.

7. (II) If a slit diffracts 580-nm light so that the diffraction maximum is 6.0 cm wide on a screen 2.50 m away, what will be the width of the diffraction maximum for light with a wavelength of 460 nm?

8. (II) (*a*) For a given wavelength λ, what is the minimum slit width for which there will be no diffraction minimum? (*b*) What is the minimum slit width so that no visible light exhibits a diffraction minimum?

9. (II) When blue light of wavelength 440 nm falls on a single slit, the first dark bands on either side of center are separated by 51.0°. Determine the width of the slit.

10. (II) A single slit 0.10 mm wide is illuminated by 450-nm light. What is the width of the central maximum in the diffraction pattern on a screen 5.0 m away?

11. (II) Coherent light from a laser diode is emitted through a rectangular area $3.0\,\mu\text{m} \times 1.5\,\mu\text{m}$ (horizontal-by-vertical). If the laser light has a wavelength of 680 nm, determine the angle between the first diffraction minima on either side, (*a*) above and below the central maximum, (*b*) to the left and right of the central maximum.

*35–2 Intensity, Single-Slit Diffraction Pattern

***12.** (II) If you double the width of a single slit, the intensity of the light passing through the slit is doubled. (*a*) Show, however, that the intensity at the center of the screen increases by a factor of 4. (*b*) Explain why this does not violate conservation of energy.

***13.** (II) Light of wavelength 750 nm passes through a slit 1.0 μm wide and a single-slit diffraction pattern is formed vertically on a screen 25 cm away. Determine the light intensity I 18 cm above the central maximum, expressed as a fraction of the central maximum's intensity I_0.

***14.** (III) (*a*) Explain why the secondary maxima in the single-slit diffraction pattern do not occur precisely at $\beta/2 = (m + \frac{1}{2})\pi$ where $m = 1, 2, 3, \cdots$. (*b*) By differentiating Eq. 35–7 with respect to β show that the secondary maxima occur when $\beta/2$ satisfies the relation $\tan(\beta/2) = \beta/2$. (*c*) Carefully and precisely plot the curves $y = \beta/2$ and $y = \tan\beta/2$. From their intersections, determine the values of β for the first and second secondary maxima. What is the percent difference from $\beta/2 = (m + \frac{1}{2})\pi$?

*35–3 Intensity, Double-Slit Diffraction

***15.** (II) If a double-slit pattern contains exactly eleven fringes in the central diffraction peak, what can you say about the slit width and separation? Assume the first diffraction minimum occurs at an interference minimum.

***16.** (II) Design a double-slit apparatus so that the central diffraction peak contains precisely fifteen fringes. Assume the first diffraction minimum occurs at (*a*) an interference minimum, (*b*) an interference maximum.

***17.** (II) In a double-slit experiment, if the central diffraction peak contains 13 interference fringes (Fig. 35–11c), how many fringes are contained within each secondary diffraction peak (between $m = +1$ and $+2$ in Eq. 35–2a)? Assume the first diffraction minimum occurs at an interference minimum.

***18.** (II) Two 0.010-mm-wide slits are 0.030 mm apart (center to center). Determine (*a*) the spacing between interference fringes for 520-nm light on a screen 1.0 m away and (*b*) the distance between the two diffraction minima on either side of the central maximum of the envelope.

***19.** (II) Suppose $d = D$ in a double-slit apparatus, so that the two slits merge into one slit of width $2D$. Show that Eq. 35–9 reduces to the correct equation for single-slit diffraction.

***20.** (II) In a double-slit experiment, let $d = 5.00D = 40.0\lambda$. Compare (as a ratio) the intensity of the third-order interference maximum with that of the zero-order maximum.

***21.** (II) How many fringes are contained in the central diffraction peak for a double-slit pattern if (*a*) $d = 2.00D$, (*b*) $d = 4.50D$, (*c*) $d = 7.80D$, (*d*) $d = 16.0D$?

***22.** (III) (*a*) Derive an expression for the intensity in the interference pattern for three equally spaced slits. Express in terms of $\delta = 2\pi d \sin\theta/\lambda$ where d is the distance between adjacent slits and assume the slit width $D \approx \lambda$. (*b*) Show that there is only one secondary maximum between principal peaks.

35–5 to 35–7 Resolution Limits

23. (I) What is the angular resolution limit (degrees) set by diffraction for the 100-inch (254-cm mirror diameter) Mt. Wilson telescope ($\lambda = 560$ nm)?

24. (I) What is the resolving power of a microscope ($\lambda = 550$ nm) with a 5-mm-diameter objective which has $f = 8$ mm?

25. (II) Two stars 18 light-years away are barely resolved by a 66-cm (mirror diameter) telescope. How far apart are the stars? Assume $\lambda = 550$ nm and that the resolution is limited by diffraction.

26. (II) The nearest neighboring star to the Sun is about 4 light-years away. If a planet happened to be orbiting this star at an orbital radius equal to that of the Earth–Sun distance, what minimum diameter would an Earth-based telescope's aperture have to be in order to obtain an image that resolved this star–planet system? Assume the light emitted by the star and planet has a wavelength of 550 nm.

27. (II) If you could shine a very powerful flashlight beam toward the Moon, estimate the diameter of the beam when it reaches the Moon. Assume that the beam leaves the flashlight through a 5.0-cm aperture, that its white light has an average wavelength of 550 nm, and that the beam spreads due to diffraction only.

28. (II) Suppose that you wish to construct a telescope that can resolve features 6.5 km across on the Moon, 384,000 km away. You have a 2.0-m-focal-length objective lens whose diameter is 11.0 cm. What focal-length eyepiece is needed if your eye can resolve objects 0.10 mm apart at a distance of 25 cm? What is the resolution limit set by the size of the objective lens (that is, by diffraction)? Use $\lambda = 560$ nm.

29. (II) The normal lens on a 35-mm camera has a focal length of 50.0 mm. Its aperture diameter varies from a maximum of 25 mm $(f/2)$ to a minimum of 3.0 mm $(f/16)$. Determine the resolution limit set by diffraction for $(f/2)$ and $(f/16)$. Specify as the number of lines per millimeter resolved on the detector or film. Take $\lambda = 560$ nm.

30. (II) A microscope 14.0 cm long has an objective lens of diameter 6.0 mm and focal length 8.25 mm. Its magnifying power is approximately 275×. What is its resolving power?

31. (II) Two ants stand 2 cm apart. (*a*) Assuming diffraction-limited vision, at what distance could a human eye with a pupil diameter of 4 mm tell that there are two separate ants? Assume light of 550-nm wavelength. (*b*) Does this result agree with your everyday experience? Explain.

32. (II) When driving at night, your eyes' pupils have dilated to a 7.5-mm diameter. If your vision is diffraction limited, what would be the greatest distance at which you could resolve the two headlights of an oncoming car, which are spaced 1.5 m apart? Assume a wavelength of 550 nm for the light.

35–8 and 35–9 Diffraction Grating, Spectroscopy

33. (I) At what angle will 480-nm light produce a second-order maximum when falling on a grating whose slits are 1.25×10^{-3} cm apart?

34. (I) A source produces first-order lines when incident normally on a 12,000-slit/cm diffraction grating at angles 28.8°, 36.7°, 38.6°, and 47.9°. What are the wavelengths?

35. (I) A 3800-slit/cm grating produces a third-order fringe at a 35.0° angle. What wavelength of light is being used?

36. (I) A grating has 6300 slits/cm. How many spectral orders can be seen (400 to 700 nm) when it is illuminated by white light?

37. (II) How many slits per centimeter does a grating have if the third order occurs at an 18.0° angle for 650-nm light?

38. (II) White light containing wavelengths from 410 nm to 750 nm falls on a grating with 7800 slits/cm. How wide is the first-order spectrum on a screen 2.50 m away?

39. (II) A diffraction grating has 6.5×10^5 slits/m. Find the angular spread in the second-order spectrum between red light of wavelength 7.0×10^{-7} m and blue light of wavelength 4.5×10^{-7} m.

40. (II) A tungsten–halogen bulb emits a continuous spectrum of ultraviolet, visible, and infrared light in the wavelength range 360 nm to 2000 nm. Assume that the light from a tungsten–halogen bulb is incident on a diffraction grating with slit spacing d and that the first-order brightness maximum for the wavelength of 1200 nm occurs at angle θ. What other wavelengths within the spectrum of incident light will produce a brightness maximum at this same angle θ? [Optical filters are used to deal with this bothersome effect when a continuous spectrum of light is measured by a spectrometer.]

41. (II) Show that the second- and third-order spectra of white light produced by a diffraction grating always overlap. What wavelengths overlap?

42. (II) Two first-order spectrum lines are measured by a 9350-slit/cm spectrometer at angles, on each side of center, of $+26°38'$, $+41°02'$ and $-26°18'$, $-40°27'$. Calculate the wavelengths based on these data.

43. (II) Suppose the angles measured in Problem 42 were produced when the spectrometer (but not the source) was submerged in water. What then would be the wavelengths (in air)?

44. (II) The first-order line of 589-nm light falling on a diffraction grating is observed at a 15.5° angle. How far apart are the slits? At what angle will the third order be observed?

45. (II) Red laser light from a He–Ne laser $(\lambda = 632.8\text{ nm})$ creates a second-order fringe at 53.2° after passing through a grating. What is the wavelength λ of light that creates a first-order fringe at 21.2°?

46. (II) Two (and only two) full spectral orders can be seen on either side of the central maximum when white light is sent through a diffraction grating. What is the maximum number of slits per cm for the grating?

47. (II) White light passes through a 640-slit/mm diffraction grating. First-order and second-order visible spectra ("rainbows") appear on the wall 32 cm away as shown in Fig. 35–40. Determine the widths ℓ_1 and ℓ_2 of the two "rainbows" (400 nm to 700 nm). In which order is the "rainbow" dispersed over a larger distance?

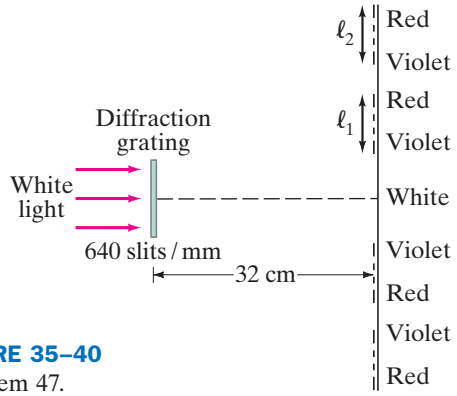


FIGURE 35–40
Problem 47.

48. (II) **Missing orders** occur for a diffraction grating when a diffraction minimum coincides with an interference maximum. Let D be the width of each slit and d the separation of slits. (*a*) Show that if $d = 2D$, all even orders $(m = 2, 4, 6, \cdots)$ are missing. (*b*) Show there will be missing orders whenever

$$\frac{d}{D} = \frac{m_1}{m_2}$$

where m_1 and m_2 are integers. (*c*) Discuss the case $d = D$, the limit in which the space between slits becomes negligible.

49. (II) Monochromatic light falls on a transmission diffraction grating at an angle ϕ to the normal. (*a*) Show that Eq. 35–13 for diffraction maxima must be replaced by

$$d(\sin\phi + \sin\theta) = \pm m\lambda. \qquad m = 0, 1, 2, \cdots.$$

(*b*) Explain the $\pm$ sign. (*c*) Green light with a wavelength of 550 nm is incident at an angle of 15° to the normal on a diffraction grating with 4500 slits/cm. Find the angles at which the first-order maxima occur.

*35–10 Grating, Peak Widths, Resolving Power

***50.** (II) A 6500-slit/cm diffraction grating is 3.26 cm wide. If light with wavelengths near 624 nm falls on the grating, how close can two wavelengths be if they are to be resolved in any order? What order gives the best resolution?

***51.** (II) A diffraction grating has 15,000 rulings in its 1.9 cm width. Determine (*a*) its resolving power in first and second orders, and (*b*) the minimum wavelength resolution ($\Delta\lambda$) it can yield for $\lambda = 410$ nm.

***52.** (II) Let 580-nm light be incident normally on a diffraction grating for which $d = 3.00D = 1050$ nm. (*a*) How many orders (principal maxima) are present? (*b*) If the grating is 1.80 cm wide, what is the full angular width of each principal maximum?

35–11 X-Ray Diffraction

53. (II) X-rays of wavelength 0.138 nm fall on a crystal whose atoms, lying in planes, are spaced 0.315 nm apart. At what angle ϕ (relative to the surface, Fig. 35–28) must the X-rays be directed if the first diffraction maximum is to be observed?

54. (II) First-order Bragg diffraction is observed at 23.6° relative to the crystal surface, with spacing between atoms of 0.24 nm. (*a*) At what angle will second order be observed? (*b*) What is the wavelength of the X-rays?

55. (II) If X-ray diffraction peaks corresponding to the first three orders ($m = 1, 2,$ and 3) are measured, can both the X-ray wavelength λ and lattice spacing d be determined? Prove your answer.

*35–12 X-Ray Imaging and CT

***56.** (II) (*a*) Suppose for a conventional X-ray image that the X-ray beam consists of parallel rays. What would be the magnification of the image? (*b*) Suppose, instead, that the X-rays come from a point source (as in Fig. 35–31) that is 15 cm in front of a human body which is 25 cm thick, and the film is pressed against the person's back. Determine and discuss the range of magnifications that result.

General Problems

57. The wings of a certain beetle have a series of parallel lines across them. When normally incident 520-nm light is reflected from the wing, the wing appears bright when viewed at an angle of 56°. How far apart are the lines?

58. A teacher stands well back from an outside doorway 0.88 m wide, and blows a whistle of frequency 820 Hz. Ignoring reflections, estimate at what angle(s) it is *not* possible to hear the whistle clearly on the playground outside the doorway. Assume 340 m/s for the speed of sound.

59. Light is incident on a diffraction grating with 8200 slits/cm and the pattern is viewed on a screen located 2.5 m from the grating. The incident light beam consists of two wavelengths, $\lambda_1 = 4.4 \times 10^{-7}$ m and $\lambda_2 = 6.8 \times 10^{-7}$ m. Calculate the linear distance between the first-order bright fringes of these two wavelengths on the screen.

60. When yellow sodium light, $\lambda = 589$ nm, falls on a diffraction grating, its first-order peak on a screen 62.0 cm away falls 3.32 cm from the central peak. Another source produces a line 3.71 cm from the central peak. What is its wavelength? How many slits/cm are on the grating?

61. Two of the lines of the atomic hydrogen spectrum have wavelengths of 656 nm and 410 nm. If these fall at normal incidence on a grating with 8500 slits/cm, what will be the angular separation of the two wavelengths in the first-order spectrum?

62. (*a*) How far away can a human eye distinguish two car headlights 2.0 m apart? Consider only diffraction effects and assume an eye pupil diameter of 6.0 mm and a wavelength of 560 nm. (*b*) What is the minimum angular separation an eye could resolve when viewing two stars, considering only diffraction effects? In reality, it is about 1′ of arc. Why is it not equal to your answer in (*b*)?

63. A laser beam passes through a slit of width 1.0 cm and is pointed at the Moon, which is approximately 380,000 km from the Earth. Assume the laser emits waves of wavelength 633 nm (the red light of a He–Ne laser). Estimate the width of the beam when it reaches the Moon due to diffraction.

64. A He–Ne gas laser which produces monochromatic light of wavelength $\lambda = 6.328 \times 10^{-7}$ m is used to calibrate a reflection grating in a spectroscope. The first-order diffraction line is found at an angle of 18.5° to the incident beam. How many lines per meter are imprinted on this reflecting diffraction grating?

65. A 1.0-cm-diameter lens with a focal length of 35 cm uses blue light (450 nm) to image two objects 18 m away that are very close together. What is the closest those objects can be to each other and still be imaged as separate objects?

66. Red light from three separate sources passes through a diffraction grating with 3.30×10^5 slits/m. The wavelengths of the three lines are 6.56×10^{-7} m (hydrogen), 6.50×10^{-7} m (neon), and 6.97×10^{-7} m (argon). Calculate the angles for the first-order diffraction line of each source.

67. The Earth and Moon are separated by about 400×10^6 m. When Mars is 8×10^{10} m from Earth, could a person standing on Mars resolve the Earth and its Moon as two separate objects without a telescope? Assume a pupil diameter of 5 mm and $\lambda = 550$ nm.

68. You want to design a spy satellite to photograph license plate numbers. Assuming it is necessary to resolve points separated by 2 cm with 550-nm light, and that the satellite orbits at a height of 130 km, what minimum lens aperture (diameter) is required?

69. What is the highest spectral order that can be seen if a grating with 6800 slits per cm is illuminated with 633-nm laser light? Assume normal incidence.

70. A parallel beam of light containing two wavelengths, 420 nm and 680 nm, enters a borate flint glass equilateral prism (Fig. 35–41). (*a*) What is the angle between the two beams leaving the prism? (See Section 32–6.) (*b*) Repeat part (*a*) for a diffraction grating with 5800 slits/cm. (*c*) Discuss two advantages of a diffraction grating, including one that you see from your results.

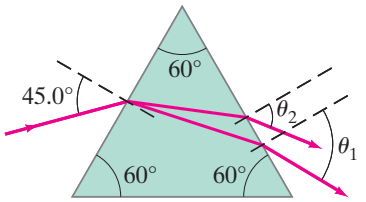


FIGURE 35–41
Problem 70.

71. Mizar, the second star from the end of the Big Dipper's handle, appears to have a companion star, Alcor. From Earth, Mizar and Alcor have an angular separation of 12 arc minutes ($1 \text{ arc min} = \frac{1}{60}$ of $1°$). Using Example 35–6, estimate the angular resolution of the human eye (in arc min). From your estimate, explain if these two stars can be resolved by the naked eye.

72. The entrance to a boy's bedroom consists of two doorways, each 1.0 m wide, which are separated by a distance of 3.0 m. The boy's mother yells at him through the two doors as shown in Fig. 35–42, telling him to clean up his room. Her voice has a frequency of 400 Hz. Later, when the mother discovers the room is still a mess, the boy says he never heard her telling him to clean his room. The velocity of sound is 340 m/s. (*a*) Find all of the angles θ (Fig. 35–42) at which no sound will be heard within the bedroom when the mother yells. Assume sound is fully absorbed when it strikes a bedroom wall. (*b*) If the boy was at the position shown when his mother yelled, does he have a good explanation for not having heard her? Explain.

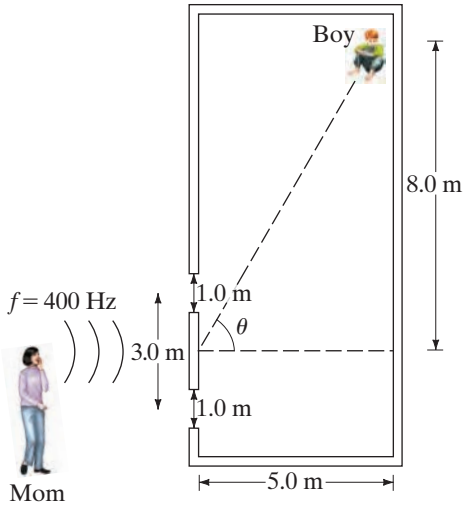


FIGURE 35–42
Problem 72.

73. How many slits per centimeter must a grating have if there is to be no second-order spectrum for any visible wavelength?

74. Spy planes fly at extremely high altitudes (25 km) to avoid interception. If their cameras are to discern features as small as 5 cm, what is the minimum aperture of the camera lens to afford this resolution? (Use $\lambda = 580$ nm.)

75. X-rays of wavelength 0.0973 nm are directed at an unknown crystal. The second diffraction maximum is recorded when the X-rays are directed at an angle of 21.8° relative to the crystal surface. What is the spacing between crystal planes parallel to the surface?

76. X-rays of wavelength 0.10 nm fall on a microcrystalline powder sample. The sample is located 15 cm from a photographic sensor. The crystal structure of the sample has an atomic spacing of 0.22 nm. Calculate the radii of the diffraction rings corresponding to first- and second-order scattering. Note in Fig. 35–28 that the X-ray beam is deflected through an angle 2ϕ.

77. The Hubble Space Telescope, with an objective diameter of 2.4 m, is viewing the Moon. Estimate the minimum distance between two objects on the Moon that the Hubble can distinguish. Consider diffraction of light with wavelength 550 nm. Assume the Hubble is near the Earth.

78. A slit of width $D = 22\ \mu$m is cut through a thin aluminum plate. Light with wavelength $\lambda = 620$ nm passes through this slit and forms a single-slit diffraction pattern on a screen a distance $\ell = 2.0$ m away. Defining x to be the distance between the two first minima on either side of the center in this diffraction pattern ($m = +1$ and $m = -1$), find the change Δx in this distance when the temperature T of the metal plate is changed by an amount $\Delta T = 55$ C°. [*Hint*: Since $\lambda \ll D$, the first minima occur at a small angle.]

79. A satellite dish is about 0.5 m in diameter. According to the user's manual, the dish has to be pointed in the direction of the satellite, but an error of about 2° to either side is allowed without loss of reception. Estimate the wavelength of the electromagnetic waves received by the dish.

ANSWERS TO EXERCISES

A: Narrower.
B: (*b*).
C: 2 m.
D: (*c*).
E: (*c*) By the equation just above; (*c*) by the λ rule.
F: A.
G: 0.28 nm.



(a) (b)

(a) That is me standing on the corner, with messy hair and nerdy eyeglasses. I am talking to my girlfriend sitting on her Vespa. She is glamorous; she could be a model. As we chat, I notice a UPS truck zooming by which looks quite a bit shorter than when it is parked. (b) My girlfriend takes off, turning right and going around the block, and then comes speeding past me on her scooter at 60 or 70 km/hr. She looks so skinny now. I hate it when she drives that fast. Oh, I need to mention that in my fantasy universe the speed limit of light is only 100 km/h.

One of the predictions of Einstein's special theory of relativity is that objects moving at high speed, near the speed of light, would look shorter than when they are at rest. This effect is called "length contraction". Special relativity predicts other strange and wonderful effects as well.

CHAPTER 36

The Special Theory of Relativity

CHAPTER-OPENING QUESTION—Guess now!

A rocket is headed away from Earth at a speed of 0.80c. The rocket fires a small payload at a speed of 0.70c (relative to the rocket) aimed away from Earth. How fast is the payload moving relative to Earth?

(a) 1.50c.
(b) a little less than 1.50c.
(c) a little over c.
(d) a little under c.
(e) 0.75c.

Physics at the end of the nineteenth century looked back on a period of great progress. The theories developed over the preceding three centuries had been very successful in explaining a wide range of natural phenomena. Newtonian mechanics beautifully explained the motion of objects on Earth and in the heavens. Furthermore, it formed the basis for successful treatments of fluids, wave motion, and sound. Kinetic theory explained the behavior of gases and other materials. Maxwell's theory of electromagnetism embodied all of electric and magnetic phenomena, and it predicted the existence of electromagnetic waves that would behave just like light—so light came to be thought of as an electromagnetic wave. Indeed, it seemed that the natural world, as seen through the eyes of physicists, was very well explained. A few puzzles remained, but it was felt that these would soon be explained using already known principles.

CONTENTS

36 – 1 Galilean–Newtonian Relativity
36 – 2 The Michelson–Morley Experiment
36 – 3 Postulates of the Special Theory of Relativity
36 – 4 Simultaneity
36 – 5 Time Dilation and the Twin Paradox
36 – 6 Length Contraction
36 – 7 Four-Dimensional Space–Time
36 – 8 Galilean and Lorentz Transformations
36 – 9 Relativistic Momentum
36–10 The Ultimate Speed
36–11 $E = mc^2$; Mass and Energy
36–12 Doppler Shift for Light
36–13 The Impact of Special Relativity

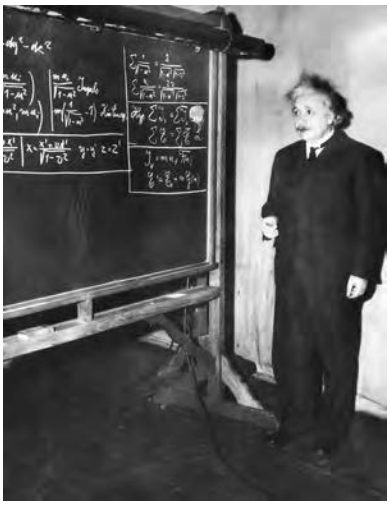


FIGURE 36–1 Albert Einstein (1879–1955), one of the great minds of the twentieth century, was the creator of the special and general theories of relativity.

It did not turn out so simply. Instead, these puzzles were to be solved only by the introduction, in the early part of the twentieth century, of two revolutionary new theories that changed our whole conception of nature: the *theory of relativity* and *quantum theory*.

Physics as it was known at the end of the nineteenth century (what we've covered up to now in this book) is referred to as **classical physics**. The new physics that grew out of the great revolution at the turn of the twentieth century is now called **modern physics**. In this Chapter, we present the special theory of relativity, which was first proposed by Albert Einstein (1879–1955; Fig. 36–1) in 1905. In Chapter 37, we introduce the equally momentous quantum theory.

36–1 Galilean–Newtonian Relativity

Einstein's special theory of relativity deals with how we observe events, particularly how objects and events are observed from different frames of reference.[†] This subject had already been explored by Galileo and Newton.

The special theory of relativity deals with events that are observed and measured from so-called **inertial reference frames** (Sections 4–2 and 11–8), which are reference frames in which Newton's first law is valid: if an object experiences no net force, the object either remains at rest or continues in motion with constant speed in a straight line. It is usually easiest to analyze events when they are observed and measured by observers at rest in an inertial frame. The Earth, though not quite an inertial frame (it rotates), is close enough that for most purposes we can approximate it as an inertial frame. Rotating or otherwise accelerating frames of reference are noninertial frames,[‡] and won't concern us in this Chapter (they are dealt with in Einstein's general theory of relativity).

PROBLEM SOLVING
A reference frame moving at constant velocity relative to an inertial frame is itself an inertial frame

A reference frame that moves with constant velocity with respect to an inertial frame is itself also an inertial frame, since Newton's laws hold in it as well. When we say that we observe or make measurements from a certain reference frame, it means that we are at rest in that reference frame.

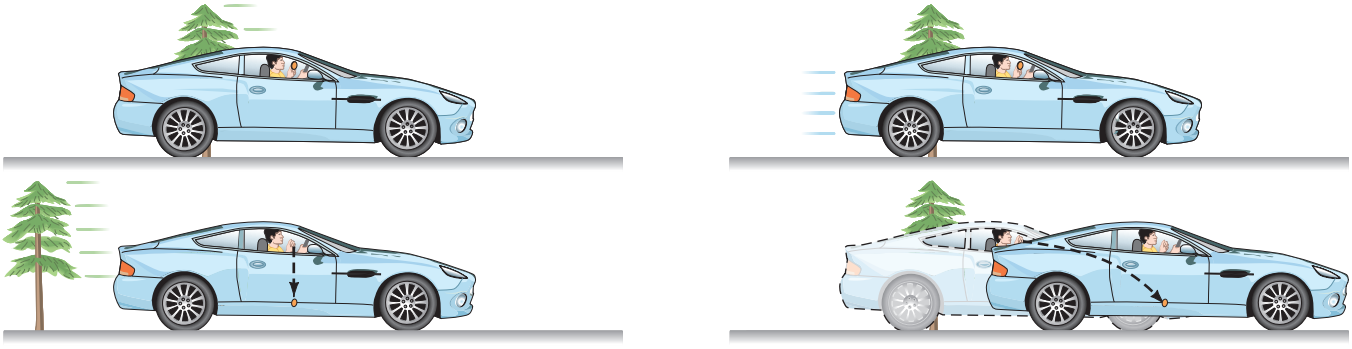
Both Galileo and Newton were aware of what we now call the **relativity principle** applied to mechanics: that *the basic laws of physics are the same in all inertial reference frames*. You may have recognized its validity in everyday life. For example, objects move in the same way in a smoothly moving (constant-velocity) train or airplane as they do on Earth. (This assumes no vibrations or rocking which would make the reference frame noninertial.) When you walk, drink a cup of soup, play pool, or drop a pencil on the floor while traveling in a train, airplane, or ship moving at *constant velocity*, the objects move just as they do when you are at rest on Earth. Suppose you are in a car traveling rapidly at constant velocity. If you drop a coin from above your head inside the car, how will it fall? It falls straight downward with respect to the car, and hits the floor directly below the point of release, Fig. 36–2a. This is just how objects fall on the Earth—straight down—and thus our experiment in the moving car is in accord with the relativity principle. (If you drop the coin out the car's window, this won't happen because the moving air drags the coin backward relative to the car.)

CAUTION
Laws are the same, but paths may be different in different reference frames

Note in this example, however, that to an observer on the Earth, the coin follows a curved path, Fig. 36–2b. The actual path followed by the coin is different as viewed from different frames of reference. This does not violate the relativity principle because this principle states that the *laws* of physics are the same in all inertial frames. The same law of gravity, and the same laws of motion, apply in both reference frames. The acceleration of the coin is the same in both reference frames.

[†]On a rotating platform (say a merry-go-round), for example, a ball at rest starts moving outward even though no object exerts a force on it. This is therefore not an inertial frame. See Section 11–8.

[‡]A reference frame is a set of coordinate axes fixed to some object such as the Earth, a train, or the Moon. See Section 2–1.



(a)
Reference frame = car

(b)
Reference frame = Earth

FIGURE 36–2 A coin is dropped by a person inside a moving car. The upper views show the moment of the coin's release, the lower views are a short time later. (a) In the reference frame of the car, the coin falls straight down (and the tree moves to the left). (b) In a reference frame fixed on the Earth, the coin has an initial velocity (= to car's) and follows a curved (parabolic) path.

The difference in Figs. 36–2a and b is that in the Earth's frame of reference, the coin has an initial velocity (equal to that of the car). The laws of physics therefore predict it will follow a parabolic path like any projectile (Chapter 3). In the car's reference frame, there is no initial velocity, and the laws of physics predict that the coin will fall straight down. The laws are the same in both reference frames, although the specific paths are different.

Galilean–Newtonian relativity involves certain unprovable assumptions that make sense from everyday experience. It is assumed that the lengths of objects are the same in one reference frame as in another, and that time passes at the same rate in different reference frames. In classical mechanics, then, space and time intervals are considered to be **absolute**: their measurement does not change from one reference frame to another. The mass of an object, as well as all forces, are assumed to be unchanged by a change in inertial reference frame.

CAUTION
Length and time intervals are absolute (pre-relativity)

The position of an object, however, is different when specified in different reference frames, and so is velocity. For example, a person may walk inside a bus toward the front with a speed of 2 m/s. But if the bus moves 10 m/s with respect to the Earth, the person is then moving with a speed of 12 m/s with respect to the Earth. The acceleration of an object, however, is the same in any inertial reference frame according to classical mechanics. This is because the change in velocity, and the time interval, will be the same. For example, the person in the bus may accelerate from 0 to 2 m/s in 1.0 seconds, so $a = 2\,\text{m/s}^2$ in the reference frame of the bus. With respect to the Earth, the velocity changes from 10 m/s to 12 m/s so the acceleration is $(12\,\text{m/s} - 10\,\text{m/s})/(1.0\,\text{s}) = 2\,\text{m/s}^2$, which is the same.

CAUTION
Position and velocity are different in different reference frames, but length is the same (classical)

Since neither F, m, nor a changes from one inertial frame to another, then Newton's second law, $F = ma$, does not change. Thus Newton's second law satisfies the relativity principle. The other laws of mechanics also satisfy the relativity principle.

That the laws of mechanics are the same in all inertial reference frames implies that no one inertial frame is special in any sense. We express this important conclusion by saying that **all inertial reference frames are equivalent** for the description of mechanical phenomena. No one inertial reference frame is any better than another. A reference frame fixed to a car or an aircraft traveling at constant velocity is as good as one fixed on the Earth. When you travel smoothly at constant velocity in a car or airplane, it is just as valid to say you are at rest and the Earth is moving, as it is to say the reverse.[†] There is no experiment you can do to tell which frame is "really" at rest and which is moving. Thus, there is no way to single out one particular reference frame as being at absolute rest.

CAUTION
There is no reference frame truly at rest

A complication arose, however, in the last half of the nineteenth century. Maxwell's comprehensive and successful theory of electromagnetism (Chapter 31) predicted that light is an electromagnetic wave. Maxwell's equations gave the velocity of light c as $3.00 \times 10^8\,\text{m/s}$; and this is just what is measured. The question then arose: in what reference frame does light have precisely the value predicted by Maxwell's theory? It was assumed that light would have a different speed in different frames of reference.

[†]We are ignoring the rotation and curvature of the Earth, and using the reasonable approximation that Earth is an inertial reference frame.

For example, if observers could travel on a rocket ship at a speed of 1.0×10^8 m/s away from a source of light, we might expect them to measure the speed of the light reaching them to be $(3.0 \times 10^8 \text{ m/s}) - (1.0 \times 10^8 \text{ m/s}) = 2.0 \times 10^8$ m/s. But Maxwell's equations have no provision for relative velocity. They predicted the speed of light to be $c = 3.0 \times 10^8$ m/s, which seemed to imply that there must be some preferred reference frame where c would have this value.

We discussed in Chapters 15 and 16 that waves can travel on water and along ropes or strings, and sound waves travel in air and other materials. Nineteenth-century physicists viewed the material world in terms of the laws of mechanics, so it was natural for them to assume that light too must travel in some *medium*. They called this transparent medium the **ether** and assumed it permeated all space.[†] It was therefore assumed that the velocity of light given by Maxwell's equations must be with respect to the ether.

At first it appeared that Maxwell's equations did *not* satisfy the relativity principle. They were simplest in the frame where $c = 3.00 \times 10^8$ m/s in a reference frame at rest in the ether. In any other reference frame, extra terms were needed to account for the relative velocity. Thus, although other laws of physics obeyed the relativity principle, the laws of electricity and magnetism apparently did not. Einstein's second postulate (Section 36–3) resolved this problem: Maxwell's equations do satisfy relativity.

Scientists soon set out to determine the speed of the Earth relative to this absolute frame, whatever it might be. A number of clever experiments were designed. The most direct were performed by A. A. Michelson and E. W. Morley in the 1880s. They measured the difference in the speed of light in different directions using Michelson's interferometer (Section 34–6). They expected to find a difference depending on the orientation of their apparatus with respect to the ether. For just as a boat has different speeds relative to the land when it moves upstream, downstream, or across the stream, so too light would be expected to have different speeds depending on the velocity of the ether past the Earth.

Strange as it may seem, they detected no difference at all. This was a great puzzle. A number of explanations were put forth over a period of years, but they led to contradictions or were otherwise not generally accepted. This **null result** was one of the great puzzles at the end of the nineteenth century.

Then in 1905, Albert Einstein proposed a radical new theory that reconciled these many problems in a simple way. But at the same time, as we shall see, it completely changed our ideas of space and time.

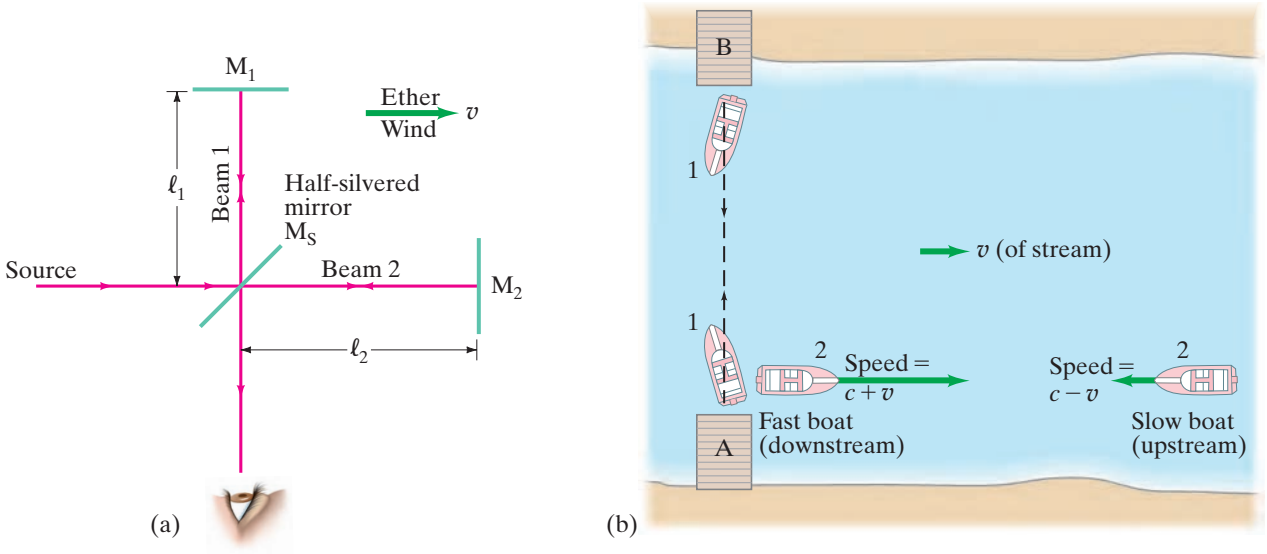
36–2 The Michelson–Morley Experiment

The Michelson–Morley experiment was designed to measure the speed of the *ether*—the medium in which light was assumed to travel—with respect to the Earth. The experimenters thus hoped to find an absolute reference frame, one that could be considered to be at rest.

One of the possibilities nineteenth-century scientists considered was that the ether is fixed relative to the Sun, for even Newton had taken the Sun as the center of the universe. If this were the case (there was no guarantee, of course), the Earth's speed of about 3×10^4 m/s in its orbit around the Sun could produce a change of 1 part in 10^4 in the speed of light $(3.0 \times 10^8 \text{ m/s})$. Direct measurement of the speed of light to this precision was not possible. But A. A. Michelson, later with the help of E. W. Morley, was able to use his interferometer (Section 34–6) to measure the difference in the speed of light in different directions to this precision.

[†]The medium for light waves could not be air, since light travels from the Sun to Earth through nearly empty space. Therefore, another medium was postulated, the ether. The ether was not only transparent but, because of difficulty in detecting it, was assumed to have zero density.

This famous experiment is based on the principle shown in Fig. 36–3. Part (a) is a diagram of the Michelson interferometer, and it is assumed that the "ether wind" is moving with speed v to the right. (Alternatively, the Earth is assumed to move to the left with respect to the ether at speed v.) The light from a source is split into two beams by a half-silvered mirror M_S. One beam travels to mirror M_1 and the other to mirror M_2. The beams are reflected by M_1 and M_2 and are joined again after passing through M_S. The now superposed beams interfere with each other and the resultant is viewed by the observer's eye as an interference pattern (discussed in Section 34–6).



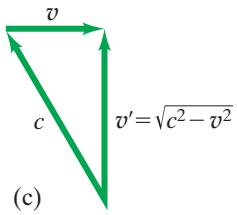


FIGURE 36–3 The Michelson–Morley experiment. (a) Michelson interferometer. (b) Boat analogy: boat 1 goes across the stream and back; boat 2 goes downstream and back upstream (boat has speed c relative to the water). (c) Calculation of the velocity of boat (or light beam) traveling perpendicular to the current (or ether wind).

Whether constructive or destructive interference occurs at the center of the interference pattern depends on the relative phases of the two beams after they have traveled their separate paths. Let us consider an analogy of a boat traveling up and down, and across, a river, whose current moves with speed v, as shown in Fig. 36–3b. In still water, the boat can travel with speed c (not the speed of light in this case).

First we consider beam 2 in Fig. 36–3a, which travels parallel to the "ether wind." In its journey from M_S to M_2, the light would travel with speed $c + v$, according to classical physics, just as for a boat traveling downstream (see Fig. 36–3b) we add the speed of the river water to the boat's own speed (relative to the water) to get the boat's speed relative to the shore. Since the beam travels a distance ℓ_2, the time it takes to go from M_S to M_2 would be $t = \ell_2/(c + v)$. To make the return trip from M_2 to M_S, the light moves against the ether wind (like the boat going upstream), so its relative speed is expected to be $c - v$. The time for the return trip would be $\ell_2/(c - v)$. The total time for beam 2 to travel from M_S to M_2 and back to M_S is

$$t_2 = \frac{\ell_2}{c + v} + \frac{\ell_2}{c - v} = \frac{\ell_2(c - v) + \ell_2(c + v)}{(c + v)(c - v)} = \frac{2\ell_2}{c(1 - v^2/c^2)}.$$

Now let us consider beam 1, which travels crosswise to the ether wind. Here the boat analogy (Fig. 36–3b) is especially helpful. The boat is to go from wharf A to wharf B directly across the stream. If it heads directly across, the stream's current will drag it downstream. To reach wharf B, the boat must head at an angle upstream. The precise angle depends on the magnitudes of c and v, but is of no interest to us in itself. Part (c) of Fig. 36–3 shows how to calculate the velocity v' of the boat relative to Earth as it crosses the stream. Since c, v, and v' form a right triangle, we have that $v' = \sqrt{c^2 - v^2}$. The boat has the same speed when it returns.

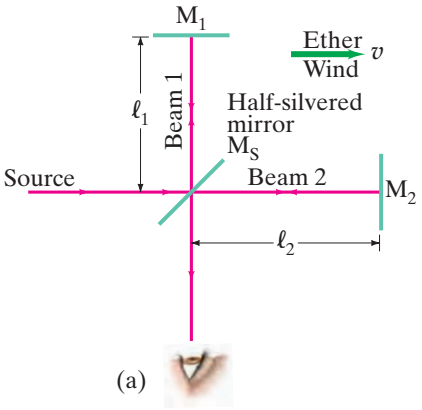


FIGURE 36–3a (Repeated.)

If we now apply these principles to light beam 1 in Fig. 36–3a, we expect the beam to travel with speed $\sqrt{c^2 - v^2}$ in going from M_S to M_1 and back again. The total distance traveled is $2\ell_1$, so the time required for beam 1 to make the round trip would be $2\ell_1/\sqrt{c^2 - v^2}$, or

$$t_1 = \frac{2\ell_1}{c\sqrt{1 - v^2/c^2}}.$$

Notice that the denominator in this equation for t_1 involves a square root, whereas that for t_2 does not.

If $\ell_1 = \ell_2 = \ell$, we see that beam 2 will lag behind beam 1 by an amount

$$\Delta t = t_2 - t_1 = \frac{2\ell}{c}\left(\frac{1}{1 - v^2/c^2} - \frac{1}{\sqrt{1 - v^2/c^2}}\right).$$

If $v = 0$, then $\Delta t = 0$, and the two beams will return in phase since they were initially in phase. But if $v \neq 0$, then $\Delta t \neq 0$, and the two beams will return out of phase. If this change of phase, from the condition $v = 0$ to that for $v \neq 0$, could be measured, then v could be determined. But the Earth cannot be stopped. Furthermore, we should not be too quick to assume that lengths are not affected by motion and therefore we should not assume $\ell_1 = \ell_2$.

Michelson and Morley realized that they could detect the difference in phase (assuming that $v \neq 0$) if they rotated their apparatus by 90°, for then the interference pattern between the two beams should change. In the rotated position, beam 1 would now move parallel to the ether and beam 2 perpendicular to it. Thus the roles could be reversed, and in the rotated position the times (designated by primes, $'$) would be

$$t_1' = \frac{2\ell_1}{c(1 - v^2/c^2)} \quad \text{and} \quad t_2' = \frac{2\ell_2}{c\sqrt{1 - v^2/c^2}}.$$

The time lag between the two beams in the nonrotated position (unprimed) would be

$$\Delta t = t_2 - t_1 = \frac{2\ell_2}{c(1 - v^2/c^2)} - \frac{2\ell_1}{c\sqrt{1 - v^2/c^2}}.$$

In the rotated position, the time difference would be

$$\Delta t' = t_2' - t_1' = \frac{2\ell_2}{c\sqrt{1 - v^2/c^2}} - \frac{2\ell_1}{c(1 - v^2/c^2)}.$$

When the rotation is made, the fringes of the interference pattern that reach the eye (Section 34–6) will shift an amount determined by the difference:

$$\Delta t - \Delta t' = \frac{2}{c}(\ell_1 + \ell_2)\left(\frac{1}{1 - v^2/c^2} - \frac{1}{\sqrt{1 - v^2/c^2}}\right).$$

This expression can be considerably simplified if we assume that $v/c \ll 1$. In this case we can use the binomial expansion (Appendix A), so

$$\frac{1}{1 - v^2/c^2} \approx 1 + \frac{v^2}{c^2} \quad \text{and} \quad \frac{1}{\sqrt{1 - v^2/c^2}} \approx 1 + \frac{1}{2}\frac{v^2}{c^2}.$$

Then

$$\Delta t - \Delta t' \approx \frac{2}{c}(\ell_1 + \ell_2)\left(1 + \frac{v^2}{c^2} - 1 - \frac{1}{2}\frac{v^2}{c^2}\right)$$

$$\approx (\ell_1 + \ell_2)\frac{v^2}{c^3}.$$

We assume $v = 3.0 \times 10^4$ m/s, the speed of the Earth in its orbit around the Sun, in this equation. In Michelson and Morley's experiments, the arms ℓ_1 and ℓ_2 were about 11 m long so $\ell_1 + \ell_2 = 22$ m. The time difference would then be about

$$(22\text{ m})(3.0 \times 10^4\text{ m/s})^2/(3.0 \times 10^8\text{ m/s})^3 \approx 7.3 \times 10^{-16}\text{ s}.$$

For visible light of wavelength $\lambda = 5.5 \times 10^{-7}$ m, say, the frequency would be $f = c/\lambda = (3.0 \times 10^8\text{ m/s})/(5.5 \times 10^{-7}\text{ m}) = 5.5 \times 10^{14}$ Hz, which means that wave

crests pass by a point every $1/(5.5\times10^{14}\,\text{Hz}) = 1.8\times10^{-15}\,\text{s}$. Thus, with a time difference of $7.3\times10^{-16}\,\text{s}$, Michelson and Morley should have noted a movement in the interference pattern of $(7.3\times10^{-16}\,\text{s})/(1.8\times10^{-15}\,\text{s}) = 0.4$ fringe. They could easily have detected this, since their apparatus was capable of observing a fringe shift as small as 0.01 fringe.

But they found *no significant fringe shift whatever*! They set their apparatus at various orientations. They made observations day and night so that they would be at various orientations with respect to the Sun (due to the Earth's rotation). They tried at different seasons of the year (the Earth at different locations due to its orbit around the Sun). Never did they observe a significant fringe shift.

This **null result** was one of the great puzzles of physics at the end of the nineteenth century. To explain it was a difficult challenge. One possibility to explain the null result was put forth independently by G. F. Fitzgerald and H. A. Lorentz (in the 1890s) in which they proposed that any length (including the arm of an interferometer) contracts by a factor $\sqrt{1 - v^2/c^2}$ in the direction of motion through the ether. According to Lorentz, this could be due to the ether affecting the forces between the molecules of a substance, which were assumed to be electrical in nature. This theory was eventually replaced by the far more comprehensive theory proposed by Albert Einstein in 1905—the special theory of relativity.

36–3 Postulates of the Special Theory of Relativity

The problems that existed at the start of the twentieth century with regard to electromagnetic theory and Newtonian mechanics were beautifully resolved by Einstein's introduction of the special theory of relativity in 1905. Unaware of the Michelson–Morley null result, Einstein was motivated by certain questions regarding electromagnetic theory and light waves. For example, he asked himself: "What would I see if I rode a light beam?" The answer was that instead of a traveling electromagnetic wave, he would see alternating electric and magnetic fields at rest whose magnitude changed in space, but did not change in time. Such fields, he realized, had never been detected and indeed were not consistent with Maxwell's electromagnetic theory. He argued, therefore, that it was unreasonable to think that the speed of light relative to any observer could be reduced to zero, or in fact reduced at all. This idea became the second postulate of his theory of relativity.

In his famous 1905 paper, Einstein proposed doing away with the idea of the ether and the accompanying assumption of a preferred or absolute reference frame at rest. This proposal was embodied in two postulates. The first was an extension of the Galilean–Newtonian relativity principle to include not only the laws of mechanics but also those of the rest of physics, including electricity and magnetism:

***First postulate (the relativity principle)*: The laws of physics have the same form in all inertial reference frames.**

RELATIVITY PRINCIPLE

The first postulate can also be stated as: *There is no experiment you can do in an inertial reference frame to determine if you are at rest or moving uniformly at constant velocity*. The second postulate is consistent with the first:

***Second postulate (constancy of the speed of light)*: Light propagates through empty space with a definite speed *c* independent of the speed of the source or observer.**

SPEED OF LIGHT PRINCIPLE

These two postulates form the foundation of Einstein's **special theory of relativity**. It is called "special" to distinguish it from his later "general theory of relativity," which deals with noninertial (accelerating) reference frames (Chapter 44). The special theory, which is what we discuss here, deals only with inertial frames.

The second postulate may seem hard to accept, for it seems to violate common sense. First of all, we have to think of light traveling through empty space. Giving up the ether is not too hard, however, since it had never been detected. But the second postulate also tells us that the speed of light in vacuum is always the same, $3.00\times10^8\,\text{m/s}$, no matter what the speed of the observer or the source. Thus, a person traveling toward or away from a source of light will measure the same speed for that light as someone at rest with respect to the source.

This conflicts with our everyday experience: we would expect to have to add in the velocity of the observer. On the other hand, perhaps we can't expect our everyday experience to be helpful when dealing with the high velocity of light. Furthermore, the null result of the Michelson–Morley experiment is fully consistent with the second postulate.

Einstein's proposal has a certain beauty. By doing away with the idea of an absolute reference frame, it was possible to reconcile classical mechanics with Maxwell's electromagnetic theory. The speed of light predicted by Maxwell's equations *is* the speed of light in vacuum in *any* reference frame.

Einstein's theory required us to give up common sense notions of space and time, and in the following Sections we will examine some strange but interesting consequences of special relativity. Our arguments for the most part will be simple ones. We will use a technique that Einstein himself did: we will imagine very simple experimental situations in which little mathematics is needed. In this way, we can see many of the consequences of relativity theory without getting involved in detailed calculations. Einstein called these **thought experiments**.

36–4 Simultaneity

An important consequence of the theory of relativity is that we can no longer regard time as an absolute quantity. No one doubts that time flows onward and never turns back. But according to relativity the time interval between two events, and even whether or not two events are simultaneous, depends on the observer's reference frame. By an **event**, which we use a lot here, we mean something that happens at a particular place and at a particular time.

CAUTION
Event = at one place, at one time

Two events are said to occur simultaneously if they occur at exactly the same time. But how do we know if two events occur precisely at the same time? If they occur at the same point in space—such as two apples falling on your head at the same time—it is easy. But if the two events occur at widely separated places, it is more difficult to know whether the events are simultaneous since we have to take into account the time it takes for the light from each event to reach us. Because light travels at finite speed, a person who sees two events must calculate back to find out when they actually occurred. For example, if two events are *observed* to occur at the same time, but one actually took place farther from the observer than the other, then the more distant one must have occurred earlier, and the two events were not simultaneous.

We now imagine a simple thought experiment. Assume an observer, called O, is located exactly halfway between points A and B where two events occur, Fig. 36–4. Suppose the two events are lightning that strikes the points A and B, as shown. For brief events like lightning, only short pulses of light (blue in Fig. 36–4) will travel outward from A and B and reach O. Observer O "sees" the events when the pulses of light reach point O. If the two pulses reach O at the same time, then the two events had to be simultaneous. This is because (i) the two light pulses travel at the same speed (postulate 2), and (ii) the distance OA equals OB, so the time for the light to travel from A to O and from B to O must be the same. Observer O can then definitely state that the two events occurred simultaneously. On the other hand, if O sees the light from one event earlier than from the other, then the former event occurred first.

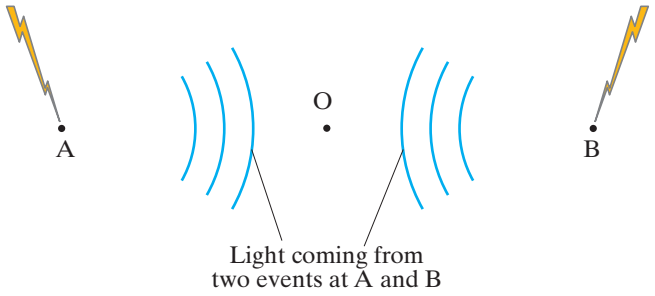


FIGURE 36–4 A moment after lightning strikes at points A and B, the pulses of light (shown as blue waves) are traveling toward the observer O, but O "sees" the lightning only when the light reaches O. Note: OA = OB.

The question we really want to examine is this: if two events are simultaneous to an observer in one reference frame, are they also simultaneous to another observer moving with respect to the first? Let us call the observers O_1 and O_2 and assume they are fixed in reference frames 1 and 2 that move with speed v relative to one another. These two reference frames can be thought of as two rockets or two trains, Fig. 36–5. In Fig. 36–5a, O_2 says that O_1 is moving to the right with speed v. But O_1 sees that O_2 is moving to the left with speed v, as in Fig. 36–5b. Both viewpoints are legitimate according to the relativity principle. **There is no third point of view that will tell us which one is "really" moving**.

Now suppose that observers O_1 and O_2 observe and measure two lightning strikes. The lightning bolts mark both trains where they strike: at A_1 and B_1 on O_1's train, and at A_2 and B_2 on O_2's train, Fig. 36–6a. For simplicity, we assume that O_1 is exactly halfway between A_1 and B_1, and O_2 is halfway between A_2 and B_2. Let us first put ourselves in O_2's reference frame, so we observe O_1 moving to the right with speed v. Let us also assume that the two events occur *simultaneously* in O_2's frame, and just at the instant when O_1 and O_2 are opposite each other, Fig. 36–6a. A short time later, Fig. 36–6b, the light from A_2 and from B_2 reaches O_2 at the same time (we assumed this). Since O_2 knows (or measures) the distances O_2A_2 and O_2B_2 as equal, O_2 knows the two events are simultaneous in the O_2 reference frame.

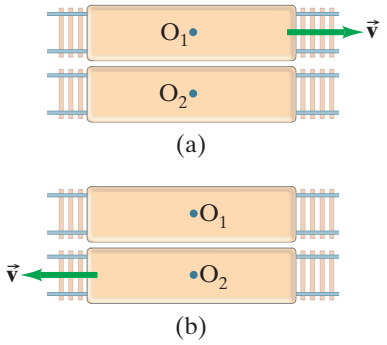


FIGURE 36–5 Observers O_1 and O_2, on two different trains (two different reference frames), are moving with relative speed v. (a) O_2 says that O_1 is moving to the right. (b) O_1 says O_2 is moving to the left. Both viewpoints are legitimate: it all depends on your reference frame.

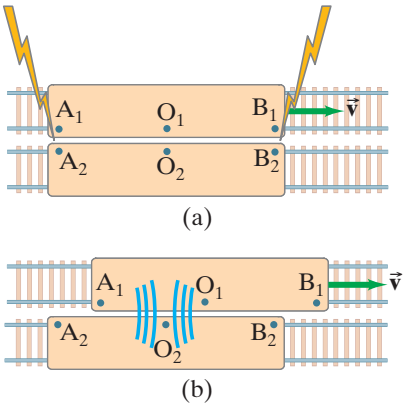


FIGURE 36–6 Thought experiment on simultaneity. In both (a) and (b) we are in the reference frame of observer O_2, who sees the reference frame of O_1 moving to the right. In (a), one lightning bolt strikes the two reference frames at A_1 and A_2, and a second lightning bolt strikes at B_1 and B_2. (b) A moment later, the light (shown in blue) from the two events reaches O_2 at the same time. So according to observer O_2, the two bolts of lightning struck simultaneously. But in O_1's reference frame, the light from B_1 has already reached O_1, whereas the light from A_1 has not yet reached O_1. So in O_1's reference frame, the event at B_1 must have preceded the event at A_1. Simultaneity in time is not absolute.

But what does observer O_1 observe and measure? From our (O_2) reference frame, we can predict what O_1 will observe. We see that O_1 moves to the right during the time the light is traveling to O_1 from A_1 and B_1. As shown in Fig. 36–6b, we can see from our O_2 reference frame that the light from B_1 has already passed O_1, whereas the light from A_1 has not yet reached O_1. That is, O_1 observes the light coming from B_1 before observing the light coming from A_1. Given (i) that light travels at the same speed c in any direction and in any reference frame, and (ii) that the distance O_1A_1 equals O_1B_1, then observer O_1 can only conclude that the event at B_1 occurred before the event at A_1. The two events are *not* simultaneous for O_1, even though they are simultaneous for O_2.

We thus find that two events which take place at different locations and are simultaneous to one observer, are actually not simultaneous to a second observer who moves relative to the first.

It may be tempting to ask: "Which observer is right, O_1 or O_2?" The answer, according to relativity, is that they are *both* right. There is no "best" reference frame we can choose to determine which observer is right. Both frames are equally good. We can only conclude that *simultaneity is not an absolute concept*, but is relative. We are not aware of this lack of agreement on simultaneity in everyday life because the effect would be noticeable only when the relative speed of the two reference frames is very large (near c), or the distances involved are very large.

EXERCISE A Examine the experiment of Fig. 36–6 from O_1's reference frame. In this case, O_1 will be at rest and will see event B_1 occur before A_1. Will O_1 recognize that O_2, who is moving with speed v to the left, will see the two events as simultaneous? [*Hint*: Draw a diagram equivalent to Fig. 36–6.]

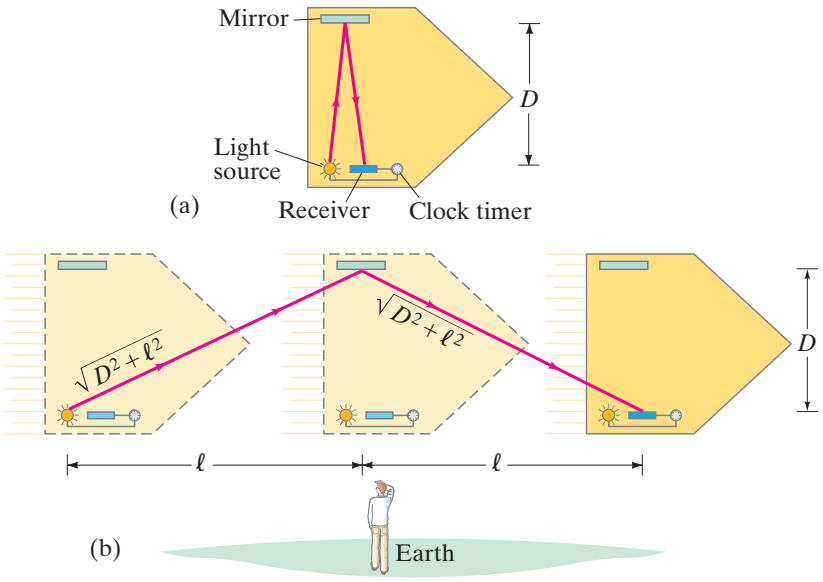


FIGURE 36–7 Time dilation can be shown by a thought experiment: the time it takes for light to travel across a spaceship and back is longer for the observer on Earth (b) than for the observer on the spaceship (a).

36–5 Time Dilation and the Twin Paradox

The fact that two events simultaneous to one observer may not be simultaneous to a second observer suggests that time itself is not absolute. Could it be that time passes differently in one reference frame than in another? This is, indeed, just what Einstein's theory of relativity predicts, as the following thought experiment shows.

Figure 36–7 shows a spaceship traveling past Earth at high speed. The point of view of an observer on the spaceship is shown in part (a), and that of an observer on Earth in part (b). Both observers have accurate clocks. The person on the spaceship (Fig. 36–7a) flashes a light and measures the time it takes the light to travel directly across the spaceship and return after reflecting from a mirror that is a distance D away (the rays are drawn at a slight angle for clarity). In the reference frame of the spaceship, the light travels a distance $2D$ at speed c. So the time required to go across and back, Δt_0, is

$$\Delta t_0 = \frac{2D}{c}.$$

The observer on Earth, Fig. 36–7b, observes the same process. But to this observer, the spaceship is moving. So the light travels the diagonal path shown going across the spaceship, reflecting off the mirror, and returning to the sender. Although the light travels at the same speed to this observer (the second postulate), it travels a greater distance. Hence the time required, as measured by the observer on Earth, will be *greater* than that measured by the observer on the spaceship.

Let us determine the time interval Δt measured by the observer on Earth between sending and receiving the light. In time Δt, the spaceship travels a distance $2\ell = v\,\Delta t$ where v is the speed of the spaceship (Fig. 36–7b). The light travels a total distance on its diagonal path (Pythagorean theorem) of $2\sqrt{D^2 + \ell^2}$, where $\ell = v\,\Delta t/2$. Therefore

$$c\,\Delta t = 2\sqrt{D^2 + \ell^2} = 2\sqrt{D^2 + v^2(\Delta t)^2/4}.$$

We square both sides, to find

$$c^2(\Delta t)^2 = 4D^2 + v^2(\Delta t)^2,$$

and solve for $(\Delta t)^2$:

$$(\Delta t)^2 = \frac{4D^2}{c^2 - v^2}.$$

We take the square root to get

$$\Delta t = \frac{2D}{c\sqrt{1 - v^2/c^2}}.$$

We combine this equation for Δt with the formula for Δt_0 above,

$$\Delta t_0 = \frac{2D}{c},$$

and obtain

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - v^2/c^2}}. \qquad \textbf{(36–1a)}$$

TIME DILATION

Since $\sqrt{1 - v^2/c^2}$ is always less than 1, we see that $\Delta t > \Delta t_0$. That is, the time interval between the two events (the sending of the light, and its reception on the spaceship) is *greater* for the observer on Earth than for the observer on the spaceship. This is a general result of the theory of relativity, and is known as **time dilation**. The time dilation effect can be stated as

clocks moving relative to an observer are measured to run more slowly, as compared to clocks at rest.

However, we should not think that the clocks are somehow at fault.

Time is actually measured to pass more slowly in any moving reference frame as compared to your own.

This remarkable result is an inevitable outcome of the two postulates of the special theory of relativity.

The factor $1/\sqrt{1 - v^2/c^2}$ occurs so often in relativity that we often give it the shorthand symbol γ (the Greek letter "gamma"), and write Eq. 36–1a as

$$\Delta t = \gamma\, \Delta t_0 \qquad \textbf{(36–1b)}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}. \qquad \textbf{(36–2)}$$

Note that γ is never less than one, and has no units. At speeds we normally experience, $\gamma = 1$ to many decimal places, and exceeds 1.00 significantly only at very high speeds, much above about 10^6 m/s (for which $\gamma = 1.000006$).

The concept of time dilation may be hard to accept, for it contradicts our experience. We can see from Eqs. 36–1 that the time dilation effect is indeed negligible unless v is reasonably close to c. If v is much less than c, then the term v^2/c^2 is much smaller than the 1 in the denominator of Eq. 36–1a, and then $\Delta t \approx \Delta t_0$ (see Example 36–2). The speeds we experience in everyday life are much smaller than c, so it is little wonder we don't ordinarily notice time dilation. But experiments that have tested the time dilation effect have confirmed Einstein's predictions. In 1971, for example, extremely precise atomic clocks were flown around the Earth in jet planes. The speed of the planes $(10^3\ \text{km/h})$ was much less than c, so the clocks had to be accurate to nanoseconds $(10^{-9}\ \text{s})$ in order to detect any time dilation. They were this accurate, and they confirmed Eqs. 36–1 to within experimental error. Time dilation had been confirmed decades earlier, however, by observations on "elementary particles" which have very small masses (typically 10^{-30} to 10^{-27} kg) and so require little energy to be accelerated to speeds close to the speed of light, c. Many of these elementary particles are not stable and decay after a time into lighter particles. One example is the muon, whose mean lifetime is 2.2 μs when at rest. Careful experiments showed that when a muon is traveling at high speeds, its lifetime is measured to be longer than when it is at rest, just as predicted by the time dilation formula.

EXAMPLE 36–1 Lifetime of a moving muon. (*a*) What will be the mean lifetime of a muon as measured in the laboratory if it is traveling at $v = 0.60c = 1.80 \times 10^8\,\text{m/s}$ with respect to the laboratory? A muon's mean lifetime at rest is $2.20\,\mu\text{s} = 2.20 \times 10^{-6}\,\text{s}$. (*b*) How far does a muon travel in the laboratory, on average, before decaying?

APPROACH If an observer were to move along with the muon (the muon would be at rest to this observer), the muon would have a mean life of $2.20 \times 10^{-6}\,\text{s}$. To an observer in the lab, the muon lives longer because of time dilation. We find the mean lifetime using Eq. 36–1a and the average distance using $d = v\,\Delta t$.

SOLUTION (*a*) From Eq. 36–1a with $v = 0.60c$, we have

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - v^2/c^2}} = \frac{2.20 \times 10^{-6}\,\text{s}}{\sqrt{1 - 0.36c^2/c^2}} = \frac{2.20 \times 10^{-6}\,\text{s}}{\sqrt{0.64}} = 2.8 \times 10^{-6}\,\text{s}.$$

(*b*) Relativity predicts that a muon with speed $1.80 \times 10^8\,\text{m/s}$ would travel an average distance $d = v\,\Delta t = (1.80 \times 10^8\,\text{m/s})(2.8 \times 10^{-6}\,\text{s}) = 500\,\text{m}$ before decaying and disappearing into lighter particles. This is the average distance that is measured experimentally in the laboratory.

NOTE At a speed of $1.8 \times 10^8\,\text{m/s}$, classical physics would tell us that with a mean life of $2.2\,\mu\text{s}$, an average muon would on average travel $d = vt = (1.8 \times 10^8\,\text{m/s})(2.2 \times 10^{-6}\,\text{s}) = 400\,\text{m}$. This is shorter than the average distance measured.

EXERCISE B What is the muon's mean lifetime (Example 36–1) if it is traveling at $v = 0.90c$? (*a*) 0.42 μs; (*b*) 2.3 μs; (*c*) 5.0 μs; (*d*) 5.3 μs; (*e*) 12.0 μs.

CAUTION
Proper time Δt_0 is for two events at the same point in space

We need to clarify how to use Eq. 36–1, $\Delta t = \gamma\,\Delta t_0$, and the meaning of Δt and Δt_0. The equation is true only when Δt_0 represents the time interval between the two events *in a reference frame where an observer at rest sees the two events occur at the same point in space*. (This is the case in Fig. 36–7a where the two events are the light flash being sent and being received at the same point in the spaceship.) This time interval, Δt_0, is called the **proper time**. Then Δt in Eqs. 36–1 represents the time interval between the two events as measured in a reference frame *moving* with speed v with respect to the first. In Example 36–1 above, Δt_0 (and not Δt) was set equal to $2.2 \times 10^{-6}\,\text{s}$ because it is only in the rest frame of the muon that the two events ("birth" and "decay") occur at the same point in space. The proper time Δt_0 is the shortest time between two events any observer can measure. In any other moving reference frame, the time Δt is greater.

CAUTION
Proper time is shortest: $\Delta t > \Delta t_0$

EXAMPLE 36–2 Time dilation at 100 km/h. Let us check time dilation for everyday speeds. A car traveling 100 km/h covers a certain distance in 10.00 s according to the driver's watch. What does an observer at rest on Earth measure for the time interval?

APPROACH The car's speed relative to Earth, written in meters per second, is $100\,\text{km/h} = (1.00 \times 10^5\,\text{m})/(3600\,\text{s}) = 27.8\,\text{m/s}$. The driver is at rest in the reference frame of the car, so we set $\Delta t_0 = 10.00\,\text{s}$ in the time dilation formula.

SOLUTION We use Eq. 36–1a:

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - \dfrac{v^2}{c^2}}} = \frac{10.00\,\text{s}}{\sqrt{1 - \left(\dfrac{27.8\,\text{m/s}}{3.00 \times 10^8\,\text{m/s}}\right)^2}}$$

$$= \frac{10.00\,\text{s}}{\sqrt{1 - (8.59 \times 10^{-15})}}.$$

If you put these numbers into a calculator, you will obtain $\Delta t = 10.00\,\text{s}$, because the denominator differs from 1 by such a tiny amount. The time measured by an observer fixed on Earth would show no difference from that measured by the driver, even with the best instruments. A computer that could calculate to a large number of decimal places would reveal a slight difference between Δt and Δt_0.

We can estimate the difference using the binomial expansion (Appendix A–2),

PROBLEM SOLVING
Use of the binomial expansion

$$(1 \pm x)^n \approx 1 \pm nx. \qquad [\text{for } x \ll 1]$$

In our time dilation formula, we have the factor $\gamma = \left(1 - v^2/c^2\right)^{-\frac{1}{2}}$. Thus

$$\Delta t = \gamma\,\Delta t_0 = \Delta t_0\left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \approx \Delta t_0\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right)$$

$$\approx 10.00\,\text{s}\left[1 + \frac{1}{2}\left(\frac{27.8\,\text{m/s}}{3.00 \times 10^8\,\text{m/s}}\right)^2\right]$$

$$\approx 10.00\,\text{s} + 4 \times 10^{-14}\,\text{s}.$$

So the difference between Δt and Δt_0 is predicted to be $4 \times 10^{-14}\,\text{s}$, an extremely small amount.

EXERCISE C A certain atomic clock keeps precise time on Earth. If the clock is taken on a spaceship traveling at a speed $v = 0.60c$, does this clock now run slow according to the people (*a*) on the spaceship, (*b*) on Earth?

EXAMPLE 36–3 **Reading a magazine on a spaceship.** A passenger on a fictional high-speed spaceship traveling between Earth and Jupiter at a steady speed of $0.75c$ reads a magazine which takes 10.0 min according to her watch. (*a*) How long does this take as measured by Earth-based clocks? (*b*) How much farther is the spaceship from Earth at the end of reading the magazine than it was at the beginning?

APPROACH (*a*) The time interval in one reference frame is related to the time interval in the other by Eq. 36–1a or b. (*b*) At constant speed, distance is speed × time. Because there are two time intervals (Δt and Δt_0) we will get two distances, one for each reference frame. [This surprising result is explored in the next Section (36–6).]

SOLUTION (*a*) The given 10.0-min time interval is the proper time Δt_0—starting and finishing the magazine happen at the same place on the spaceship. Earth clocks measure

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - (v^2/c^2)}} = \frac{10.00\,\text{min}}{\sqrt{1 - (0.75)^2}} = 15.1\,\text{min}.$$

(*b*) In the Earth frame, the spaceship travels a distance $D = v\,\Delta t = (0.75c)\,(15.1\,\text{min}) = (0.75)\,(3.00 \times 10^8\,\text{m/s})\,(15.1\,\text{min} \times 60\,\text{s/min}) = 2.04 \times 10^{11}\,\text{m}$. In the spaceship's frame, the Earth is moving away from the spaceship at $0.75c$, but the time is only 10.0 min, so the distance is measured to be $D_0 = v\,\Delta t_0 = (0.75)\,(3.00 \times 10^8\,\text{m/s})\,(600\,\text{s}) = 1.35 \times 10^{11}\,\text{m}$.

Values for $\gamma = 1/\sqrt{1 - v^2/c^2}$ at a few speeds v are given in Table 36–1.

TABLE 36–1 Values of γ

v	γ
0	1.00000. . .
$0.01c$	1.00005
$0.10c$	1.005
$0.50c$	1.15
$0.90c$	2.3
$0.99c$	7.1

Space Travel?

Time dilation has aroused interesting speculation about space travel. According to classical (Newtonian) physics, to reach a star 100 light-years away would not be possible for ordinary mortals (1 light-year is the distance light can travel in 1 year $= 3.0 \times 10^8\ \text{m/s} \times 3.16 \times 10^7\ \text{s} = 9.5 \times 10^{15}\ \text{m}$). Even if a spaceship could travel at close to the speed of light, it would take over 100 years to reach such a star. But time dilation tells us that the time involved could be less. In a spaceship traveling at $v = 0.999c$, the time for such a trip would be only about $\Delta t_0 = \Delta t\sqrt{1 - v^2/c^2} = (100\ \text{yr})\sqrt{1 - (0.999)^2} = 4.5\ \text{yr}$. Thus time dilation allows such a trip, but the enormous practical problems of achieving such speeds may not be possible to overcome, certainly not in the near future.

When we talk in this Chapter and in the Problems about spaceships moving at speeds close to c, it is for understanding and for fun, but not realistic, although for tiny elementary particles such high speeds *are* realistic.

In this example, 100 years would pass on Earth, whereas only 4.5 years would pass for the astronaut on the trip. Is it just the clocks that would slow down for the astronaut? No.

> *All processes, including aging and other life processes, run more slowly for the astronaut as measured by an Earth observer. But to the astronaut, time would pass in a normal way.*

The astronaut would experience 4.5 years of normal sleeping, eating, reading, and so on. And people on Earth would experience 100 years of ordinary activity.

Twin Paradox

Not long after Einstein proposed the special theory of relativity, an apparent paradox was pointed out. According to this **twin paradox**, suppose one of a pair of 20-year-old twins takes off in a spaceship traveling at very high speed to a distant star and back again, while the other twin remains on Earth. According to the Earth twin, the astronaut twin will age less. Whereas 20 years might pass for the Earth twin, perhaps only 1 year (depending on the spacecraft's speed) would pass for the traveler. Thus, when the traveler returns, the earthbound twin could expect to be 40 years old whereas the traveling twin would be only 21.

This is the viewpoint of the twin on the Earth. But what about the traveling twin? If all inertial reference frames are equally good, won't the traveling twin make all the same claims the Earth twin does, only in reverse? From the reference frame of the spacecraft, the Earth is moving away at high speed, so can't the astronaut twin claim that time passes more slowly on Earth and the twin on Earth will age less? This is the opposite of what the Earth twin predicts. They cannot both be right, for after all the spacecraft returns to Earth and a direct comparison of ages and clocks can be made.

There is, however, no contradiction here. One of the viewpoints is indeed incorrect. The consequences of the special theory of relativity—in this case, time dilation—can be applied only by observers in an *inertial reference frame*. The Earth is such a frame (or nearly so), whereas the spacecraft is not. The spacecraft accelerates at the start and end of its trip and when it turns around at the far point of its journey. Part of the time, the astronaut twin may be in an inertial frame (and is justified in saying the Earth twin's clocks run slow). But during the accelerations, the twin on the spacecraft is not in an inertial frame. So she cannot use special relativity to predict their relative ages when she returns to Earth. The Earth twin stays in the same inertial frame, and we can thus trust her predictions based on special relativity. Thus, there is no paradox. The prediction of the Earth twin that the traveling twin ages less is the correct one.

*Global Positioning System (GPS)

PHYSICS APPLIED
Global positioning system (GPS)

Airplanes, cars, boats, and hikers use **global positioning system (GPS)** receivers to tell them quite accurately where they are at a given moment (Fig. 36–8). There are more than 30 global positioning system satellites that send out precise time signals using atomic clocks. Your receiver compares the times received from at least four satellites, all of whose times are carefully synchronized to within 1 part in 10^{13}. By comparing the time differences with the known satellite positions and the fixed speed of light, the receiver can determine how far it is from each satellite and thus where it is on the Earth. It can do this to an accuracy of a few meters, if it has been constructed to make corrections due to relativity, such as the one we discuss next.

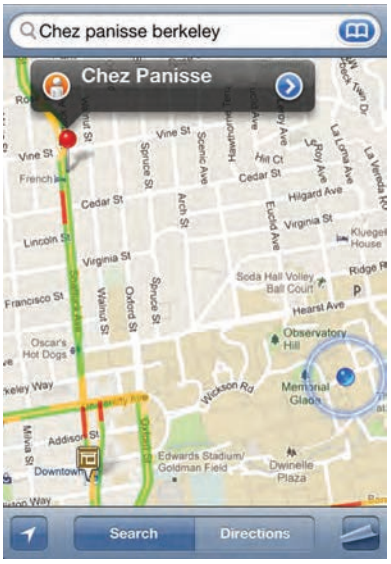


FIGURE 36–8 A visiting professor of physics uses the GPS on her smart phone to find a restaurant (red dot). Her location in the physics department is the blue dot. Traffic on some streets is also shown (green = good, orange = slow, red = heavy traffic) which comes in part by tracking cell phone movements.

CONCEPTUAL EXAMPLE 36–4 **A relativity correction to GPS.** GPS satellites move at about $4\,\text{km/s} = 4000\,\text{m/s}$. Show that a good GPS receiver needs to correct for time dilation if it is to produce results consistent with atomic clocks accurate to 1 part in 10^{13}.

RESPONSE Let us calculate the magnitude of the time dilation effect by inserting $v = 4000\,\text{m/s}$ into Eq. 36–1a:

$$\Delta t = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\Delta t_0 = \frac{1}{\sqrt{1 - \left(\frac{4\times10^3\,\text{m/s}}{3\times10^8\,\text{m/s}}\right)^2}}\Delta t_0$$

$$= \frac{1}{\sqrt{1 - 1.8\times10^{-10}}}\Delta t_0.$$

We use the binomial expansion: $(1 \pm x)^n \approx 1 \pm nx$ for $x \ll 1$ (see Appendix A–2) which here is $(1-x)^{-\frac{1}{2}} \approx 1 + \frac{1}{2}x$. That is

$$\Delta t = \left(1 + \tfrac{1}{2}(1.8\times10^{-10})\right)\Delta t_0 = \left(1 + 9\times10^{-11}\right)\Delta t_0.$$

The time "error" divided by the time interval is

$$\frac{(\Delta t - \Delta t_0)}{\Delta t_0} = 1 + 9\times10^{-11} - 1 = 9\times10^{-11} \approx 1\times10^{-10}.$$

Time dilation, if not accounted for, would introduce an error of about 1 part in 10^{10}, which is 1000 times greater than the precision of the atomic clocks. Not good! By not correcting for time dilation means a receiver could give much poorer position accuracy.

NOTE GPS devices must make other corrections as well, including effects associated with general relativity.

36–6 Length Contraction

Time intervals are not the only things different in different reference frames. Space intervals—lengths and distances—are different as well, according to the special theory of relativity, and we illustrate this with a thought experiment.

Observers on Earth watch a spacecraft traveling at speed v away from Earth to Neptune, Fig. 36–9a. The distance between the planets, as measured by the Earth observers, is ℓ_0. The time required for the trip, measured from Earth, is

$$\Delta t = \frac{\ell_0}{v}. \qquad \text{[Earth observer]}$$

In Fig. 36–9b we see the point of view of observers on the spacecraft. In this frame of reference, the spaceship is at rest; Earth and Neptune move† with speed v (but opposite direction).

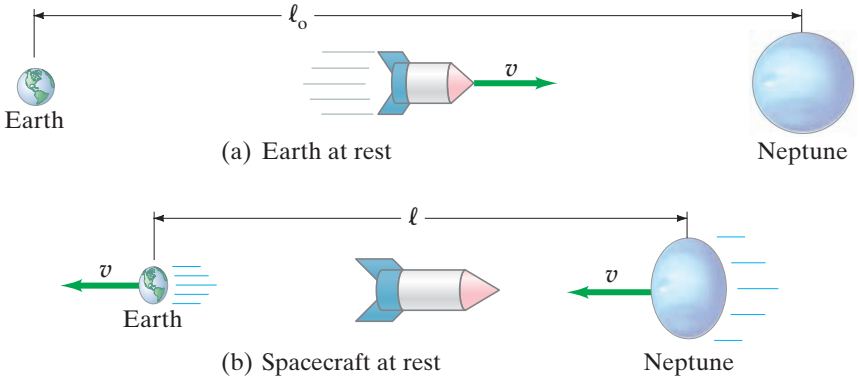


FIGURE 36–9 (a) A spaceship traveling at very high speed v from Earth to the planet Neptune, as seen from Earth's frame of reference. (b) According to an observer on the spaceship, Earth and Neptune are moving at the very high speed v: Earth leaves the spaceship, and a time Δt_0 later Neptune arrives at the spaceship.

The time between departure of Earth and arrival of Neptune, as observed from the spacecraft, is the "proper time" Δt_0 (page 1066) because these two events occur at the same point in space (i.e., at the spacecraft). Therefore the time interval is less for the spacecraft observers than for the Earth observers. That is, because of time dilation (Eq. 36–1a), the time for the trip as viewed by the spacecraft is

$$\Delta t_0 = \Delta t\sqrt{1 - v^2/c^2} = \Delta t/\gamma. \qquad \text{[spacecraft observer]}$$

Because the spacecraft observers measure the same speed but less time between these two events, they also measure the distance as less. If we let ℓ be the distance between the planets as viewed by the spacecraft observers (Fig. 36–9b), then $\ell = v\,\Delta t_0$, which we can rewrite as $\ell = v\,\Delta t_0 = v\,\Delta t\sqrt{1 - v^2/c^2} = \ell_0\sqrt{1 - v^2/c^2}$. Thus we have the important result that

LENGTH CONTRACTION

$$\ell = \ell_0\sqrt{1 - v^2/c^2} \qquad \textbf{(36–3a)}$$

or

$$\ell = \ell_0/\gamma. \qquad \textbf{(36–3b)}$$

This is a general result of the special theory of relativity and applies to lengths of objects as well as to distance between objects. The result can be stated most simply in words as:

> **the length of an object moving relative to an observer is measured to be shorter along its direction of motion than when it is at rest.**

CAUTION

Proper length is measured in reference frame where the two positions are at rest

This is called **length contraction**. The length ℓ_0 in Eqs. 36–3 is called the **proper length**. It is the length of the object (or distance between two points whose positions are measured at the same time) as determined by *observers at rest* with respect to the object. Equations 36–3 give the length ℓ that will be measured by observers when the object travels past them at speed v.

It is important to note that length contraction occurs *only along the direction of motion*. For example, the moving spaceship in Fig. 36–9a is shortened in length, but its height is the same as when it is at rest.

Length contraction, like time dilation, is not noticeable in everyday life because the factor $\sqrt{1 - v^2/c^2}$ in Eq. 36–3a differs significantly from 1.00 only when v is very large (Table 36–1 on page 1067).

†We assume v is much greater than the relative speed of Neptune and Earth (which we thus ignore).

EXAMPLE 36–5 **Painting's contraction.** A rectangular painting measures 1.00 m tall and 1.50 m wide, Fig. 36–10a. It is hung on the side wall of a spaceship which is moving past the Earth at a speed of $0.90c$. (*a*) What are the dimensions of the picture according to the captain of the spaceship? (*b*) What are the dimensions as seen by an observer on the Earth?

APPROACH We apply the length contraction formula in part (*b*) to the dimension parallel to the motion; v is the speed of the painting relative to the Earth observer.

SOLUTION (*a*) The painting is at rest $(v = 0)$ on the spaceship so it (as well as everything else in the spaceship) looks perfectly normal to everyone on the spaceship. The captain sees a 1.00-m by 1.50-m painting.

(*b*) Only the dimension in the direction of motion is shortened, so the height is unchanged at 1.00 m, Fig. 36–10b. The length, however, is contracted to

$$\ell = \ell_0\sqrt{1 - v^2/c^2} = (1.50\,\text{m})\sqrt{1 - (0.90)^2} = 0.65\,\text{m}.$$

So the picture has dimensions $1.00\,\text{m} \times 0.65\,\text{m}$ to an observer on Earth.

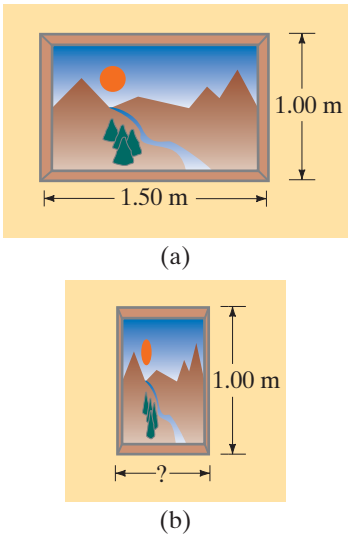


FIGURE 36–10 Example 36–5.

EXAMPLE 36–6 **A fantasy supertrain.** A very fast train with a "proper length" of $\ell_0 = 500\,\text{m}$ (measured by people at rest on the train) is passing through a tunnel that is 200 m long according to observers on the ground. Let us imagine the train's speed to be so great that the train fits completely within the tunnel as seen by observers on the ground. That is, the engine is just about to emerge from one end of the tunnel at the time the last car disappears into the other end. What is the train's speed?

APPROACH Since the train just fits inside the tunnel, its length measured by the person on the ground is $\ell = 200\,\text{m}$. The length contraction formula, Eq. 36–3a or b, can thus be used to solve for v.

SOLUTION Substituting $\ell = 200\,\text{m}$ and $\ell_0 = 500\,\text{m}$ into Eq. 36–3a gives

$$200\,\text{m} = 500\,\text{m}\sqrt{1 - v^2/c^2}.$$

Dividing both sides by 500 m and squaring, we get

$$(0.40)^2 = 1 - v^2/c^2$$

or $v = c\sqrt{1 - (0.40)^2} = 0.92c$.

NOTE No real train could go this fast. But it is fun to think about.

NOTE An observer on the *train* would *not* see the two ends of the train inside the tunnel at the same time. Recall that observers moving relative to each other do not agree about simultaneity. (See Example 36–7, next.)

EXERCISE D What is the length of the tunnel as measured by observers on the train in Example 36–6?

CONCEPTUAL EXAMPLE 36–7 **Resolving the train and tunnel length.** Observers at rest on the Earth see a very fast 200-m-long train pass through a 200-m-long tunnel (as in Example 36–6) so that the train momentarily disappears from view inside the tunnel. Observers on the train measure the train's length to be 500 m and the tunnel's length to be only 80 m (Exercise D, using Eq. 36–3a). Clearly a 500-m-long train cannot fit inside an 80-m-long tunnel. How is this apparent inconsistency explained?

RESPONSE Events simultaneous in one reference frame may not be simultaneous in another. Let the engine emerging from one end of the tunnel be "event A," and the last car disappearing into the other end of the tunnel "event B." To observers in the Earth frame, as we stated in Example 36–6, events A and B are simultaneous. To observers on the train, however, the events are not simultaneous. In the train's frame, event A occurs before event B. As the engine emerges from the tunnel, observers on the train observe the last car as still $500\,\text{m} - 80\,\text{m} = 420\,\text{m}$ from the entrance to the tunnel. What a kick—not our usual way of thinking.

36–7 Four-Dimensional Space–Time

Let us imagine a person is on a train moving at a very high speed, say 0.65c, Fig. 36–11. This person begins a meal at 7:00 and finishes at 7:15, according to a clock on the train. The two events, beginning and ending the meal, take place at the same point on the train, so the "proper time" between these two events is 15 min. To observers on Earth, the plate is moving and the meal will take longer—20 min according to Eqs. 36–1. Let us assume that the meal was served on a 20-cm-diameter plate (its "proper length"). To observers on the Earth, the plate is moving and is only 15 cm wide (length contraction). Thus, to observers on the Earth, the meal looks smaller but lasts longer.

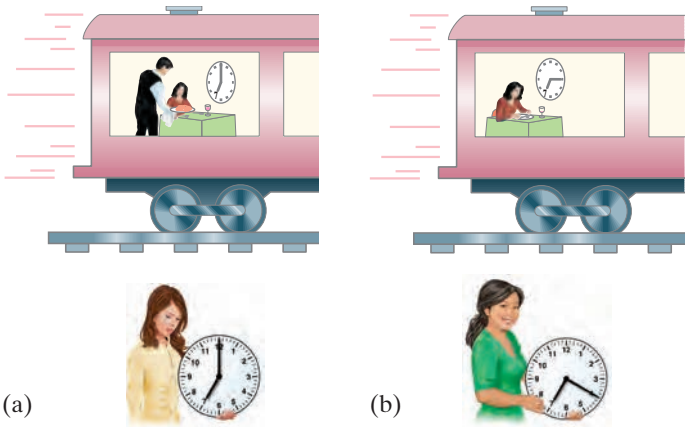


FIGURE 36–11 According to an accurate clock on a fast-moving train, a person (a) begins dinner at 7:00 and (b) finishes at 7:15. At the beginning of the meal, two observers on Earth set their watches to correspond with the clock on the train. These observers measure the eating time as 20 minutes.

CAUTION
Space and time "exchanged"

In a sense the two effects, time dilation and length contraction, balance each other. When viewed from the Earth, what an object seems to lose in size it gains in length of time it lasts. Space, or length, is exchanged for time.

Considerations like this led to the idea of **four-dimensional space–time**: space takes up three dimensions and time is a fourth dimension. Space and time are intimately connected. Just as when we squeeze a balloon we make one dimension larger and another smaller, so when we examine objects and events from different reference frames, a certain amount of space is exchanged for time, or vice versa.

Although the idea of four dimensions may seem strange, it refers to the idea that any object or event is specified by four quantities—three to describe where in space, and one to describe when in time. The really unusual aspect of four-dimensional space–time is that space and time can intermix: a little of one can be exchanged for a little of the other when the reference frame is changed.

[In Galilean–Newtonian relativity, the time interval between two events, Δt, and the distance between two events or points, Δx, are invariant quantities no matter what inertial reference frame they are viewed from. Neither of these quantities is invariant according to Einstein's relativity. But there is an invariant quantity in four-dimensional space–time, called the **space–time interval**, which is $(\Delta s)^2 = (c\,\Delta t)^2 - (\Delta x)^2$.]

36–8 Galilean and Lorentz Transformations

We examine in detail the mathematics of relating quantities in one inertial reference frame to the equivalent quantities in another. In particular, we will see how positions and velocities *transform* (that is, change) from one inertial reference frame to the other.

We begin with the classical or Galilean viewpoint. Consider two inertial reference frames S and S′ which are each characterized by a set of coordinate axes, Fig. 36–12. The axes x and y (z is not shown) refer to S, and x' and y' refer to S′. The x' and x axes overlap one another, and we assume that frame S′ moves to the right in the x direction at constant speed v with respect to S.

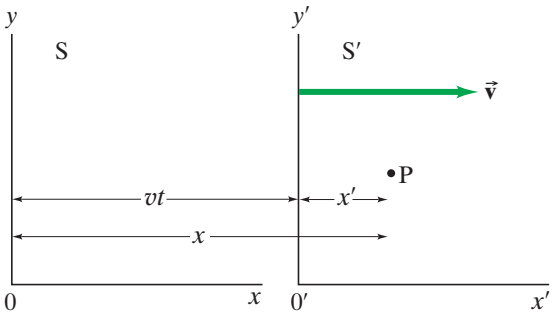


FIGURE 36–12 Inertial reference frame S′ moves to the right at constant speed v with respect to frame S.

We assume that we, as observers, are at rest in S. For simplicity let us assume the origins 0 and 0′ of the two reference frames are superimposed at time $t = 0$. That is, $x = x'$ at $t = 0$ for all positions.

Now consider an event that occurs at some point P (Fig. 36–12) represented by the coordinates x', y', z' in reference frame S′ at the time t'. What will be the coordinates of P in S? Since S and S′ overlap precisely initially, after a time t', S′ will have moved a distance vt'. Therefore, at time t', $x = x' + vt'$. The y and z coordinates, on the other hand, are not altered by motion along the x axis, as we saw on page 1070. Thus $y = y'$ and $z = z'$. Finally, since time is assumed to be absolute in Galilean–Newtonian physics, clocks in the two frames will agree with each other, so $t = t'$. We summarize these in the following **Galilean transformation equations**:

$$\begin{aligned} x &= x' + vt' \\ y &= y' \\ z &= z' \\ t &= t'. \end{aligned} \qquad \text{[Galilean]} \quad \textbf{(36–4)}$$

These equations give the coordinates of an event in the S frame when those in the S′ frame are known. If those in the S frame are known, then the S′ coordinates are obtained from

$$x' = x - vt, \qquad y' = y, \qquad z' = z, \qquad t' = t. \qquad \text{[Galilean]}$$

These four equations are the "inverse" transformation. They can be obtained from Eqs. 36–4 (1) by solving Eqs. 36–4 for x', and (2) by exchanging primed and unprimed quantities and replacing v by $-v$. This makes sense because looking from the S′ frame, S moves to the left (negative x direction) with speed v.

Next suppose the point P in Fig. 36–12 represents a particle that is moving. Let the components of its velocity vector in S′ be u'_x, u'_y, u'_z. (We use u to distinguish it from the relative velocity of the two frames, v.) Now $u'_x = dx'/dt'$, $u'_y = dy'/dt'$ and $u'_z = dz'/dt'$. The velocity of P as seen from S will have components u_x, u_y, and u_z. We can show how these are related to the velocity components in S′ by differentiating Eqs. 36–4. For u_x we get

$$u_x = \frac{dx}{dt} = \frac{d(x' + vt')}{dt'} = u'_x + v$$

since v is assumed constant. For the other components, $u'_y = u_y$ and $u'_z = u_z$, so we have

$$\begin{aligned} u_x &= u'_x + v \\ u_y &= u'_y \\ u_z &= u'_z. \end{aligned} \qquad \text{[Galilean]} \quad \textbf{(36–5)}$$

These are known as the **Galilean velocity transformation equations**. We see that the y and z components of velocity are unchanged, but the x components differ by v. This is just what we have used before when dealing with relative velocity (Section 3–9). For example, if S′ is a train and S the Earth, and the train moves with speed v with respect to Earth, a person walking toward the front of the train with speed u'_x will have a speed with respect to the Earth of $u_x = u'_x + v$.

The Galilean transformations, Eqs. 36–4 and 36–5, give quite accurate results only when the velocities involved are much less than the speed of light, c. We can see, for example, that the first of Eqs. 36–5 will not work for the speed of light, c, which is the same in all inertial reference frames (the second postulate of special relativity). That is, light traveling in S′ with speed $u'_x = c$ will have speed $c + v$ in S, according to Eqs. 36–5, whereas the theory of relativity insists it must be c also in S. Clearly, then, a new set of transformation equations is needed to deal with relativistic velocities.

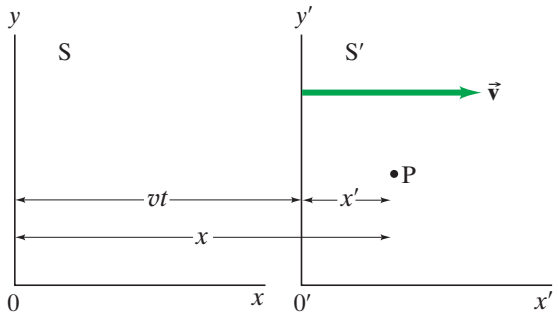


FIGURE 36–12 (Repeated.) Inertial reference frame S′ moves to the right at constant speed v with respect to frame S.

We can derive the required equations, again looking at Fig. 36–12. We hope it will work out to be simple, and assume the transformation is linear and for x is of the form

$$x = \gamma(x' + vt'). \qquad \textbf{(i)}$$

[We are NOT assuming γ is given by Eq. 36–2, $1/\sqrt{1 - v^2/c^2}$, as in Section 36–5. Our minds are open. Let's see.] That is, we modify the first of Eqs. 36–4 by multiplying by a factor γ which is yet to be determined. We assume the y and z equations are unchanged

$$y = y', \qquad z = z'$$

because there is no length contraction in these directions. We will not assume a form for t, but will derive it. The inverse equations must have the same form with v replaced by $-v$. (The principle of relativity demands it, since S′ moving to the right with respect to S is equivalent to S moving to the left with respect to S′.) Therefore

$$x' = \gamma(x - vt). \qquad \textbf{(ii)}$$

Suppose a light pulse leaves the common origin of S and S′ at time $t = t' = 0$. Then after a time t it will have traveled along the x axis a distance $x = ct$ (in S), or $x' = ct'$ (in S′). Therefore, from Eqs. (i) and (ii) above,

$$ct = x = \gamma(ct' + vt') = \gamma(c + v)t', \qquad \textbf{(iii)}$$

$$ct' = x' = \gamma(ct - vt) = \gamma(c - v)t. \qquad \textbf{(iv)}$$

From Eq. (iv), $t' = \gamma(c - v)(t/c)$, and we substitute this into Eq. (iii) and find $ct = \gamma(c + v)\gamma(c - v)(t/c) = \gamma^2(c^2 - v^2)t/c$. We cancel out the t on each side and solve for γ to find

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

We have found that γ is, in fact, the value for γ we used before, Eq. 36–2.

Now that we have found γ, we need only find the relation between t and t'. To do so, we combine $x' = \gamma(x - vt)$ with $x = \gamma(x' + vt')$:

$$x' = \gamma(x - vt) = \gamma[\gamma(x' + vt') - vt].$$

We solve for t, doing some algebra, and find $t = \gamma(t' + vx'/c^2)$. In summary,

$$\begin{aligned} x &= \frac{1}{\sqrt{1 - v^2/c^2}}(x' + vt') \\ y &= y' \\ z &= z' \\ t &= \frac{1}{\sqrt{1 - v^2/c^2}}\left(t' + \frac{vx'}{c^2}\right). \end{aligned} \qquad \textbf{(36–6)}$$

LORENTZ TRANSFORMATIONS

These are the **Lorentz transformation equations**. They were first proposed, in a slightly different form, by Lorentz in 1904 to explain the null result of the Michelson–Morley experiment and to make Maxwell's equations take the same form in all inertial reference frames. A year later, Einstein derived them independently based on his theory of relativity. Notice that not only is the x equation modified as compared to the Galilean transformation, but so is the t equation. Indeed, we see directly in this last equation how the space and time coordinates mix.

Deriving Length Contraction

We now derive the length contraction formula, Eq. 36–3, from the Lorentz transformation equations. We consider two reference frames S and S′ as in Fig. 36–12.

Let an object of length ℓ_0 be at rest on the x axis in S. The coordinates of its two end points are x_1 and x_2, so that $x_2 - x_1 = \ell_0$. At any instant in S′, the end points will be at x_1' and x_2' as given by the Lorentz transformation equations. The length measured in S′ is $\ell = x_2' - x_1'$. An observer in S′ measures this length by measuring x_2' and x_1' at the same time (in the S′ reference frame), so $t_2' = t_1'$. Then, from the first of Eqs. 36–6,

$$\ell_0 = x_2 - x_1 = \frac{1}{\sqrt{1 - v^2/c^2}}(x_2' + vt_2' - x_1' - vt_1'),$$

and with $t_2' = t_1'$, we have

$$\ell_0 = \frac{1}{\sqrt{1 - v^2/c^2}}(x_2' - x_1') = \frac{\ell}{\sqrt{1 - v^2/c^2}},$$

or

$$\ell = \ell_0\sqrt{1 - v^2/c^2},$$

which is Eq. 36–3a, the length contraction formula.

Deriving Time Dilation

We now derive the time-dilation formula, Eq. 36–1a, from the Lorentz transformation equations.

The time Δt_0 between two events that occur at the same place $(x_2' = x_1')$ in S′ is measured to be $\Delta t_0 = t_2' - t_1'$. Since $x_2' = x_1'$, then from the last of Eqs. 36–6, the time Δt between the events as measured in S is

$$\begin{aligned} \Delta t = t_2 - t_1 &= \frac{1}{\sqrt{1 - v^2/c^2}}\left(t_2' + \frac{vx_2'}{c^2} - t_1' - \frac{vx_1'}{c^2}\right) \\ &= \frac{1}{\sqrt{1 - v^2/c^2}}(t_2' - t_1') = \frac{\Delta t_0}{\sqrt{1 - v^2/c^2}}, \end{aligned}$$

which is Eq. 36–1a, the time-dilation formula. Note that we chose S′ to be the frame in which the two events occur at the same place, so that $x_1' = x_2'$ and the terms containing x_1' and x_2' cancel out.

Relativistic Addition of Velocities

The relativistically correct velocity equations are obtained by differentiating Eqs. 36–6 with respect to time. For example (using $\gamma = 1/\sqrt{1 - v^2/c^2}$ and the chain rule for derivatives):

$$u_x = \frac{dx}{dt} = \frac{d}{dt}\left[\gamma(x' + vt')\right]$$
$$= \frac{d}{dt'}\left[\gamma(x' + vt')\right]\frac{dt'}{dt} = \gamma\left[\frac{dx'}{dt'} + v\right]\frac{dt'}{dt}.$$

But $dx'/dt' = u'_x$ and we differentiate the last of Eqs. 36–6 with respect to time to get

$$\frac{dt'}{dt} = \frac{1}{(dt/dt')} = \frac{1}{\gamma(1 + vu'_x/c^2)}.$$

Then

$$u_x = \frac{[\gamma(u'_x + v)]}{[\gamma(1 + vu'_x/c^2)]} = \frac{u'_x + v}{1 + vu'_x/c^2}.$$

The others are obtained in the same way and we collect them here:

LORENTZ VELOCITY TRANSFORMATIONS

$$u_x = \frac{u'_x + v}{1 + vu'_x/c^2} \qquad \textbf{(36–7a)}$$

$$u_y = \frac{u'_y\sqrt{1 - v^2/c^2}}{1 + vu'_x/c^2} \qquad \textbf{(36–7b)}$$

$$u_z = \frac{u'_z\sqrt{1 - v^2/c^2}}{1 + vu'_x/c^2}. \qquad \textbf{(36–7c)}$$

⚠ CAUTION
Velocity components "mix"

Note that even though the relative velocity $\vec{\mathbf{v}}$ is in the x direction, if the object has y or z components of velocity, they too are affected by v and the x' component of the object's velocity. This was not true for the Galilean transformation, Eqs. 36–5.

EXAMPLE 36–8 **Adding velocities.** Calculate the speed of rocket 2 in Fig. 36–13 with respect to Earth.

APPROACH Consider Earth as reference frame S, and rocket 1 as reference frame S′. Rocket 2 moves with speed $u' = 0.60c$ with respect to rocket 1. Rocket 1 has speed $v = 0.60c$ with respect to Earth. The velocities are along the same straight line which we take to be the x (and x') axis. We need use only the first of Eqs. 36–7, and drop the subscripts x and x'.

SOLUTION The speed of rocket 2 with respect to Earth is

$$u = \frac{u' + v}{1 + \dfrac{vu'}{c^2}} = \frac{0.60c + 0.60c}{1 + \dfrac{(0.60c)(0.60c)}{c^2}} = \frac{1.20c}{1.36} = 0.88c.$$

NOTE The Galilean transformation would have given $u = 1.20c$.

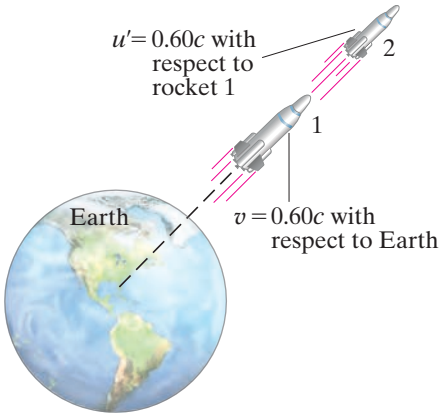


FIGURE 36–13 Rocket 1 moves away from Earth at speed $v = 0.60c$. Rocket 2 is fired from rocket 1 with speed $u' = 0.60c$. What is the speed of rocket 2 with respect to Earth?

For one-dimensional motion, we can rewrite Eq. 36–7a as we did in Example 36–8:

$$u = \frac{u' + v}{1 + vu'/c^2}. \qquad \text{[1–D]} \qquad \textbf{(36–7d)}$$

EXERCISE E Calculate the speed of rocket 2 in Fig. 36–13 relative to Earth if it was shot from rocket 1 at a speed $u' = 3000\ \text{km/s} = 0.010c$. Assume rocket 1 had a speed $v = 6000\ \text{km/s} = 0.020c$.

EXERCISE F Return to the Chapter-Opening Question, page 1055, and answer it again now. Try to explain why you may have answered differently the first time.

We can see that Eqs. 36–7 reduce to the classical (Galilean) forms for velocities small compared to the speed of light, since $1 + vu'/c^2 \approx 1$ for v and $u' \ll c$. Thus, $u \approx v + u'$, as in classical physics (Chapter 3).

At the other extreme, let rocket 1 in Fig. 36–13 send out a beam of light, so that $u' = c$. Then Eq. 36–7a tells us the speed of light relative to Earth is

$$u = \frac{0.60c + c}{1 + \dfrac{(0.60c)(c)}{c^2}} = c,$$

which is fully consistent with the second postulate of relativity.

36–9 Relativistic Momentum

So far in this Chapter, we have seen that two basic mechanical quantities, length and time intervals, need modification because they are relative—their value depends on the reference frame from which they are measured. We might expect that other physical quantities might need some modification according to the theory of relativity, such as momentum and energy.

The analysis of collisions between two particles shows that if we want to preserve the law of conservation of momentum in relativity, we must redefine momentum as

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}} = \gamma mv. \qquad \textbf{(36–8)}$$

Here γ is shorthand for $1/\sqrt{1 - v^2/c^2}$ as before (Eq. 36–2). For speeds much less than the speed of light, Eq. 36–8 gives the classical momentum, $p = mv$.

Relativistic momentum has been tested many times on tiny elementary particles (such as muons), and it has been found to behave in accord with Eq. 36–8. We derive Eq. 36–8 in the optional Subsection starting below.

EXAMPLE 36–9 **Momentum of moving electron.** Compare the momentum of an electron to its classical value when it has a speed of (*a*) 4.00×10^7 m/s in the CRT of an old TV set, and (*b*) $0.98c$ in an accelerator used for cancer therapy.

APPROACH We use Eq. 36–8 for the momentum of a moving electron.

SOLUTION (*a*) At $v = 4.00 \times 10^7$ m/s, the electron's momentum is

$$p = \frac{mv}{\sqrt{1 - \dfrac{v^2}{c^2}}} = \frac{mv}{\sqrt{1 - \dfrac{(4.00 \times 10^7\ \text{m/s})^2}{(3.00 \times 10^8\ \text{m/s})^2}}} = 1.01mv.$$

The factor $\gamma = 1/\sqrt{1 - v^2/c^2} \approx 1.01$, so the momentum is only about 1% greater than the classical value.

(*b*) With $v = 0.98c$, the momentum is

$$p = \frac{mv}{\sqrt{1 - \dfrac{v^2}{c^2}}} = \frac{mv}{\sqrt{1 - \dfrac{(0.98c)^2}{c^2}}} = \frac{mv}{\sqrt{1 - (0.98)^2}} = 5.0mv.$$

An electron traveling at 98% the speed of light has $\gamma = 5.0$ and a momentum 5.0 times its classical value.

Newton's second law, stated in its most general form, is

$$\vec{\mathbf{F}} = \frac{d\vec{\mathbf{p}}}{dt} = \frac{d}{dt}(\gamma m\vec{\mathbf{v}}) = \frac{d}{dt}\left(\frac{m\vec{\mathbf{v}}}{\sqrt{1 - v^2/c^2}}\right) \qquad \textbf{(36–9)}$$

and is valid relativistically.

*Derivation of Relativistic Momentum

Classically, momentum is a conserved quantity. We hope to find a formula for momentum that will also be valid relativistically. To do so, let us assume it has the general form given by $p = fmv$ where f is some function of v: $f(v)$. We consider a hypothetical collision between two objects—a thought experiment—and see what form $f(v)$ must take if momentum is to be conserved.

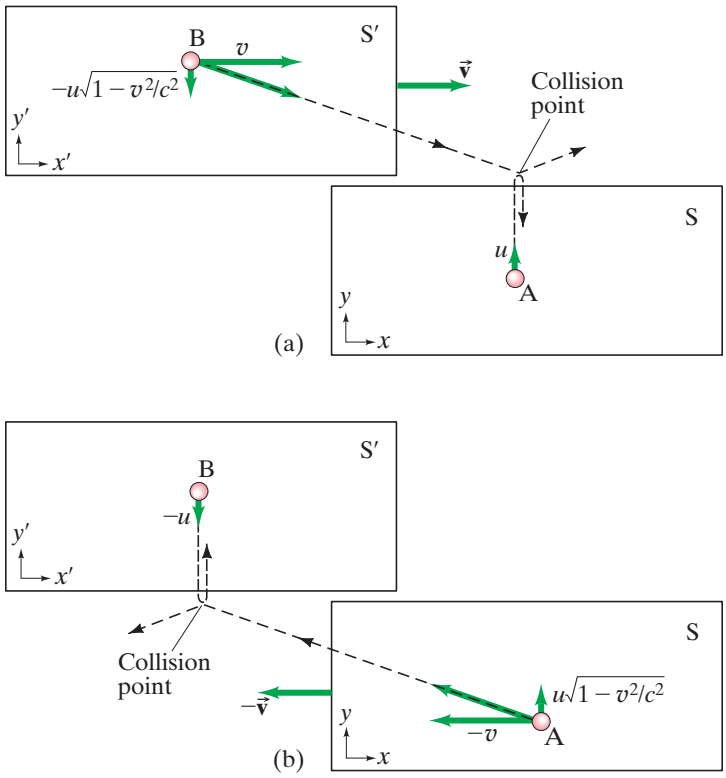


FIGURE 36–14 Deriving the momentum formula. Collision as seen by observers (a) in reference frame S, (b) in reference frame S′.

Our thought experiment involves the elastic collision of two identical balls, A and B. We consider two inertial reference frames, S and S′, moving along the x axis with a speed v with respect to each other, Fig. 36–14. That is, frame S′ moves to the right with velocity $\vec{\mathbf{v}}$ as seen by observers in frame S (Fig. 36–14a); and frame S moves to the left with $-\vec{\mathbf{v}}$ as seen by observers in S′ (Fig. 36–14b). In reference frame S, ball A is thrown with speed u in the $+y$ direction. In reference frame S′, ball B is thrown with speed u in the negative y' direction. The two balls are thrown at just the right time so that they collide. We assume that they rebound elastically and that each moves with the same speed u back in the opposite direction in its thrower's reference frame. In reference frame S, Fig. 36–14a, ball A has $u_x = 0$ both before and after the collision; it has $u_y = +u$ before the collision and $u_y = -u$ after the collision. In frame S′, Fig. 36–14b, ball A has x component of velocity $u'_x = -v$ both before and after the collision, and a y' component (solving Eq. 36–7b for u'_y and putting $u'_x = -v$) of magnitude

$$u'_y = \frac{u_y(1 + vu'_x/c^2)}{\sqrt{1 - v^2/c^2}} = \frac{u(1 + v(-v)/c^2)}{\sqrt{1 - v^2/c^2}}$$

or

$$u'_y = u\sqrt{1 - v^2/c^2}.$$

The same holds true for ball B, except in reverse: $u'_y = -u\sqrt{1 - v^2/c^2}$. These velocity components are indicated in Fig. 36–14.

We now apply the law of conservation of momentum, which we hope remains valid in relativity, even if momentum has to be redefined. That is, we assume that the total momentum before the collision is equal to the total momentum after the collision. We apply conservation of momentum to the y component of momentum in reference frame S (Fig. 36–14a). To make our task easier, let us assume $u \ll v$ so that the speed of ball B as seen in reference frame S is essentially v. Then B's y component of momentum in S before collision is $-f(v)mu\sqrt{1 - v^2/c^2}$ and after the collision B's momentum in S is $+f(v)mu\sqrt{1 - v^2/c^2}$. Ball A in S has y component $f(u)mu$ before and $-f(u)mu$ after the collision. (We use $f(u)$ for A because its speed in S is only u.)

Conservation of momentum in S for the y component is

$$(p_{Ay} + p_{By})_{before} = (p_{Ay} + p_{By})_{after}$$

$$f(u)mu - f(v)mu\sqrt{1 - v^2/c^2} = -f(u)mu + f(v)mu\sqrt{1 - v^2/c^2}.$$

We solve this for $f(v)$ and obtain

$$f(v) = \frac{f(u)}{\sqrt{1 - v^2/c^2}}.$$

To simplify this relation so we can solve for f, let us allow u to become very small so that it approaches zero (this corresponds to a glancing collision with one of the balls essentially at rest and the other moving with speed v). Then the momentum terms $f(u)mu$ are in the nonrelativistic realm and take on the classical form, simply mu, meaning that $f(u) = 1$. So the previous equation becomes

$$f(v) = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

We see that $f(v)$ comes out to be the factor we used before and called γ, and here has been shown to be valid for ball A. Using Fig. 36–14b we can derive the same relation for ball B. Thus we can conclude that we need to define the relativistic momentum of a particle moving with velocity $\vec{\mathbf{v}}$ as

$$\vec{\mathbf{p}} = \frac{m\vec{\mathbf{v}}}{\sqrt{1 - v^2/c^2}} = \gamma m\vec{\mathbf{v}}. \qquad \textbf{(36–8, again)}$$

With this definition the law of conservation of momentum will remain valid even in the relativistic realm. This relativistic momentum formula (Eq. 36–8) has been tested countless times on tiny elementary particles and been found valid.

*Rest Mass and Relativistic Mass

CAUTION
Most physicists prefer to consider the mass of a particle as fixed

The relativistic definition of momentum, Eq. 36–8, has sometimes been interpreted as an increase in the mass of an object. In this interpretation, a particle can have a **relativistic mass**, m_{rel}, which increases with speed according to

$$m_{rel} = \frac{m}{\sqrt{1 - v^2/c^2}}.$$

In this "mass-increase" formula, m is referred to as the **rest mass** of the object. With this interpretation, *the mass of an object appears to increase as its speed increases*. But there are problems with relativistic mass. If we plug it into formulas like $F = ma$ or $K = \frac{1}{2}mv^2$, we obtain formulas that do not agree with experiment. If we write Newton's second law in its more general form, $\vec{\mathbf{F}} = d\vec{\mathbf{p}}/dt$, we do get a correct result, Eq. 36–9.

Also, be careful *not* to think a mass acquires more particles or more molecules as its speed becomes very large. It doesn't. Today, most physicists prefer not to use relativistic mass, so an object has only one mass, its rest mass, and it is only the momentum that increases with speed.

Whenever we talk about the mass of an object, we will always mean its rest mass (a fixed value). [But see Problem 103.]

36–10 The Ultimate Speed

A basic result of the special theory of relativity is that the speed of an object cannot equal or exceed the speed of light. That the speed of light is a natural speed limit in the universe can be seen from any of Eqs. 36–1, 36–3, 36–8, or the addition of velocities formula. It is perhaps easiest to see from Eq. 36–8. As an object is accelerated to greater and greater speeds, its momentum becomes larger and larger. Indeed, if v were to equal c, the denominator in this equation would be zero, and the momentum would be infinite. To accelerate an object up to $v = c$ would thus require infinite energy, and so is not possible.

36–11 $E = mc^2$; Mass and Energy

If momentum needs to be modified to fit with relativity as we just saw in Eq. 36–8, then we might expect that energy would also need to be rethought. Indeed, Einstein not only developed a new formula for kinetic energy to replace $K = \frac{1}{2}mv^2$, but also found a new relation between mass and energy, and the startling idea that mass is a form of energy.

Deriving Relativistic Energy

We start with the work-energy principle (Chapter 7), hoping it is still valid in relativity and will give verifiable results. That is, we assume the net work done on a particle is equal to its change in kinetic energy (K), and we take the motion of the particle to be along the x axis. The work done to increase a particle's speed from zero to v is

$$W = \int_i^f F\,dx = \int_i^f \frac{dp}{dt}\,dx = \int_i^f \frac{dp}{dt}\,v\,dt = \int_i^f v\,dp$$

where i and f refer to the initial $(v = 0)$ and final $(v = v)$ states. Since $d(pv) = p\,dv + v\,dp$ we can write

$$v\,dp = d(pv) - p\,dv$$

so

$$W = \int_i^f d(pv) - \int_i^f p\,dv.$$

The first term on the right of the equal sign is

$$\int_i^f d(pv) = pv\Big|_i^f = (\gamma mv)v = \frac{mv^2}{\sqrt{1 - v^2/c^2}}.$$

The second term in our equation for W above is integrated using

$$\frac{d}{dv}\left(\sqrt{1 - v^2/c^2}\right) = -(v/c^2)/\sqrt{1 - v^2/c^2},$$

and so the second term in the formula for W becomes

$$-\int_i^f p\,dv = -\int_0^v \frac{mv}{\sqrt{1 - v^2/c^2}}\,dv = mc^2\sqrt{1 - v^2/c^2}\Big|_0^v$$

$$= mc^2\sqrt{1 - v^2/c^2} - mc^2.$$

Combining these results, we have for W:

$$W = \frac{mv^2}{\sqrt{1 - v^2/c^2}} + mc^2\sqrt{1 - v^2/c^2} - mc^2.$$

We multiply the second term on the right by $\sqrt{1 - v^2/c^2}/\sqrt{1 - v^2/c^2} = 1$, and obtain

$$W = \frac{mv^2 + mc^2 - mv^2}{\sqrt{1 - v^2/c^2}} - mc^2$$

or

$$W = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2.$$

By the work-energy theorem, the work W done on the particle must equal its final kinetic energy K since the particle started from rest. Therefore

$$K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 \qquad \textbf{(36–10a)}$$

$$= \gamma mc^2 - mc^2 = (\gamma - 1)mc^2. \qquad \textbf{(36–10b)}$$

These Eqs. 36–10 replace the nonrelativistic form, $K = \frac{1}{2}mv^2$.

Using Relativistic Energy

Equations 36–10 require some interpretation. The first term increases with the speed v of the particle. The second term, mc^2, is constant; it is called the **rest energy** of the particle, and represents a form of energy that a particle has even when at rest. Note that if a particle is at rest $(v = 0)$ the first term in Eq. 36–10a becomes mc^2, so $K = 0$ as it should.

We can rearrange Eq. 36–10b to get

$$\gamma mc^2 = mc^2 + K.$$

We call γmc^2 the *total energy* E of the particle (assuming no potential energy), because it equals the rest energy plus the kinetic energy:

$$E = K + mc^2. \qquad \textbf{(36–11a)}$$

The total energy[†] can also be written, using Eqs. 36–10, as

$$E = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}}. \qquad \textbf{(36–11b)}$$

For a particle at rest in a given reference frame, K is zero in Eq. 36–11a, so the total energy is its rest energy:

$$E = mc^2. \qquad \textbf{(36–12)}$$

MASS RELATED TO ENERGY

Here we have Einstein's famous formula, $E = mc^2$. This formula mathematically relates the concepts of energy and mass. But if this idea is to have any physical meaning, then mass ought to be convertible to other forms of energy and vice versa. Einstein suggested that this might be possible, and indeed changes of mass to other forms of energy, and vice versa, have been experimentally confirmed countless times in nuclear and elementary particle physics. For example, an electron and a positron (= a positive electron, see Section 37–5) have often been observed to collide and disappear, producing pure electromagnetic radiation. The amount of electromagnetic energy produced is found to be exactly equal to that predicted by Einstein's formula, $E = mc^2$. The reverse process is also observed in the laboratory: electromagnetic radiation under certain conditions can be converted into material particles such as electrons (see Section 37–5 on pair production). On a larger scale, the energy produced in nuclear power plants is a result of the loss in mass of the uranium fuel as it undergoes the process called fission (Chapter 42). Even the radiant energy we receive from the Sun is an example of $E = mc^2$; the Sun's mass is continually decreasing as it radiates electromagnetic energy outward.

The relation $E = mc^2$ is now believed to apply to all processes, although the changes are often too small to measure. That is, when the energy of a system changes by an amount ΔE, the mass of the system changes by an amount Δm given by

$$\Delta E = (\Delta m)(c^2). \qquad \textbf{(36–13)}$$

In a nuclear reaction where an energy E is required or released, the masses of the reactants and the products will be different by $\Delta m = \Delta E/c^2$.

[†]This is for a "free particle," without forces and potential energy. Potential energy terms can be added.

EXAMPLE 36–10 **Pion's kinetic energy.** A π^0 meson $(m = 2.4 \times 10^{-28}\,\text{kg})$ travels at a speed $v = 0.84c = 2.5 \times 10^8\,\text{m/s}$. What is its kinetic energy? Compare to a classical calculation.

APPROACH We use Eq. 36–10 and compare to $\frac{1}{2}mv^2$.

SOLUTION We substitute values into Eq. 36–10a or b.

$$K = (\gamma - 1)mc^2$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{1}{\sqrt{1 - (0.84)^2}} = 1.84.$$

Then

$$K = (1.84 - 1)(2.4 \times 10^{-28}\,\text{kg})(3.0 \times 10^8\,\text{m/s})^2$$
$$= 1.8 \times 10^{-11}\,\text{J}.$$

Notice that the units of mc^2 are $\text{kg} \cdot \text{m}^2/\text{s}^2$, which is the joule.

NOTE Classically, $K = \frac{1}{2}mv^2 = \frac{1}{2}(2.4 \times 10^{-28}\,\text{kg})(2.5 \times 10^8\,\text{m/s})^2 = 7.5 \times 10^{-12}\,\text{J}$, less than half as much, but this is not a correct result. Note that $\frac{1}{2}\gamma mv^2$ also does not work.

EXERCISE G A proton is traveling in an accelerator with a speed of $1.0 \times 10^8\,\text{m/s}$. By what factor does the proton's kinetic energy increase if its speed is doubled? (*a*) 1.3, (*b*) 2.0, (*c*) 4.0, (*d*) 5.6.

EXAMPLE 36–11 **Energy from nuclear decay.** The energy required or released in nuclear reactions and decays comes from a change in mass between the initial and final particles. In one type of radioactive decay (Chapter 41), an atom of uranium $(m = 232.03716\,\text{u})$ decays to an atom of thorium $(m = 228.02874\,\text{u})$ plus an atom of helium $(m = 4.00260\,\text{u})$ where the masses given are in atomic mass units $(1\,\text{u} = 1.6605 \times 10^{-27}\,\text{kg})$. Calculate the energy released in this decay.

APPROACH The initial mass minus the total final mass gives the mass loss in atomic mass units (u); we convert that to kg, and multiply by c^2 to find the energy released, $\Delta E = \Delta mc^2$.

SOLUTION The initial mass is 232.03716 u, and after the decay the mass is 228.02874 u + 4.00260 u = 232.03134 u, so there is a loss of mass of 0.00582 u. This mass, which equals $(0.00582\,\text{u})(1.66 \times 10^{-27}\,\text{kg/u}) = 9.66 \times 10^{-30}\,\text{kg}$, is changed into energy. By $\Delta E = \Delta mc^2$, we have

$$\Delta E = (9.66 \times 10^{-30}\,\text{kg})(3.0 \times 10^8\,\text{m/s})^2$$
$$= 8.70 \times 10^{-13}\,\text{J}.$$

Since $1\,\text{MeV} = 1.60 \times 10^{-13}\,\text{J}$ (Section 23–8), the energy released is 5.4 MeV.

In the tiny world of atoms and nuclei, it is common to quote energies in eV (electron volts) or multiples such as MeV $(10^6\,\text{eV})$. Momentum (see Eq. 36–8) can be quoted in units of eV/c (or MeV/c). And mass can be quoted (from $E = mc^2$) in units of eV/c^2 (or MeV/c^2).

Note the use of c to keep the units correct. The masses of the electron and the proton are $0.511\,\text{MeV}/c^2$ and $938\,\text{MeV}/c^2$, respectively. For example, for the electron, $mc^2 = (9.11\times10^{-31}\,\text{kg})(2.998\times10^8\,\text{m/s})^2/(1.602\times10^{-13}\,\text{J/MeV}) = 0.511\,\text{MeV}$. See also the Table inside the front cover.

EXAMPLE 36–12 **A 7.0-TeV proton.** If a proton in the LHC accelerator in Switzerland is accelerated to 7.0 TeV, what is its speed?

APPROACH We solve the kinetic energy formula, Eq. 36–10a, for v.

SOLUTION The rest energy of a proton is $mc^2 = 938\,\text{MeV}$ or $9.38\times10^8\,\text{eV}$. Compared to the given kinetic energy of $7.0\times10^{12}\,\text{eV}$, the rest energy can be neglected, so we simplify Eq. 36–10a to

$$K \approx \frac{mc^2}{\sqrt{1 - v^2/c^2}}.$$

We solve this for v in the following steps:

$$\sqrt{1 - \frac{v^2}{c^2}} = \frac{mc^2}{K};$$

$$1 - \frac{v^2}{c^2} = \left(\frac{mc^2}{K}\right)^2;$$

$$\frac{v^2}{c^2} = 1 - \left(\frac{mc^2}{K}\right)^2 = 1 - \left(\frac{9.38\times10^8\,\text{eV}}{7.0\times10^{12}\,\text{eV}}\right)^2;$$

$$v = \sqrt{1 - (1.34\times10^{-4})^2}\,c = 0.99999998c.$$

So the proton is traveling at a speed very nearly equal to c.

At low speeds, $v \ll c$, the relativistic formula for kinetic energy reduces to the classical one, as we now show by using the binomial expansion (Appendix A–2): $(1 \pm x)^n = 1 \pm nx + n(n-1)x^2/2! + \cdots$. We keep only the first two terms, because when $x = v/c$ is very much less than 1, each term in the sequence is very much smaller than the previous one. With $n = -\frac{1}{2}$ we expand the square root in Eq. 36–10a

$$K = mc^2\left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right)$$

so that

$$K \approx mc^2\left(1 + \frac{1}{2}\frac{v^2}{c^2} + \cdots - 1\right) \approx \tfrac{1}{2}mv^2.$$

The dots represent very small terms in the expansion which we neglect since we assumed that $v \ll c$. Thus at low speeds, the relativistic form for kinetic energy reduces to the classical form, $K = \frac{1}{2}mv^2$. This makes relativity a viable theory in that it can predict accurate results at low speed as well as at high. Indeed, the other equations of special relativity also reduce to their classical equivalents at ordinary speeds: length contraction, time dilation, and modifications to momentum as well as kinetic energy, all become negligibly small for $v \ll c$ since $\sqrt{1 - v^2/c^2} \approx 1$.

A useful relation between the total energy E of a particle and its momentum p can also be derived. The momentum of a particle of mass m and speed v is given by Eq. 36–8

$$p = \gamma mv = \frac{mv}{\sqrt{1 - v^2/c^2}}.$$

The total energy is (Eqs. 36–11)

$$E = K + mc^2$$

or

$$E = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}}.$$

We square this equation (and we insert "$v^2 - v^2$" which is zero, but will help us):

$$E^2 = \frac{m^2c^2c^2}{1 - v^2/c^2} = \frac{m^2c^2(c^2 - v^2 + v^2)}{1 - v^2/c^2} = \frac{m^2c^2v^2}{1 - v^2/c^2} + \frac{m^2c^2(c^2 - v^2)}{1 - v^2/c^2}$$

$$= p^2c^2 + \frac{m^2c^4(1 - v^2/c^2)}{1 - v^2/c^2}$$

or

$$E^2 = p^2c^2 + m^2c^4. \qquad \textbf{(36–14)}$$

Thus, the total energy can be written in terms of the momentum p, or in terms of the kinetic energy (Eq. 36–11a), where we have assumed there is no potential energy.

Invariant Energy–Momentum

We can rewrite Eq. 36–14 as $E^2 - p^2c^2 = m^2c^4$. Since the mass m of a given particle is the same in any reference frame, we see that the quantity $E^2 - p^2c^2$ must also be the same in any reference frame. Thus, at any given moment the total energy E and momentum p of a particle will be different in different reference frames, but the quantity $E^2 - p^2c^2$ will have the same value in all inertial reference frames. We say that the quantity $E^2 - p^2c^2$ is **invariant** under a Lorentz transformation.

When Do We Use Relativistic Formulas?

From a practical point of view, we do not have much opportunity in our daily lives to use the mathematics of relativity. For example, the γ factor, $\gamma = 1/\sqrt{1 - v^2/c^2}$ has a value of 1.005 when $v = 0.10c$. Thus, for speeds even as high as $0.10c = 3.0 \times 10^7$ m/s, the factor $\sqrt{1 - v^2/c^2}$ in relativistic formulas gives a numerical correction of less than 1%. For speeds less than $0.10c$, or unless mass and energy are interchanged, we don't usually need the more complicated relativistic formulas, and can use the simpler classical formulas.

If you are given a particle's mass m and its kinetic energy K, you can do a quick calculation to determine if you need to use relativistic formulas or if classical ones are good enough. You simply compute the ratio K/mc^2 because (Eq. 36–10b)

$$\frac{K}{mc^2} = \gamma - 1 = \frac{1}{\sqrt{1 - v^2/c^2}} - 1.$$

If this ratio comes out to be less than, say, 0.01, then $\gamma \leq 1.01$ and relativistic equations will correct the classical ones by about 1%. If your expected precision is no better than 1%, classical formulas are good enough. But if your precision is 1 part in 1000 (0.1%) then you would want to use relativistic formulas. If your expected precision is only 10%, you need relativity if $(K/mc^2) \gtrsim 0.1$.

EXERCISE H For 1% accuracy, does an electron with $K = 100$ eV need to be treated relativistically? [*Hint*: The mass of an electron is 0.511 MeV.]

EXERCISE I For what value of v will $K/mc^2 = 0.10$?

36–12 Doppler Shift for Light

In Section 16–7 we discussed how the frequency and wavelength of sound are altered if the source of the sound and the observer are moving toward or away from each other. When a source is moving toward us, the frequency is higher than when the source is at rest. If the source moves away from us, the frequency is lower. We obtained four different equations for the Doppler shift (Eqs. 16–9a and b, Eqs. 16–10a and b), depending on the direction of the relative motion and whether the source or the observer is moving. The Doppler effect occurs also for light; but the shifted frequency or wavelength is given by slightly different equations, and there are only two of them, because for light—according to special relativity—we can make no distinction between motion of the source and motion of the observer. (Recall that sound must travel in a medium such as air, whereas light does not need a medium—there is no evidence for an ether.)

To derive the Doppler shift for light, let us consider a light source and an observer that move toward each other, and let their relative velocity be v as measured in the reference frame of either the source or the observer. Figure 36–15a shows a source at rest emitting light waves of frequency f_0 and wavelength $\lambda_0 = c/f_0$. Two wavecrests are shown, a distance λ_0 apart, the second crest just having been emitted. In Fig. 36–15b, the source is shown moving at speed v toward a stationary observer who will see the wavelength λ being somewhat less than λ_0. (This is much like Fig. 16–19 for sound.) Let Δt represent the time between crests as detected by the observer, whose reference frame is shown in Fig. 36–15b. From Fig. 36–15b we see that

$$\lambda = c\,\Delta t - v\,\Delta t,$$

where $c\,\Delta t$ is the distance crest 1 has moved in the time Δt after it was emitted, and $v\,\Delta t$ is the distance the source has moved in time Δt. So far our derivation has not differed from that for sound (Section 16–7). Now we invoke the theory of relativity. The time between emission of wavecrests has undergone time dilation:

$$\Delta t = \Delta t_0/\sqrt{1 - v^2/c^2}$$

where Δt_0 is the time between emissions of wavecrests in the reference frame where the source is at rest (the "proper" time). In the source's reference frame (Fig. 36–15a), the time between emissions of the two crests is

$$\Delta t_0 = \frac{1}{f_0} = \frac{\lambda_0}{c}$$

(Eqs. 5–2 and 31–14). Thus

$$\lambda = (c - v)\,\Delta t = (c - v)\,\frac{\Delta t_0}{\sqrt{1 - v^2/c^2}} = \frac{(c - v)}{\sqrt{c^2 - v^2}}\,\lambda_0$$

or

$$\lambda = \lambda_0\sqrt{\frac{c - v}{c + v}}. \qquad \left[\begin{array}{c}\text{source and observer}\\ \text{moving toward}\\ \text{each other}\end{array}\right] \qquad \textbf{(36–15a)}$$

The frequency f is (recall $\lambda_0 = c/f_0$)

$$f = \frac{c}{\lambda} = f_0\sqrt{\frac{c + v}{c - v}}. \qquad \left[\begin{array}{c}\text{source and observer}\\ \text{moving toward}\\ \text{each other}\end{array}\right] \qquad \textbf{(36–15b)}$$

Here f_0 is the frequency of the light as seen in the source's reference frame, and f is the frequency as measured by an observer moving toward the source or toward whom the source is moving. Equations 36–15 depend only on the relative velocity v.

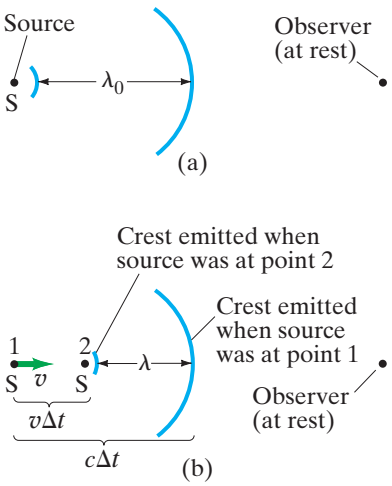


FIGURE 36–15 Doppler shift for light. (a) Source and observer at rest. (b) Source moving toward stationary observer.

For relative motion *away* from each other we set $v < 0$ in Eqs. 36–15, and obtain

$$\lambda = \lambda_0 \sqrt{\frac{c + v}{c - v}} \qquad \text{(36–16a)}$$

$$f = f_0 \sqrt{\frac{c - v}{c + v}}. \qquad \text{(36–16b)}$$

[source and observer moving away from each other]

From Eqs. 36–15 and 36–16 we see that light from a source moving *toward* us will have a higher frequency and shorter wavelength, whereas if a light source moves *away* from us, we will see a lower frequency and a longer wavelength. In the latter case, visible light will have its wavelength lengthened toward the red end of the visible spectrum (Fig. 32–26), an effect called a **redshift**. As we will see in the next Chapter, all atoms have their own distinctive signature in terms of the frequencies of the light they emit. More than a century ago it was observed that radiation from atoms in distant galaxies is redshifted. That is, the frequencies of light observed are lower than those emitted by stationary atoms on Earth, suggesting that the galaxies are receding from us. This is the origin of the idea that the universe is expanding, first proposed by the Belgian physics professor and Catholic priest George Lemaître (1894–1966) in 1927.

EXAMPLE 36–13 **Speeding through a red light.** A driver claims that he did not go through a red light because the light was Doppler shifted and appeared green. Calculate the speed of a driver in order for a red light to appear green.

APPROACH We apply the Doppler shift equation for red light ($\lambda_0 \approx 650\,\text{nm}$) and green light ($\lambda \approx 500\,\text{nm}$).

SOLUTION Equation 36–15a holds for the source and the object moving toward each other:

$$\lambda = \lambda_0 \sqrt{\frac{c - v}{c + v}}.$$

We square this equation:

$$\frac{c - v}{c + v} = \left(\frac{\lambda}{\lambda_0}\right)^2$$

where $(\lambda/\lambda_0)^2 = (500\,\text{nm}/650\,\text{nm})^2 = 0.59$. We solve for v:

$$v = c\left[\frac{1 - (\lambda/\lambda_0)^2}{1 + (\lambda/\lambda_0)^2}\right] = 0.26c.$$

With this defense, the driver might not be guilty of running a red light, but he would clearly be guilty of speeding.

36–13 The Impact of Special Relativity

A great many experiments have been performed to test the predictions of the special theory of relativity. Within experimental error, no contradictions have been found. Scientists have therefore accepted relativity as an accurate description of nature.

At speeds much less than the speed of light, the relativistic formulas reduce to the old classical ones, as we have discussed. We would, of course, hope—or rather, insist—that this be true since Newtonian mechanics works so well for objects moving with speeds $v \ll c$. This insistence that a more general theory (such as relativity) give the same results as a more restricted theory (such as classical mechanics which works for $v \ll c$) is called the **correspondence principle**. The two theories must correspond where their realms of validity overlap. Relativity thus does not contradict classical mechanics. Rather, it is a more general theory, of which classical mechanics is now considered to be a limiting case.

The importance of relativity is not simply that it gives more accurate results, especially at very high speeds. Much more than that, it has changed the way we view the world. The concepts of space and time are now seen to be relative, and intertwined with one another, whereas before they were considered absolute and separate. Even our concepts of matter and energy have changed: either can be converted to the other. The impact of relativity extends far beyond physics. It has influenced the other sciences, and even the world of art and literature; it has, indeed, entered the general culture.

From a practical point of view, we do not have much opportunity in our daily lives to use the mathematics of relativity. For example, the γ factor $1/\sqrt{1 - v^2/c^2}$, which appears in relativistic formulas, has a value of only 1.005 even for a speed as high as $0.10c = 3.0 \times 10^7$ m/s, giving a correction of less than 1%. For speeds less than $0.10c$, or unless mass and energy are interchanged, we don't usually need to use the more complicated relativistic formulas, and can use the simpler classical formulas.

The special theory of relativity we have studied in this Chapter deals with inertial (nonaccelerating) reference frames. In Chapter 44 we will briefly discuss the more complicated "general theory of relativity" which deals with noninertial reference frames.

Summary

An **inertial reference frame** is one in which Newton's law of inertia holds. Inertial reference frames move at constant velocity relative to one another. Accelerating reference frames are **noninertial**.

The **special theory of relativity** is based on two principles: the **relativity principle**, which states that the laws of physics are the same in all inertial reference frames, and the principle of the **constancy of the speed of light**, which states that the speed of light in empty space has the same value in all inertial reference frames.

One consequence of relativity theory is that two events that are simultaneous in one reference frame may not be simultaneous in another. Other effects are **time dilation**: moving clocks are measured to run slow; and **length contraction**: the length of a moving object is measured to be shorter (in its direction of motion) than when it is at rest. Quantitatively,

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - v^2/c^2}} = \gamma \, \Delta t_0 \qquad \textbf{(36–1)}$$

$$\ell = \ell_0 \sqrt{1 - v^2/c^2} = \frac{\ell_0}{\gamma} \qquad \textbf{(36–3)}$$

where ℓ and Δt are the length and time interval of objects (or events) observed as they move by at the speed v. The quantities ℓ_0 and Δt_0 are the **proper length** and **proper time**—that is, the same quantities as measured in the rest frame of the objects or events. The quantity γ is shorthand for

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}. \qquad \textbf{(36–2)}$$

The theory of relativity has changed our notions of space and time, and of momentum, energy, and mass. Space and time are seen to be intimately connected, with time being the fourth dimension in addition to the three dimensions of space.

The **Lorentz transformations** relate the positions and times (x, t) of events in one inertial reference frame to their positions and times (x', t') in a second inertial reference frame which moves at constant speed v in the positive x direction:

$$\begin{aligned} x &= \gamma(x' + vt') \\ y &= y' \\ z &= z' \\ t &= \gamma\left(t' + \frac{vx'}{c^2}\right) \end{aligned} \qquad \textbf{(36–6)}$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$.

Velocity addition also must be done in a special way. All these relativistic effects are significant only at high speeds, close to the speed of light, which itself is the ultimate speed in the universe.

The **momentum** of an object is given by

$$p = \gamma m v = \frac{mv}{\sqrt{1 - v^2/c^2}}. \qquad \textbf{(36–8)}$$

Mass and energy are interconvertible. The equation

$$E = mc^2 \qquad \textbf{(36–12)}$$

tells how much energy E is needed to create a mass m, or vice versa. Said another way, $E = mc^2$ is the amount of energy an object has because of its mass m. The law of conservation of energy must include mass as a form of energy.

The **kinetic energy** K of an object moving at speed v is given by

$$K = (\gamma - 1)mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 \qquad \textbf{(36–10)}$$

where m is the mass of the object. The total energy E, if there is no potential energy, is

$$\begin{aligned} E &= K + mc^2 \\ &= \gamma mc^2. \end{aligned} \qquad \textbf{(36–11)}$$

The momentum p of an object is related to its total energy E (assuming no potential energy) by

$$E^2 = p^2c^2 + m^2c^4. \qquad \textbf{(36–14)}$$

Because m^2c^4 is constant, independent of reference frame, $E^2 - p^2c^2$ is **invariant** under a Lorentz transformation.

Questions

1. You are in a windowless car in an exceptionally smooth train moving at constant velocity. Is there any physical experiment you can do in the train car to determine whether you are moving? Explain.
2. You might have had the experience of being in a car at a red light when, out of the corner of your eye, you see the car beside you creep forward. Instinctively you stomp on the brake pedal, thinking that you are rolling backward. What does this say about absolute and relative motion?
3. A worker stands on top of a railroad car moving at constant velocity and throws a heavy ball straight up (from his point of view). Ignoring air resistance, explain whether the ball will land back in his hand or behind him.
4. Does the Earth really go around the Sun? Or is it also valid to say that the Sun goes around the Earth? Discuss in view of the relativity principle (that there is no best reference frame). Explain. See Section 6–5.
5. If you were on a spaceship traveling at $0.6c$ away from a star, at what speed would the starlight pass you?
6. The time dilation effect is sometimes expressed as "moving clocks run slowly." Actually, this effect has nothing to do with motion affecting the functioning of clocks. What then does it deal with?
7. Does time dilation mean that time actually passes more slowly in moving reference frames or that it only *seems* to pass more slowly?
8. A young-looking woman astronaut has just arrived home from a long trip. She rushes up to an old gray-haired man and in the ensuing conversation refers to him as her son. How might this be possible?
9. If you were traveling away from Earth at speed $0.6c$, would you notice a change in your heartbeat? Would your mass, height, or waistline change? What would observers on Earth using telescopes say about you?
10. Do time dilation and length contraction occur at ordinary speeds, say 90 km/h?
11. What if the speed of light was infinite. What would happen to the relativistic predictions of length contraction and time dilation?
12. Explain how the length contraction and time dilation formulas might be used to indicate that c is the limiting speed in the universe.
13. Discuss how our everyday lives would be different if the speed of light was only 25 m/s.
14. An electron is limited to travel at speeds less than c. Does this put an upper limit on the momentum of an electron? If so, what is this upper limit? If not, explain.
15. Can a particle of nonzero mass attain the speed of light? Explain.
16. Does the equation $E = mc^2$ conflict with the conservation of energy principle? Explain.
17. You are traveling in a spaceship at a speed of $0.70c$ away from Earth. You send a laser beam toward the Earth traveling at velocity c relative to you. What do observers on the Earth measure for the speed of the laser beam?
18. If mass is a form of energy, does this mean that a spring has more mass when compressed than when relaxed? Explain.
19. It is not correct to say that "matter can neither be created nor destroyed." What must we say instead?
20. Is our intuitive notion that velocities simply add, as in Section 3–9, completely wrong?
21. The drawing at the start of this Chapter (repeated here) shows the street and my girlfriend as I saw them, with the speed of light pretended to be $c = 100$ km/h. (a) Describe what I and the streetscape look like to my girlfriend as she speeds by at 70 km/h. (b) Also, make a drawing showing the shops from her reference frame. Include both of us.



FIGURE 36–16 Question 21. Chapter-Opening Figure on page 1055, part b, repeated here.

MisConceptual Questions

1. Which statement is true concerning the speed of light relative to its source?
   (*a*) Light moves more slowly if emitted by a source moving in the direction of emission.
   (*b*) Light moves faster if emitted by a source moving in the direction of emission.
   (*c*) Light moves more slowly if emitted by a source moving in the direction opposite the emission.
   (*d*) Light moves faster if emitted by a source moving in the direction opposite the emission.
   (*e*) Light moves at the same speed independent of the motion of the source.
2. Which of the following are noninertial frames of reference?
   (*a*) A reference frame that is not moving relative to another inertial frame.
   (*b*) A reference frame that is moving at constant velocity relative to another inertial frame.
   (*c*) A reference frame that is moving at constant acceleration relative to another inertial frame.
   (*d*) A reference frame that is rotating at constant angular velocity relative to another inertial frame.

3. The fictional rocket ship *Endurance* is measured to be 50 m long by the ship's captain inside the rocket. When the rocket moves past a space dock at 0.5*c*, space-dock personnel measure the rocket ship to be 43.3 m long. What is its proper length?
(*a*) 50 m.
(*b*) 43.3 m.
(*c*) 93.3 m.
(*d*) 13.3 m.

4. As rocket ship *Endurance* (MisConceptual Question 3) passes by the space dock, the ship's captain flashes a flashlight at 1.00-s intervals as measured by space-dock personnel. How far apart in time are the flashes relative to the captain?
(*a*) More than 1.00 s.
(*b*) 1.00 s.
(*c*) Less than 1.00 s.
(*d*) We need to know the distance between the ship and the space dock.

5. The rocket ship of MisConceptual Question 3 travels to a star many light-years away, then turns around and returns at the same speed. When it returns to the space dock, who would have aged less: the space-dock personnel or the ship's crew?
(*a*) The space-dock personnel.
(*b*) The ship's crew.
(*c*) Both the same amount, because both sets of people were moving relative to each other.
(*d*) We need to know how far away the star is.

6. Relativistic formulas for time dilation, length contraction, and momentum are valid
(*a*) only for speeds less than 0.10*c*.
(*b*) only for speeds greater than 0.10*c*.
(*c*) only for speeds very close to *c*.
(*d*) for all speeds.

7. Reference frames A and B are in relative motion. A mass *m* is at rest in B. Observers in both reference frames calculate the particle's momentum and energy, getting different values. This is an example of
(*a*) a situation where momentum is not conserved.
(*b*) a situation where energy is not conserved.
(*c*) Both (*a*) and (*b*) are true.
(*d*) Neither (*a*) nor (*b*) is true.

8. In the theory of special relativity, which of the following are not relative?
(*a*) The speed of light in a vacuum.
(*b*) The time interval between two events.
(*c*) The length of an object in its direction of motion.
(*d*) The length of an object perpendicular to its direction of motion.
(*e*) The laws of physics.
(*f*) The simultaneity of events.

9. Spaceships A and B are traveling directly toward each other at a speed 0.5*c* relative to the Earth, and each has a headlight aimed toward the other ship. What value do technicians on ship B get by measuring the speed of the light emitted by ship A's headlight?
(*a*) 0.5*c*.
(*b*) 0.80*c*.
(*c*) 1.0*c*.
(*d*) 1.5*c*.

10. An Earth observer notes that clocks on a passing spacecraft run slowly. The person on the spacecraft
(*a*) agrees her clocks move slower than those on Earth.
(*b*) feels normal, and her heartbeat and eating habits are normal.
(*c*) observes that Earth clocks are moving slowly.
(*d*) The real time is in between the times measured by the two observers.
(*e*) Both (*a*) and (*b*).
(*f*) Both (*b*) and (*c*).

11. Which of the following will two observers in inertial reference frames that move relative to each other always agree on?
(*a*) The time an event occurred.
(*b*) The distance between two events.
(*c*) The time interval between the occurence of two events.
(*d*) The speed of light.
(*e*) The validity of the laws of physics.
(*f*) The simultaneity of two events.

12. Two observers in different inertial reference frames moving relative to each other at nearly the speed of light see the same two events but, using precise equipment, record different time intervals between the two events. Which of the following is true of their measurements?
(*a*) One observer is incorrect, but it is impossible to tell which one.
(*b*) One observer is incorrect, and it is possible to tell which one.
(*c*) Both observers are incorrect.
(*d*) Both observers can be correct.

13. You are in a spaceship with no windows, radios, or other means to check outside. How could you determine whether your spaceship is at rest or moving at constant velocity?
(*a*) By determining the apparent velocity of light in the spaceship.
(*b*) By checking your precision watch. If it's running slow, then the ship is moving.
(*c*) By measuring the lengths of objects in the spaceship. If they are shortened, then the ship is moving.
(*d*) Give up, because you can't tell.

14. The period of a pendulum swinging in a spaceship is 2 s while the spaceship is parked on Earth. What is the period to an observer on Earth when the spaceship moves at 0.6*c* with respect to the Earth?
(*a*) Less than 2 s.
(*b*) More than 2 s.
(*c*) 2 s.

15. Two spaceships, each moving at a speed 0.75*c* relative to the Earth, are headed directly toward each other. What do occupants of one ship measure the speed of the other ship to be?
(*a*) 0.96*c*.
(*b*) 1.0*c*.
(*c*) 1.5*c*.
(*d*) 1.75*c*.
(*e*) 0.75*c*.

Problems

36–5 and 36–6 Time Dilation, Length Contraction

1. (I) A certain type of elementary particle travels at a speed of 2.70×10^8 m/s. At this speed, the average lifetime is measured to be 5.16×10^{-6} s. What is the particle's lifetime at rest?
2. (I) A spaceship passes you at a speed of $0.750c$. You measure its length to be 112 m. How long would it be when at rest?
3. (II) You travel to a star 115 light-years from Earth at a speed of 2.90×10^8 m/s. What do you measure this distance to be?
4. (II) What is the speed of a pion if its average lifetime is measured to be 4.80×10^{-8} s? At rest, its average lifetime is 2.60×10^{-8} s.
5. (II) In an Earth reference frame, a star is 49 light-years away. How fast would you have to travel so that to you the distance would be only 32 light-years?
6. (II) At what speed v will the length of a 1.00-m stick look 10.0% shorter (90.0 cm)?
7. (II) At what speed do the relativistic formulas for (*a*) length and (*b*) time intervals differ from classical values by 1.00%? (This is a reasonable way to estimate when to use relativistic calculations rather than classical.)
8. (II) You decide to travel to a star 68 light-years from Earth at a speed that tells you the distance is only 25 light-years. How many years would it take you to make the trip?
9. (II) A friend speeds by you in her spacecraft at a speed of $0.670c$. It is measured in your frame to be 4.80 m long and 1.35 m high. (*a*) What will be its length and height at rest? (*b*) How many seconds elapsed on your friend's watch when 20.0 s passed on yours? (*c*) How fast did you appear to be traveling according to your friend? (*d*) How many seconds elapsed on your watch when she saw 20.0 s pass on hers?
10. (II) A star is 23.5 light-years from Earth. How long would it take a spacecraft traveling $0.950c$ to reach that star as measured by observers: (*a*) on Earth, (*b*) on the spacecraft? (*c*) What is the distance traveled according to observers on the spacecraft? (*d*) What will the spacecraft occupants compute their speed to be from the results of (*b*) and (*c*)?
11. (II) A fictional news report stated that starship *Enterprise* had just returned from a 5-year voyage while traveling at $0.80c$. (*a*) If the report meant 5.0 years of *Earth time*, how much time elapsed on the ship? (*b*) If the report meant 5.0 years of *ship time*, how much time passed on Earth?
12. (II) A box at rest has the shape of a cube 2.4 m on a side. This box is loaded onto the flat floor of a spaceship and the spaceship then flies past us with a horizontal speed of $0.80c$. What is the volume of the box as we observe it?
13. (II) Escape velocity from the Earth is 11.2 km/s. What would be the percent decrease in length of a 73.6-m-long spacecraft traveling at that speed as seen from Earth?
14. (II) An unstable particle produced in an accelerator experiment travels at constant velocity, covering 1.00 m in 3.80 ns in the lab frame before changing ("decaying") into other particles. In the rest frame of the particle, determine (*a*) how long it lived before decaying, (*b*) how far it moved before decaying.
15. (II) How fast must a pion be moving on average to travel 28 m before it decays? The average lifetime, at rest, is 2.6×10^{-8} s.
16. (III) When at rest, a spaceship has the form of an isosceles triangle whose two equal sides have length ℓ and whose base has length 1.5ℓ. At what speed is this spaceship traveling past you (parallel to its base) if it appears to be an equilateral triangle?
17. (III) When it is stationary, the half-life of a certain subatomic particle is T_0. That is, if N of these particles are present at a certain time, then a time T_0 later only $N/2$ particles will be present, assuming the particles are at rest. A beam carrying N such particles per second is created at position $x = 0$ in a high-energy physics laboratory. This beam travels along the x axis at speed v in the laboratory reference frame and it is found that only $N/2$ particles per second travel in the beam at $x = 2cT_0$, where c is the speed of light. Find the speed v of the particles within the beam.

36–8 Lorentz Transformations

18. (I) An observer on Earth sees an alien vessel approach at a speed of $0.70c$. The fictional starship *Enterprise* comes to the rescue (Fig. 36–17), overtaking the aliens while moving directly toward Earth at a speed of $0.90c$ relative to Earth. What is the relative speed of one vessel as seen by the other?

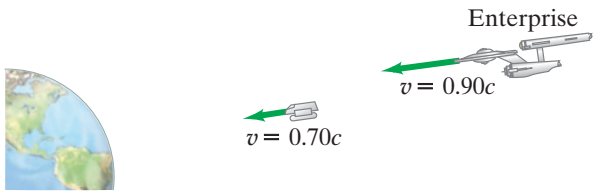


FIGURE 36–17 Problem 18.

19. (I) Suppose in Fig. 36–12 that the origins of S and S′ overlap at $t = t' = 0$ and that S′ moves at speed $v = 30$ m/s with respect to S. In S′, a person is resting at a point whose coordinates are $x' = 25$ m, $y' = 20$ m, and $z' = 0$. Calculate this person's coordinates in S (x, y, z) at (*a*) $t = 3.5$ s, (*b*) $t = 10.0$ s. Use the Galilean transformation.
20. (I) Repeat Problem 19 using the Lorentz transformation and a relative speed $v = 1.60 \times 10^8$ m/s, but choose the time t to be (*a*) 3.5 μs and (*b*) 10.0 μs.
21. (II) In Problem 20, suppose that the person moves with a velocity whose components are $u'_x = u'_y = 1.30 \times 10^8$ m/s. What will be her velocity with respect to S? (Give magnitude and direction.)
22. (II) Two spaceships leave Earth in opposite directions, each with a speed of $0.50c$ with respect to Earth. (*a*) What is the velocity of spaceship 1 relative to spaceship 2? (*b*) What is the velocity of spaceship 2 relative to spaceship 1?
23. (II) A spaceship leaves Earth traveling at $0.66c$. A second spaceship leaves the first at a speed of $0.87c$ with respect to the first. Calculate the speed of the second ship with respect to Earth if it is fired (*a*) in the same direction the first spaceship is already moving, (*b*) directly backward toward Earth.

24. (II) Reference frame S′ moves at speed $v = 0.88c$ in the $+x$ direction with respect to reference frame S. The origins of S and S′ overlap at $t = t' = 0$. An object is stationary in S′ at position $x' = 100\,\text{m}$. What is the position of the object in S when the clock in S reads $1.00\,\mu\text{s}$ according to the (*a*) Galilean and (*b*) Lorentz transformation equations?

25. (II) A spaceship traveling at $0.76c$ away from Earth fires a module with a speed of $0.85c$ at right angles to its own direction of travel (as seen by the spaceship). What is the speed of the module, and its direction of travel (relative to the spaceship's direction), seen by an observer on Earth?

26. (II) Your spaceship, traveling at $0.85c$, needs to launch a probe out the forward hatch so that its speed relative to the planet that you are approaching is $0.95c$. With what speed must it leave your ship?

27. (II) A stick of length ℓ_0, at rest in reference frame S, makes an angle θ with the x axis. In reference frame S′, which moves to the right with velocity $\vec{\mathbf{v}} = v\hat{\mathbf{i}}$ with respect to S, determine (*a*) the length ℓ of the stick, and (*b*) the angle θ' it makes with the x' axis.

28. (III) An observer in reference frame S notes that two events are separated in space by 180 m and in time by $0.80\,\mu\text{s}$. How fast must reference frame S′ be moving relative to S in order for an observer in S′ to detect the two events as occurring at the same location in space?

29. (III) In the old West, a marshal riding on a train traveling 35.0 m/s sees a duel between two men standing on the Earth 55.0 m apart parallel to the train. The marshal's instruments indicate that in his reference frame the two men fired simultaneously. (*a*) Which of the two men, the first one the train passes (A) or the second one (B) should be arrested for firing the first shot? That is, in the gunfighter's frame of reference, who fired first? (*b*) How much earlier did he fire? (*c*) Who was struck first?

30. (III) If a particle moves in the xy plane of system S (Fig. 36–12) with speed u in a direction that makes an angle θ with the x axis, show that it makes an angle θ' in S′ given by $\tan\theta' = (\sin\theta)\sqrt{1 - v^2/c^2}/(\cos\theta - v/u)$.

31. (III) Two lightbulbs, A and B, are placed at rest on the x axis at positions $x_\text{A} = 0$ and $x_\text{B} = +\ell$. In this reference frame, the bulbs are turned on simultaneously. Use the Lorentz transformations to find an expression for the time interval between when the bulbs are turned on as measured by an observer moving at velocity v in the $+x$ direction. According to this observer, which bulb is turned on first?

32. (III) A farm boy studying physics believes that he can fit a 13.0-m-long pole into a 10.0-m-long barn if he runs fast enough, carrying the pole. Can he do it? Explain in detail. How does this fit with the idea that when he is running the barn looks even shorter than 10.0 m?

36–9 Relativistic Momentum

33. (I) What is the momentum of a proton traveling at $v = 0.72c$?

34. (II) (*a*) A particle travels at $v = 0.10c$. By what percentage will a calculation of its momentum be wrong if you use the classical formula? (*b*) Repeat for $v = 0.50c$.

35. (II) A particle of mass m travels at a speed $v = 0.32c$. At what speed will its momentum be doubled?

36. (II) An unstable particle is at rest and suddenly decays into two fragments. No external forces act on the particle or its fragments. One of the fragments has a speed of $0.53c$ and a mass of $6.68 \times 10^{-27}\,\text{kg}$, while the other has a mass of $1.67 \times 10^{-27}\,\text{kg}$. What is the speed of the less massive fragment?

37. (II) What is the percent change in momentum of a proton that accelerates (*a*) from $0.45c$ to $0.80c$, (*b*) from $0.80c$ to $0.98c$?

36–11 $E = mc^2$; Mass and Energy

38. (I) Calculate the rest energy of an electron $(m = 9.11 \times 10^{-31}\,\text{kg})$ in joules and in MeV $(1\,\text{MeV} = 1.60 \times 10^{-13}\,\text{J})$.

39. (I) When a uranium nucleus at rest breaks apart in the process known as fission in a nuclear reactor, the resulting fragments have a total kinetic energy of about 200 MeV. How much mass (kg) was lost in the process?

40. (I) The total annual energy consumption in the United States is about $1 \times 10^{20}\,\text{J}$. How much mass would have to be converted to energy to fuel this need?

41. (II) Suppose there was a process by which two photons, each with momentum $0.75\,\text{MeV}/c$, could collide and make a single particle. (Photons are particles of light with mass $m = 0$; see next Chapter.) What is the maximum mass (kg) that the particle could possess?

42. (II) (*a*) How much work is required to accelerate a proton from rest up to a speed of $0.995c$? (*b*) What would be the momentum of this proton?

43. (II) A proton is traveling in an accelerator with a speed of $1.0 \times 10^8\,\text{m/s}$. By what factor does the proton's kinetic energy increase if its speed is doubled?

44. (II) What is the speed of an electron after being accelerated from rest by 28,000 V?

45. (II) To accelerate a particle of mass m from rest to speed $0.90c$ requires work W_1. To accelerate the particle from speed $0.90c$ to $0.99c$, requires work W_2. Determine the ratio W_2/W_1.

46. (II) How much energy can be obtained from conversion of 1.0 gram of mass? How much mass could this energy raise to a height of 1.0 km above the Earth's surface?

47. (II) What is the momentum of a 950-MeV proton (that is, its kinetic energy is 950 MeV)?

48. (II) What is the speed of a particle when its kinetic energy equals its rest energy? Does the mass of the particle affect the result?

49. (II) What is the speed of an electron whose kinetic energy is 1.45 MeV?

50. (II) Two identical particles of mass m approach each other at equal and opposite speeds, v. The collision is completely inelastic and results in a single particle at rest. What is the mass of the new particle? How much energy was lost in the collision? How much kinetic energy was lost in this collision?

51. (II) The kinetic energy of a particle is 45 MeV. If its momentum is $121\,\text{MeV}/c$, what is the particle's mass?

52. (II) Calculate the speed of a proton $(m = 1.67 \times 10^{-27}\,\text{kg})$ whose kinetic energy is exactly half (*a*) its total energy, (*b*) its rest energy.

53. (II) Calculate the kinetic energy and momentum of a proton $(m = 1.67 \times 10^{-27}\,\text{kg})$ traveling $7.85 \times 10^7\,\text{m/s}$. By what percentages would your calculations have been in error if you had used classical formulas?

54. (II) Suppose a spacecraft of mass 17,000 kg was accelerated to $0.22c$. (*a*) How much kinetic energy would it have? (*b*) If you used the classical formula for kinetic energy, by what percentage would you be in error?

55. (II) The americium nucleus, $^{241}_{95}\text{Am}$, decays to a neptunium nucleus, $^{237}_{93}\text{Np}$, by emitting an alpha particle of mass 4.00260 u and kinetic energy 5.5 MeV. Estimate the mass of the neptunium nucleus, ignoring its recoil, given that the americium mass is 241.05682 u.

56. (II) Make a graph of the kinetic energy versus momentum for (*a*) a particle of nonzero mass, and (*b*) a particle with zero mass.

57. (II) A negative muon traveling at 38% the speed of light collides head on with a positive muon traveling at 55% the speed of light. The two muons (each of mass $105.7\,\text{MeV}/c^2$) annihilate, and produce how much electromagnetic energy?

58. (II) Show that the kinetic energy K of a particle of mass m is related to its momentum p by the equation

$$p = \sqrt{K^2 + 2Kmc^2}/c.$$

59. (II) The fictional starship *Enterprise* obtains its power by combining matter and antimatter, achieving complete conversion of mass into energy. If the mass of the *Enterprise* is approximately $4 \times 10^9\,\text{kg}$, how much mass must be converted into kinetic energy to accelerate it from rest to one-tenth the speed of light?

60. (III) (*a*) In reference frame S, a particle has momentum $\vec{\mathbf{p}} = p_x\,\hat{\mathbf{i}}$ along the positive x axis. Show that in frame S′, which moves with speed v as in Fig. 36–12, the momentum has components

$$p'_x = \frac{p_x - vE/c^2}{\sqrt{1 - v^2/c^2}}$$
$$p'_y = p_y$$
$$p'_z = p_z$$
$$E' = \frac{E - p_x v}{\sqrt{1 - v^2/c^2}}.$$

(These transformation equations hold, actually, for any direction of $\vec{\mathbf{p}}$, as long as the motion of S′ is along the x axis.) (*b*) Show that $p_x, p_y, p_z, E/c$ transform according to the Lorentz transformation in the same way as x, y, z, ct.

36–12 Doppler Shift for Light

61. (II) A certain galaxy has a Doppler shift given by $f_0 - f = 0.1015 f_0$. Estimate how fast it is moving away from us.

62. (II) A spaceship moving toward Earth at $0.65c$ transmits radio signals at 95.0 MHz. At what frequency should Earth receivers be tuned?

63. (II) Starting from Eq. 36–16a, show that the Doppler shift in wavelength is, if $v \ll c$,

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c}.$$

64. (III) A radar "speed gun" emits microwaves of frequency $f_0 = 36.0\,\text{GHz}$. When the gun is pointed at an object moving toward it at speed v, the object senses the microwaves at the Doppler-shifted frequency f. The moving object reflects these microwaves at this same frequency f. The stationary radar apparatus detects these reflected waves at a (double) Doppler-shifted frequency f'. The gun combines its emitted wave at f_0 and its detected wave at f'. These waves interfere, creating a beat pattern whose beat frequency is $f_{\text{beat}} = f' - f_0$. (*a*) Show that

$$v \approx \frac{c f_{\text{beat}}}{2 f_0},$$

if $f_{\text{beat}} \ll f_0$. If $f_{\text{beat}} = 6670\,\text{Hz}$, what is v (km/h)? (*b*) If the object's speed is different by Δv, show that the difference in beat frequency Δf_{beat} is given by

$$\Delta f_{\text{beat}} = \frac{2 f_0\,\Delta v}{c}.$$

If the accuracy of the speed gun is to be 1 km/h, to what accuracy must the beat frequency be measured?

65. (III) A certain atom emits light of frequency f_0 when at rest. A monatomic gas composed of these atoms is at temperature T. Some of the gas atoms move toward, and others away from, an observer due to their random thermal motion. Using the rms speed of thermal motion, (*a*) show that the fractional difference between the Doppler-shifted frequencies for atoms moving directly toward the observer and directly away from the observer is $\Delta f/f_0 \approx 2\sqrt{3kT/mc^2}$. Assume $mc^2 \gg 3kT$. (*b*) Evaluate $\Delta f/f_0$ for a gas of hydrogen atoms at 650 K. [This "Doppler-broadening" effect is commonly used to measure gas temperature, such as in astronomy.]

General Problems

66. An atomic clock is taken to the North Pole, while another stays at the Equator. How far will they be out of synchronization after 1.5 years has elapsed? [*Hint*: Use the binomial expansion, Appendix A–2.]

67. An airplane travels 1500 km/h around the Earth in a circle of radius essentially equal to that of the Earth, returning to the same place. Using special relativity, estimate the difference in time to make the trip as seen by Earth and airplane observers. [*Hint*: Use the binomial expansion, Appendix A–2.]

68. The nearest star to Earth is Proxima Centauri, 4.3 light-years away. (*a*) At what constant velocity must a spacecraft travel from Earth if it is to reach the star in 5.5 years, as measured by travelers on the spacecraft? (*b*) How long does the trip take according to Earth observers?

69. An electron $(m = 9.11 \times 10^{-31}\,\text{kg})$ is accelerated from rest to speed v by a conservative force. In this process, its potential energy decreases by $7.20 \times 10^{-14}\,\text{J}$. Determine the electron's speed, v.

70. A healthy astronaut's heart rate is 60 beats/min. Flight doctors on Earth can monitor an astronaut's vital signs remotely while in flight. How fast would an astronaut be flying away from Earth if the doctor measured her heart rate as 52 beats/min?

71. A spacecraft (reference frame S′) moves past Earth (reference frame S) at velocity $\vec{\mathbf{v}}$, which points along the x and x' axes. The spacecraft emits a light beam (speed c) along its y' axis as shown in Fig. 36–18. (*a*) What angle θ does this light beam make with the x axis in the Earth's reference frame? (*b*) Use velocity transformations to show that the light moves with speed c also in the Earth's reference frame. (*c*) Compare these relativistic results to what you would have obtained classically (Galilean transformations).

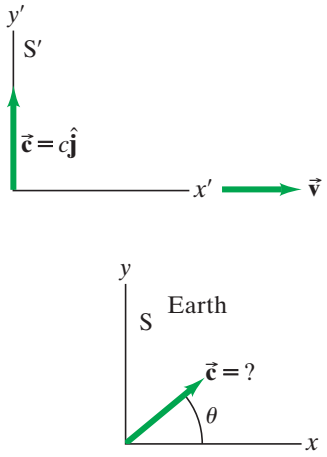


FIGURE 36–18 Problem 71.

72. Rocket A passes Earth at a speed of $0.65c$. At the same time, rocket B passes Earth moving $0.90c$ relative to Earth in the same direction. How fast is B moving relative to A when it passes A?

73. (*a*) What is the speed v of an electron whose kinetic energy is 14,000 times its rest energy? You can state the answer as the difference $c - v$. Such speeds have been reached in the Stanford Linear Accelerator, SLAC. (*b*) If the electrons travel in the lab through a tube 3.0 km long (as at SLAC), how long is this tube in the electrons' reference frame? [*Hint*: Use the binomial expansion.]

74. As a rough rule, anything traveling faster than about $0.1c$ is called *relativistic*—that is, special relativity is a significant effect. Determine the speed of an electron in a hydrogen atom (radius 0.53×10^{-10} m) and state whether or not it is relativistic. (Treat the electron as though it were in a circular orbit around the proton.)

75. What minimum amount of electromagnetic energy is needed to produce an electron and a positron together? A positron is a particle with the same mass as an electron, but has the opposite charge. (Note that electric charge is conserved in this process. See Section 37–5.)

76. How many grams of matter would have to be totally destroyed to run a 15-W lightbulb for 1.0 year?

77. If E is the total energy of a particle with zero potential energy, show that $dE/dp = v$, where p and v are the momentum and velocity of the particle, respectively.

78. A free neutron can decay into a proton, an electron, and a neutrino. Assume the neutrino's mass is zero; the other masses can be found in the Table inside the front cover. Determine the total kinetic energy shared among the three particles when a neutron decays at rest.

79. The Sun radiates energy at a rate of about 4×10^{26} W. (*a*) At what rate is the Sun's mass decreasing? (*b*) How long does it take for the Sun to lose a mass equal to that of Earth? (*c*) Estimate how long the Sun could last if it radiated constantly at this rate.

80. An unknown particle is measured to have a negative charge and a speed of 2.24×10^8 m/s. Its momentum is determined to be 3.07×10^{-22} kg · m/s. Identify the particle by finding its mass. See inside the front cover.

81. How much energy would be required to break a helium nucleus into its constituents, two protons and two neutrons? The rest masses of a proton (including an electron), a neutron, and neutral helium are, respectively, 1.00783 u, 1.00867 u, and 4.00260 u. (This energy difference is called the *total binding energy* of the ${}^4_2\text{He}$ nucleus.)

82. Show analytically that a particle with momentum p and energy E has a speed given by

$$v = \frac{pc^2}{E} = \frac{pc}{\sqrt{m^2c^2 + p^2}}.$$

83. Two protons, each having a speed of $0.945c$ in the laboratory, are moving toward each other. Determine (*a*) the momentum of each proton in the laboratory, (*b*) the total momentum of the two protons in the laboratory, and (*c*) the momentum of one proton as seen by the other proton.

84. Suppose a 14,500-kg spaceship left Earth at a speed of $0.80c$. What is the spaceship's kinetic energy? Compare with the total U.S. annual energy consumption (about 10^{20} J).

85. A pi meson of mass m_π decays at rest into a muon (mass m_μ) and a neutrino of negligible or zero mass. Show that the kinetic energy of the muon is $K_\mu = (m_\pi - m_\mu)^2c^2/(2m_\pi)$.

86. When two moles of hydrogen molecules (H_2) and one mole of oxygen molecules (O_2) react to form two moles of water (H_2O), the energy released is 484 kJ. How much does the mass decrease in this reaction? What % of the total original mass is this?

87. A quasar emits familiar hydrogen lines whose wavelengths are 8.5% longer than what we measure in the laboratory. (*a*) Using the Doppler formula for light, estimate the speed of this quasar. (*b*) What result would you obtain if you used the "classical" Doppler shift discussed in Chapter 16?

88. A spaceship and its occupants have a total mass of 160,000 kg. The occupants would like to travel to a star that is 32 light-years away at a speed of $0.70c$. To accelerate, the engine of the spaceship changes mass directly to energy. (*a*) Estimate how much mass will be converted to energy to accelerate the spaceship to this speed. (*b*) Assuming the acceleration is rapid, so the speed for the entire trip can be taken to be $0.70c$, determine how long the trip will take according to the astronauts on board.

89. A 42,000-kg spaceship is to travel to the vicinity of a star 6.6 light-years from Earth. Passengers on the ship want the (one-way) trip to take no more than 1.0 year. How much work must be done on the spaceship to bring it to the speed necessary for this trip?

90. For a 1.0-kg mass, make a plot of the kinetic energy as a function of speed for speeds from 0 to $0.9c$, using both the classical formula $\left(K = \frac{1}{2}mv^2\right)$ and the correct relativistic formula $\left(K = (\gamma - 1)mc^2\right)$.

91. Examine the experiment of Fig. 36–6 from O_1's reference frame. In this case, O_1 will be at rest and will see the lightning bolt at B_1 and B_2 before the lightning bolt at A_1 and A_2. Will O_1 recognize that O_2, who is moving with speed v to the left, will see the two events as simultaneous? Explain in detail, drawing diagrams equivalent to Fig. 36–6. [*Hint*: Include length contraction.]

92. Using Example 36–2 as a guide, show that for objects that move slowly in comparison to c, the length contraction formula is roughly $\ell \approx \ell_0\left(1 - \frac{1}{2}v^2/c^2\right)$. Use this approximation to find the "length shortening" $\Delta\ell = \ell_0 - \ell$ of the train in Example 36–6 if the train travels at 100 km/h (rather than $0.92c$).

93. In Example 36–5, the spaceship is moving at $0.90c$ in the horizontal direction relative to an observer on the Earth. If instead the spaceship moved at $0.90c$ directed at 35° above the horizontal, what would be the painting's dimensions as seen by the observer on Earth?

94. Protons from outer space crash into the Earth's atmosphere at a high rate and create particles that eventually decay into other particles called *muons*. These "cosmic rays" travel through the atmosphere. Every second, dozens of muons pass through your body. If a muon is created 30 km above the Earth's surface, what minimum speed and kinetic energy must the muon have in order to hit Earth's surface? A muon's mean lifetime (at rest) is 2.20 μs and its mass is 105.7 MeV/c^2.

95. An astronaut on a spaceship traveling at $0.75c$ relative to Earth measures his ship to be 35 m long. On the ship, he eats his lunch in 25 min. (*a*) What length is the spaceship according to observers on Earth? (*b*) How long does the astronaut's lunch take to eat according to observers on Earth?

96. At what rate is the Sun losing mass? Energy reaches Earth 1.5×10^{11} m from the Sun at the rate of 1300 W/m^2.

97. A 0.30-kg meter stick moving parallel to its length passes you at high speed. You measure its length to be 48.0 cm. What is its kinetic energy?

98. Astronomers measure the distance to a particular star to be 6.0 light-years (1 ly = distance light travels in 1 year). A spaceship travels from Earth to the vicinity of this star at steady speed, arriving in 3.25 years as measured by clocks on the spaceship. (*a*) How long does the trip take as measured by clocks in Earth's reference frame? (*b*) What distance does the spaceship travel as measured in its own reference frame?

99. An electron is accelerated so that its kinetic energy is greater than its rest energy mc^2 by a factor of (*a*) 5.00, (*b*) 999. What is the speed of the electron in each case?

100. A 2.42-kg mass oscillates on the end of a spring whose spring stiffness constant is $k = 84.2$ N/m. If this system is in a spaceship moving past Earth at $0.900c$, what is its period of oscillation according to (*a*) observers on the ship, and (*b*) observers on Earth?

101. Show that the space-time interval, $(c\,\Delta t)^2 - (\Delta x)^2$, is invariant, meaning that all observers in all inertial reference frames calculate the same number for this quantity for any pair of events.

102. A slab of glass with index of refraction n moves to the right with speed v. A flash of light is emitted at point A (Fig. 36–19) and passes through the glass arriving at point B a distance ℓ away. The glass has thickness d in the reference frame where it is at rest, and the speed of light in the glass is c/n. How long does it take the light to go from point A to point B according to an observer at rest with respect to points A and B? Check your answer for the cases $v = c$, $v = 0$, and $n = 1$.

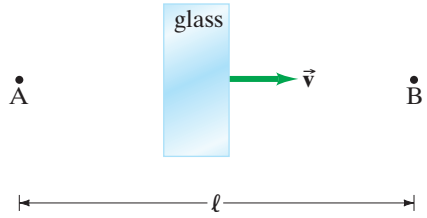


FIGURE 36–19 Problem 102.

***103.** What magnetic field B is needed to keep 998-GeV protons revolving in a circle of radius 1.0 km? Use the relativistic mass. The proton's "rest mass" is 0.938 GeV/c^2. (1 GeV = 10^9 eV.) [*Hint*: In relativity, $m_{\text{rel}}v^2/r = qvB$ is still valid in a magnetic field, where $m_{\text{rel}} = \gamma m$.]

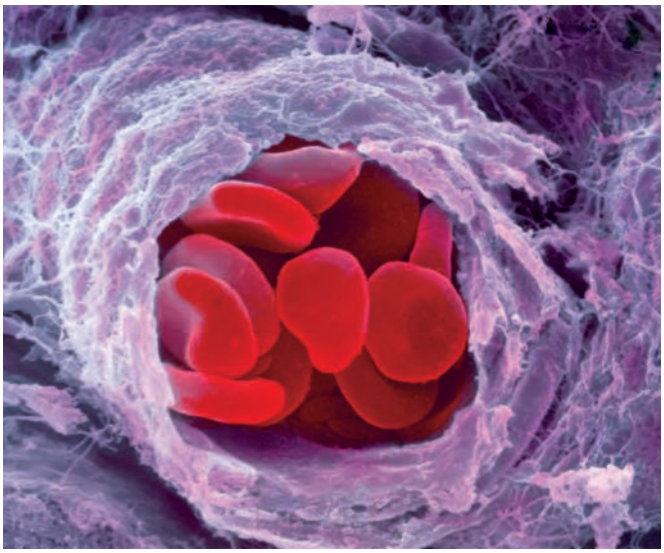
ANSWERS TO EXERCISES

A: Yes.
B: (*c*).
C: (*a*) No; (*b*) yes.
D: 80 m.
E: $0.030c$, same as classical, to an accuracy of better than 0.1%.
F: (*d*).
G: (*d*).
H: No.
I: $0.41c$.

Electron microscopes (EM) produce images using electrons which have wave properties just as light does. Because the wavelength of electrons can be much smaller than that of visible light, much greater resolution and magnification can be obtained. A scanning electron microscope (SEM) can produce images with a three-dimensional quality.

All EM images are monochromatic (black and white). Artistic coloring has been added here, as is common. On the left is an SEM image of a blood clot forming (yellow-color web) due to a wound. White blood cells are colored green here for visibility. On the right, red blood cells in a small artery. A red blood cell travels about 15 km a day inside our bodies and lives roughly four months before damage or rupture. Humans contain 4 to 6 liters of blood, and 2 to 3×10^{13} red blood cells.





CHAPTER 37

Early Quantum Theory and Models of the Atom

CHAPTER-OPENING QUESTION—Guess now!

It has been found experimentally that

(a) light behaves as a wave.
(b) light behaves as a particle.
(c) electrons behave as particles.
(d) electrons behave as waves.
(e) all of the above are true.
(f) only (a) and (b) are true.
(g) only (a) and (c) are true.
(h) none of the above are true.

The second aspect of the revolution that shook the world of physics in the early part of the twentieth century was the quantum theory (the other was Einstein's theory of relativity). Unlike the special theory of relativity, the revolution of quantum theory required almost three decades to unfold, and many scientists contributed to its development. It began in 1900 with Planck's quantum hypothesis, used to explain the radiation emitted by hot objects; it culminated in the mid-1920s with the theory of quantum mechanics of Schrödinger and Heisenberg which has been so effective in explaining the structure of matter.

CONTENTS

37 – 1 Blackbody Radiation; Planck's Quantum Hypothesis
37 – 2 Photon Theory of Light and the Photoelectric Effect
37 – 3 Energy, Mass, and Momentum of a Photon
37 – 4 Compton Effect
37 – 5 Photon Interactions; Pair Production
37 – 6 Wave–Particle Duality; the Principle of Complementarity
37 – 7 Wave Nature of Matter
37 – 8 Electron Microscopes
37 – 9 Early Models of the Atom
37–10 Atomic Spectra: Key to the Structure of the Atom
37–11 The Bohr Model
37–12 de Broglie's Hypothesis Applied to Atoms

37–1 Blackbody Radiation; Planck's Quantum Hypothesis

One of the observations that was unexplained at the end of the nineteenth century was the spectrum of light emitted by hot objects. We saw in Section 19–10 that all objects emit radiation whose total intensity is proportional to the fourth power of the Kelvin (absolute) temperature (T^4). At normal temperatures $(\approx 300\,\text{K})$, we are not aware of this electromagnetic radiation because of its low intensity. At higher temperatures, there is sufficient infrared radiation that we can feel heat if we are close to the object. At still higher temperatures (on the order of 1000 K), objects actually glow, such as a red-hot electric stove burner or the heating element in a toaster. At temperatures above 2000 K, objects glow with a yellow or whitish color, such as white-hot iron and the filament of a lightbulb. The light emitted contains a continuous range of wavelengths or frequencies, and the *spectrum* is a plot of intensity vs. wavelength or frequency. As the temperature increases, the electromagnetic radiation emitted by objects not only increases in total intensity but has its peak intensity at higher and higher frequencies.

The spectrum of light emitted by a hot dense object is shown in Fig. 37–1 for an idealized **blackbody**. A blackbody is a body that would absorb all the radiation falling on it (and so, when cool, would appear black under reflection when illuminated by other sources). The radiation such an idealized blackbody would emit when hot and luminous, called **blackbody radiation** (though not necessarily black in color), approximates that from many real objects. The 6000-K curve in Fig. 37–1, corresponding to the temperature of the surface of the Sun, peaks in the visible part of the spectrum. At lower temperatures, the total intensity drops considerably and the peak occurs at longer wavelengths (or lower frequencies). (This is why objects glow with a red color at around 1000 K.) It is found experimentally that the wavelength at the peak of the spectrum, λ_P, is related to the Kelvin temperature T by

$$\lambda_P T = 2.90 \times 10^{-3}\,\text{m}\cdot\text{K}. \qquad \textbf{(37–1)}$$

This is known as **Wien's law**.

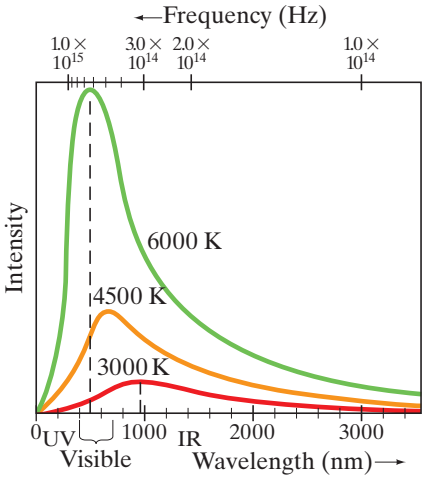


FIGURE 37–1 Measured spectra of wavelengths and frequencies emitted by a blackbody at three different temperatures.

EXAMPLE 37–1 **The Sun's surface temperature.** Estimate the temperature of the surface of our Sun, given that the Sun emits light whose peak intensity occurs in the visible spectrum at around 500 nm.

APPROACH We assume the Sun acts as a blackbody, and use $\lambda_P = 500\,\text{nm}$ in Wien's law (Eq. 37–1).

SOLUTION Wien's law gives

$$T = \frac{2.90 \times 10^{-3}\,\text{m}\cdot\text{K}}{\lambda_P} = \frac{2.90 \times 10^{-3}\,\text{m}\cdot\text{K}}{500 \times 10^{-9}\,\text{m}} \approx 6000\,\text{K}.$$

EXAMPLE 37–2 **Star color.** Suppose a star has a surface temperature of 32,500 K. What color would this star appear?

APPROACH We assume the star emits radiation as a blackbody, and solve for λ_P in Wien's law, Eq. 37–1.

SOLUTION From Wien's law we have

$$\lambda_P = \frac{2.90 \times 10^{-3}\,\text{m}\cdot\text{K}}{T} = \frac{2.90 \times 10^{-3}\,\text{m}\cdot\text{K}}{3.25 \times 10^{4}\,\text{K}} = 89.2\,\text{nm}.$$

The peak is in the UV range of the spectrum, and will be way to the left in Fig. 37–1. In the visible region, the curve will be descending, so the shortest visible wavelengths will be strongest. Hence the star will appear bluish (or blue-white).

NOTE This example helps us to understand why stars have different colors (reddish for the coolest stars, orangish, yellow, white, bluish for "hotter" stars.)

EXERCISE A What is the color of an object at 4000 K?

A major problem facing scientists in the 1890s was to explain blackbody radiation. Maxwell's electromagnetic theory had predicted that oscillating electric charges produce electromagnetic waves, and the radiation emitted by a hot object could be due to the oscillations of electric charges in the molecules of the material. Although this would explain where the radiation came from, it did not correctly predict the observed spectrum of emitted light. Two important theoretical curves based on classical ideas were those proposed by W. Wien (in 1896) and by Lord Rayleigh (in 1900). The latter was modified later by J. Jeans and since then has been known as the Rayleigh–Jeans theory. As experimental data came in, it became clear that neither Wien's nor the Rayleigh–Jeans formulations were in accord with experiment (see Fig. 37–2).

In the year 1900 Max Planck (1858–1947) proposed an empirical formula that nicely fit the data (now often called *Planck's radiation formula*):

$$I(\lambda, T) = \frac{2\pi hc^2\lambda^{-5}}{e^{hc/\lambda kT} - 1}.$$

$I(\lambda, T)$ is the radiation intensity as a function of wavelength λ at the temperature T; k is Boltzman's constant, c is the speed of light, and h is a new constant, now called **Planck's constant**. The value of h was estimated by Planck by fitting his formula for the blackbody radiation curve to experiment (and the fit was excellent). The value accepted today is

$$h = 6.626 \times 10^{-34}\,\text{J}\cdot\text{s}.$$

To provide a theoretical basis for his formula, Planck made a new and radical assumption: that the energy of the oscillations of atoms within molecules cannot have just any value; instead each has energy which is a multiple of a minimum value related to the frequency of oscillation by

$$E = hf.$$

Planck's assumption suggests that the energy of any molecular vibration could be only a whole number multiple of hf:

$$E = nhf, \qquad n = 1, 2, 3, \cdots, \tag{37–2}$$

where n is called a **quantum number** ("quantum" means "discrete amount" as opposed to "continuous"). This idea is often called **Planck's quantum hypothesis**, although little attention was brought to this point at the time. In fact, it appears that Planck considered it more as a mathematical device to get the "right answer" rather than as an important discovery. Planck himself continued to seek a classical explanation for the introduction of h. The recognition that this was an important and radical innovation did not come until later, after about 1905 when others, particularly Einstein, entered the field.

The quantum hypothesis, Eq. 37–2, states that the energy of an oscillator can be $E = hf$, or $2hf$, or $3hf$, and so on, but there cannot be vibrations with energies between these values. That is, energy would not be a continuous quantity as had been believed for centuries; rather it is **quantized**—it exists only in discrete amounts. The smallest amount of energy possible (hf) is called the **quantum of energy**. Recall from Chapter 14 that the energy of an oscillation is proportional to the amplitude squared. Another way of expressing the quantum hypothesis is that not just any amplitude of vibration is possible. The possible values for the amplitude are related to the frequency f.

A simple analogy may help. Compare a ramp, on which a box can be placed at any height, to a flight of stairs on which the box can have only certain discrete amounts of potential energy, as shown in Fig. 37–3. The stairs are "quantized", but the ramp is continuous.

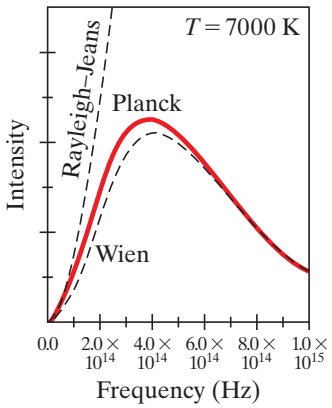
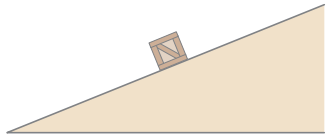


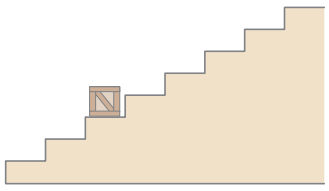
FIGURE 37–2 Comparison of the Wien and the Rayleigh–Jeans theories to that of Planck, which closely follows experiment. The dashed lines show lack of agreement of older theories.

CAUTION
Quantum vs. continuous

FIGURE 37–3 Ramp versus stair analogy. (a) On a ramp, a box can have continuous values of potential energy. (b) But on stairs, the box can have only discrete (quantized) values of energy.



(a)



(b)

37–2 Photon Theory of Light and the Photoelectric Effect

In 1905, the same year that he introduced the special theory of relativity, Einstein made a bold extension of the quantum idea by proposing a new theory of light. Planck's work had suggested that the vibrational energy of molecules in a radiating object is quantized with energy $E = nhf$, where n is an integer and f is the frequency of molecular vibration. Einstein argued that when light is emitted by a molecular oscillator, the molecule's vibrational energy of nhf must decrease by an amount hf (or by $2hf$, etc.) to another integer times hf, such as $(n-1)hf$. Then to conserve energy, the light ought to be emitted in packets, or *quanta*, each with an energy

Photon energy

$$E = hf, \tag{37–3}$$

where f is here the frequency of the emitted light.

Again h is Planck's constant. Because all light ultimately comes from a radiating source, this idea suggests that *light is transmitted as tiny particles*, or **photons** as they are now called, as well as via the waves predicted by Maxwell's electromagnetic theory. The photon theory of light was a radical departure from classical ideas. Einstein proposed a test of the quantum theory of light: quantitative measurements on the photoelectric effect.

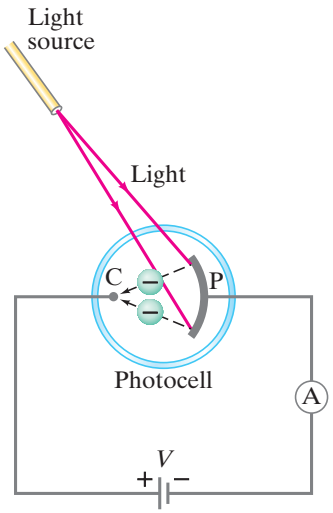


FIGURE 37–4 The photoelectric effect.

When light shines on a metal surface, electrons are found to be emitted from the surface. This effect is called the **photoelectric effect** and it occurs in many materials, but is most easily observed with metals. It can be observed using the apparatus shown in Fig. 37–4. A metal plate P and a smaller electrode C are placed inside an evacuated glass tube, called a **photocell**. The two electrodes are connected to an ammeter and a source of emf, as shown. When the photocell is in the dark, the ammeter reads zero. But when light of sufficiently high frequency illuminates the plate P, the ammeter indicates a current flowing in the circuit. We explain completion of the circuit by imagining that electrons, ejected from the plate by the impinging radiation, flow across the tube from the plate to the "collector" C as indicated in Fig. 37–4.

That electrons should be emitted when light shines on a metal is consistent with the electromagnetic (EM) wave theory of light: the electric field of an EM wave could exert a force on electrons in the metal and eject some of them. Einstein pointed out, however, that the wave theory and the photon theory of light give very different predictions on the details of the photoelectric effect. For example, one thing that can be measured with the apparatus of Fig. 37–4 is the maximum kinetic energy $(K_{\max})$ of the emitted electrons. This can be done by using a variable voltage source V in Fig. 37–4, and reversing the terminals so that electrode C is negative and P is positive. The electrons emitted from P will be repelled by the negative electrode, but if this reverse voltage is small enough, the fastest electrons will still reach C and there will be a current in the circuit. If the reversed voltage is increased, a point is reached where the current reaches zero—no electrons have sufficient kinetic energy to reach C. This is called the *stopping potential*, or **stopping voltage**, V_0, and from its measurement, $K_{\max}$ can be determined using conservation of energy (loss of kinetic energy = gain in potential energy):

$$K_{\max} = eV_0.$$

Now let us examine the details of the photoelectric effect from the point of view of the wave theory versus Einstein's particle theory.

First the wave theory, assuming monochromatic light. The two important properties of a light wave are its intensity and its frequency (or wavelength). When these two quantities are varied, the wave theory makes the following predictions:

Wave theory predictions

1. If the light intensity is increased, the number of electrons ejected and their maximum kinetic energy should be increased because the higher intensity means a greater electric field amplitude, and the greater electric field should eject electrons with higher speed.
2. The frequency of the light should not affect the kinetic energy of the ejected electrons. Only the intensity of the incoming light should affect K_{max}.

The photon theory makes completely different predictions. First we note that in a monochromatic beam, all photons have the same energy $(= hf)$. Increasing the intensity of the light beam means increasing the number of photons in the beam, but does not affect the energy of each photon as long as the frequency is not changed. According to Einstein's theory, an electron is ejected from the metal by a collision with a single photon. In the process, all the photon energy is transferred to the electron and the photon ceases to exist. Since electrons are held in the metal by attractive forces, some minimum energy W_0 is required just to get an electron out through the surface. W_0 is called the **work function**, and is a few electron volts $(1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J})$ for most metals. If the frequency f of the incoming light is so low that hf is less than W_0, then the photons will not have enough energy to eject any electrons at all. If $hf > W_0$, then electrons will be ejected and energy will be conserved in the process. That is, the input energy (of the photon), hf, will equal the outgoing kinetic energy K of the electron plus the energy required to get it out of the metal, W:

$$hf = K + W. \qquad \textbf{(37–4a)}$$

The least tightly held electrons will be emitted with the most kinetic energy (K_{max}), in which case W in this equation is the work function W_0, and K becomes K_{max}:

$$hf = K_{max} + W_0. \qquad \text{[least bound electrons]} \quad \textbf{(37–4b)}$$

Many electrons will require more energy than the bare minimum (W_0) to get out of the metal, and thus the kinetic energy of such electrons will be less than the maximum.

From these considerations, the photon theory makes the following predictions:

Photon theory predictions

1. An increase in intensity of the light beam means more photons are incident, so more electrons will be ejected; but since the energy of each photon is not changed, the maximum kinetic energy of electrons is not changed by an increase in intensity.
2. If the frequency of the light is increased, the maximum kinetic energy of the electrons increases linearly, according to Eq. 37–4b. That is,

$$K_{max} = hf - W_0.$$

This relationship is plotted in Fig. 37–5.

3. If the frequency f is less than the "cutoff" frequency f_0, where $hf_0 = W_0$, no electrons will be ejected, no matter how great the intensity of the light.

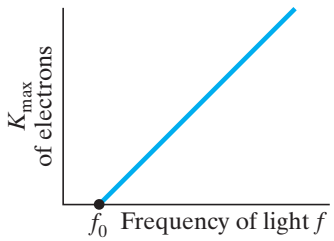


FIGURE 37–5 Photoelectric effect: the maximum kinetic energy of ejected electrons increases linearly with the frequency of incident light. No electrons are emitted if $f < f_0$.

These predictions of the photon theory are very different from the predictions of the wave theory. In 1913–1914, careful experiments were carried out by R. A. Millikan. The results were fully in agreement with Einstein's photon theory.

One other aspect of the photoelectric effect also confirmed the photon theory. If extremely low light intensity is used, the wave theory predicts a time delay before electron emission so that an electron can absorb enough energy to exceed the work function. The photon theory predicts no such delay—it only takes one photon (if its frequency is high enough) to eject an electron—and experiments showed no delay. This too confirmed Einstein's photon theory.

EXAMPLE 37–3 **Photon energy.** Calculate the energy of a photon of blue light, $\lambda = 450\,\text{nm}$ in air (or vacuum).

APPROACH The photon has energy $E = hf$ (Eq. 37–3) where $f = c/\lambda$ (Eq. 31–14).

SOLUTION Since $f = c/\lambda$, we have

$$E = hf = \frac{hc}{\lambda} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.0 \times 10^{8}\,\text{m/s})}{(4.5 \times 10^{-7}\,\text{m})} = 4.4 \times 10^{-19}\,\text{J},$$

or $(4.4 \times 10^{-19}\,\text{J})/(1.60 \times 10^{-19}\,\text{J/eV}) = 2.8\,\text{eV}$. (See definition of eV in Section 23–8, $1\,\text{eV} = 1.60 \times 10^{-19}\,\text{J}$.)

EXAMPLE 37–4 **ESTIMATE** **Photons from a lightbulb.** Estimate how many visible light photons a 100-W incandescent lightbulb emits per second. Assume the bulb has a typical efficiency of about 3% (that is, 97% of the energy goes to heat).

APPROACH Let's assume an average wavelength in the middle of the visible spectrum, $\lambda \approx 500\,\text{nm}$. The energy of each photon is $E = hf = hc/\lambda$. Only 3% of the 100-W power is emitted as visible light, or $3\,\text{W} = 3\,\text{J/s}$. The number of photons emitted per second equals the light output of 3 J/s divided by the energy of each photon.

SOLUTION The energy emitted in one second $(= 3\,\text{J})$ is $E = Nhf$ where N is the number of photons emitted per second and $f = c/\lambda$. Hence

$$N = \frac{E}{hf} = \frac{E\lambda}{hc} = \frac{(3\,\text{J})(500 \times 10^{-9}\,\text{m})}{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.0 \times 10^{8}\,\text{m/s})} \approx 8 \times 10^{18}$$

per second, or almost 10^{19} photons emitted per second, an enormous number.

EXERCISE B A beam contains infrared light of a single wavelength, 1000 nm, and monochromatic UV at 100 nm, both of the same intensity. Are there more 100-nm photons or more 1000-nm photons?

EXAMPLE 37–5 **Photoelectron speed and energy.** What is the kinetic energy and speed of an electron ejected from a sodium surface whose work function is $W_0 = 2.28\,\text{eV}$ when illuminated by light of wavelength (*a*) 410 nm, (*b*) 550 nm?

APPROACH We first find the energy of the photons $(E = hf = hc/\lambda)$. If the energy is greater than W_0, then electrons will be ejected with varying amounts of kinetic energy, with a maximum of $K_{\text{max}} = hf - W_0$.

SOLUTION (*a*) For $\lambda = 410\,\text{nm}$,

$$hf = \frac{hc}{\lambda} = 4.85 \times 10^{-19}\,\text{J} \quad \text{or} \quad 3.03\,\text{eV}.$$

The maximum kinetic energy an electron can have is given by Eq. 37–4b, $K_{\text{max}} = 3.03\,\text{eV} - 2.28\,\text{eV} = 0.75\,\text{eV}$, or $(0.75\,\text{eV})(1.60 \times 10^{-19}\,\text{J/eV}) = 1.2 \times 10^{-19}\,\text{J}$. Since $K = \frac{1}{2}mv^2$ where $m = 9.11 \times 10^{-31}\,\text{kg}$,

$$v_{\text{max}} = \sqrt{\frac{2K}{m}} = 5.1 \times 10^{5}\,\text{m/s}.$$

Most ejected electrons will have less kinetic energy and less speed than these maximum values.

(*b*) For $\lambda = 550\,\text{nm}$, $hf = hc/\lambda = 3.61 \times 10^{-19}\,\text{J} = 2.26\,\text{eV}$. Since this photon energy is less than the work function, no electrons are ejected.

NOTE In (*a*) we used the nonrelativistic equation for kinetic energy. If v had turned out to be more than about $0.1c$, our calculation would have been inaccurate by more than a percent or so, and we would probably prefer to redo it using the relativistic form (Eq. 36–10).

EXERCISE C Determine the lowest frequency and the longest wavelength needed to emit electrons from sodium.

By converting units, we can show that the energy of a photon in electron volts, when given the wavelength λ in nm, is

$$E\,(\text{eV}) = \frac{1.240 \times 10^3\,\text{eV}\cdot\text{nm}}{\lambda\,(\text{nm})}. \qquad \text{[photon energy in eV]}$$

Applications of the Photoelectric Effect

The photoelectric effect, besides playing an important historical role in confirming the photon theory of light, also has many practical applications. Burglar alarms and automatic doors often make use of the photocell circuit of Fig. 37–4. When a person interrupts the beam of light, the sudden drop in current in the circuit activates a switch—often a solenoid—which operates a bell or opens the door. UV or IR light is sometimes used in burglar alarms because of its invisibility. Many smoke detectors use the photoelectric effect to detect tiny amounts of smoke that interrupt the flow of light and so alter the electric current. Photographic light meters use this circuit as well. Photocells are used in many other devices, such as absorption spectrophotometers, to measure light intensity. One type of film sound track is a variably shaded narrow section at the side of the film, Fig. 37–6. Light passing through the film is thus "modulated," and the output electrical signal of the photocell detector follows the frequencies on the sound track. For many applications today, the vacuum-tube photocell of Fig. 37–4 has been replaced by a semiconductor device known as a **photodiode** (Section 40–9). In these semiconductors, the absorption of a photon liberates a bound electron so it can move freely, which changes the conductivity of the material and the current through a photodiode is altered.

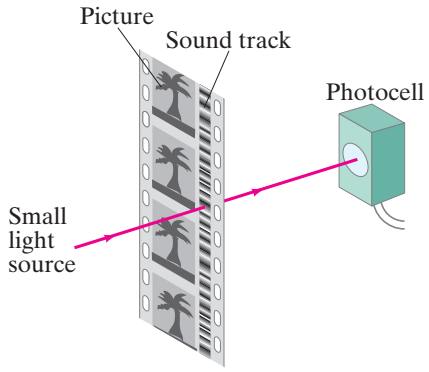


FIGURE 37–6 Optical sound track on movie film. In the projector, light from a small source (different from that for the picture) passes through the sound track on the moving film. The light and dark areas on the sound track vary the intensity of the transmitted light which reaches the photocell, whose output current is then a replica of the original sound. This output is amplified and sent to the loudspeakers. High-quality projectors can show movies containing several parallel sound tracks to go to different speakers around the theater.

37–3 Energy, Mass, and Momentum of a Photon

We have seen that the total energy of a single photon is given by $E = hf$ (Eq. 37–3). But because a photon always travels at the speed of light, it is truly a relativistic particle. Thus we must use relativistic formulas for dealing with its energy and momentum. The momentum of any particle of mass m is given by $p = mv/\sqrt{1 - v^2/c^2}$. Since $v = c$ for a photon, the denominator is zero. To avoid having an infinite momentum, we conclude that the photon's mass must be zero: $m = 0$. This makes sense too because a photon can never be at rest (it always moves at the speed of light). A photon's kinetic energy is its total energy (rest energy $= 0$):

$$K = E = hf. \qquad \text{[photon]}$$

The momentum of a photon can be obtained from the relativistic formula (Eq. 36–14) $E^2 = p^2c^2 + m^2c^4$ where we set $m = 0$, so $E^2 = p^2c^2$ or

$$p = \frac{E}{c}. \qquad \text{[photon]}$$

CAUTION
Momentum of photon is not *mv*

Since $E = hf$ for a photon, its momentum is related to its wavelength by

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda}. \qquad \textbf{(37–5)}$$

EXAMPLE 37–6 **Photosynthesis.** In *photosynthesis*, pigments such as chlorophyll in plants capture the energy of sunlight to change CO_2 to useful carbohydrate. About nine photons are needed to transform one molecule of CO_2 to carbohydrate and O_2. Assuming light of wavelength $\lambda = 670\,\text{nm}$ (chlorophyll absorbs most strongly in the range 650 nm to 700 nm), how efficient is the photosynthetic process? The reverse chemical reaction releases an energy of 4.9 eV/molecule of CO_2, so 4.9 eV is needed to transform one molecule of CO_2 to carbohydrate.

PHYSICS APPLIED
Photosynthesis

APPROACH The efficiency is the minimum energy required (4.9 eV) divided by the actual energy absorbed, nine times the energy (hf) of one photon.

SOLUTION The energy of nine photons, each of energy $hf = hc/\lambda$, is $(9)(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.0 \times 10^8\,\text{m/s})/(6.7 \times 10^{-7}\,\text{m}) = 2.7 \times 10^{-18}\,\text{J}$ or 17 eV. Thus the process is about $(4.9\,\text{eV}/17\,\text{eV}) = 29\%$ efficient.

EXAMPLE 37–7 **ESTIMATE** **Photon momentum and force.** Suppose the 10^{19} photons emitted per second from the 100-W lightbulb in Example 37–4 were all focused onto a piece of black paper and absorbed. (*a*) Calculate the momentum of one photon and (*b*) estimate the force all these photons could exert on the paper.

APPROACH Each photon's momentum is obtained from Eq. 37–5, $p = h/\lambda$. Next, each absorbed photon's momentum changes from $p = h/\lambda$ to zero. We use Newton's second law, $F = \Delta p/\Delta t$, to get the force. Let $\lambda = 500\,\text{nm}$.

SOLUTION (*a*) Each photon has a momentum

$$p = \frac{h}{\lambda} = \frac{6.63 \times 10^{-34}\,\text{J}\cdot\text{s}}{500 \times 10^{-9}\,\text{m}} = 1.3 \times 10^{-27}\,\text{kg}\cdot\text{m/s}.$$

(*b*) Using Newton's second law for $N = 10^{19}$ photons (Example 37–4) whose momentum changes from h/λ to 0, we obtain

$$F = \frac{\Delta p}{\Delta t} = \frac{Nh/\lambda - 0}{1\,\text{s}} = N\frac{h}{\lambda} \approx (10^{19}\,\text{s}^{-1})(10^{-27}\,\text{kg}\cdot\text{m/s}) \approx 10^{-8}\,\text{N}.$$

NOTE This is a tiny force, but we can see that a much stronger light source could exert a measurable force, and near the Sun or a star the force due to photons in electromagnetic radiation could be considerable. See Section 31–9.

37–4 Compton Effect

Besides the photoelectric effect, a number of other experiments were carried out in the early twentieth century which also supported the photon theory. One of these was the **Compton effect** (1923) named after its discoverer, A. H. Compton (1892–1962). Compton aimed short-wavelength light (actually X-rays) at various materials, and detected light scattered at various angles. He found that the scattered light had a slightly longer wavelength than did the incident light, and therefore a slightly lower frequency indicating a loss of energy. He explained this result on the basis of the photon theory as incident photons colliding with electrons of the material, Fig. 37–7.

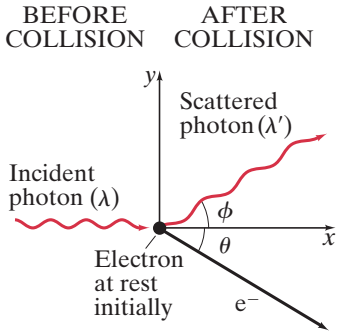


FIGURE 37–7 The Compton effect. A single photon of wavelength λ strikes an electron in some material, knocking it out of its atom. The scattered photon has less energy (some energy is given to the electron) and hence has a longer wavelength λ' (shown exaggerated). Experiments found scattered X-rays of just the wavelengths predicted by conservation of energy and momentum using the photon model.

Using Eq. 37–5 for momentum of a photon, Compton applied the laws of conservation of momentum and energy to the collision of Fig. 37–7 and derived the following equation for the wavelength of the scattered photons:

$$\lambda' = \lambda + \frac{h}{m_e c}(1 - \cos\phi), \qquad \textbf{(37–6a)}$$

where m_e is the mass of the electron. For $\phi = 0$, the wavelength is unchanged (there is no collision for this case of the photon passing straight through).

At any other angle, λ' is longer than λ. The difference in wavelength,

$$\Delta\lambda = \lambda' - \lambda = \frac{h}{m_e c}(1 - \cos\phi), \qquad \textbf{(37–6b)}$$

is called the **Compton shift**. The quantity $h/m_e c$, which has the dimensions of length, is called the **Compton wavelength** λ_C of a free electron,

$$\lambda_C = \frac{h}{m_e c} = 2.43 \times 10^{-3}\,\text{nm} = 2.43\,\text{pm}. \qquad \text{[electron]}$$

Equations 37–6 predict that λ' depends on the angle ϕ at which the photons are detected. Compton's measurements of 1923 were consistent with this formula, confirming the value of λ_C and the dependence of λ' on ϕ. See Fig. 37–8.

In contrast, the wave theory of light predicts no wavelength shift: an incoming electromagnetic wave of frequency f should set electrons into oscillation at the same frequency f, and such oscillating electrons should reemit EM waves of this same frequency f (Chapter 31), which would not change with the angle ϕ. Hence the Compton effect adds to the firm experimental foundation for the photon theory of light.

EXERCISE D When a photon scatters off an electron by the Compton effect, which of the following increases: its energy, frequency, or wavelength?

EXAMPLE 37–8 **X-ray scattering.** X-rays of wavelength 0.140 nm are scattered from a very thin slice of carbon. What will be the wavelengths of X-rays scattered at (*a*) 0°, (*b*) 90°, (*c*) 180°?

APPROACH This is an example of the Compton effect, and we use Eq. 37–6a to find the wavelengths.

SOLUTION (*a*) For $\phi = 0°$, $\cos\phi = 1$ and $1 - \cos\phi = 0$. Then Eq. 37–6a gives $\lambda' = \lambda = 0.140\,\text{nm}$. This makes sense since for $\phi = 0°$, there really isn't any collision as the photon goes straight through without interacting.

(*b*) For $\phi = 90°$, $\cos\phi = 0$, and $1 - \cos\phi = 1$. So

$$\lambda' = \lambda + \frac{h}{m_e c} = 0.140\,\text{nm} + \frac{6.63 \times 10^{-34}\,\text{J}\cdot\text{s}}{(9.11 \times 10^{-31}\,\text{kg})(3.00 \times 10^{8}\,\text{m/s})}$$

$$= 0.140\,\text{nm} + 2.4 \times 10^{-12}\,\text{m} = 0.142\,\text{nm};$$

that is, the wavelength is longer by one Compton wavelength ($= h/m_e c^2 = 0.0024\,\text{nm}$ for an electron).

(*c*) For $\phi = 180°$, which means the photon is scattered backward, returning in the direction from which it came (a direct "head-on" collision), $\cos\phi = -1$, and $1 - \cos\phi = 2$. So

$$\lambda' = \lambda + 2\frac{h}{m_e c} = 0.140\,\text{nm} + 2(0.0024\,\text{nm}) = 0.145\,\text{nm}.$$

NOTE The maximum shift in wavelength occurs for backward scattering, and it is twice the Compton wavelength.

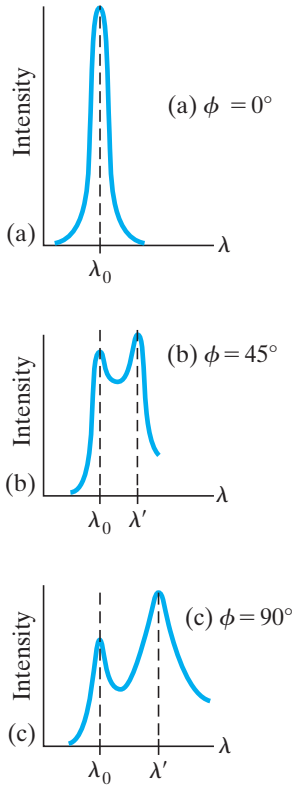


FIGURE 37–8 Plots of intensity of radiation scattered from a target such as graphite (carbon), for three different angles. The values for λ' match Eq. 37–6. For (a) $\phi = 0°$, $\lambda' = \lambda_0$. In (b) and (c) a peak is found at 45° or 90° at λ' (Eq. 37–6a) due to photons scattered from free electrons (or very nearly free). But there is also a peak at just slightly more than λ_0, which is due to scattering from electrons very tightly bound to their atoms so the mass in Eqs. 37–6 becomes very large (mass of the atom) and $\Delta\lambda$ becomes very small.

PHYSICS APPLIED
Measuring bone density

The Compton effect has been used to diagnose bone disease such as osteoporosis. Gamma rays (γ), which are photons of even shorter wavelength than X-rays, coming from a radioactive source are scattered off bone material. The total intensity of the scattered radiation is proportional to the density of electrons, which is in turn proportional to the bone density. A low bone density may indicate osteoporosis.

PHYSICS APPLIED
How visible light comes from the Sun: Compton scattering

The energy output of the Sun comes from nuclear reactions (Chapters 41 and 42) that release high-energy γ rays. Created near the center of the Sun, they may take tens of thousands of years to escape the Sun, undergoing repeated Compton scattering at small angles, slowly losing energy and increasing their wavelengths to be visible light, UV, or IR.

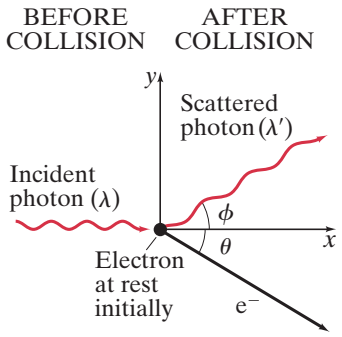


FIGURE 37–7 (Repeated.) The Compton effect. A single photon of wavelength λ strikes an electron in some material, knocking it out of its atom. The scattered photon has less energy (some energy is given to the electron) and hence has a longer wavelength λ' (shown exaggerated). Experiments found scattered X-rays of just the wavelengths predicted by conservation of energy and momentum using the photon model.

*Derivation of Compton Shift

If the incoming photon in Fig. 37–7 has wavelength λ, then its total energy and momentum are

$$E = hf = \frac{hc}{\lambda} \quad \text{and} \quad p = \frac{h}{\lambda}.$$

After the collision of Fig. 37–7, the photon scattered at the angle ϕ has a wavelength which we call λ'. Its energy and momentum are

$$E' = \frac{hc}{\lambda'} \quad \text{and} \quad p' = \frac{h}{\lambda'}.$$

The electron, assumed at rest before the collision but free to move when struck, is scattered at an angle θ. The electron's kinetic energy after the photon collides with it is (see Eq. 36–10):

$$K_e = \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right) m_e c^2$$

where m_e is the mass of the electron and v is its speed. The electron's momentum is

$$p_e = \frac{1}{\sqrt{1 - v^2/c^2}} m_e v.$$

We apply conservation of energy, before and after the collision of photon and electron (Fig. 37–7):

$$\text{incoming photon} \longrightarrow \text{scattered photon} + \text{electron}$$

$$\frac{hc}{\lambda} = \frac{hc}{\lambda'} + \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right) m_e c^2.$$

We apply conservation of momentum, before and after collision, to the x and y components of momentum:

$$(x) \qquad \frac{h}{\lambda} = \frac{h}{\lambda'} \cos\phi + \frac{m_e v \cos\theta}{\sqrt{1 - v^2/c^2}}$$

$$(y) \qquad 0 = \frac{h}{\lambda'} \sin\phi - \frac{m_e v \sin\theta}{\sqrt{1 - v^2/c^2}}.$$

We can combine the last three equations to eliminate v and θ, and we obtain, as Compton did, an equation for the wavelength of the scattered photon in terms of its scattering angle ϕ:

$$\lambda' = \lambda + \frac{h}{m_e c}(1 - \cos\phi),$$

which is Eq. 37–6a, and gives us the Compton shift (previous page).

37–5 Photon Interactions; Pair Production

When a photon passes through matter, it interacts with the atoms and electrons. There are four important types of interactions that a photon can undergo:

1. The *photoelectric effect*: A photon may knock an electron out of an atom and in the process the photon disappears.
2. The photon may knock an atomic electron to a higher energy state in the atom if its energy is not sufficient to knock the electron out altogether. In this process the photon also disappears, and all its energy is given to the atom. Such an atom is then said to be in an *excited state*, and we shall discuss it more later.
3. The photon can be scattered from an electron (or a nucleus) and in the process lose some energy; this is the *Compton effect* (Fig. 37–7). But notice that the photon is not slowed down. It still travels with speed c, but its frequency will be lower because its energy is less.
4. *Pair production*: A photon can actually create matter, such as the production of an electron and a positron, Fig. 37–9. (A positron has the same mass as an electron, but the opposite charge, $+e$, and is called its *antiparticle*.)

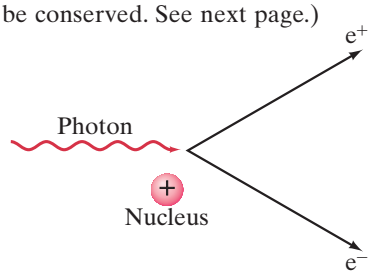


FIGURE 37–9 Pair production: a photon disappears and produces an electron and a positron. (This process requires an additional massive object, often an atomic nucleus, so that energy and momentum can be conserved. See next page.)

In process 4, **pair production**, the photon disappears in the process of creating the electron–positron pair. This is an example of mass being created from pure energy, and it occurs in accord with Einstein's equation $E = mc^2$. Notice that a photon cannot create an electron alone since electric charge would not then be conserved. The inverse of pair production also occurs: if a positron comes close to an electron, the two quickly **annihilate** each other and their energy, including their mass, appears as electromagnetic energy of photons. Positrons are not nearly as plentiful in nature as electrons, and they usually do not last long before annihilating with an electron.

Electron–positron annihilation is the basis for the type of medical imaging known as PET, as discussed in Section 42–9.

EXAMPLE 37–9 **Pair production.** (*a*) What is the minimum energy of a photon that can produce an electron–positron pair? (*b*) What is this photon's wavelength?

APPROACH The minimum photon energy E equals the rest energy (mc^2) of the two particles created, via Einstein's famous equation $E = mc^2$ (Eq. 36–12). There is no energy left over, so the particles produced will have zero kinetic energy. The wavelength is $\lambda = c/f$ where $E = hf$ for the original photon.

SOLUTION (*a*) Because $E = mc^2$, and the mass created is equal to two electron masses, the photon must have energy of at least

$$E = 2(9.11 \times 10^{-31}\,\text{kg})(3.0 \times 10^8\,\text{m/s})^2 = 1.64 \times 10^{-13}\,\text{J} = 1.02\,\text{MeV}$$

$(1\,\text{MeV} = 10^6\,\text{eV} = 1.60 \times 10^{-13}\,\text{J})$. A photon with less energy cannot undergo pair production.

(*b*) Since $E = hf = hc/\lambda$, the wavelength of a 1.02-MeV photon is

$$\lambda = \frac{hc}{E} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.0 \times 10^8\,\text{m/s})}{(1.64 \times 10^{-13}\,\text{J})} = 1.2 \times 10^{-12}\,\text{m},$$

which is 0.0012 nm. Such photons are in the gamma-ray (or very short X-ray) region of the electromagnetic spectrum (Fig. 31–12).

NOTE Photons of higher energy (shorter wavelength) can also create an electron–positron pair, with the excess energy becoming kinetic energy of the particles.

Pair production cannot occur in empty space, for momentum could not be conserved. In Example 37–9, for instance, energy is conserved, but only enough energy was provided to create the electron–positron pair at rest, and thus with zero momentum, which could not equal the initial momentum of the photon. Indeed, it is necessary that an additional massive object, such as an atomic nucleus, must take part in the interaction to carry off some of the momentum (Fig. 37–9).

37–6 Wave–Particle Duality; the Principle of Complementarity

The photoelectric effect, the Compton effect, and other experiments have placed the particle theory of light on a firm experimental basis. But what about the classic experiments of Young and others (Chapters 34 and 35) on interference and diffraction which showed that the *wave theory* of light also rests on a firm experimental basis?

We seem to be in a dilemma. Some experiments indicate that light behaves like a wave; others indicate that it behaves like a stream of particles. These two theories seem to be incompatible, but both have been shown to have validity. Physicists finally came to the conclusion that this duality of light must be accepted as a fact of life. It is referred to as the **wave–particle duality**. Apparently, light is a more complex phenomenon than just a simple wave or a simple beam of particles.



FIGURE 37–10 Niels Bohr (right), walking with Enrico Fermi along the Appian Way outside Rome. This photo shows one important way physics is done.

CAUTION

Not correct to say light is a wave and/or a particle. Light can **act** *like a wave or like a particle.*

To clarify the situation, the great Danish physicist Niels Bohr (1885–1962, Fig. 37–10) proposed his famous **principle of complementarity**. It states that to understand an experiment, sometimes we find an explanation using wave theory and sometimes using particle theory. Yet we must be aware of both the wave and particle aspects of light if we are to have a full understanding of light. Therefore these two aspects of light complement one another.

It is not easy to "visualize" this duality. We cannot readily picture a combination of wave and particle. Instead, we must recognize that the two aspects of light are different "faces" that light shows to experimenters.

Part of the difficulty stems from how we think. Visual pictures (or models) in our minds are based on what we see in the everyday world. We apply the concepts of waves and particles to light because in the macroscopic world we see that energy is transferred from place to place by these two methods. We cannot see directly whether light is a wave or particle, so we do indirect experiments. To explain the experiments, we apply the models of waves or of particles to the nature of light. But these are abstractions of the human mind. When we try to conceive of what light really "is," we insist on a visual picture. Yet there is no reason why light should conform to these models (or visual images) taken from the macroscopic world. The "true" nature of light—if that means anything—is not possible to visualize. The best we can do is recognize that our knowledge is limited to the indirect experiments, and that in terms of everyday language and images, light reveals both wave and particle properties.

It is worth noting that Einstein's equation $E = hf$ itself links the particle and wave properties of a light beam. In this equation, E refers to the energy of a particle; and on the other side of the equals sign we have the frequency f of the corresponding wave.

37–7 Wave Nature of Matter

In 1923, Louis de Broglie (1892–1987) extended the idea of the wave–particle duality. He appreciated the *symmetry* in nature, and argued that if light sometimes behaves like a wave and sometimes like a particle, then perhaps those things in nature thought to be particles—such as electrons and other material objects—might also have wave properties. De Broglie proposed that the wavelength of a material particle would be related to its momentum in the same way as for a photon, Eq. 37–5, $p = h/\lambda$. That is, for a particle having linear momentum $p = mv$, the wavelength λ is given by

$$\lambda = \frac{h}{p}, \tag{37–7}$$

and is valid classically ($p = mv$ for $v \ll c$) and relativistically ($p = \gamma mv = mv/\sqrt{1 - v^2/c^2}$). This is sometimes called the **de Broglie wavelength** of a particle.

EXAMPLE 37–10 **Wavelength of a ball.** Calculate the de Broglie wavelength of a 0.20-kg ball moving with a speed of 15 m/s.

APPROACH We use Eq. 37–7.

SOLUTION $\lambda = \dfrac{h}{p} = \dfrac{h}{mv} = \dfrac{(6.6 \times 10^{-34}\,\text{J}\cdot\text{s})}{(0.20\,\text{kg})\,(15\,\text{m/s})} = 2.2 \times 10^{-34}\,\text{m}.$

Ordinary size objects, such as the ball of Example 37–10, have unimaginably small wavelengths. Even if the speed is extremely small, say 10^{-4} m/s, the wavelength would be about 10^{-29} m. Indeed, the wavelength of any ordinary object is much too small to be measured and detected. The problem is that the properties of waves, such as interference and diffraction, are significant only when the size of objects or slits is not much larger than the wavelength. And there are no known objects or slits to diffract waves only 10^{-30} m long, so the wave properties of ordinary objects go undetected.

But tiny elementary particles, such as electrons, are another matter. Since the mass m appears in the denominator of Eq. 37–7, a very small mass should have a much larger wavelength.

EXAMPLE 37–11 **Wavelength of an electron.** Determine the wavelength of an electron that has been accelerated through a potential difference of 100 V.

APPROACH If the kinetic energy is much less than the rest energy, we can use the classical formula, $K = \frac{1}{2}mv^2$ (see Section 36–11). For an electron, $mc^2 = 0.511\,\text{MeV}$. We then apply conservation of energy: the kinetic energy acquired by the electron equals its loss in potential energy. After solving for v, we use Eq. 37–7 to find the de Broglie wavelength.

SOLUTION The gain in kinetic energy equals the loss in potential energy: $\Delta U = eV - 0$. Thus $K = eV$, so $K = 100\,\text{eV}$. The ratio $K/mc^2 = 100\,\text{eV}/(0.511 \times 10^6\,\text{eV}) \approx 10^{-4}$, so relativity is not needed. Thus

$$\frac{1}{2}mv^2 = eV$$

and

$$v = \sqrt{\frac{2eV}{m}} = \sqrt{\frac{(2)(1.6 \times 10^{-19}\,\text{C})(100\,\text{V})}{(9.1 \times 10^{-31}\,\text{kg})}} = 5.9 \times 10^6\,\text{m/s}.$$

Then

$$\lambda = \frac{h}{mv} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})}{(9.1 \times 10^{-31}\,\text{kg})(5.9 \times 10^6\,\text{m/s})} = 1.2 \times 10^{-10}\,\text{m},$$

or 0.12 nm.

EXERCISE E As a particle travels faster, does its de Broglie wavelength decrease, increase, or remain the same?

EXERCISE F Return to the Chapter-Opening Question, page 1095, and answer it again now. Try to explain why you may have answered differently the first time.

Electron Diffraction; Wave–Particle Duality of Matter

From Example 37–11, we see that electrons can have wavelengths on the order of $10^{-10}\,\text{m}$, and even smaller. Although small, this wavelength can be detected: the spacing of atoms in a crystal is on the order of $10^{-10}\,\text{m}$ and the orderly array of atoms in a crystal could be used as a type of diffraction grating, as was done earlier for X-rays (Section 35–11, especially Fig. 35–28, and Fig. 37–12 on next page.) C. J. Davisson and L. H. Germer performed the crucial experiment; they scattered electrons from the surface of a metal crystal and, in early 1927, observed that the electrons were scattered into a pattern of regular peaks. When they interpreted these peaks as a diffraction pattern, the wavelength of the diffracted electron wave was found to be just that predicted by de Broglie, Eq. 37–7. In the same year, G. P. Thomson (son of J. J. Thomson) used different materials and also detected diffraction of electrons. See Fig. 37–11 and further discussion on next page, Example 37–12 and Fig. 37–12. Compare it to X-ray diffraction, Section 35–11.

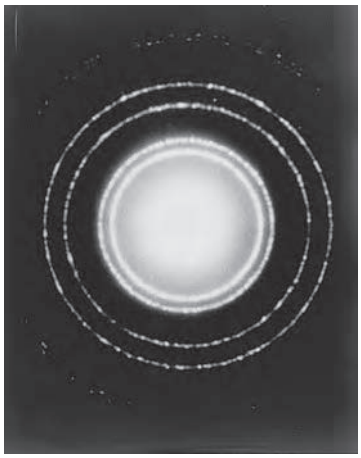


FIGURE 37–11 Diffraction pattern of electrons scattered from aluminum foil, as recorded on film.

Later experiments showed that protons, neutrons, and other particles also have wave properties.

Thus the wave–particle duality applies to material objects as well as to light. The principle of complementarity applies to matter as well. That is, we must be aware of both the particle and wave aspects in order to have an understanding of matter, including electrons. But again we must recognize that a visual picture of a "wave–particle" is not possible. The principle of complementarity (Section 37–6) applies also to matter.

PHYSICS APPLIED
Electron diffraction

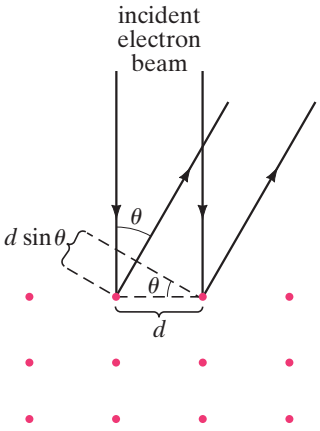


FIGURE 37–12 Example 37–12. The red dots represent atoms in an orderly array in a solid.

EXAMPLE 37–12 **Electron diffraction.** The wave nature of electrons is manifested in experiments where an electron beam interacts with the atoms on the surface of a solid, especially crystals. By studying the angular distribution of the diffracted electrons, one can indirectly measure the geometrical arrangement of atoms. Assume that the electrons strike perpendicular to the surface of a solid (see Fig. 37–12), and that their energy is low, $K = 100\,\text{eV}$, so that they interact only with the surface layer of atoms. If the smallest angle at which a diffraction maximum occurs is at 24°, what is the separation d between the atoms on the surface?

SOLUTION Treating the electrons as waves, we need to determine the condition where the difference in path traveled by the wave diffracted from adjacent atoms is an integer multiple of the de Broglie wavelength, so that constructive interference occurs. The path length difference is $d\sin\theta$; so for the smallest value of θ we must have

$$d\sin\theta = \lambda.$$

However, λ is related to the (non-relativistic) kinetic energy K by

$$K = \frac{p^2}{2m_e} = \frac{h^2}{2m_e\lambda^2}.$$

Thus

$$\lambda = \frac{h}{\sqrt{2m_eK}}$$

$$= \frac{(6.63\times10^{-34}\,\text{J}\cdot\text{s})}{\sqrt{2(9.11\times10^{-31}\,\text{kg})(100\,\text{eV})(1.6\times10^{-19}\,\text{J/eV})}} = 0.123\,\text{nm}.$$

The surface inter-atomic spacing is

$$d = \frac{\lambda}{\sin\theta} = \frac{0.123\,\text{nm}}{\sin 24^\circ} = 0.30\,\text{nm}.$$

NOTE Experiments of this type verify both the wave nature of electrons and the orderly array of atoms in crystalline solids.

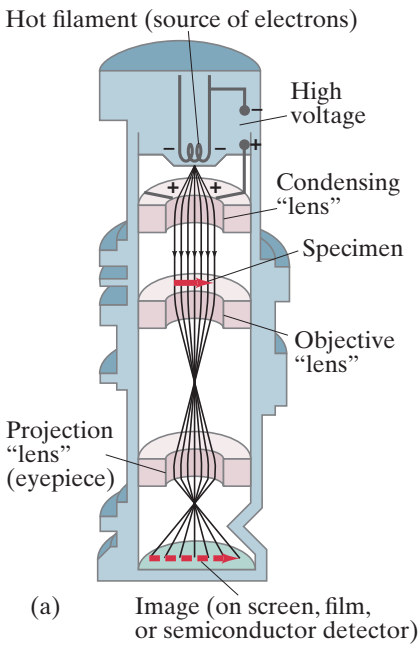
What Is an Electron?

We might ask ourselves: "What is an electron?" The early experiments of J. J. Thomson (Section 27–7) indicated a glow in a tube, and that glow moved when a magnetic field was applied. The results of these and other experiments were best interpreted as being caused by tiny negatively charged particles which we now call electrons. No one, however, has actually seen an electron directly. The drawings we sometimes make of electrons as tiny spheres with a negative charge on them are merely convenient pictures (now recognized to be inaccurate). Again we must rely on experimental results, some of which are best interpreted using the particle model and others using the wave model. These models are mere pictures that we use to extrapolate from the macroscopic world to the tiny microscopic world of the atom. And there is no reason to expect that these models somehow reflect the reality of an electron. We thus use a wave or a particle model (whichever works best in a situation) so that we can talk about what is happening. But we should not be led to believe that an electron *is* a wave or a particle. Instead we could say that an electron is the set of its properties that we can measure. Bertrand Russell said it well when he wrote that an electron is "a logical construction."

37–8 Electron Microscopes

PHYSICS APPLIED
Electron microscope

The idea that electrons have wave properties led to the development of the **electron microscope** (EM), which can produce images of much greater magnification than a light microscope because electron wavelengths can be much shorter (Section 35–6). Figures 37–13 (a) and (b) are diagrams of two types, developed around the middle of the twentieth century: the **transmission electron microscope** (TEM), which produces a two-dimensional image, and the **scanning electron microscope** (SEM), which produces images with a three-dimensional quality.



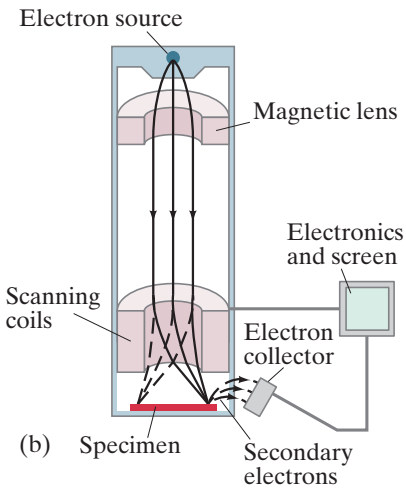


FIGURE 37–13 (a) Transmission electron microscope. The magnetic field coils are designed to be "magnetic lenses," which bend the electron paths and bring them to a focus, as shown. The sensors of the image measure electron intensity only, not color. (b) Scanning electron microscope. Scanning coils move an electron beam back and forth across the specimen. Secondary electrons produced when the beam strikes the specimen are collected and their intensity affects the brightness of pixels in a monitor to produce a picture.

In both types, the objective and eyepiece lenses are actually magnetic fields that exert forces on the electrons to bring them to a focus. The fields are produced by carefully designed current-carrying coils of wire. Photographs using each type are shown in Fig. 37–14. EMs measure the intensity of electrons, producing monochromatic photos. Color is often added artificially to highlight.

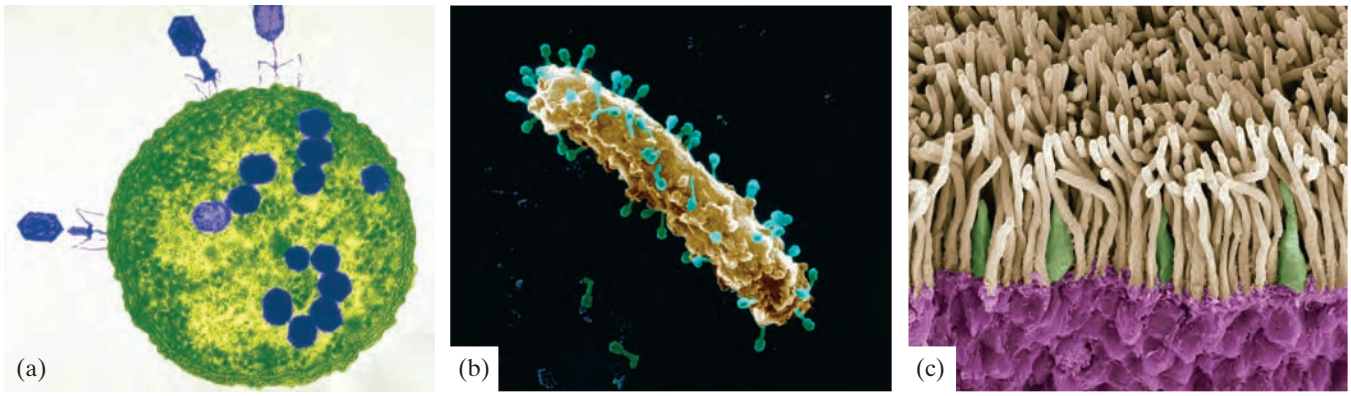


FIGURE 37–14 Electron micrographs, in false color, of (a) viruses attacking a bacterium (TEM, ≈ 50,000×). (b) Same subject by an SEM (≈ 35,000×). (c) SEM image of an eye's retina (Section 33–6); rods and cones colored beige and green, respectively.

As discussed in Section 35–6, the maximum resolution of details on an object is about the size of the wavelength of the radiation used to view it. Electrons accelerated by voltages on the order of 10^5 V have wavelengths of about 0.004 nm. The maximum resolution obtainable would be on this order, but in practice, aberrations in the magnetic lenses limit the resolution in transmission electron microscopes to about 0.1 to 0.5 nm. This is still 1000 times better than a visible-light microscope, and corresponds to a useful magnification of about a million. Such magnifications are difficult to achieve, and more common magnifications are 10^4 to 10^5. The maximum resolution of a scanning electron microscope is less, typically 5 to 10 nm, although new high-resolution SEMs approach 1 nm.

The **scanning tunneling electron microscope** (STM), developed in the 1980s, contains a tiny probe, whose tip may be only one (or a few) atoms wide, that is moved across the specimen to be examined in a series of linear passes. The tip, as it scans, remains very close to the surface of the specimen, about 1 nm above it, Fig. 37–15. A small voltage applied between the probe and the surface causes electrons to leave the surface and pass through the vacuum to the probe, by a quantum process known as *tunneling* (discussed in Section 38–10). This "tunneling" current is very sensitive to the gap width, so a feedback mechanism can be used to raise and lower the probe to maintain a constant electron current.

PHYSICS APPLIED
STM and AFM

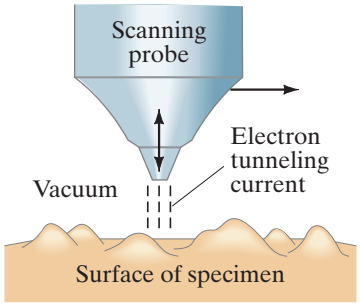


FIGURE 37–15 The probe tip of a scanning tunneling electron microscope, as it is moved horizontally, automatically moves up and down to maintain a constant tunneling current, and this motion is translated into an image of the surface.

The probe's vertical motion, following the surface of the specimen, is then plotted as a function of position, scan after scan, producing a three-dimensional image of the surface. Surface features as fine as the size of an atom can be resolved: a resolution better than 50 pm (0.05 nm) laterally and 0.01 to 0.001 nm vertically. This resolution allows the study of the surface structure of materials. The "topographic" image of a surface represents the distribution of electron charge.

The **atomic force microscope** (AFM) is similar to an STM, but can be used on a wider range of materials. An AFM measures the force between a cantilevered tip and the sample, a force which depends strongly on the tip-sample separation at each point.

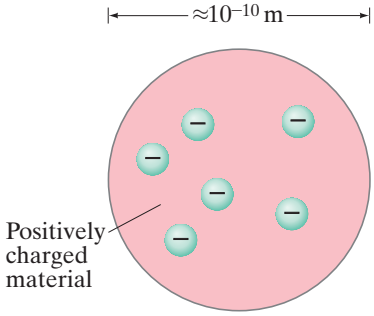


FIGURE 37–16 Plum-pudding model of the atom.

37–9 Early Models of the Atom

The idea that matter is made up of atoms was accepted by most scientists by 1900. With the discovery of the electron in the 1890s, scientists began to think of the atom itself as having a structure with electrons as part of that structure. We now discuss how our modern view of the atom developed, and the quantum theory with which it is intertwined.†

A typical model of the atom in the 1890s visualized the atom as a homogeneous sphere of positive charge inside of which there were tiny negatively charged electrons, a little like plums in a pudding, Fig. 37–16.

Around 1911, Ernest Rutherford (1871–1937) and his colleagues performed experiments whose results contradicted the plum-pudding model of the atom. In these experiments a beam of positively charged alpha (α) particles was directed at a thin sheet of metal foil such as gold, Fig. 37–17a. (These newly discovered α particles were emitted by certain radioactive materials and were soon shown to be doubly ionized helium atoms—that is, having a charge of $+2e$.) It was expected from the plum-pudding model that the alpha particles would not be deflected significantly because electrons are so much lighter than alpha particles, and the alpha particles should not have encountered any massive concentration of positive charge to strongly repel them. The experimental results completely contradicted these predictions. It was found that most of the alpha particles passed through the foil unaffected, as if the foil was mostly empty space.

FIGURE 37–17 (a) Experimental setup for Rutherford's experiment: α particles emitted by radon are deflected by the atoms of a thin metal foil and a few rebound backward. (b) Backward rebound of α particles explained as the repulsion from a heavy positively charged nucleus.

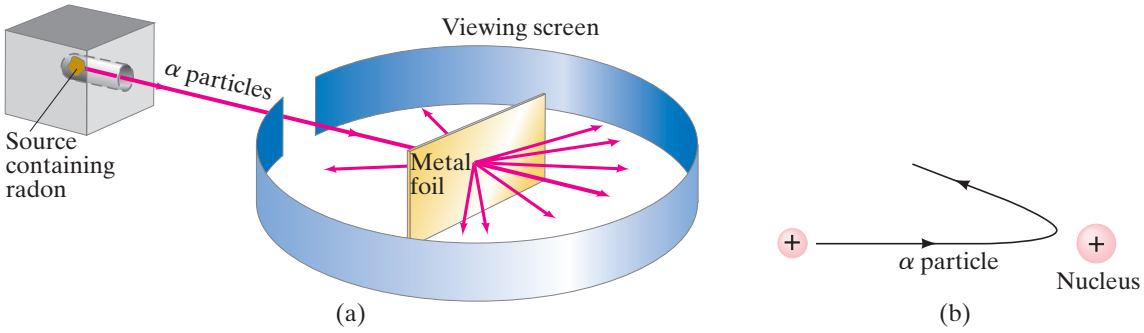
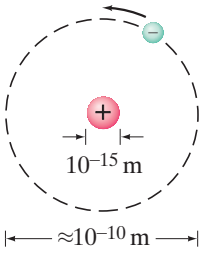


FIGURE 37–18 Rutherford's model of the atom: electrons orbit a tiny positive nucleus (not to scale). The atom is visualized as mostly empty space.



And of those deflected, a few were deflected at very large angles—some even backward, nearly in the direction from which they had come. This could happen, Rutherford reasoned, only if the positively charged alpha particles were being repelled by a massive positive charge concentrated in a very small region of space (see Fig. 37–17b). He hypothesized that the atom must consist of a tiny but massive positively charged nucleus, containing over 99.9% of the mass of the atom, surrounded by much lighter electrons some distance away. The electrons would be moving in orbits about the nucleus—much as the planets move around the Sun—because if they were at rest, they would fall into the nucleus due to electrical attraction. See Fig. 37–18.

†Some readers may say: "Tell us the facts as we know them today, and don't bother us with the historical background and its outmoded theories." Such an approach would ignore the creative aspect of science and give a false impression of how science develops. Moreover, it is not really possible to understand today's view of the atom without insight into the concepts that led to it.

Rutherford's experiments suggested that the nucleus must have a radius of about 10^{-15} to 10^{-14} m. From kinetic theory, and especially Einstein's analysis of Brownian motion (see Section 17–1), the radius of atoms was estimated to be about 10^{-10} m. Thus the electrons would seem to be at a distance from the nucleus of about 10,000 to 100,000 times the radius of the nucleus itself. (If the nucleus were the size of a baseball, the atom would have the diameter of a big city several kilometers across.) So an atom would be mostly empty space.

Atom diameters $\approx 10^{-10}$ m
Nucleus diameters $\approx 10^{-15}$ m

Rutherford's **planetary model** of the atom (also called the **nuclear model** of the atom) was a major step toward how we view the atom today. It was not, however, a complete model and presented some major problems, as we shall see.

37–10 Atomic Spectra: Key to the Structure of the Atom

Earlier in this Chapter we saw that heated solids (as well as liquids and dense gases) emit light with a continuous spectrum of wavelengths. This radiation is assumed to be due to oscillations of atoms and molecules, which are largely governed by the interaction of each atom or molecule with its neighbors.

Rarefied gases can also be excited to emit light. This is done by intense heating, or more commonly by applying a high voltage to a "discharge tube" containing the gas at low pressure, Fig. 37–19. The radiation from excited gases had been observed early in the nineteenth century, and it was found that the spectrum was not continuous. Rather, excited gases emit light of only certain wavelengths, and when this light is analyzed through the slit of a spectroscope or spectrometer (Section 35–9), a **line spectrum** is seen rather than a continuous spectrum. The line spectra emitted by a number of elements in the visible region are shown below in Fig. 37–20, and in Chapter 35, Fig. 35–22. The **emission spectrum** is characteristic of the material and can serve as a type of "fingerprint" for identification of the gas.

We also saw (Section 35–9 and Fig. 35–22) that if a continuous spectrum passes through a rarefied gas, dark lines are observed in the emerging spectrum, at wavelengths corresponding to lines normally emitted by the gas. This is called an **absorption spectrum** (Fig. 37–20c), and it became clear that gases can absorb light at the same frequencies at which they emit. Using film sensitive to ultraviolet and to infrared light, it was found that gases emit and absorb discrete frequencies in these regions as well as in the visible.

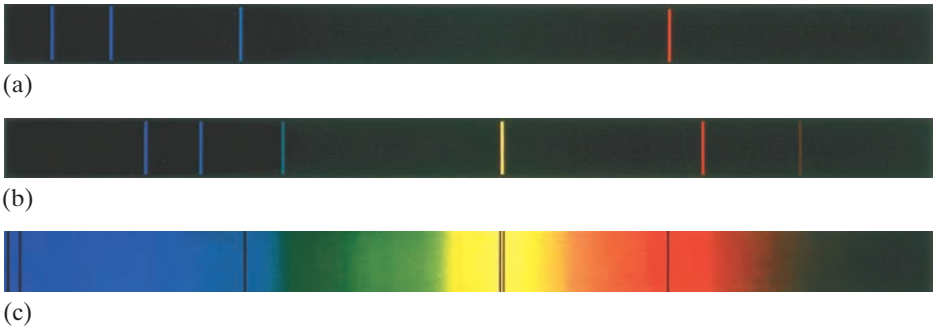
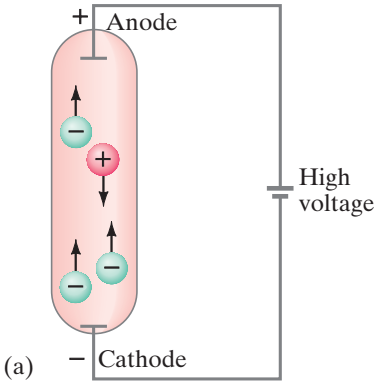


FIGURE 37–20 Emission spectra of the gases (a) atomic hydrogen, (b) helium, and (c) the *solar absorption* spectrum.

In low-density gases, the atoms are far apart on the average and hence the light emitted or absorbed is assumed to be by *individual atoms* rather than through interactions between atoms, as in a solid, liquid, or dense gas. Thus the line spectra serve as a key to the structure of the atom: any theory of atomic structure must be able to explain why atoms emit light only of discrete wavelengths, and it should be able to predict what these wavelengths are.



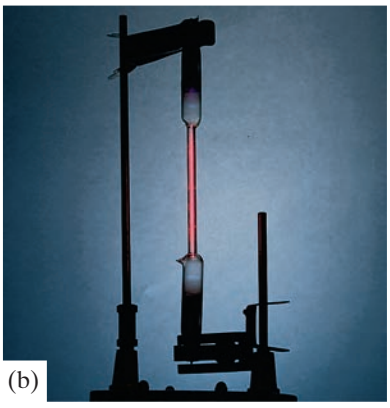


FIGURE 37–19 Gas-discharge tube: (a) diagram; (b) photo of an actual discharge tube for hydrogen.

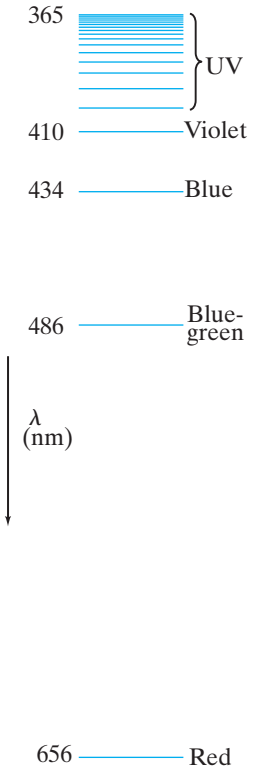


FIGURE 37–21 Balmer series of lines for hydrogen.

Hydrogen is the simplest atom—it has only one electron orbiting its nucleus. It also has the simplest spectrum. The spectrum of most atoms shows little apparent regularity. But the spacing between lines in the hydrogen spectrum decreases in a regular way, Fig. 37–21. Indeed, in 1885, J. J. Balmer (1825–1898) showed that the four lines in the visible portion of the hydrogen spectrum (with measured wavelengths 656 nm, 486 nm, 434 nm, and 410 nm) have wavelengths that fit the formula

$$\frac{1}{\lambda} = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right), \qquad n = 3, 4, \cdots. \tag{37–8}$$

Here n takes on the values 3, 4, 5, 6 for the four visible lines, and R, called the **Rydberg constant**, has the value $R = 1.0974 \times 10^7\ \text{m}^{-1}$. Later it was found that this **Balmer series** of lines extended into the UV region, ending at $\lambda = 365$ nm, as shown in Fig. 37–21. Balmer's formula, Eq. 37–8, also worked for lines with higher integer values of n. The lines near 365 nm became too close together to distinguish, but the limit of the series at 365 nm corresponds to $n = \infty$ (so $1/n^2 = 0$ in Eq. 37–8).

Later experiments on hydrogen showed that there were similar series of lines in the UV and IR regions, and each series had a pattern just like the Balmer series, but at different wavelengths, Fig. 37–22. Each of these series was found to fit a formula with the same form as Eq. 37–8 but with the $1/2^2$ replaced by $1/1^2$, $1/3^2$, $1/4^2$, and so on. For example, the **Lyman series** contains lines with wavelengths from 91 nm to 122 nm (in the UV region) and fits the formula

$$\frac{1}{\lambda} = R\left(\frac{1}{1^2} - \frac{1}{n^2}\right), \qquad n = 2, 3, \cdots.$$

The wavelengths of the **Paschen series** (in the IR region) fit

$$\frac{1}{\lambda} = R\left(\frac{1}{3^2} - \frac{1}{n^2}\right), \qquad n = 4, 5, \cdots.$$

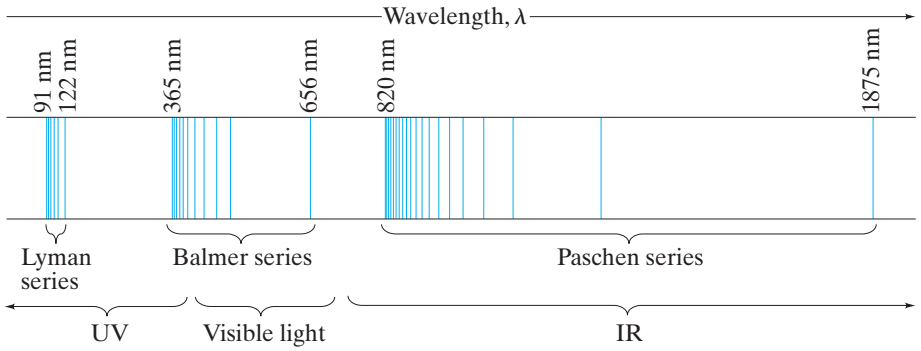


FIGURE 37–22 Line spectrum of atomic hydrogen. Each series fits the formula $\frac{1}{\lambda} = R\left(\frac{1}{n'^2} - \frac{1}{n^2}\right)$ where $n' = 1$ for the Lyman series, $n' = 2$ for the Balmer series, $n' = 3$ for the Paschen series, and so on; n can take on all integer values from $n = n' + 1$ up to infinity. The only lines in the visible region of the electromagnetic spectrum are part of the Balmer series.

The Rutherford model was unable to explain why atoms emit line spectra. It had other difficulties as well. According to the Rutherford model, electrons orbit the nucleus, and since their paths are curved the electrons are accelerating. Hence they should give off light like any other accelerating electric charge (Section 31–4). Since light carries off energy and energy is conserved, the electron's own energy must decrease to compensate. Hence electrons would be expected to spiral into the nucleus. As they spiraled inward, their frequency would increase in a short time and so too would the frequency of the light emitted. Thus the two main difficulties of the Rutherford model are these: (1) it predicts that light of a continuous range of frequencies would be emitted, whereas experiment shows line spectra; (2) it predicts that atoms are unstable—electrons would quickly spiral into the nucleus—but we know that atoms in general are stable, because there is stable matter all around us.

Clearly Rutherford's model was not sufficient. Some sort of modification was needed, and Niels Bohr provided it in a model that included the quantum hypothesis. Although the Bohr model has been superseded, it did provide a crucial stepping stone to our present understanding. And some aspects of the Bohr model are still useful today, so we examine it in detail in the next Section.

37–11 The Bohr Model

Bohr had studied in Rutherford's laboratory for several months in 1912 and was convinced that Rutherford's planetary model of the atom had validity. But in order to make it work, he felt that the newly developing quantum theory would somehow have to be incorporated in it. The work of Planck and Einstein had shown that in heated solids, the energy of oscillating electric charges must change discontinuously—from one discrete energy state to another, with the emission of a quantum of light. Perhaps, Bohr argued, the electrons in an atom also cannot lose energy continuously, but must do so in quantum "jumps." In working out his model during the next year, Bohr postulated that electrons move about the nucleus in circular orbits, but that only certain orbits are allowed. He further postulated that an electron in each orbit would have a definite energy and would move in the orbit *without radiating energy* (even though this violated classical ideas since accelerating electric charges are supposed to emit EM waves; see Section 31–4). He thus called the possible orbits **stationary states**. In this **Bohr model**, light is emitted only when an electron jumps from a higher (upper) stationary state to another of lower energy, Fig. 37–23. When such a transition occurs, a single photon of light is emitted whose energy, by energy conservation, is given by

$$hf = E_u - E_\ell, \tag{37–9}$$

where E_u refers to the energy of the upper state and E_ℓ the energy of the lower state.

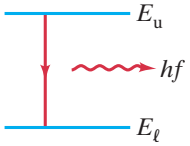


FIGURE 37–23 An atom emits a photon (energy = hf) when its energy changes from E_u to a lower energy E_ℓ.

In 1912–13, Bohr set out to determine what energies these orbits would have in the simplest atom, hydrogen; the spectrum of light emitted could then be predicted from Eq. 37–9. In the Balmer formula he had the key he was looking for. Bohr quickly found that his theory would agree with the Balmer formula if he assumed that the electron's angular momentum L is quantized and equal to an integer n times $h/2\pi$. As we saw in Chapter 11 angular momentum is given by $L = I\omega$, where I is the moment of inertia and ω is the angular velocity. For a single particle of mass m moving in a circle of radius r with speed v, $I = mr^2$ and $\omega = v/r$; hence, $L = I\omega = (mr^2)(v/r) = mvr$. Bohr's **quantum condition** is

$$L = mvr_n = n\frac{h}{2\pi}, \qquad n = 1, 2, 3, \cdots, \tag{37–10}$$

where n is an integer and r_n is the radius of the n^{th} possible orbit. The allowed orbits are numbered $1, 2, 3, \cdots$, according to the value of n, which is called the **principal quantum number** of the orbit.

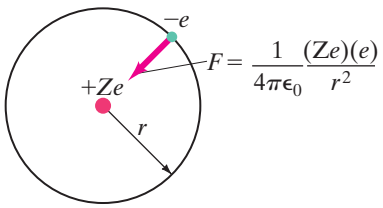
Equation 37–10 did not have a firm theoretical foundation. Bohr had searched for some "quantum condition," and such tries as $E = hf$ (where E represents the energy of the electron in an orbit) did not give results in accord with experiment. Bohr's reason for using Eq. 37–10 was simply that it worked; and we now look at how. In particular, let us determine what the Bohr theory predicts for the measurable wavelengths of emitted light.

An electron in a circular orbit of radius r (Fig. 37–24) would have a centripetal acceleration v^2/r produced by the electrical force of attraction between the negative electron and the positive nucleus. This force is given by Coulomb's law,

$$F = k\frac{(Ze)(e)}{r^2} = \frac{1}{4\pi\epsilon_0}\frac{(Ze)(e)}{r^2}$$

where $k = 1/4\pi\epsilon_0 = 8.99 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2$. The charge on the electron is $q_1 = -e$, and the charge on the nucleus is $q_2 = +Ze$: $+e$ is the charge on a proton, Z is the number of protons in the nucleus (called "atomic number,").[†] For the hydrogen atom, $Z = +1$.

FIGURE 37–24 Electric force (Coulomb's law) keeps the negative electron in orbit around the positively charged nucleus.



[†]We include Z in our derivation so that we can treat other single-electron ("hydrogenlike") atoms such as the ions He^+ $(Z = 2)$ and Li^{2+} $(Z = 3)$. Helium in the neutral state has two electrons: if one electron is missing, the remaining He^+ ion consists of one electron revolving around a nucleus of charge $+2e$. Similarly, doubly ionized lithium, Li^{2+}, also has a single electron, and in this case $Z = 3$.

In Newton's second law, $F = ma$, we substitute Coulomb's law for F and $a = v^2/r_n$ for a particular allowed orbit of radius r_n, and obtain

$$F = ma$$

$$\frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r_n^2} = \frac{mv^2}{r_n}.$$

We solve this for r_n,

$$r_n = \frac{Ze^2}{4\pi\epsilon_0 mv^2},$$

and then substitute for v from Eq. 37–10 (which says $v = nh/2\pi mr_n$):

$$r_n = \frac{Ze^2}{4\pi\epsilon_0 mv^2} = \frac{Ze^2 4\pi^2 mr_n^2}{4\pi\epsilon_0 n^2 h^2}.$$

We solve for r_n (it appears on both sides, so we cancel one of them) and find

$$r_n = \frac{n^2h^2\epsilon_0}{\pi mZe^2} = \frac{n^2}{Z}r_1 \qquad \textbf{(37–11)}$$

where n is an integer, the principal quantum number (Eq. 37–10), and

$$r_1 = \frac{h^2\epsilon_0}{\pi me^2}.$$

Equation 37–11 gives the radii of all possible orbits. The smallest orbit is for $n = 1$, and for hydrogen $(Z = 1)$ has the value

$$r_1 = \frac{(1)^2(6.626 \times 10^{-34}\,\text{J}\cdot\text{s})^2(8.85 \times 10^{-12}\,\text{C}^2/\text{N}\cdot\text{m}^2)}{(3.14)(9.11 \times 10^{-31}\,\text{kg})(1.602 \times 10^{-19}\,\text{C})^2}$$

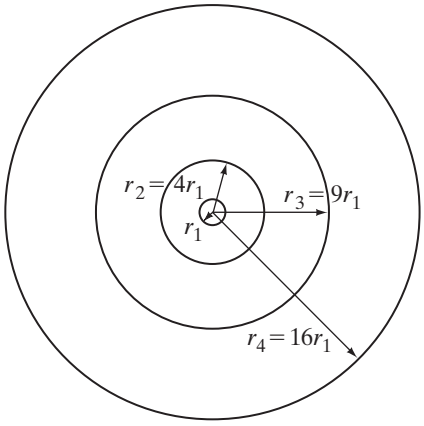
or

$$r_1 = 0.529 \times 10^{-10}\,\text{m}. \qquad \textbf{(37–12)}$$

The radius of the smallest orbit in hydrogen, r_1, is often called the **Bohr radius**. From Eq. 37–11, we see that the radii of the larger orbits† increase as n^2, so

$$r_2 = 4r_1 = 2.12 \times 10^{-10}\,\text{m},$$
$$r_3 = 9r_1 = 4.76 \times 10^{-10}\,\text{m},$$
$$\vdots$$
$$r_n = n^2 r_1. \qquad n = 1, 2, 3, \cdots.$$

FIGURE 37–25 The four smallest orbits in the Bohr model of hydrogen; $r_1 = 0.529 \times 10^{-10}\,\text{m}$.



The first four orbits are shown in Fig. 37–25. Notice that, according to Bohr's model, an electron can exist only in the orbits given by Eq. 37–11. There are no allowable orbits in between.

For an atom with $Z > 1$, we can write the orbital radii, r_n, using Eq. 37–11:

$$r_n = \frac{n^2}{Z}(0.529 \times 10^{-10}\,\text{m}), \qquad n = 1, 2, 3, \cdots. \qquad \textbf{(37–13)}$$

In each of its possible orbits, the electron in a Bohr model atom would have a definite energy, as the following calculation shows. The total energy equals the sum of the kinetic and potential energies. The potential energy of the electron is given by $U = qV = -eV$, where V is the potential due to a point charge $+Ze$ as given by Eq. 23–5: $V = (1/4\pi\epsilon_0)(Q/r) = (1/4\pi\epsilon_0)(Ze/r)$. So

$$U = -eV = -\frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r}.$$

†Be careful not to believe that these well-defined orbits actually exist. Today electrons are better thought of as forming "clouds," as discussed in Chapter 39.

The total energy E_n for an electron in the n^{th} orbit of radius r_n is the sum of the kinetic and potential energies:

$$E_n = \tfrac{1}{2}mv^2 - \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r_n}.$$

When we substitute v from Eq. 37–10 and r_n from Eq. 37–11 into this equation, we obtain

$$E_n = -\left(\frac{Z^2e^4m}{8\epsilon_0^2h^2}\right)\left(\frac{1}{n^2}\right), \qquad n = 1, 2, 3, \cdots. \tag{37–14a}$$

We evaluate the constant term in Eq. 37–14a and convert it to electron volts, as is customary in atomic physics, recalling (Section 23–8) that $1\,\text{eV} = 1.60 \times 10^{-19}\,\text{J}$. Then we get

$$E_n = -(13.6\,\text{eV})\frac{Z^2}{n^2}, \qquad n = 1, 2, 3, \cdots. \tag{37–14b}$$

The lowest energy level $(n = 1)$ for hydrogen $(Z = 1)$ is

$$E_1 = -13.6\,\text{eV}.$$

Since n^2 appears in the denominator of Eq. 37–14b, the energies of the larger orbits in hydrogen $(Z = 1)$ are given by

$$E_n = \frac{-13.6\,\text{eV}}{n^2}.$$

For example,

$$E_2 = \frac{-13.6\,\text{eV}}{4} = -3.40\,\text{eV},$$

$$E_3 = \frac{-13.6\,\text{eV}}{9} = -1.51\,\text{eV}.$$

We see that not only are the orbit radii quantized, but from Eqs. 37–14 so is the energy. The quantum number n that labels the orbit radii also labels the energy levels. The lowest **energy level** or **energy state** has energy E_1, and is called the **ground state**. The higher states, E_2, E_3, and so on, are called **excited states**. The fixed energy levels are also called **stationary states**.

Notice that although the energy for the larger orbits has a smaller numerical value, all the energies are less than zero. Thus, $-3.4\,\text{eV}$ is a higher energy than $-13.6\,\text{eV}$. Hence the orbit closest to the nucleus (r_1) has the lowest energy (the most negative). The reason the energies have negative values has to do with the way we defined the zero for potential energy (U). For two point charges, $U = (1/4\pi\epsilon_0)(q_1q_2/r)$ corresponds to zero potential energy when the two charges are infinitely far apart. Thus, an electron that can just barely be free from the atom by reaching $r = \infty$ (or, at least, far from the nucleus) with zero kinetic energy will have $E = K + U = 0 + 0 = 0$, corresponding to $n = \infty$ in Eqs. 37–14. If an electron is free and has kinetic energy, then $E > 0$. To remove an electron that is part of an atom requires an energy input (otherwise atoms would not be stable). Since $E \geq 0$ for a free electron, then an electron bound to an atom needs to have $E < 0$. That is, energy must be added to bring its energy up, from a negative value, to at least zero in order to free the electron.

The minimum energy required to remove an electron from an atom initially in the ground state is called the **binding energy** or **ionization energy**. The ionization energy for hydrogen has been measured to be 13.6 eV, and this corresponds precisely to removing an electron from the lowest state, $E_1 = -13.6\,\text{eV}$, up to $E = 0$ where it can be free.

At room temperature, nearly all atoms are in the ground state with $r_1 = 0.5 \times 10^{-10}\,\text{m}$. Thus we can expect that atoms typically have a diameter of about $10^{-10}\,\text{m}$, consistent with Brownian motion (Section 17–1).

Atom diameter $\approx 10^{-10}\,\text{m}$

Spectra Lines Explained

It is useful to show the various possible energy values as horizontal lines on an energy-level diagram. This is shown for hydrogen in Fig. 37–26.[†] The electron in a hydrogen atom can be in any one of these levels according to Bohr's theory. But it could never be in between, say at $-9.0\,\text{eV}$. At room temperature, nearly all H atoms will be in the ground state $(n = 1)$. At higher temperatures, or during an electric discharge when there are many collisions between free electrons and atoms, many atoms can be in excited states $(n > 1)$. Once in an excited state, an atom's electron can jump down to a lower state, and give off a photon in the process. This is, according to the Bohr model, the origin of the emission spectra of excited gases.

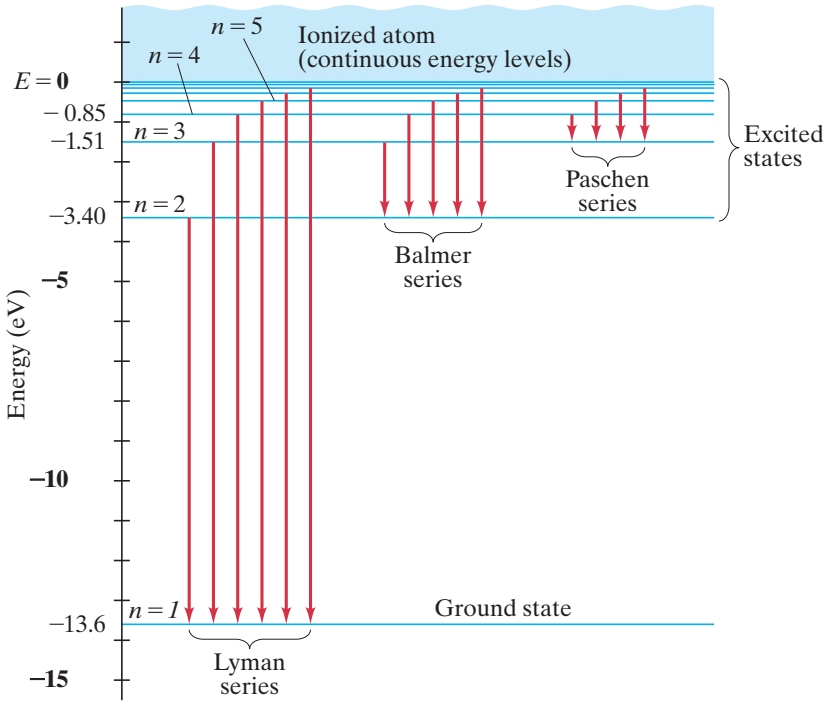


FIGURE 37–26 Energy-level diagram for the hydrogen atom, showing the transitions for the spectral lines of the Lyman, Balmer, and Paschen series (Fig. 37–22). Each vertical arrow represents an atomic transition that gives rise to the photons of one spectral line (a single wavelength or frequency).

The vertical arrows in Fig. 37–26 represent the transitions or jumps that correspond to the various observed spectral lines. For example, an electron jumping from the level $n = 3$ to $n = 2$ would give rise to the 656-nm line in the Balmer series, and the jump from $n = 4$ to $n = 2$ would give rise to the 486-nm line (see Fig. 37–21). We can predict wavelengths of the spectral lines emitted according to Bohr by combining Eq. 37–9 $(hf = E_u - E_\ell)$ with Eq. 37–14a. Since $hf = hc/\lambda$, we have from Eq. 37–9

$$\frac{1}{\lambda} = \frac{hf}{hc} = \frac{1}{hc}(E_n - E_{n'}),$$

where n refers to the upper state and n' to the lower state. Then using Eq. 37–14a,

$$\frac{1}{\lambda} = \frac{Z^2e^4m}{8\epsilon_0^2h^3c}\left(\frac{1}{(n')^2} - \frac{1}{(n)^2}\right). \qquad \textbf{(37–15)}$$

This theoretical formula has the same form as the experimental Balmer formula, Eq. 37–8, with $n' = 2$. Thus we see that the Balmer series of lines corresponds to transitions or "jumps" that bring the electron down to the second energy level from higher energy levels. Similarly, $n' = 1$ corresponds to the Lyman series and $n' = 3$ to the Paschen series (see Fig. 37–26).

[†]Note that if $E \geq 0$, an electron is free and can have any energy (E is not quantized). Thus there is a continuum of energy states above $E = 0$, as indicated as pale blue in the energy-level diagram of Fig. 37–26.

When the constant in Eq. 37–15 is evaluated with $Z = 1$, it is found to have the measured value of the Rydberg constant, $R = 1.0974 \times 10^7\,\text{m}^{-1}$, as in Eq. 37–8, in accord with experiment.

The great success of Bohr's model is that it gives an explanation for why atoms emit line spectra, and accurately predicts the wavelengths of emitted light for hydrogen. The Bohr model also explains absorption spectra: photons of just the right wavelength can knock an electron from one energy level to a higher one. To conserve energy, only photons that have just the right energy will be absorbed. This explains why a continuous spectrum of light entering a gas will emerge with dark (absorption) lines at frequencies that correspond to emission lines (Fig. 37–20c).

The Bohr theory also ensures the stability of atoms. It establishes stability by decree: the ground state is the lowest state for an electron and there is no lower energy level to which it can go and emit more energy. Finally, as we saw above, the Bohr theory accurately predicts the ionization energy of 13.6 eV for hydrogen. However, the Bohr model was not so successful for other atoms, and has been superseded as we shall discuss in the next Chapter. We discuss the Bohr model because it *was* an important start, and because we still use the concept of stationary states, the ground state, and transitions between states. Also, the terminology used in the Bohr model is still used by chemists and spectroscopists.

EXAMPLE 37–13 **Wavelength of a Lyman line.** Use Fig. 37–26 to determine the wavelength of the first Lyman line, the transition from $n = 2$ to $n = 1$ in a hydrogen atom. In what region of the electromagnetic spectrum does this lie?

APPROACH We use Eq. 37–9, $hf = E_\text{u} - E_\ell$, with the energies obtained from Fig. 37–26 to find the energy and the wavelength of the transition. The region of the electromagnetic spectrum is found using the EM spectrum in Fig. 31–12.

SOLUTION For the transition from $n = 2$ to $n = 1$ in hydrogen, $hf = E_2 - E_1 = \{-3.4\,\text{eV} - (-13.6\,\text{eV})\} = 10.2\,\text{eV} = (10.2\,\text{eV})(1.60 \times 10^{-19}\,\text{J/eV}) = 1.63 \times 10^{-18}\,\text{J}$. Since $\lambda = c/f$, we have

$$\lambda = \frac{c}{f} = \frac{hc}{E_2 - E_1} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.00 \times 10^8\,\text{m/s})}{1.63 \times 10^{-18}\,\text{J}} = 1.22 \times 10^{-7}\,\text{m},$$

or 122 nm, which is in the UV region of the EM spectrum, Fig. 31–12. See also Fig. 37–22, where this value is confirmed experimentally.

NOTE An alternate approach: use Eq. 37–15 to find λ, and get the same result.

EXAMPLE 37–14 **Wavelength of a Balmer line.** Use the Bohr model to calculate the wavelength of light emitted when a hydrogen atom makes a transition from the $n = 6$ to the $n = 2$ energy level.

APPROACH We can use Eq. 37–15 or its equivalent, Eq. 37–8, with $R = 1.097 \times 10^7\,\text{m}^{-1}$. (We could also use Fig. 37–26.)

SOLUTION We find

$$\frac{1}{\lambda} = (1.097 \times 10^7\,\text{m}^{-1})\left(\frac{1}{4} - \frac{1}{36}\right) = 2.44 \times 10^6\,\text{m}^{-1}.$$

So $\lambda = 1/(2.44 \times 10^6\,\text{m}^{-1}) = 4.10 \times 10^{-7}\,\text{m}$ or 410 nm. This is the fourth line in the Balmer series, Fig. 37–21, and is violet in color.

Alternate Solution We can also use Fig. 37–26, but first we need Eq. 37–14b to give us $E_6 = 13.6\,\text{eV}/(6)^2 = 0.38\,\text{eV}$. So $hf = \Delta E = 3.40\,\text{eV} - 0.38\,\text{eV} = 3.02\,\text{eV}$. Then $\lambda = c/f = hc/hf = (6.626 \times 10^{-34}\,\text{J}\cdot\text{s})(3.00 \times 10^8\,\text{m/s})/(3.02\,\text{eV})(1.602 \times 10^{-19}\text{eV/J}) = 411\,\text{nm}$.

EXERCISE G The energy of the photon emitted when a hydrogen atom goes from the $n = 6$ state to the $n = 3$ state is (*a*) 0.378 eV; (*b*) 0.503 eV; (*c*) 1.13 eV; (*d*) 3.06 eV; (*e*) 13.6 eV.

EXAMPLE 37–15 Absorption wavelength. Use Fig. 37–26 to determine the maximum wavelength that hydrogen in its ground state can absorb. What would be the next smaller wavelength that would work?

APPROACH Maximum wavelength corresponds to minimum energy, and this would be the jump from the ground state up to the first excited state (Fig. 37–26). The next smaller wavelength occurs for the jump from the ground state to the second excited state. In each case, the energy difference can be used to find the wavelength.

SOLUTION The energy needed to jump from the ground state to the first excited state is $13.6\,\text{eV} - 3.4\,\text{eV} = 10.2\,\text{eV}$; the required wavelength, as we saw in Example 37–13, is 122 nm. The energy to jump from the ground state to the second excited state is $13.6\,\text{eV} - 1.5\,\text{eV} = 12.1\,\text{eV}$, which corresponds to a wavelength

$$\lambda = \frac{c}{f} = \frac{hc}{hf} = \frac{hc}{E_3 - E_1}$$
$$= \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.00 \times 10^{8}\,\text{m/s})}{(12.1\,\text{eV})(1.60 \times 10^{-19}\,\text{J/eV})} = 103\,\text{nm}.$$

EXAMPLE 37–16 He^+ ionization energy. (*a*) Use the Bohr model to determine the ionization energy of the He^+ ion, which has a single electron. (*b*) Also calculate the maximum wavelength a photon can have to cause ionization. The helium atom is the second atom, after hydrogen, in the Periodic Table (see inside rear cover).

APPROACH We want to determine the minimum energy required to lift the electron from its ground state and to barely reach the free state at $E = 0$. The ground state energy of He^+ is given by Eq. 37–14b with $n = 1$ and $Z = 2$.

SOLUTION (*a*) Since all the symbols in Eq. 37–14b are the same as for the calculation for hydrogen, except that Z is 2 instead of 1, we see that E_1 will be $Z^2 = 2^2 = 4$ times the E_1 for hydrogen:

$$E_1 = 4(-13.6\,\text{eV}) = -54.4\,\text{eV}.$$

Thus, to ionize the He^+ ion should require 54.4 eV, and this value agrees with experiment.

(*b*) The maximum wavelength photon that can cause ionization will have energy $hf = 54.4\,\text{eV}$ and wavelength

$$\lambda = \frac{c}{f} = \frac{hc}{hf} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.00 \times 10^{8}\,\text{m/s})}{(54.4\,\text{eV})(1.60 \times 10^{-19}\,\text{J/eV})} = 22.8\,\text{nm}.$$

If $\lambda > 22.8\,\text{nm}$, ionization can not occur.

NOTE If the atom absorbed a photon of greater energy (wavelength shorter than 22.8 nm), the atom could be ionized and the freed electron would have kinetic energy of its own.

In Example 37–16, we saw that E_1 for the He^+ ion is four times more negative than that for hydrogen. Indeed, the energy-level diagram for He^+ looks just like that for hydrogen, Fig. 37–26, except that the numerical values for each energy level are four times larger. Note, however, that we are talking here about the He^+ *ion*. Normal (neutral) helium has two electrons and its energy level diagram is entirely different.

CONCEPTUAL EXAMPLE 37–17 **Hydrogen at 20°C.** (*a*) Estimate the average kinetic energy of whole hydrogen atoms (not just the electrons) at room temperature. (*b*) Use the result to explain why, at room temperature, very few H atoms are in excited states and nearly all are in the ground state, and hence emit no light.

RESPONSE According to kinetic theory (Chapter 18), the average kinetic energy of atoms or molecules in a gas is given by Eq. 18–4:

$$\overline{K} = \tfrac{3}{2}kT,$$

where $k = 1.38 \times 10^{-23}\,\text{J/K}$ is Boltzmann's constant, and T is the kelvin (absolute) temperature. Room temperature is about $T = 300\,\text{K}$, so

$$\overline{K} = \tfrac{3}{2}(1.38 \times 10^{-23}\,\text{J/K})(300\,\text{K}) = 6.2 \times 10^{-21}\,\text{J},$$

or, in electron volts:

$$\overline{K} = \frac{6.2 \times 10^{-21}\,\text{J}}{1.6 \times 10^{-19}\,\text{J/eV}} = 0.04\,\text{eV}.$$

The average kinetic energy of an atom as a whole is thus very small compared to the energy between the ground state and the next higher energy state ($13.6\,\text{eV} - 3.4\,\text{eV} = 10.2\,\text{eV}$). Any atoms in excited states quickly fall to the ground state and emit light. Once in the ground state, collisions with other atoms can transfer energy of only 0.04 eV on the average. A small fraction of atoms can have much more energy (see Section 18–2 on the distribution of molecular speeds), but even a kinetic energy that is 10 times the average is not nearly enough to excite atoms into states above the ground state. Thus, at room temperature, practically all atoms are in the ground state. Atoms can be excited to upper states by very high temperatures, or by applying a high voltage so a current of high-energy electrons passes through the gas as in a discharge tube (Fig. 37–19).

Correspondence Principle

We should note that Bohr made some radical assumptions that were at variance with classical ideas. He assumed that electrons in fixed orbits do not radiate light even though they are accelerating (moving in a circle), and he assumed that angular momentum is quantized. Furthermore, he was not able to say how an electron moved when it made a transition from one energy level to another. On the other hand, there is no real reason to expect that in the tiny world of the atom electrons would behave as ordinary-sized objects do. Nonetheless, he felt that where quantum theory overlaps with the macroscopic world, it should predict classical results. This is the **correspondence principle**, already mentioned in regard to relativity (Section 36–13). This principle does work for Bohr's theory of the hydrogen atom. The orbit sizes and energies are quite different for $n = 1$ and $n = 2$, say. But orbits with $n = 100{,}000{,}000$ and 100,000,001 would be very close in radius and energy. Indeed, transitions between such large orbits (which would approach macroscopic sizes) would be imperceptible. Such orbits would thus appear to be continuously spaced, which is what we expect in the everyday world. [Don't confuse the correspondence principle with the principle of complementarity, page 1106.]

Finally, it must be emphasized that the well-defined orbits of the Bohr model do not actually exist. The Bohr model is only a model, not reality. The idea of electron orbits was rejected a few years later, and today electrons are thought of (Chapter 39) as forming "probability clouds."

37–12 de Broglie's Hypothesis Applied to Atoms

Bohr's theory of 1913 was largely of an *ad hoc* nature. Assumptions were made so that theory would agree with experiment. But Bohr could give no reason why the orbits were quantized, nor why there should be a stable ground state. Finally, ten years later, a reason was proposed by Louis de Broglie. We saw in Section 37–7 that in 1923, de Broglie proposed that material particles, such as electrons, have a wave nature; and that this hypothesis was confirmed by experiment several years later.

One of de Broglie's original arguments in favor of the wave nature of electrons was that it provided an explanation for Bohr's theory of the hydrogen atom. According to de Broglie, a particle of mass m moving with a nonrelativistic speed v would have a wavelength (Eq. 37–7) of

$$\lambda = \frac{h}{mv}.$$

Each electron orbit in an atom, he proposed, is actually a standing wave. As we saw in Chapter 15, when a violin or guitar string is plucked, a vast number of wavelengths are excited. But only certain ones—those that have nodes at the ends—are sustained (Fig. 15–26). These are the *resonant* modes of the string. Waves with other wavelengths interfere with themselves upon reflection and their amplitudes quickly drop to zero. With electrons moving in circles, according to Bohr's theory, de Broglie argued that the electron wave was a *circular* standing wave that closes on itself, Fig. 37–27. If the wavelength of a wave does not close on itself, as in Fig. 37–28, destructive interference takes place as the wave travels repeatedly around the loop, and the wave quickly dies out. Thus, the only waves that persist are those for which the circumference of the circular orbit contains a whole number of wavelengths, Fig. 37–29. The circumference of a Bohr orbit of radius r_n is $2\pi r_n$, so to have constructive interference, we need

$$2\pi r_n = n\lambda, \qquad n = 1, 2, 3, \cdots.$$

When we substitute $\lambda = h/mv$, we get $2\pi r_n = nh/mv$, or

$$mvr_n = \frac{nh}{2\pi}.$$

This is just the *quantum condition* proposed by Bohr on an *ad hoc* basis, Eq. 37–10. It is from this equation that the discrete orbits and energy levels were derived. Thus we have a first explanation for the quantized orbits and energy states in the Bohr model: they are due to the wave nature of the electron, and only resonant "standing" waves can persist.[†] This implies that the *wave–particle duality* is at the root of atomic structure.

In viewing the circular electron waves of Fig. 37–29, the electron is not to be thought of as following the oscillating wave pattern. In the Bohr model of hydrogen, the electron moves in a circle. The circular wave, on the other hand, represents the *amplitude* of the electron "matter wave," and in Fig. 37–29 the wave amplitude is shown superimposed on the circular path of the particle orbit for convenience.

Bohr's theory worked well for hydrogen and for one-electron ions. But it did not prove successful for multi-electron atoms. Bohr's theory could not predict line spectra even for the next simplest atom, helium. It could not explain why some emission lines are brighter than others, nor why some lines are split into two or more closely spaced lines ("fine structure"). A new theory was needed and was indeed developed in the 1920s. This new and radical theory is called *quantum mechanics*. It finally solved the problem of atomic structure, and it gives us a very different view of the atom: the idea of electrons in well-defined orbits was replaced with the idea of electron "clouds." This new theory of quantum mechanics has given us a wholly different view of the basic mechanisms underlying physical processes.

[†]We note, however, that Eq. 37–10 is no longer considered valid, as discussed in Chapter 39.

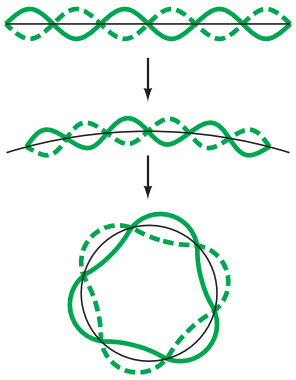


FIGURE 37–27 An ordinary standing wave compared to a circular standing wave.

FIGURE 37–28 When a wave does not close (and hence interferes destructively with itself), it rapidly dies out.

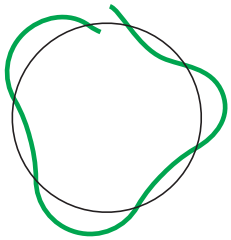
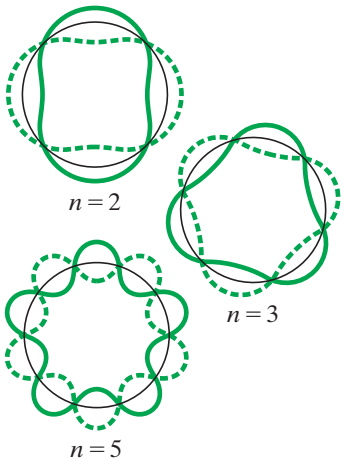


FIGURE 37–29 Standing circular waves for two, three, and five wavelengths on the circumference; n, the number of wavelengths, is also the quantum number.



Summary

Quantum theory has its origins in **Planck's quantum hypothesis**, that the energy contained in molecular oscillations is **quantized**: the energy E can only be integer (n) multiples of hf, where h is Planck's constant and f is the natural frequency of oscillation:

$$E = nhf. \quad \textbf{(37–2)}$$

This hypothesis explained the spectrum of radiation emitted by a **blackbody** at high temperature.

Einstein proposed that for some experiments, light could be pictured as being emitted and absorbed as **quanta** (particles), which we now call **photons**, each with energy

$$E = hf \quad \textbf{(37–3)}$$

and momentum

$$p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda}. \quad \textbf{(37–5)}$$

He proposed the photoelectric effect as a test for the photon theory of light. In the **photoelectric effect**, the photon theory says that each incident photon can strike an electron in a material and eject it if the photon has sufficient energy. The maximum energy of ejected electrons is then linearly related to the frequency of the incident light.

The photon theory is also supported by the **Compton effect** and the observation of electron–positron **pair production**.

The **wave–particle duality** refers to the idea that light and matter (such as electrons) have both wave and particle properties. The wavelength of a material particle is given by

$$\lambda = \frac{h}{p}, \quad \textbf{(37–7)}$$

where p is the momentum of the object ($p = mv$ for a particle of mass m and speed v).

The **principle of complementarity** states that we must be aware of both the particle and wave properties of light and of matter for a complete understanding of them.

Electron microscopes (EM) make use of the wave properties of electons to form an image: their "lenses" are magnetic. Various types of EM exist: some can magnify 100,000× (1000× better than a light mciroscope); others can give a 3-D image.

Early models of the atom include the plum-pudding model, and Rutherford's planetary (or nuclear) model of an atom which consists of a tiny but massive positively charged nucleus surrounded (at a relatively great distance) by electrons.

To explain the **line spectra** emitted by atoms, as well as the stability of atoms, **Bohr's theory** postulated that: (1) electrons bound in an atom can only occupy orbits for which the angular momentum is quantized, which results in discrete values for the radius and energy; (2) an electron in such a **stationary state** emits no radiation; (3) if an electron jumps to a lower state, it emits a photon whose energy equals the difference in energy between the two states; (4) the angular momentum L of atomic electrons is quantized by the rule

$$L = \frac{nh}{2\pi}, \quad \textbf{(37–10)}$$

where n is an integer called the **principal quantum number**. The $n = 1$ state is the **ground state**, which in hydrogen has an energy $E_1 = -13.6\,\text{eV}$. Higher values of n correspond to **excited states**, and their energies are

$$E_n = -(13.6\,\text{eV})\frac{Z^2}{n^2}, \quad \textbf{(37–14b)}$$

where Ze is the electric charge on the nucleus.

Atoms are excited to these higher states by collisions with other atoms or electrons, or by absorption of a photon of just the right frequency.

De Broglie's hypothesis that electrons (and other matter) have a wavelength $\lambda = h/mv$ gave an explanation for Bohr's quantized orbits by bringing in the wave–particle duality: the orbits correspond to circular standing waves in which the circumference of the orbit equals a whole number (n) of wavelengths.

Questions

1. What can be said about the relative temperatures of whitish-yellow, reddish, and bluish stars? Explain.
2. If energy is radiated by all objects, why can we not see most of them in the dark?
3. Does a lightbulb at a temperature of 2500 K produce as white a light as the Sun at 6000 K? Explain.
4. Explain why the existence of a cutoff frequency in the photoelectric effect more strongly favors a particle theory rather than a wave theory of light.
5. If the threshold wavelength in the photoelectric effect increases when the emitting metal is changed to a different metal, what can you say about the work functions of the two metals?
6. The work functions for sodium and cesium are 2.28 eV and 2.14 eV, respectively. For incident photons of a given frequency, which metal will give a higher maximum kinetic energy for the electrons?
7. UV light causes sunburn, whereas visible light does not. Suggest a reason.
8. Light of 450-nm wavelength strikes a metal surface, and a stream of electrons emerges from the metal. If light of the same intensity but of wavelength 400 nm strikes the surface, are more electrons emitted? Does the energy of the emitted electrons change? Explain.
9. (*a*) Does a beam of infrared photons always have less energy than a beam of ultraviolet photons? Explain. (*b*) Does a single photon of infrared light always have less energy than a single photon of ultraviolet light?
10. If an X-ray photon is scattered by an electron at rest, does the photon's wavelength change? If so, does it increase or decrease?
11. In both the photoelectric effect and in the Compton effect, a photon collides with an electron causing the electron to fly off. What then, is the difference between the two processes?
12. Consider a point source of light. How would the intensity of light vary with distance from the source according to (*a*) wave theory, (*b*) particle (photon) theory? Would this help to distinguish the two theories?

13. If an electron and a proton travel at the same speed, which has the shorter de Broglie wavelength? Explain.

14. Why do we say that light has wave properties? Why do we say that light has particle properties?

15. Why do we say that electrons have wave properties? Why do we say that electrons have particle properties?

16. What are the differences between a photon and an electron? Be specific: make a list.

17. In Rutherford's planetary model of the atom, what keeps the electrons from flying off into space?

18. How can you tell if there is oxygen near the surface of the Sun?

19. When a wide spectrum of light passes through hydrogen gas at room temperature, absorption lines are observed that correspond only to the Lyman series. Why don't we observe the other series?

20. Explain how the closely spaced energy levels for hydrogen near the top of Fig. 37–26 correspond to the closely spaced spectral lines at the top of Fig. 37–21.

21. (*a*) List at least three successes of the Bohr model of the atom, according to Section 37–11. (*b*) List at least two observations that the Bohr model could not explain, according to Section 37–12.

22. Is it possible for the de Broglie wavelength of a "particle" to be greater than the dimensions of the particle? To be smaller? Is there any direct connection? Explain.

23. How can the spectrum of hydrogen contain so many lines when hydrogen contains only one electron?

24. An electron and a proton are accelerated through the same voltage. Which has the longer wavelength? Explain why.

25. The Lyman series is brighter than the Balmer series, because this series of transitions ends up in the most common state for hydrogen, the ground state. Why then was the Balmer series discovered first?

26. Use conservation of momentum to explain why photons emitted by hydrogen atoms have slightly less energy than that predicted by Eq. 37–9.

27. State if a continuous or a line spectrum is produced by each of the following: (*a*) a hot solid object; (*b*) an excited, rarefied gas; (*c*) a hot liquid; (*d*) light from a hot solid that passes through a cooler rarefied gas; (*e*) a hot dense gas. For each, if a line spectrum is produced, is it an emission or an absorption spectrum?

28. Compare the emission spectrum for hydrogen at very high temperature (when some of the atoms are in excited states) with the absorption spectrum of hydrogen at room temperature (when all atoms are in the ground state). Will the two spectra contain identical lines? Explain.

29. In a helium atom, which contains two electrons, do you think that on average the electrons are closer to the nucleus or farther away than in a hydrogen atom? Why?

30. Does each of the following support the wave nature or the particle nature of light? (*a*) The existence of the cutoff frequency in the photoelectric effect; (*b*) Young's double-slit experiment; (*c*) the shift in the photon frequency in Compton scattering; (*d*) the diffraction of light.

MisConceptual Questions

1. Which of the following statements is true regarding how blackbody radiation changes as the temperature of the radiating object increases?
(*a*) Both the maximum intensity and the peak wavelength increase.
(*b*) The maximum intensity increases, and the peak wavelength decreases.
(*c*) Both the maximum intensity and the peak wavelength decrease.
(*d*) The maximum intensity decreases, and the peak wavelength increases.

2. A beam of red light and a beam of blue light have equal intensities. Which statement is true?
(*a*) There are more photons in the blue beam.
(*b*) There are more photons in the red beam.
(*c*) Both beams contain the same number of photons.
(*d*) The number of photons is not related to intensity.

3. Which of the following is necessarily true?
(*a*) Red light has more energy than violet light.
(*b*) Violet light has more energy than red light.
(*c*) A single photon of red light has more energy than a single photon of violet light.
(*d*) A single photon of violet light has more energy than a single photon of red light.
(*e*) None of the above.
(*f*) A combination of the above (specify).

4. If a photon of energy E ejects electrons from a metal with kinetic energy K, then a photon with energy $E/2$
(*a*) will eject electrons with kinetic energy $K/2$.
(*b*) will eject electrons with an energy greater than $K/2$.
(*c*) will eject electrons with an energy less than $K/2$.
(*d*) might not eject any electrons.

5. Monochromatic light is incident on a metal surface, and electrons are ejected. If the intensity of that light increases, what will happen to the maximum energy of the electrons?
(*a*) The maximum energy will increase.
(*b*) The maximum energy will decrease.
(*c*) The maximum energy will remain the same.
(*d*) Changes in the maximum energy cannot be determined without additional information.

6. Monochromatic light is incident on a metal surface, but no photoelectrons are emitted. If electrons are to be ejected from the surface, how should the incident light be adjusted?
(*a*) Light of a shorter wavelength should be used.
(*b*) The intensity of the light should be increased.
(*c*) Light of a longer wavelength should be used.
(*d*) The intensity of the light should be decreased.

7. If the momentum of an electron is doubled, how would its wavelength change?
(*a*) No change.
(*b*) It would be halved.
(*c*) It would double.
(*d*) It would be quadrupled.
(*e*) It would be reduced to one-fourth.

8. Which of the following can be thought of as either a wave or a particle?
(*a*) Light.
(*b*) An electron.
(*c*) A proton.
(*d*) All of the above.

9. An electron
(*a*) is a particle.
(*b*) is a wave.
(*c*) is both a particle and a wave.
(*d*) exhibits particle properties, or wave properties, depending on the experiment being performed.

10. When you throw a baseball, its de Broglie wavelength is
(*a*) the same size as the ball.
(*b*) about the same size as an atom $(10^{-10}\,\text{m})$.
(*c*) about the same size as an atom's nucleus $(10^{-15}\,\text{m})$.
(*d*) much smaller than the size of an atom's nucleus.

11. Electrons and photons of light are similar in that
(*a*) both have momentum given by h/λ.
(*b*) both exhibit wave–particle duality.
(*c*) both are used in diffraction experiments to explore structure.
(*d*) All of the above.
(*e*) None of the above.

12. In Rutherford's famous set of experiments described in Section 37–9, the fact that some alpha particles were deflected at large angles indicated that
(*a*) the nucleus was positive.
(*b*) charge was quantized.
(*c*) the nucleus was concentrated in a small region of space.
(*d*) most of the atom is empty space.
(*e*) None of the above.

13. Which of the following is the currently accepted model of the atom?
(*a*) The plum-pudding model.
(*b*) The Rutherford atom.
(*c*) The Bohr atom.
(*d*) None of the above.

14. Which of the following electron transitions between two energy states (n) in the hydrogen atom corresponds to the emission of a photon with the longest wavelength?
(*a*) $2 \rightarrow 5$.
(*b*) $5 \rightarrow 2$.
(*c*) $5 \rightarrow 8$.
(*d*) $8 \rightarrow 5$.

15. If we set the potential energy of an electron and a proton to be zero when they are an infinite distance apart, then the lowest energy a bound electron in a hydrogen atom can have is
(*a*) 0.
(*b*) −13.6 eV.
(*c*) any possible value.
(*d*) any value between −13.6 eV and 0.

Problems

37–1 Blackbodies; Planck's Quantum Hypothesis

1. (I) What is the peak wavelength for radiation from (*a*) ice at 273 K, (*b*) a floodlamp at 3500 K, (*c*) helium at 4.2 K, (*d*) for the universe at $T = 2.725$ K, assuming blackbody emission? In what region of the EM spectrum is each?

2. (I) How hot is metal being welded if it radiates most strongly at 490 nm?

3. (I) An HCl molecule vibrates with a natural frequency of 8.1×10^{13} Hz. What is the difference in energy (in joules and electron volts) between successive values of the oscillation energy?

4. (I) (*a*) What is the temperature if the peak of a blackbody spectrum is at 16.0 nm? (*b*) What is the wavelength at the peak of a blackbody spectrum if the body is at a temperature of 2200 K?

5. (II) The steps of a flight of stairs are 20.0 cm high (vertically). If a 68.0-kg person stands with both feet on the same step, what is the gravitational potential energy of this person, relative to the ground, on (*a*) the first step, (*b*) the second step, (*c*) the third step, (*d*) the n^{th} step? (*e*) What is the change in energy as the person descends from step 6 to step 2?

6. (II) Estimate the peak wavelength of light issuing from the pupil of the human eye (which approximates a blackbody) assuming normal body temperature.

7. (III) Planck's radiation law is given by:

$$I(\lambda, T) = \frac{2\pi hc^2\lambda^{-5}}{e^{hc/\lambda kT} - 1}$$

where $I(\lambda, T)$ is the rate energy is radiated per unit surface area per unit wavelength interval at wavelength λ and Kelvin temperature T. (*a*) Show that Wien's displacement law follows from this relationship. (*b*) Determine the value of h from the experimental value of $\lambda_P T$ given in the text. [You may want to use graphing techniques.] (*c*) Derive the T^4 dependence of the rate at which energy is radiated (as in the Stefan-Boltzmann law, Eq. 19–18), by integrating Planck's formula over all wavelengths; that is, show that

$$\int I(\lambda, T)\, d\lambda \propto T^4.$$

37–2 and 37–3 Photons and the Photoelectric Effect

8. (I) What is the energy of photons (in joules) emitted by an 88.5-MHz FM radio station?

9. (I) What is the energy range (in joules and eV) of photons in the visible spectrum, of wavelength 410 nm to 750 nm?

10. (I) What minimum frequency of light is needed to eject electrons from a metal whose work function is 5.2×10^{-19} J?

11. (I) What is the longest wavelength of light that will emit electrons from a metal whose work function is 3.55 eV?

12. (I) A gamma ray emitted from a nucleus during radioactive decay has an energy of 420 keV. What is its wavelength? Would we expect significant diffraction of this type of light when it passes through an everyday opening, such as a door?

13. (II) About 0.1 eV is required to break a "hydrogen bond" in a protein molecule. Calculate the minimum frequency and maximum wavelength of a photon that can accomplish this.

14. (II) What wavelength photon would have the same energy as a 145-gram baseball moving 27.0 m/s?

15. (II) The human eye can respond to as little as 10^{-18} J of light energy. For a wavelength at the peak of visual sensitivity, 550 nm, how many photons lead to an observable flash?

16. (II) The work functions for sodium, cesium, copper, and iron are 2.3, 2.1, 4.7, and 4.5 eV, respectively. Which of these metals will not emit electrons when visible light shines on it?

17. (II) In a photoelectric-effect experiment it is observed that no current flows unless the wavelength is less than 540 nm. (*a*) What is the work function of this material? (*b*) What is the stopping voltage required if light of wavelength 480 nm is used?

18. (II) What is the maximum kinetic energy of electrons ejected from barium $(W_0 = 2.48\,\text{eV})$ when illuminated by white light, $\lambda = 410$ to 750 nm?

19. (II) When UV light of wavelength 285 nm falls on a metal surface, the maximum kinetic energy of emitted electrons is 1.55 eV. What is the work function of the metal?

20. (II) The threshold wavelength for emission of electrons from a given surface is 330 nm. What will be the maximum kinetic energy of ejected electrons when the wavelength is changed to (*a*) 290 nm, (*b*) 370 nm?

21. (II) When 230-nm light falls on a metal, the current through a photoelectric circuit (Fig. 37–4) is brought to zero at a stopping voltage of 1.74 V. What is the work function of the metal?

22. (II) A certain type of film is sensitive only to light whose wavelength is less than 670 nm. What is the energy (eV and kcal/mol) needed for the chemical reaction to occur which causes the film to change?

23. (II) The range of visible light wavelengths extends from about 410 nm to 700 nm. (*a*) Estimate the minimum energy (eV) required for the retina to respond and initiate vision. (*b*) Speculate as to why, at the other end of the visible range, there is a threshold photon energy beyond which the eye registers no sensation of sight. Determine this threshold photon energy (eV).

24. (II) At the atomic scale, the electron volt and nanometer are well-suited units for energy and distance, respectively. (*a*) Show that the energy E (in electron volts) of a photon whose wavelength is λ (nm) is given by

$$E = \frac{1.240 \times 10^3\,\text{eV}\cdot\text{nm}}{\lambda\,(\text{nm})}.$$

Use at least four significant figures for values of h, c, e (see inside front cover). (*b*) How much energy (eV) does a 610-nm photon have?

25. (II) A **photomultiplier tube** (a very sensitive light sensor), is based on the photoelectric effect: incident photons strike a metal surface and the resulting ejected electrons are collected. By counting the number of collected electrons, the number of incident photons (i.e., the incident light intensity) can be determined. (*a*) If a photomultiplier tube is to respond properly for incident wavelengths throughout the visible range (410 nm to 750 nm), what is the maximum value for the work function W_0 (eV) of its metal surface? (*b*) If W_0 for its metal surface is above a certain threshold value, the photomultiplier will only function for incident ultraviolet wavelengths and be unresponsive to visible light. Determine this threshold value (eV).

26. (III) A group of atoms is confined to a very small (point-like) volume in a laser-based **atom trap**. The incident laser light causes each atom to emit 1.3×10^6 photons of wavelength 780 nm every second. A sensor of area 1.0 cm^2 measures the power emanating from the trap to be 1.6 nW when placed 25 cm away from the trapped atoms. Assuming each atom emits photons with equal probability in all directions, determine the number of trapped atoms.

27. (III) Assume light of wavelength λ is incident on a metal surface, whose work function is known precisely (i.e., its uncertainty is better than 0.1% and can be ignored). Show that if the stopping voltage can be determined to an accuracy of ΔV_0, the fractional uncertainty (magnitude) in wavelength is

$$\frac{\Delta\lambda}{\lambda} = \frac{\lambda e}{hc}\Delta V_0.$$

Determine this fractional uncertainty if $\Delta V_0 = 0.01$ V and $\lambda = 550$ nm.

37–4 Compton Effect

28. (I) A high-frequency photon is scattered off of an electron and experiences a change of wavelength of 1.2×10^{-4} nm. At what angle must a detector be placed to detect the scattered photon (relative to the direction of the incoming photon)?

29. (II) Determine the Compton wavelength for (*a*) an electron, (*b*) a proton. (*c*) Show that if a photon has wavelength equal to the Compton wavelength of a particle, the photon's energy is equal to the rest energy of the particle.

30. (II) X-rays of wavelength $\lambda = 0.135$ nm are scattered from carbon. What is the expected Compton wavelength shift for photons detected at angles (relative to the incident beam) of exactly (*a*) 60°, (*b*) 90°, (*c*) 180°?

31. (II) In the Compton effect, determine the ratio $(\Delta\lambda/\lambda)$ of the maximum change $\Delta\lambda$ in a photon's wavelength to the photon's initial wavelength λ, if the photon is (*a*) a visible-light photon with $\lambda = 550$ nm, (*b*) an X-ray photon with $\lambda = 0.10$ nm.

32. (II) A 1.0-MeV gamma-ray photon undergoes a sequence of Compton-scattering events. If the photon is scattered at an angle of 0.50° in each event, estimate the number of events required to convert the photon into a visible-light photon with wavelength 555 nm. You can use an expansion for small θ; see Appendix A. [Gamma rays created near the center of the Sun are transformed to visible wavelengths as they travel to the Sun's surface through a sequence of small-angle Compton scattering events.]

33. (III) In the Compton effect, a 0.140-nm photon strikes a free electron in a head-on collision and knocks it into the forward direction. The rebounding photon recoils directly backward. Use conservation of (relativistic) energy and momentum to determine (*a*) the kinetic energy of the electron, and (*b*) the wavelength of the recoiling photon. Use Eq. 37–5, but not Eqs. 37–6.

34. (III) In the Compton effect (see Fig. 37–7), use the relativistic equations for conservation of energy and of linear momentum to derive, showing all details, the Compton shift in wavelength as given by Eqs. 37–6.

37–5 Pair Production

35. (I) How much total kinetic energy will an electron–positron pair have if produced by a 2.85-MeV photon?

36. (II) What is the longest wavelength photon that could produce a proton–antiproton pair? (Each has a mass of 1.67×10^{-27} kg.)

37. (II) What is the minimum photon energy needed to produce a $\mu^+ - \mu^-$ pair? The mass of each μ (muon) is 207 times the mass of an electron. What is the wavelength of such a photon?

38. (II) An electron and a positron, each moving at 2.5×10^5 m/s, collide head on, disappear, and produce two photons moving in opposite directions, each with the same energy and momentum. Determine the energy and momentum of each photon.

39. (II) A gamma-ray photon produces an electron and a positron, each with a kinetic energy of 345 keV. Determine the energy and wavelength of the photon.

37–7 Wave Nature of Matter

40. (I) Calculate the wavelength of a 0.21-kg ball traveling at 0.10 m/s.

41. (I) What is the wavelength of a neutron $(m = 1.67\times10^{-27}\,\text{kg})$ traveling at 1.6×10^5 m/s?

42. (II) The speed of an electron in a particle accelerator is 0.95*c*. Find its de Broglie wavelength. (Use relativistic momentum.)

43. (II) Calculate the ratio of the kinetic energy of an electron to that of a proton if their wavelengths are equal. Assume that the speeds are nonrelativistic.

44. (II) A Ferrari with a mass of 1400 kg approaches a freeway underpass that is 12 m across. At what speed must the car be moving, in order for it to have a wavelength such that it might somehow "diffract" after passing through this "single slit"? How do these conditions compare to normal freeway speeds of 30 m/s?

45. (II) Neutrons can be used in diffraction experiments to probe the lattice structure of crystalline solids. Since the neutron's wavelength needs to be on the order of the spacing between atoms in the lattice, about 0.3 nm, what should the speed of the neutrons be?

46. (II) An electron has a de Broglie wavelength $\lambda = 7.0\times10^{-10}$ m. (*a*) What is its momentum? (*b*) What is its speed? (*c*) What voltage was needed to accelerate it to this speed?

47. (II) What is the wavelength of an electron of energy (*a*) 20 eV, (*b*) 200 eV, (*c*) 2.0 keV?

48. (II) Show that if an electron and a proton have the same nonrelativistic kinetic energy, the proton has the shorter wavelength.

49. (II) After passing through two slits separated by a distance of 3.0 μm, a beam of electrons creates an interference pattern with its second-order maximum at an angle of 45°. Find the speed of the electrons in this beam.

37–8 Electron Microscope

50. (II) What voltage is needed to produce electron wavelengths of 0.33 nm? (Assume that the electrons are nonrelativistic.)

51. (II) Electrons are accelerated by 6450 V in an electron microscope. Estimate the maximum possible resolution of the microscope.

37–10 and 37–11 Spectra and the Bohr Model

52. (I) For the three hydrogen transitions indicated below, with n being the initial state and n' being the final state, is the transition an absorption or an emission? Which is higher, the initial state energy or the final state energy of the atom? Finally, which of these transitions involves the largest energy photon? (*a*) $n = 1$, $n' = 3$; (*b*) $n = 6$, $n' = 2$; (*c*) $n = 4$, $n' = 5$.

53. (I) How much energy is needed to ionize a hydrogen atom in the $n = 3$ state?

54. (I) (*a*) Determine the wavelength of the second Balmer line ($n = 4$ to $n = 2$ transition) using Fig. 37–26. Determine likewise (*b*) the wavelength of the third Lyman line.

55. (I) Calculate the ionization energy of doubly ionized lithium, Li^{2+}, which has $Z = 3$.

56. (I) Evaluate the Rydberg constant R using the Bohr model (compare Eqs. 37–8 and 37–15) and show that its value is $R = 1.0974\times10^7\,\text{m}^{-1}$. (Use values inside front cover to 5 or 6 significant figures.)

57. (II) What is the longest wavelength light capable of ionizing a hydrogen atom in the ground state?

58. (II) In the Sun, an ionized helium (He^+) atom makes a transition from the $n = 6$ state to the $n = 2$ state, emitting a photon. Can that photon be absorbed by hydrogen atoms present in the Sun? If so, between what energy states will the hydrogen atom transition occur?

59. (II) What wavelength photon would be required to ionize a hydrogen atom in the ground state and give the ejected electron a kinetic energy of 17.5 eV?

60. (II) For what maximum kinetic energy is a collision between an electron and a hydrogen atom in its ground state definitely elastic?

61. (II) Construct the energy-level diagram for the He^+ ion (like Fig. 37–26).

62. (II) Determine the electrostatic potential energy and the kinetic energy of an electron in the ground state of the hydrogen atom, according to the Bohr model.

63. (II) An excited hydrogen atom could, in principle, have a diameter of 0.10 mm. What would be the approximate value of n for a Bohr orbit of this size? What would its energy be?

64. (II) Is the use of nonrelativistic formulas justified in the Bohr atom? To check, calculate the electron's velocity, v, in terms of c, for the ground state of hydrogen, and then calculate $\sqrt{1 - v^2/c^2}$.

65. (II) Assume hydrogen atoms in a gas are all initially in their ground state. If free electrons with kinetic energy 12.75 eV collide with these atoms, what photon wavelengths will be emitted by the gas?

66. (II) Suppose an electron was bound to a proton, as in the hydrogen atom, but by the gravitational force rather than by the electric force. What would be the radius, and energy, of the first Bohr orbit?

67. (II) *Correspondence principle*: Show that for large values of n, the difference in radius Δr between two adjacent orbits (with quantum numbers n and $n - 1$) is given by

$$\Delta r = r_n - r_{n-1} \approx \frac{2r_n}{n},$$

so $\Delta r/r_n \rightarrow 0$ as $n \rightarrow \infty$ in accordance with the correspondence principle. [Note that we can check the correspondence principle by either considering large values of n $(n \rightarrow \infty)$ or by letting $h \rightarrow 0$. Are these equivalent?]

General Problems

68. If a lightbulb emits 4.5 W as visible light (average wavelength 550 nm) uniformly in all directions, estimate how many photons per second of visible light will strike the pupil (4.0 mm diameter) of the eye of an observer (*a*) 1.0 m away, (*b*) 1.0 km away.

69. At low temperatures, nearly all the atoms in hydrogen gas will be in the ground state. What minimum frequency photon is needed if the photoelectric effect is to be observed?

70. A beam of 125-eV electrons is scattered from a crystal, as in X-ray diffraction, and a first-order peak is observed at $\theta = 43°$. What is the spacing between planes in the diffracting crystal? (See Section 35–11.)

71. A microwave oven produces electromagnetic radiation at $\lambda = 12.2$ cm and produces a power of 830 W. Calculate the number of microwave photons produced by the microwave oven each second.

72. Sunlight reaching the Earth's atmosphere has an intensity of about 1350 W/m^2. Estimate how many photons per square meter per second this represents. Take the average wavelength to be 550 nm.

73. Estimate the number of photons emitted by the Sun in a year. (Take the average wavelength to be 550 nm and the intensity of sunlight reaching the Earth (outer atmosphere) as 1350 W/m^2.)

74. The Big Bang theory (Chapter 44) states that the beginning of the universe was accompanied by a huge burst of photons. Those photons are still present today and make up the so-called cosmic *microwave background radiation*. The universe radiates like a blackbody with a temperature today of about 2.7 K. Calculate the peak wavelength of this radiation.

75. In some of Rutherford's experiments (Fig. 37–17) the α particles $(\text{mass} = 6.64 \times 10^{-27}\,\text{kg})$ had a kinetic energy of 4.8 MeV. How close could they get to the surface of a gold nucleus $(\text{radius} \approx 7 \times 10^{-15}\,\text{m}, Z = 79)$? Ignore the recoil motion of the nucleus.

76. Show that the magnitude of the electrostatic potential energy of an electron in any Bohr orbit of a hydrogen atom is twice the magnitude of its kinetic energy in that orbit.

77. Calculate the ratio of the gravitational force to the electric force for the electron in a hydrogen atom. Can the gravitational force be safely ignored?

78. In an X-ray tube (see Fig. 35–26 and discussion in Section 35–11), the high voltage between filament and target is V. After being accelerated through this voltage, an electron strikes the target where it is decelerated (by positively charged nuclei) and in the process one or more X-ray photons are emitted. (*a*) Show that the photon of shortest wavelength will have

$$\lambda_0 = \frac{hc}{eV}.$$

(*b*) What is the shortest wavelength of X-ray emitted when accelerated electrons strike a target at 31-kV?

79. Visible light incident on a diffraction grating with slit spacing of 0.010 mm has the first maximum at an angle of 3.5° from the central peak. If electrons could be diffracted by the same grating, what electron velocity would produce the same diffraction pattern as the visible light?

80. (*a*) Suppose an unknown element has an absorption spectrum with lines corresponding to 2.5, 4.7, and 5.1 eV above its ground state, and an ionization energy of 11.5 eV. Draw an energy level diagram for this element. (*b*) If a 5.1-eV photon is absorbed by an atom of this substance, in which state was the atom before absorbing the photon? What will be the energies of the photons that can subsequently be emitted by this atom?

81. Light of wavelength 408 nm falls on a metal which has a work function of 2.28 eV. (*a*) How much voltage should be applied to bring the current to zero? (*b*) What is the maximum speed of the emitted electrons? (*c*) What is the de Broglie wavelength of these electrons?

82. (*a*) Apply Bohr's assumptions to the Earth–Moon system to calculate the allowed energies and radii of motion. (*b*) Given the known distance between Earth and the Moon, is the quantization of the energy and radius apparent?

83. Show that the wavelength of a particle of mass m with kinetic energy K is given by the relativistic formula $\lambda = hc/\sqrt{K^2 + 2mc^2K}$.

84. What would be the theoretical limit of resolution for an electron microscope whose electrons are accelerated through 78 kV? (Relativistic formulas should be used.)

85. A flashlight emits 2.8 W of light. As the light leaves the flashlight in one direction, a reaction force is exerted on the flashlight in the opposite direction. Estimate the size of this reaction force.

*86. Three fundamental constants of nature—the gravitational constant G, Planck's constant h, and the speed of light c—have the **dimensions** of $[L^3/MT^2]$, $[ML^2/T]$, and $[L/T]$, respectively. (*a*) Find the mathematical combination of these fundamental constants that has the dimension of time. This combination is called the **Planck time** t_P and is thought to be the earliest time, after the creation of the universe, at which the currently known laws of physics can be applied. (*b*) Determine the numerical value of t_P. (*c*) Find the mathematical combination of these fundamental constants that has the dimension of length. This combination is called the **Planck length** λ_P and is thought to be the smallest length over which the currently known laws of physics can be applied. (*d*) Determine the numerical value of λ_P.

87. Imagine a free particle of mass m bouncing back and forth between two perfectly reflecting walls, separated by distance ℓ. Imagine that the two oppositely directed matter waves associated with this particle interfere to create a standing wave with a node at each of the walls. Show that the ground state (first harmonic) and first excited state (second harmonic) have (non-relativistic) kinetic energies $h^2/8m\ell^2$ and $h^2/2m\ell^2$, respectively.

88. A photon of momentum $3.53 \times 10^{-28}\,\text{kg}\cdot\text{m/s}$ is emitted from a hydrogen atom. To what spectrum series does this photon belong, and from what energy level was it ejected?

89. Imagine the following Young's double-slit experiment using matter rather than light: electrons are accelerated through a potential difference of 18 V, pass through two closely spaced slits separated by a distance d, and create an interference pattern. (*a*) Using Example 37–11 and Section 34–3 as guides, find the required value for d if the first-order interference fringe is to be produced at an angle of 10°. (*b*) Given the approximate size of atoms, would it be possible to construct the required two-slit set-up for this experiment?

90. A beam of red laser light $(\lambda = 633\,\text{nm})$ hits a black wall and is fully absorbed. If this light exerts a total force $F = 7.5\,\text{nN}$ on the wall, how many photons per second are hitting the wall?

91. At what temperature would the average kinetic energy (Chapter 18) of a molecule of hydrogen gas (H_2) be sufficient to excite a hydrogen atom out of the ground state?

92. (*a*) From Sections 31–6, 32–8, and 37–2, estimate the minimum energy (eV) that initiates the chemical process on the retina responsible for vision, (*b*) Estimate the threshold photon energy above which the eye registers no sensation of sight.

93. Suppose a particle of mass m is confined to a one-dimensional box of width L. According to quantum theory, the particle's wave (with $\lambda = h/mv$) is a standing wave with nodes at the edges of the box. (*a*) Show the possible modes of vibration on a diagram. (*b*) Show that the kinetic energy of the particle has quantized energies given by $K = n^2h^2/8mL^2$, where n is an integer. (*c*) Calculate the ground-state energy $(n = 1)$ for an electron confined to a box of width $0.50 \times 10^{-10}\,\text{m}$. (*d*) What is the ground-state energy, and speed, of a baseball $(m = 140\,\text{g})$ in a box 0.55 m wide? (*e*) An electron confined to a box has a ground-state energy of 18 eV. What is the width of the box? [*Hint*: See Sections 37–7, 37–12, and 15–9.]

94. A photon and an electron both have the same wavelength, 560 nm. What is the ratio of their speeds?

95. When a photon ionizes an atom, the event can be viewed as the photoelectric effect. What is the work function for H atoms?

96. (*a*) A rubidium atom ($m = 85$ u) is at rest with one electron in an excited energy level. When the electron jumps to the ground state, the atom emits a photon of wavelength $\lambda = 780\,\text{nm}$. Determine the resulting (nonrelativistic) recoil speed v of the atom. (*b*) The recoil speed sets the lower limit on the temperature to which an ideal gas of rubidium atoms can be cooled in a laser-based **atom trap**. Using the kinetic theory of gases (Chapter 18), estimate this "lowest achievable" temperature.

97. A rubidium atom (atomic mass 85) is initially at room temperature and has a velocity $v = 260\,\text{m/s}$ due to its thermal motion. Consider the absorption of photons by this atom from a laser beam of wavelength $\lambda = 780\,\text{nm}$. Assume the rubidium atom's initial velocity v is directed into the laser beam (the photons are moving right and the atom is moving left) and that the atom absorbs a new photon every 25 ns. How long will it take for this process to completely stop ("cool") the rubidium atom? [*Note*: A more detailed analysis predicts that the atom can be slowed to about 1 cm/s by this light absorption process, but it cannot be completely stopped.]

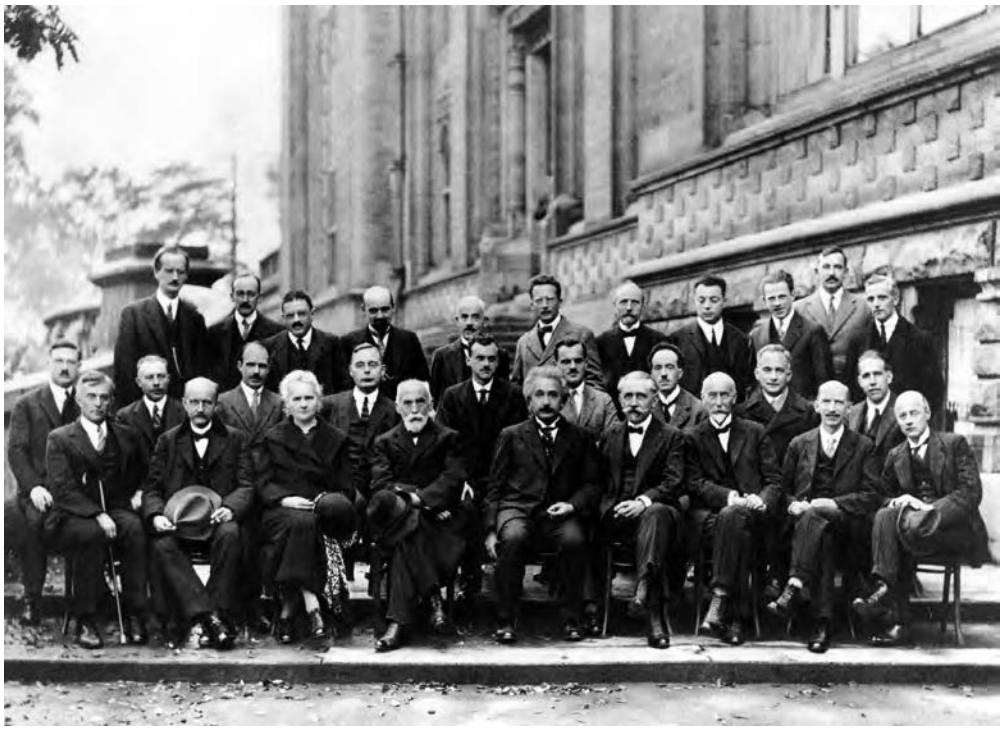
98. By what fraction does the mass of an H atom decrease when it makes an $n = 3$ to $n = 1$ transition?

99. In Chapter 31, the intensity of light striking a surface was related to the electric field of the associated electromagnetic wave. For photons, the intensity is the number of photons striking a 1-m^2 area per second. Suppose 1.0×10^{12} photons of 521-nm light are incident on a 1-m^2 surface every second. What is the intensity of the light? Using the wave model of light, what is the maximum electric field of the electromagnetic wave?

ANSWERS TO EXERCISES

A: $\lambda_P = 725\,\text{nm}$, so red.
B: More 1000-nm photons (each has lower energy).
C: $5.50 \times 10^{14}\,\text{Hz}$, 545 nm.
D: Only λ.
E: Decrease.
F: (*e*).
G: (*c*).

Participants in the 1927 Solvay Conference. Albert Einstein, seated at front center, had difficulty accepting that nature could behave according to the rules of quantum mechanics. We present a brief summary of quantum mechanics in this Chapter starting with the wave function and the uncertainty principle. We examine the Schrödinger equation and its solutions for some simple cases: free particles, the square well, and tunneling through a barrier.



A. Piccard, E. Henriot, P. Ehrenfest, Ed. Herzen, Th. De Donder, E. Schrödinger, E. Verschaffelt, W. Pauli, W. Heisenberg, R. H. Fowler, L. Brillouin, P. Debye, M. Knudsen, W. L. Bragg, H. A. Kramers, P. A. M. Dirac, A. H. Compton, L. de Broglie, M. Born, N. Bohr, I. Langmuir, M. Planck, Mme Curie, H. A. Lorentz, A. Einstein, P. Langevin, Ch. E. Guye, C. T. R. Wilson, O. W. Richardson.

CHAPTER 38

Quantum Mechanics

CONTENTS

38 – 1 Quantum Mechanics—A New Theory
38 – 2 The Wave Function and Its Interpretation; the Double-Slit Experiment
38 – 3 The Heisenberg Uncertainty Principle
38 – 4 Philosophic Implications; Probability versus Determinism
38 – 5 The Schrödinger Equation in One Dimension—Time-Independent Form
*38 – 6 Time-Dependent Schrödinger Equation
38 – 7 Free Particles; Plane Waves and Wave Packets
38 – 8 Particle in an Infinitely Deep Square Well Potential (a Rigid Box)
38 – 9 Finite Potential Well
38–10 Tunneling through a Barrier

CHAPTER-OPENING QUESTION—Guess now!

The uncertainty principle states that

(a) no measurement can be perfect because it is technologically impossible to make perfect measuring instruments.
(b) it is impossible to measure exactly where a particle is, unless it is at rest.
(c) it is impossible to simultaneously know both the position and the momentum of a particle with complete certainty.
(d) a particle cannot actually have a completely certain value of momentum.

Bohr's model of the atom gave us a first (though rough) picture of what an atom is like. It proposed explanations for why there is emission and absorption of light by atoms at only certain wavelengths. The wavelengths of the line spectra and the ionization energy for hydrogen (and one-electron ions) are in excellent agreement with experiment. But the Bohr model had important limitations. It was not able to predict line spectra for more complex atoms—atoms with more than one electron—not even for the neutral helium atom, which has only two electrons. Nor could it explain why emission lines, when viewed with great precision, consist of two or more very closely spaced lines (referred to as *fine structure*). The Bohr model also did not explain why some spectral lines were brighter than others. And it could not explain the bonding of atoms in molecules or in solids and liquids.

From a theoretical point of view, too, the Bohr model was not satisfactory: it was a strange mixture of classical and quantum ideas. Moreover, the wave–particle duality was not really resolved.

PSE5_CH38_v1.7_19-Jun-18



FIGURE 38–1 Erwin Schrödinger with Lise Meitner. (She was a codiscoverer of nuclear fission, Chapter 42.)



FIGURE 38–2 Werner Heisenberg (center) on Lake Como with Enrico Fermi (left) and Wolfgang Pauli (right).

We mention these limitations of the Bohr model not to disparage it—for it was a landmark in the history of science. Rather, we mention them to show why, in the early 1920s, it became increasingly evident that a new, more comprehensive theory was needed. It was not long in coming. Less than two years after de Broglie gave us his matter–wave hypothesis, Erwin Schrödinger (1887–1961; Fig. 38–1) and Werner Heisenberg (1901–1976; Fig. 38–2) independently developed a new comprehensive theory.

38–1 Quantum Mechanics—A New Theory

The new theory, called **quantum mechanics**, has been extremely successful. It unifies the wave–particle duality into a single consistent theory and has successfully dealt with the spectra emitted by complex atoms, even the fine details. It also explains the relative brightness of spectral lines and how atoms form molecules. It is a much more general theory that covers all quantum phenomena from blackbody radiation to atoms and molecules. It has explained a wide range of natural phenomena and from its predictions many new practical devices have become possible. Indeed, it has been so successful that it is accepted today by nearly all physicists as the fundamental theory underlying physical processes.

Quantum mechanics deals mainly with the microscopic world of atoms and light. But this new theory, when it is applied to macroscopic phenomena, must be able to produce the old classical laws. This, the **correspondence principle** (already mentioned in Section 37–11), is satisfied fully by quantum mechanics.

This doesn't mean we should throw away classical theories such as Newton's laws. In the everyday world, classical laws are far easier to apply and they give sufficiently accurate descriptions. But when we deal with high speeds, close to the speed of light, we must use the theory of relativity; and when we deal with the tiny world of the atom, we use quantum mechanics.

Although we won't go into the detailed mathematics of quantum mechanics in this Chapter, we will discuss the main ideas and how they involve the wave and particle properties of matter to explain atomic structure and other applications.

38–2 The Wave Function and Its Interpretation; the Double-Slit Experiment

The important properties of any wave are its wavelength, frequency, and amplitude. For an electromagnetic wave, the frequency (or wavelength) determines whether the wave is light in the visible spectrum or not, and if so, what color it is. We also have seen (Eq. 37–3) that the frequency is a measure of the energy of the corresponding photons ($E = hf$). The amplitude or displacement of an electromagnetic wave at any point is the strength of the electric (or magnetic) field at that point, and is related to the intensity of the wave (the brightness of the light).

For material particles such as electrons, quantum mechanics relates the wavelength to momentum according to de Broglie's formula, $\lambda = h/p$, Eq. 37–7. But what corresponds to the *amplitude* or *displacement* of a matter wave? The amplitude of an electromagnetic wave is represented by the electric and magnetic fields, E and B. In quantum mechanics, this role is played by the **wave function**, which is given the symbol Ψ (the Greek capital letter psi, pronounced "sigh" or "psigh"). Thus Ψ represents the wave displacement, as a function of time and position, of a new kind of field which we might call a "matter" field or a matter wave.

To understand how to interpret the wave function Ψ, we make an analogy with light using the wave–particle duality.

We saw in Chapter 15 that the intensity I of any wave is proportional to the square of the amplitude. This holds true for light waves as well, as we saw in Chapter 31. That is,

$$I \propto E^2,$$

where E is the electric field strength. From the *particle* point of view, the intensity of a light beam (of given frequency) is proportional to the number of photons, N, that pass through a given area per unit time. The more photons there are, the greater the intensity. Thus

$$I \propto E^2 \propto N.$$

This proportion can be turned around so that we have

$$N \propto E^2.$$

That is, the number of photons (striking a page of this book, say) is proportional to the square of the electric field strength.

If the light beam is very weak, only a few photons will be present. Indeed, it is possible to "build up" a photograph in a camera using very weak light so the effect of individual photons arriving can be seen.

If we are dealing with only one photon, the relationship above $(N \propto E^2)$ can be interpreted in a slightly different way. At any point, the square of the electric field strength E^2 is a measure of the *probability* that a photon will be at that location. At points where E^2 is large, there is a high probability the photon will be there; where E^2 is small, the probability is low.

We can interpret matter waves in the same way, as was first suggested by Max Born (1882–1970) in 1927. The wave function Ψ may vary in magnitude from point to point in space and time. If Ψ describes a collection of many electrons, then $|\Psi|^2$ at any point will be proportional to the number of electrons expected to be found at that point.[†] When dealing with small numbers of electrons we can't make very exact predictions, so $|\Psi|^2$ takes on the character of a probability. If Ψ, which depends on time and position, represents a single electron (say, in an atom), then $|\Psi|^2$ is interpreted as follows: $|\Psi|^2$ *at a certain point in space and time represents the probability of finding the electron at the given position and time*. Thus $|\Psi|^2$ is often referred to as the **probability density** or **probability distribution**.

Double-Slit Interference Experiment for Electrons

To understand this better, we take as a thought experiment the familiar double-slit experiment, and consider it both for light and for electrons.

Consider two slits whose size and separation are on the order of the wavelength of whatever we direct at them, either light or electrons, Fig. 38–3. We know very well what would happen in this case for light, since this is just Young's double-slit experiment (Section 34–3): an interference pattern would be seen on the screen behind. If light were replaced by electrons with wavelength comparable to the slit size, they too would produce an interference pattern (recall Fig. 37–11). In the case of light, the pattern would be visible to the eye or could be recorded on film, semiconductor sensor, or screen. For electrons, a fluorescent screen could be used (it glows where an electron strikes).

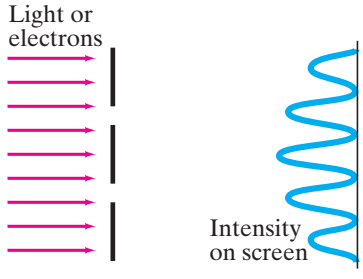


FIGURE 38–3 Parallel beam, of light or electrons, falls on two slits whose sizes are comparable to the wavelength. An interference pattern is observed.

[†]The wave function Ψ is generally a complex quantity (that is, it involves $i = \sqrt{-1}$) and hence is not directly observable. On the other hand, $|\Psi|^2$, the absolute value of Ψ squared, is always a real quantity and it is to $|\Psi|^2$ that we can give a physical interpretation.

If we reduced the flow of electrons (or photons) so they passed through the slits one at a time, we would see a flash each time one struck the screen. At first, the flashes would seem random. Indeed, there is no way to predict just where any one electron would hit the screen. If we let the experiment run for a long time, and kept track of where each electron hit the screen, we would soon see a pattern emerging—the interference pattern predicted by the wave theory; see Fig. 38–4. Thus, although we could not predict where a given electron would strike the screen, we could predict probabilities. (The same can be said for photons.) The probability, as we saw, is proportional to $|\Psi|^2$. Where $|\Psi|^2$ is zero, we would get a minimum in the interference pattern. And where $|\Psi|^2$ is a maximum, we would get a peak in the interference pattern.

The interference pattern would thus occur even when electrons (or photons) passed through the slits one at a time. So the interference pattern could not arise from the interaction of one electron with another. It is as if an electron passed through both slits at the same time, interfering with itself. This is possible because an electron is not precisely a particle. It is as much a wave as it is a particle, and a wave could travel through both slits at once. But what would happen if we covered one of the slits so we knew that the electron passed through the other slit, and a little later we covered the second slit so the electron had to have passed through the first slit? The result would be that no interference pattern would be seen. We would see, instead, two bright areas (or single-slit diffraction patterns) on the screen behind the slits.

If both slits are open, the screen shows an interference pattern as if each electron passed through both slits, like a wave. Yet each electron would make a tiny spot on the screen as if it were a particle.

The main point of this discussion is this: if we treat electrons (and other particles) as if they were waves, then Ψ represents the wave amplitude. If we treat them as particles, then we must treat them on a *probabilistic* basis. The square of the wave function, $|\Psi|^2$, gives the probability of finding a given electron at a given point. We cannot predict—or even follow—the path of a single electron precisely through space and time.

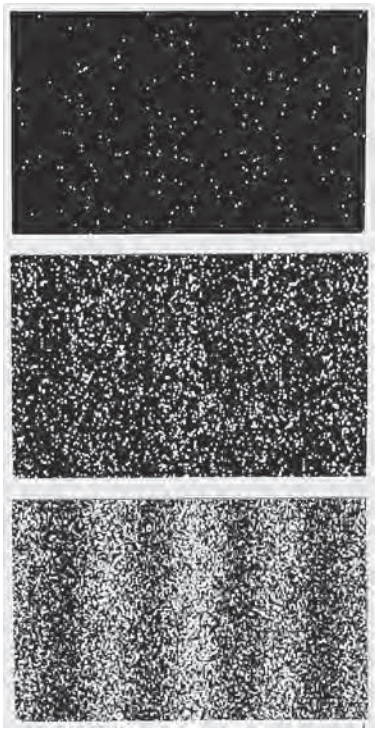


FIGURE 38–4 Young's double-slit experiment done with electrons—note that the pattern is not evident with only a few electrons (top photo), but with more and more electrons (second and third photos), the familiar double-slit interference pattern (Chapter 34) is seen.

38–3 The Heisenberg Uncertainty Principle

Whenever a measurement is made, some uncertainty is always involved. For example, you cannot make an absolutely exact measurement of the length of a table. Even with a measuring stick that has markings 1 mm apart, there will be an inaccuracy of perhaps $\frac{1}{2}$ mm or so. More precise instruments will produce more precise measurements. But there is always some uncertainty involved in a measurement, no matter how good the measuring device. We expect that by using more precise instruments, the uncertainty in a measurement can be made indefinitely small.

But according to quantum mechanics, there is actually a limit to the precision of certain measurements. This limit is not a restriction on how well instruments can be made; rather, it is inherent in nature. It is the result of two factors: the wave–particle duality, and the unavoidable interaction between the thing observed and the observing instrument. Let us look at this in more detail.

To make a measurement on an object without disturbing it, at least a little, is not possible. Consider trying to locate a Ping-Pong ball in a dark room: you could probe about with your hand or a stick; or you could shine a light and detect the photons reflecting off the ball. When you search with your hand or a stick, you find the ball's position when you touch it, but at the same time you unavoidably bump it and give it some momentum. Thus you won't know its *future* position. If you search for the Ping-Pong ball using light, in order to "see" the ball, at least one photon (actually quite a few) must scatter from it, and the reflected photon must enter your eye or some other detector. When a photon strikes an ordinary-sized object, it only slightly alters the motion or position of the object.

But a photon striking a tiny object like an electron transfers enough momentum to greatly change the electron's motion and position in an unpredictable way. The mere act of measuring the position of an object at one time makes our knowledge of its future position imprecise.

Now let us see where the wave–particle duality comes in. Imagine a thought experiment in which we are trying to measure the position of an object, say an electron, with photons, Fig. 38–5. (The arguments would be similar if we were using, instead, an electron microscope.) As we saw in Chapter 35, objects can be seen to a precision at best of about the wavelength of the radiation used due to diffraction. If we want a precise position measurement, we must use a short wavelength. But a short wavelength corresponds to high frequency and large momentum $(p = h/\lambda)$; and the more momentum the photons have, the more momentum they can give the object when they strike it. If we use photons of longer wavelength, and correspondingly smaller momentum, the object's motion when struck by the photons will not be affected as much. But the longer wavelength means lower resolution, so the object's position will be less accurately known. Thus the act of observing produces an uncertainty in both the *position* and the *momentum* of the electron. This is the essence of the *uncertainty principle* first enunciated by Heisenberg in 1927.

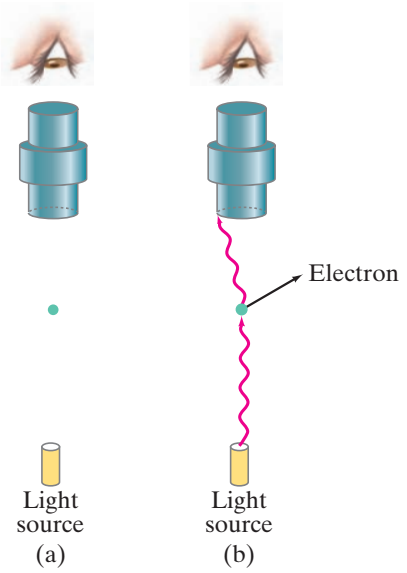


FIGURE 38–5 Thought experiment for observing an electron with a powerful light microscope. At least one photon must scatter from the electron (transferring some momentum to it) and enter the microscope.

Quantitatively, we can make an approximate calculation of the magnitude of the uncertainties. If we use light of wavelength λ, the position can be measured at best to a precision of about λ. That is, the uncertainty in the position measurement, Δx, is approximately

$$\Delta x \approx \lambda.$$

Suppose that the object can be detected by a single photon. The photon has a momentum $p_x = h/\lambda$ (Eq. 37–5). When the photon strikes our object, it will give some or all of this momentum to the object, Fig. 38–5. Therefore, the final x momentum of our object will be uncertain in the amount

$$\Delta p_x \approx \frac{h}{\lambda}$$

since we can't tell how much momentum will be transferred. The product of these uncertainties is

$$(\Delta x)(\Delta p_x) \approx (\lambda)\left(\frac{h}{\lambda}\right) \approx h.$$

The uncertainties could be larger than this, depending on the apparatus and the number of photons needed for detection. A more careful mathematical

PSE5_CH38_v1.7_19-Jun-18

calculation shows the product of the uncertainties as, at best, about

$$(\Delta x)(\Delta p_x) \gtrsim \frac{h}{2\pi} \qquad \textbf{(38–1)}$$

UNCERTAINTY PRINCIPLE (Δx and Δp)

where Δp_x is the uncertainty of the momentum in the x direction.[†] This is a mathematical statement of the **Heisenberg uncertainty principle**, or, as it is sometimes called, the **indeterminancy principle**. It tells us that we cannot measure both the position *and* momentum of an object precisely at the same time. The more accurately we try to measure the position so that Δx is small, the greater will be the uncertainty in momentum, Δp_x. If we try to measure the momentum very accurately, then the uncertainty in the position becomes large.

The uncertainty principle does not forbid individual precise measurements, however. For example, in principle we could measure the position of an object exactly. But then its momentum would be completely unknown. Thus, although we might know the position of the object exactly at one instant, we could have no idea at all where it would be a moment later. The uncertainties expressed here are inherent in nature, and reflect the best precision theoretically attainable even with the best instruments.

⚠ CAUTION

Uncertainties are not due to instrument deficiency, but are inherent in nature (wave–particle)

EXERCISE A Return to the Chapter-Opening Question, page 1128, and answer it again now. Try to explain why you may have answered differently the first time.

Another useful form of the uncertainty principle relates energy and time, and we examine this as follows. The object to be detected has an uncertainty in position $\Delta x \approx \lambda$. The photon that detects it travels with speed c, and it takes a time $\Delta t \approx \Delta x/c \approx \lambda/c$ to pass through the distance of uncertainty. Hence, the measured time when our object is at a given position is uncertain by about

$$\Delta t \approx \frac{\lambda}{c}.$$

Since the photon can transfer some or all of its energy $(= hf = hc/\lambda)$ to our object, the uncertainty in energy of our object as a result is

$$\Delta E \approx \frac{hc}{\lambda}.$$

The product of these two uncertainties is

$$(\Delta E)(\Delta t) \approx \left(\frac{hc}{\lambda}\right)\left(\frac{\lambda}{c}\right) \approx h.$$

A more careful calculation gives

$$(\Delta E)(\Delta t) \gtrsim \frac{h}{2\pi}. \qquad \textbf{(38–2)}$$

UNCERTAINTY PRINCIPLE (ΔE and Δt)

This form of the uncertainty principle tells us that the energy of an object can be uncertain (or can be interpreted as briefly nonconserved) by an amount ΔE for a time $\Delta t \approx h/(2\pi\,\Delta E)$.

The quantity $(h/2\pi)$ appears so often in quantum mechanics that for convenience it is given the symbol $\hbar$ ("h-bar"). That is,

$$\hbar = \frac{h}{2\pi} = \frac{6.626 \times 10^{-34}\,\text{J}\cdot\text{s}}{2\pi} = 1.055 \times 10^{-34}\,\text{J}\cdot\text{s}.$$

By using this notation, Eqs. 38–1 and 38–2 for the uncertainty principle can be written

$$(\Delta x)(\Delta p_x) \gtrsim \hbar \qquad \text{and} \qquad (\Delta E)(\Delta t) \gtrsim \hbar.$$

We have been discussing the position and velocity of an electron as if it were a particle. But it isn't simply a particle. Indeed, we have the uncertainty principle because an electron—and matter in general—has wave as well as particle properties.

[†]Note, however, that quantum mechanics does allow simultaneous precise measurements of p_x and y: that is, $(\Delta y)(\Delta p_x) \gtrsim 0$.

What the uncertainty principle really tells us is that if we insist on thinking of the electron as a particle, then there are certain limitations on this simplified view—namely, that the position and velocity cannot both be known precisely at the same time; and even that the electron does not *have* a precise position and momentum at the same time (because it is not simply a particle). Similarly, the energy can be uncertain (or nonconserved) by an amount ΔE for a time $\Delta t \approx \hbar/\Delta E$.

Because Planck's constant, h, is so small, the uncertainties expressed in the uncertainty principle are usually negligible on the macroscopic level. But at the level of atomic sizes, the uncertainties are significant. Because we consider ordinary objects to be made up of atoms containing nuclei and electrons, the uncertainty principle is relevant to our understanding of all of nature. The uncertainty principle expresses, perhaps most clearly, the probabilistic nature of quantum mechanics. It thus is often used as a basis for philosophic discussion.

EXAMPLE 38–1 **Position uncertainty of an electron.** An electron moves in a straight line with a constant speed $v = 1.10 \times 10^6\,\text{m/s}$ which has been measured to a precision of 0.10%. What is the maximum precision with which its position could be simultaneously measured?

APPROACH The momentum is $p = mv$, and the uncertainty in p is $\Delta p = 0.0010p$. The uncertainty principle (Eq. 38–1) gives us the smallest uncertainty in position Δx using the equals sign.

SOLUTION The momentum of the electron is

$$p = mv = (9.11 \times 10^{-31}\,\text{kg})(1.10 \times 10^6\,\text{m/s}) = 1.00 \times 10^{-24}\,\text{kg}\cdot\text{m/s}.$$

The uncertainty in the momentum is 0.10% of this, or $\Delta p = 1.0 \times 10^{-27}\,\text{kg}\cdot\text{m/s}$. From the uncertainty principle, the best simultaneous position measurement will have an uncertainty of

$$\Delta x \approx \frac{\hbar}{\Delta p} = \frac{1.055 \times 10^{-34}\,\text{J}\cdot\text{s}}{1.0 \times 10^{-27}\,\text{kg}\cdot\text{m/s}} = 1.1 \times 10^{-7}\,\text{m},$$

or 110 nm.

NOTE This is about 1000 times the diameter of an atom.

EXERCISE B An electron's position is measured with a precision of $0.50 \times 10^{-10}\,\text{m}$. Find the minimum uncertainty in its momentum and velocity.

EXAMPLE 38–2 **Position uncertainty of a baseball.** What is the uncertainty in position, imposed by the uncertainty principle, on a 150-g baseball thrown at $(93 \pm 2)\,\text{mi/h} = (42 \pm 1)\,\text{m/s}$?

APPROACH The uncertainty in the speed is $\Delta v = 1\,\text{m/s}$. We multiply Δv by m to get Δp and then use the uncertainty principle, solving for Δx.

SOLUTION The uncertainty in the momentum is

$$\Delta p = m\,\Delta v = (0.150\,\text{kg})(1\,\text{m/s}) = 0.15\,\text{kg}\cdot\text{m/s}.$$

Hence the uncertainty in a position measurement could be as small as

$$\Delta x = \frac{\hbar}{\Delta p} = \frac{1.055 \times 10^{-34}\,\text{J}\cdot\text{s}}{0.15\,\text{kg}\cdot\text{m/s}} = 7 \times 10^{-34}\,\text{m}.$$

NOTE This distance is far smaller than any we could imagine observing or measuring. It is trillions of trillions of times smaller than an atom. Indeed, the uncertainty principle sets no relevant limit on measurement for macroscopic objects.

EXAMPLE 38–3 **ESTIMATE** **J/ψ lifetime calculated.** The J/ψ meson, discovered in 1974, was measured to have an average mass of $3100\,\text{MeV}/c^2$ (note the use of energy units since $E = mc^2$) and a mass "width" of $63\,\text{keV}/c^2$. By this we mean that the masses of different J/ψ mesons were actually measured to be slightly different from one another. This mass "width" is related to the very short lifetime of the J/ψ before it decays into other particles. From the uncertainty principle, if the particle exists for only a time Δt, its mass (or rest energy) will be uncertain by $\Delta E \approx \hbar/\Delta t$. Estimate the J/ψ lifetime.

APPROACH We use the energy–time version of the uncertainty principle, Eq. 38–2.

SOLUTION The uncertainty of $63\,\text{keV}/c^2$ in the J/ψ's mass is an uncertainty in its rest energy, which in joules is

$$\Delta E = (63 \times 10^3\,\text{eV})(1.60 \times 10^{-19}\,\text{J/eV}) = 1.01 \times 10^{-14}\,\text{J}.$$

Then we expect its lifetime τ $(= \Delta t$ using Eq. 38–2) to be

$$\tau \approx \frac{\hbar}{\Delta E} = \frac{1.055 \times 10^{-34}\,\text{J}\cdot\text{s}}{1.01 \times 10^{-14}\,\text{J}} \approx 1 \times 10^{-20}\,\text{s}.$$

Lifetimes this short are difficult to measure directly, and the assignment of very short lifetimes depends on this use of the uncertainty principle. (See Chapter 43.)

The uncertainty principle applies also for angular variables:

$$(\Delta L_z)(\Delta\phi) \gtrsim \hbar$$

where L is the component of angular momentum along a given axis (z) and ϕ is the angular position in a plane perpendicular to that axis.

38–4 Philosophic Implications; Probability versus Determinism

The classical Newtonian view of the world is a deterministic one (see Section 6–5). One of its basic ideas is that once the position and velocity of an object are known at a particular time, its future position can be predicted if the forces on it are known. For example, if a stone is thrown a number of times with the same initial velocity and angle, and the forces on it remain the same, the path of the projectile will always be the same. If the forces are known (gravity and air resistance, if any), the stone's path can be precisely predicted. This mechanistic view implies that the future unfolding of the universe, assumed to be made up of particulate objects, is completely determined.

This classical deterministic view of the physical world has been radically altered by quantum mechanics. As we saw in the analysis of the double-slit experiment (Section 38–2), electrons all prepared in the same way will not all end up in the same place. According to quantum mechanics, certain probabilities exist that an electron will arrive at different points. This is very different from the classical view, in which the path of a particle is precisely predictable from the initial position and velocity and the forces exerted on it. According to quantum mechanics, the position and velocity of an object cannot even be known accurately at the same time. This is expressed in the uncertainty principle, and arises because basic entities, such as electrons, are not considered simply as particles: they have wave properties as well. Quantum mechanics allows us to calculate only the probability[†] that, say, an electron (when thought of as a particle) will be observed at various places.

[†]Note that these probabilities can be calculated precisely, just like exact predictions of probabilities at rolling dice or playing cards, but they are unlike predictions of probabilities at sporting events or for natural or man-made disasters, which are only estimates.

Quantum mechanics says there is some inherent unpredictability in nature. This is very different from the deterministic view of classical mechanics.

Because matter is considered to be made up of atoms, even ordinary-sized objects are expected to be governed by probability, rather than by strict determinism. For example, quantum mechanics predicts a finite (but negligibly small) probability that when you throw a stone, its path might suddenly curve upward instead of following the downward-curved parabola of normal projectile motion. Quantum mechanics predicts with extremely high probability that ordinary objects will behave just as the classical laws of physics predict. But these predictions are considered probabilities, not absolute certainties. The reason that macroscopic objects behave in accordance with classical laws with such high probability is due to the large number of molecules involved: when large numbers of objects are present in a statistical situation, deviations from the average (or most probable) approach zero. It is the average configuration of vast numbers of molecules that follows the so-called fixed laws of classical physics with such high probability, and gives rise to an apparent "determinism." Deviations from classical laws are observed when small numbers of molecules are dealt with. We can say, then, that although there are no precise deterministic laws in quantum mechanics, there are statistical laws based on probability.

It is important to note that there is a difference between the probability imposed by quantum mechanics and that used in the nineteenth century to understand thermodynamics and the behavior of gases in terms of molecules (Chapters 18 and 20). In thermodynamics, probability is used because there are far too many particles to keep track of. But the molecules are still assumed to move and interact in a deterministic way following Newton's laws. Probability in quantum mechanics is quite different; it is seen as *inherent* in nature, and not as a limitation on our abilities to calculate or to measure.

The view presented here is the generally accepted one and is called the **Copenhagen interpretation** of quantum mechanics in honor of Niels Bohr's home, since it was largely developed there through discussions between Bohr and other prominent physicists.

Because electrons are not simply particles, they cannot be thought of as following particular paths in space and time. This suggests that a description of matter in space and time may not be completely correct. This deep and far-reaching conclusion has been a lively topic of discussion among philosophers. Perhaps the most important and influential philosopher of quantum mechanics was Bohr. He argued that a space–time description of actual atoms and electrons is not possible. Yet a description of experiments on atoms or electrons must be given in terms of space and time and other concepts familiar to ordinary experience, such as waves and particles. We must not let our *descriptions* of experiments lead us into believing that atoms or electrons themselves actually move in space and time as classical particles.

38–5 The Schrödinger Equation in One Dimension—Time-Independent Form

In order to describe physical systems quantitatively using quantum mechanics, we must have a means of determining the wave function Ψ mathematically. The basic equation (in the nonrelativistic realm) for determining Ψ is the *Schrödinger equation*. We cannot, however, derive the Schrödinger equation from some higher principles, just as Newton's second law, for example, cannot be derived. The relation $\vec{\mathbf{F}} = m\vec{\mathbf{a}}$ was *invented* by Newton to describe how the motion of an object is related to the net applied force. As we saw early in this book, Newton's second law works exceptionally well. In the realm of classical physics it is the starting point for analytically solving a wide range of problems, and the solutions it yields are fully consistent with experiment. The validity of any fundamental equation resides in its agreement with experiment.

PSE5_CH38_v1.7_19-Jun-18

The Schrödinger equation forms part of a new theory, and it too had to be *invented*—and then checked against experiment, a test that it passed splendidly.

The Schrödinger equation can be written in two forms: the time-dependent version and the time-independent version. We will mainly be interested in steady-state situations—that is, when there is no time dependence—and so we mainly deal with the time-independent version. (We briefly discuss the time-dependent version in the optional Section 38–6.) The time-independent version involves a wave function with only spatial dependence which we represent by lowercase psi, $\psi(x)$, for the simple one-dimensional case we deal with here. In three dimensions, we write $\psi(x, y, z)$ or $\psi(r, \theta, \phi)$.

In classical mechanics, we solved problems using two approaches: via Newton's laws with the concept of force, and by using the energy concept with the conservation laws. The Schrödinger equation is based on the energy approach. Even though the Schrödinger equation cannot be derived, we can suggest what form it might take by using conservation of energy and considering a very simple case: that of a free particle on which no forces act, so that its potential energy U is constant. We assume that our particle moves along the x axis, and since no force acts on it, its momentum remains constant and its wavelength $(\lambda = h/p)$ is fixed. To describe a wave for a free particle such as an electron, we expect that its wave function will satisfy a differential equation that is akin to (but not identical to) the classical wave equation (Section 15–4 and 15–5, and Eq. 15–16). Let us see what we can infer about this equation. Consider a simple traveling wave of a single wavelength λ whose wave displacement, as we saw in Chapter 15 for mechanical waves and in Chapter 31 for electromagnetic waves, is given by $A\sin(kx - \omega t)$, or more generally as a superposition of sine and cosine: $A\sin(kx - \omega t) + B\cos(kx - \omega t)$. We are only interested in the spatial dependence, so we consider the wave at a specific moment, say $t = 0$. Thus we write as the wave function for our free particle

$$\psi(x) = A\sin kx + B\cos kx, \qquad \textbf{(38–3a)}$$

where A and B are constants† and $k = 2\pi/\lambda$ (Eq. 15–11). For a particle of mass m and velocity v, the de Broglie wavelength is $\lambda = h/p$, where $p = mv$ is the particle's momentum. Hence

$$k = \frac{2\pi}{\lambda} = \frac{2\pi p}{h} = \frac{p}{\hbar}. \qquad \textbf{(38–3b)}$$

One requirement for our wave equation, then, is that it have the wave function $\psi(x)$ as given by Eqs. 38–3 as a solution for a free particle. A second requirement is that it be consistent with the conservation of energy, which we can express as

$$\frac{p^2}{2m} + U = E,$$

where E is the total energy, U is the potential energy, and (since we are considering the nonrelativistic realm) the kinetic energy K of our particle of mass m is $K = \frac{1}{2}mv^2 = p^2/2m$. Since $p = \hbar k$ (Eq. 38–3b), we can write the conservation of energy condition as

$$\frac{\hbar^2 k^2}{2m} + U = E. \qquad \textbf{(38–4)}$$

Thus we are seeking a differential equation that satisfies conservation of energy (Eq. 38–4) when $\psi(x)$ is its solution. Now, note that if we take two derivatives of our expression for $\psi(x)$, Eq. 38–3a, we get a factor $-k^2$ multiplied by $\psi(x)$:

$$\frac{d\psi(x)}{dx} = \frac{d}{dx}(A\sin kx + B\cos kx) = k(A\cos kx - B\sin kx)$$

$$\frac{d^2\psi(x)}{dx^2} = k\frac{d}{dx}(A\cos kx - B\sin kx) = -k^2(A\sin kx + B\cos kx) = -k^2\psi(x).$$

†In quantum mechanics, constants can be complex (i.e., with a real and/or imaginary part).

Can this last term be related to the k^2 term in Eq. 38–4? Indeed, if we multiply this last relation by $-\hbar^2/2m$, we obtain

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = \frac{\hbar^2 k^2}{2m}\psi(x).$$

The right side is just the first term on the left in Eq. 38–4 multiplied by $\psi(x)$. If we multiply Eq. 38–4 through by $\psi(x)$, and make this substitution, we obtain

SCHRÖDINGER EQUATION (time-independent form)

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x). \qquad \textbf{(38–5)}$$

This is, in fact, the one-dimensional **time-independent Schrödinger equation**, where for generality we have written $U = U(x)$. It is the basis for solving problems in nonrelativistic quantum mechanics. For a particle moving in three dimensions there would be additional derivatives with respect to y and z (see Chapter 39).

Note that we have by no means derived the Schrödinger equation. Although we have made a good argument in its favor, other arguments could also be made which might or might not lead to the same equation. The Schrödinger equation as written (Eq. 38–5) is useful and valid only because it has given results in accord with experiment for a wide range of situations.

There are some requirements we impose on any wave function that is a solution of the Schrödinger equation in order that it be physically meaningful. First, we insist that it be a continuous function; after all, if $|\psi|^2$ represents the probability of finding a particle at a certain point, we expect the probability to be continuous from point to point and not to take discontinuous jumps. Second, we want the wave function to be *normalized*. By this we mean that for a single particle, the probability of finding the particle at one point or another (i.e., the probabilities summed over all space) must be exactly 1 (or 100%). For a single particle, $|\psi|^2$ represents the probability of finding the particle in unit volume. Then

CAUTION
Wave function must (1) be continuous and (2) normalized

$$|\psi|^2\, dV \qquad \textbf{(38–6a)}$$

is the probability of finding the particle within a volume dV, where ψ is the value of the wave function in this infinitesimal volume dV. For the one-dimensional case, dV becomes dx, so the probability of finding a particle within dx of position x is

$$|\psi(x)|^2\, dx. \qquad \textbf{(38–6b)}$$

Then the sum of the probabilities over all space—that is, the probability of finding the particle somewhere—becomes

$$\int_{\text{all space}} |\psi|^2\, dV = \int |\psi|^2\, dx = 1. \qquad \textbf{(38–6c)}$$

This is called the **normalization condition**, and the integral is taken over whatever region of space in which the particle has a chance of being found, which sometimes can be all of space, from $x = -\infty$ to $x = \infty$.

*38–6 Time-Dependent Schrödinger Equation

The more general form of the Schrödinger equation, including time dependence, for a particle of mass m moving in one dimension, is

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + U(x)\Psi(x,t) = i\hbar\frac{\partial\Psi(x,t)}{\partial t}. \qquad \textbf{(38–7)}$$

This is the **time-dependent Schrödinger equation**; here $U(x)$ is the potential energy of the particle as a function of position, and i is the imaginary number $i = \sqrt{-1}$.

PSE5_CH38_v1.7_19-Jun-18

For a particle moving in three dimensions, there would be additional derivatives with respect to y and z, just as for the classical wave equation discussed in Section 15–5. Indeed, it is worth noting the similarity between the Schrödinger wave equation for zero potential energy $(U = 0)$ and the classical wave equation: $\partial^2 D/\partial t^2 = v^2 \partial^2 D/\partial x^2$, where D is the wave displacement (equivalent of the wave function). In both equations there is the second derivative with respect to x; but in the Schrödinger equation there is only the first derivative with respect to time, whereas the classical wave equation has the second derivative for time.

As we pointed out in the preceding Section, we cannot derive the time-dependent Schrödinger equation. But we can show how the time-independent Schrödinger equation (Eq. 38–5) is obtained from it. For many problems in quantum mechanics, it is possible to write the wave function as a product of separate functions of space and time:

$$\Psi(x, t) = \psi(x) f(t).$$

Substituting this into the time-dependent Schrödinger equation (Eq. 38–7), we get:

$$-\frac{\hbar^2}{2m} f(t) \frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) f(t) = i\hbar\psi(x) \frac{df(t)}{dt}.$$

We divide both sides of this equation by $\psi(x) f(t)$ and obtain an equation that involves only x on one side and only t on the other:

$$-\frac{\hbar^2}{2m} \frac{1}{\psi(x)} \frac{d^2\psi(x)}{dx^2} + U(x) = i\hbar \frac{1}{f(t)} \frac{df(t)}{dt}.$$

This **separation of variables** is very convenient. Since the left side is a function only of x, and the right side is a function only of t, the equality can be valid for all values of x and all values of t only if each side is equal to a constant (the same constant), which we call C:

$$-\frac{\hbar^2}{2m} \frac{1}{\psi(x)} \frac{d^2\psi(x)}{dx^2} + U(x) = C \qquad \textbf{(38–8a)}$$

$$i\hbar \frac{1}{f(t)} \frac{df(t)}{dt} = C. \qquad \textbf{(38–8b)}$$

We multiply the first of these (Eq. 38–8a) through by $\psi(x)$ and obtain

$$-\frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = C\psi(x). \qquad \textbf{(38–8c)}$$

This we recognize immediately as the time-independent Schrödinger equation, Eq. 38–5, where the constant C equals the total energy E. Thus we have obtained the time-independent form of Schrödinger's equation from the time-dependent form.

Equation 38–8b is easy to solve. Putting $C = E$, we rewrite Eq. 38–8b as

$$\frac{df(t)}{dt} = -i\frac{E}{\hbar} f(t)$$

(note that since $i^2 = -1$, $i = -1/i$), and then as

$$\frac{df(t)}{f(t)} = -i\frac{E}{\hbar} dt.$$

We integrate both sides to obtain

$$\ln f(t) = -i\frac{E}{\hbar} t$$

or

$$f(t) = e^{-i\left(\frac{E}{\hbar}\right)t}.$$

Thus the total wave function is

$$\Psi(x,t) = \psi(x)e^{-i\left(\frac{E}{\hbar}\right)t}, \tag{38–9}$$

where $\psi(x)$ satisfies Eq. 38–5. It is, in fact, the solution of the time-independent Schrödinger equation (Eq. 38–5) that is the major task of nonrelativistic quantum mechanics. Nonetheless, we should note that in general the wave function $\Psi(x,t)$ is a complex function since it involves $i = \sqrt{-1}$. It has both a real and an imaginary part.[†] Since $\Psi(x,t)$ is not purely real, it cannot itself be physically measurable. Rather it is only $|\Psi|^2$, which *is* real, that can be measured physically.

Note also that

$$|f(t)| = |e^{-i\left(\frac{E}{\hbar}\right)t}| = 1,$$

so $|f(t)|^2 = 1$. Hence the probability density in space does not depend on time:

$$|\Psi(x,t)|^2 = |\psi(x)|^2.$$

We thus will be interested only in the time-independent Schrödinger equation, Eq. 38–5, which we will now examine for a number of simple situations.

38–7 Free Particles; Plane Waves and Wave Packets

A **free particle** is one that is not subject to any force, and we can therefore take its potential energy to be zero. (Although we dealt with the free particle in Section 38–5 in arguing for Schrödinger's equation, here we treat it directly using Schrödinger's equation as the basis.) Schrödinger's equation (Eq. 38–5) with $U(x) = 0$ becomes

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} = E\psi(x),$$

which can be written

$$\frac{d^2\psi}{dx^2} + \frac{2mE}{\hbar^2}\psi = 0.$$

This is a familiar equation that we encountered in Chapter 14 (Eq. 14–3) in connection with the simple harmonic oscillator. The solution to this equation, but with appropriate variable changes[‡] for our case here, is

$$\psi = A\sin kx + B\cos kx, \qquad \text{[free particle]} \tag{38–10}$$

where

$$k = \sqrt{\frac{2mE}{\hbar^2}}. \tag{38–11a}$$

Since $U = 0$, the total energy E of the particle is $E = \frac{1}{2}mv^2 = p^2/2m$ (where p is the momentum); thus

$$k = \frac{p}{\hbar} = \frac{h}{\lambda\hbar} = \frac{2\pi}{\lambda}. \tag{38–11b}$$

So a free particle of momentum p and energy E can be represented by a plane wave that varies sinusoidally. If we are not interested in the phase, we can choose $B = 0$ in Eq. 38–10, and we show this sine wave in Fig. 38–6a.

[†]Recall that $e^{-i\theta} = \cos\theta - i\sin\theta$.

[‡]In Eq. 14–3, t becomes x and ω becomes $k = \sqrt{2mE/\hbar^2}$. (Don't confuse this k with the spring constant k of Chapter 14.) We could also write our solution (Eq. 38–10) as $\psi = A\cos(kx + \phi)$ where ϕ is a phase constant.

PSE5_CH38_v1.7_19-Jun-18

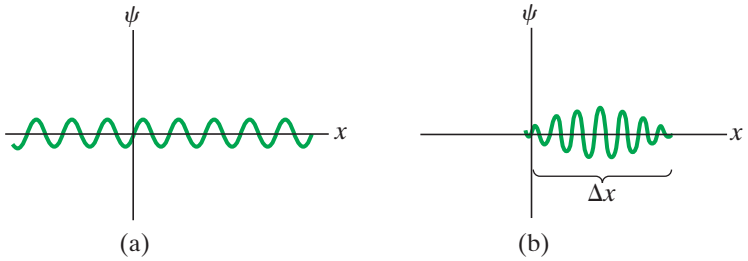


FIGURE 38–6 (a) A plane wave describing a free particle. (b) A wave packet of "width" Δx.

EXAMPLE 38–4 **Free electron.** An electron with energy $E = 6.3\,\text{eV}$ is in free space (where $U = 0$). Find (*a*) the wavelength λ and (*b*) the wave function ψ for the electron, assuming $B = 0$.

APPROACH The wavelength $\lambda = 2\pi/k$ (Eq. 38–11b) where the wave number k is given by Eq. 38–11a. The wave function is given by $\psi = A \sin kx$.

SOLUTION

$$(a)\ \lambda = \frac{2\pi}{k} = \frac{2\pi\hbar}{\sqrt{2mE}}$$

$$= \frac{2\pi(1.055\times10^{-34}\,\text{J}\cdot\text{s})}{\sqrt{2(9.11\times10^{-31}\,\text{kg})(6.3\,\text{eV})(1.60\times10^{-19}\,\text{J/eV})}}$$

$$= 4.9\times10^{-10}\,\text{m} = 0.49\,\text{nm}.$$

$$(b)\ k = \frac{2\pi}{\lambda}$$

$$= 1.28\times10^{10}\,\text{m}^{-1},$$

so

$$\psi = A\sin kx = A\sin\left[(1.28\times10^{10}\,\text{m}^{-1})(x)\right].$$

Note in Fig. 38–6a that the sine wave will extend indefinitely[†] in the $+x$ and $-x$ directions. Thus, since $|\psi|^2$ represents the probability of finding the particle, the particle could be anywhere between $x = -\infty$ and $x = \infty$. This is fully consistent with the uncertainty principle (Section 38–3): the momentum of the particle was given and hence is known precisely $(p = \hbar k)$, so the particle's position must be totally unpredictable. Mathematically, if $\Delta p = 0$, $\Delta x \gtrsim \hbar/\Delta p = \infty$.

To describe a particle whose position is well localized—that is, it is known to be within a small region of space—we can use the concept of a **wave packet**. Figure 38–6b shows an example of a wave packet whose width is about Δx as shown, meaning that the particle is most likely to be found within this region of space. A well-localized particle moving through space can thus be represented by a moving wave packet.

A wave packet can be represented mathematically as the sum of many plane waves (sine waves) of slightly different wavelengths. That this will work can be seen by looking carefully at Fig. 16–17. There we combined only two nearby frequencies (to explain why there are "beats") and found that the sum of two sine waves looked like a series of wave packets. If we add additional waves with other nearby frequencies, we can eliminate all but one of the packets and arrive at Fig. 38–6b. Thus a wave packet consists of waves of a *range* of wavelengths; hence it does not have a definite momentum $p\ (= h/\lambda)$, but rather, a range of momenta. This is consistent with the uncertainty principle: we have made Δx small, so the momentum cannot be precise; that is, Δp cannot be zero. Instead, our particle can be said to have a range of momenta, Δp, or to have an uncertainty in its momentum, Δp. It is not hard to show, even for this simple situation (see Problem 21), that $\Delta p \approx h/\Delta x$, in accordance with the uncertainty principle.

[†]Such an infinite wave makes problems for normalization since $\int_{-\infty}^{\infty}|\psi|^2\,dx = A^2\int_{-\infty}^{\infty}\sin^2 kx\,dx$ is infinite for any nonzero value for A. For practical purposes we can usually normalize the waves $(A \neq 0)$ by assuming that the particle is in a large but finite region of space. The region can be chosen large enough so that momentum is still rather precisely fixed.

38–8 Particle in an Infinitely Deep Square Well Potential (a Rigid Box)

The Schrödinger equation can be solved analytically only for a few possible forms of the potential energy U. We consider some simple cases here which at first may not seem realistic, but have simple solutions that can be used as approximations to understand a variety of phenomena.

In our first case, we assume that a particle of mass m is confined to a one-dimensional box of width ℓ whose walls are perfectly rigid. (This can serve as an approximation for an electron in a metal, for example.) The particle is trapped in this box and collisions with the walls are perfectly elastic. The potential energy for this situation, which is commonly known as an **infinitely deep square well potential** or **rigid box**, is shown in Fig. 38–7. We can write the potential energy $U(x)$ as

$$U(x) = 0 \qquad 0 < x < \ell$$
$$U(x) = \infty \qquad x \le 0 \quad \text{and} \quad x \ge \ell.$$

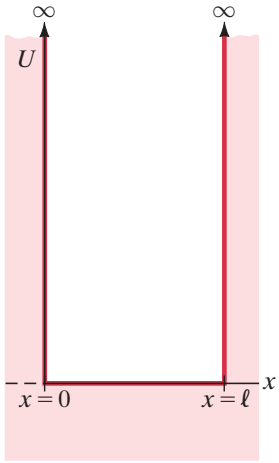


FIGURE 38–7 A plot of potential energy U vs. x for an infinitely deep square well potential.

For the region $0 < x < \ell$, where $U(x) = 0$, we already know the solution of the Schrödinger equation from our discussion in Section 38–7: it is just Eq. 38–10,

$$\psi(x) = A \sin kx + B \cos kx,$$

where (from Eq. 38–11a)

$$k = \sqrt{\frac{2mE}{\hbar^2}}.$$

(We could also use $\psi(x) = A \sin(kx + \phi)$ where ϕ is a phase constant.) Outside the well $U(x) = \infty$, so $\psi(x)$ must be zero. (If it weren't, the product $U\psi$ in the Schrödinger equation wouldn't be finite; besides, if $U = \infty$, we can't expect a particle of finite total energy to be in such a region.) So we are concerned only with the wave function within the well, and we must determine the constants A and B as well as any restrictions on the value of k (and hence on the energy E).

We insist that the wave function be continuous. Hence, if $\psi = 0$ outside the well, it must be zero at $x = 0$ and at $x = \ell$:

$$\psi(0) = 0 \quad \text{and} \quad \psi(\ell) = 0.$$

These are the **boundary conditions** for this problem. At $x = 0$, $\sin kx = 0$ but $\cos kx = 1$, so at this point Eq. 38–10 yields

$$0 = \psi(0) = A \sin 0 + B \cos 0 = 0 + B.$$

Thus B must be zero. Our solution is reduced to

$$\psi(x) = A \sin kx.$$

Now we apply the other boundary condition, $\psi = 0$ at $x = \ell$:

$$0 = \psi(\ell) = A \sin k\ell.$$

We don't want $A = 0$ or we won't have a particle at all ($|\psi|^2 = 0$ everywhere). Therefore, we set

$$\sin k\ell = 0.$$

The sine is zero for angles of $0, \pi, 2\pi, 3\pi, \cdots$ radians, which means that $k\ell = 0, \pi, 2\pi, 3\pi, \cdots$. In other words,

$$k\ell = n\pi, \qquad n = 1, 2, 3, \cdots, \tag{38–12}$$

where n is an integer. We eliminate the case $n = 0$ since that would make $\psi = 0$ everywhere. Thus k, and hence E, cannot have just any value; rather, k is limited to values

$$k = \frac{n\pi}{\ell}.$$

PSE5_CH38_v1.7_19-Jun-18

Putting this expression in Eq. 38–11a (and substituting $h/2\pi$ for $\hbar$), we find that E can have only the values

$$E = n^2 \frac{h^2}{8m\ell^2}, \qquad n = 1, 2, 3, \cdots. \tag{38–13}$$

A particle trapped in a rigid box thus can only have certain (quantized) *energies*. The lowest energy (the ground state) has $n = 1$ and is given by

$$E_1 = \frac{h^2}{8m\ell^2}. \qquad \text{[ground state]}$$

The next highest energy $(n = 2)$ is

$$E_2 = 4E_1,$$

and for higher energies (see Fig. 38–8),

$$E_3 = 9E_1$$
$$\vdots$$
$$E_n = n^2 E_1.$$

The integer n is called the **quantum number** of the state.

That the lowest energy, E_1, is not zero means that the particle in the box can never be at rest. This is contrary to classical ideas, according to which a particle can have $E = 0$. E_1 is called the **zero-point energy**. One outcome of this result is that even at a temperature of absolute zero $(0\ \text{K})$, quantum mechanics predicts that particles in a box would not be at rest but would have a zero-point energy.

We also note that both the energy E_1 and momentum $p_1 = \hbar k = \hbar\pi/\ell$ (Eq. 38–11b) in the ground state are related inversely to the width of the box. The smaller the width ℓ, the larger the momentum (and energy). This can be considered a direct result of the uncertainty principle (see Problem 26).

The wave function $\psi = A \sin kx$ for each of the quantum states is (since $k = n\pi/\ell$)

$$\psi_n = A \sin\left(\frac{n\pi}{\ell} x\right). \tag{38–14}$$

We can determine the constant A by imposing the normalization condition (Eq. 38–6c):

$$1 = \int_{-\infty}^{\infty} \psi^2\, dx = \int_0^{\ell} A^2 \sin^2\left(\frac{n\pi}{\ell} x\right) dx, \tag{38–15}$$

where the integral needs to be done only over the range $0 < x < \ell$ because outside these limits $\psi = 0$. The integral (see Example 38–6) is equal to $A^2\ell/2$, so we have

$$A = \sqrt{\frac{2}{\ell}} \qquad \text{and} \qquad \psi_n = \sqrt{\frac{2}{\ell}} \sin\left(\frac{n\pi}{\ell} x\right).$$

The amplitude A is the same for all the quantum numbers. Figure 38–9 shows the wave functions (Eq. 38–14) for $n = 1, 2, 3$, and 10. They look just like standing waves on a string—see Fig. 15–26. This is not surprising since the wave function solutions, Eq. 38–14, are the same as for the standing waves on a string, and the condition $k\ell = n\pi$ is the same in the two cases (page 448).

Figure 38–10 shows the probability distribution, $|\psi|^2$, for the same states $(n = 1, 2, 3, 10)$ for which ψ is shown in Fig. 38–9. We see immediately that the particle is more likely to be found in some places than in others. For example, in the ground state $(n = 1)$, the electron is much more likely to be found near the center of the box than near the walls. This is clearly at variance with classical ideas, which predict a uniform probability density—the particle would be as likely to be found at one point in the box as at any other. The quantum-mechanical probability densities for higher states are even more complicated, with areas of low probability not only near the walls but also at regular intervals in between.

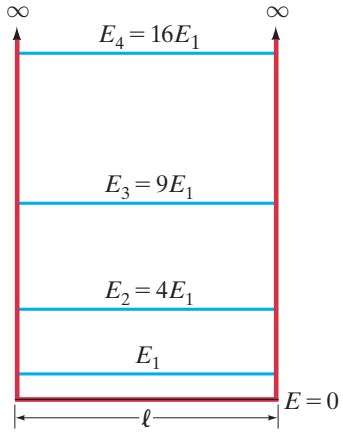


FIGURE 38–8 Possible energy levels for a particle in a box with perfectly rigid walls (infinite square well potential).

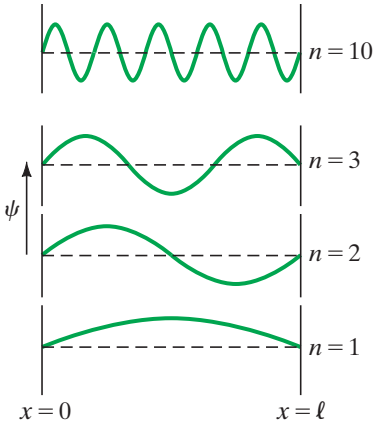
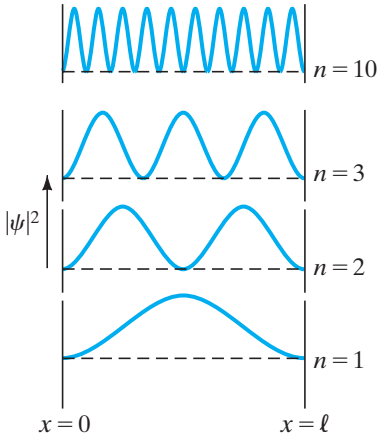


FIGURE 38–9 Wave functions corresponding to the quantum number n being 1, 2, 3, and 10 for a particle confined to a rigid box.

FIGURE 38–10 The probability distribution for a particle in a rigid box for the states with $n = 1, 2, 3$, and 10.



EXAMPLE 38–5 **Electron in an infinite potential well.** (*a*) Calculate the three lowest energy levels for an electron trapped in an infinitely deep square well potential of width $\ell = 1.00 \times 10^{-10}\,\text{m}$ (about the diameter of a hydrogen atom in its ground state). (*b*) If a photon were emitted when the electron jumps from the $n = 2$ state to the $n = 1$ state, what would its wavelength be?

APPROACH The energy levels are given by Eq. 38–13. In (*b*), $hf = hc/\lambda = E_2 - E_1$.

SOLUTION (*a*) The ground state $(n = 1)$ has energy

$$E_1 = \frac{h^2}{8m\ell^2} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})^2}{8(9.11 \times 10^{-31}\,\text{kg})(1.00 \times 10^{-10}\,\text{m})^2} = 6.03 \times 10^{-18}\,\text{J}.$$

In electron volts this is

$$E_1 = \frac{6.03 \times 10^{-18}\,\text{J}}{1.60 \times 10^{-19}\,\text{J/eV}} = 37.7\,\text{eV}.$$

Then

$$E_2 = (2)^2 E_1 = 151\,\text{eV}$$
$$E_3 = (3)^2 E_1 = 339\,\text{eV}.$$

(*b*) The energy difference is $E_2 - E_1 = 151\,\text{eV} - 38\,\text{eV} = 113\,\text{eV}$ or $1.81 \times 10^{-17}\,\text{J}$, and this would equal the energy of the emitted photon (energy conservation). Its wavelength would be

$$\lambda = \frac{c}{f} = \frac{hc}{E} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.00 \times 10^{8}\,\text{m/s})}{1.81 \times 10^{-17}\,\text{J}} = 1.10 \times 10^{-8}\,\text{m}$$

or 11.0 nm, which is in the ultraviolet region of the spectrum.

EXERCISE C The wavelength of a photon emitted in an $n = 3$ to $n = 1$ transition is (*a*) 0.062 nm, (*b*) 620 nm, (*c*) 301 nm, (*d*) $3.2 \times 10^{-15}\,\text{m}$, (*e*) 4.1 nm.

EXAMPLE 38–6 **Calculating a normalization constant.** Show that the normalization constant A for all wave functions describing a particle in an infinite potential well of width ℓ has a value of $A = \sqrt{2/\ell}$.

APPROACH The wave functions for various n are

$$\psi = A \sin\frac{n\pi x}{\ell}.$$

To normalize ψ, we must have (Eq. 38–15)

$$1 = \int_0^\ell |\psi|^2\,dx = \int_0^\ell A^2 \sin^2\frac{n\pi x}{\ell}\,dx.$$

SOLUTION We need integrate only from 0 to ℓ since $\psi = 0$ for all other values of x. To evaluate this integral we let $\theta = n\pi x/\ell$ and use the trigonometric identity $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$. Then, with $dx = \ell\,d\theta/n\pi$, we have

$$1 = A^2 \int_0^{n\pi} \sin^2\theta \left(\frac{\ell}{n\pi}\right) d\theta = \frac{A^2\ell}{2n\pi} \int_0^{n\pi} (1 - \cos 2\theta)\,d\theta$$
$$= \frac{A^2\ell}{2n\pi}\left(\theta - \tfrac{1}{2}\sin 2\theta\right)\Big|_0^{n\pi}$$
$$= \frac{A^2\ell}{2}.$$

Thus $A^2 = 2/\ell$ and

$$A = \sqrt{\frac{2}{\ell}}.$$

EXAMPLE 38–7 **ESTIMATE** **Probability near center of rigid box.** An electron is in an infinitely deep square well potential of width $\ell = 1.00 \times 10^{-10}\,\text{m}$. If the electron is in the ground state, what is the probability of finding it in a region of width $\Delta x = 1.0 \times 10^{-12}\,\text{m}$ at the center of the well (at $x = 0.50 \times 10^{-10}\,\text{m}$)?

APPROACH The probability of finding a particle in a small region of width dx is $|\psi|^2\,dx$ (Eq. 38–6b). Using A from Example 38–6, the wave function for the ground state is

$$\psi(x) = \sqrt{\frac{2}{\ell}}\sin\frac{\pi x}{\ell}.$$

SOLUTION The $n = 1$ curve in Fig. 38–9 shows that ψ is roughly constant near the center of the well. So we can avoid doing an integral over dx and just set $dx \approx \Delta x$ and find

$$|\psi|^2\,\Delta x = \frac{2}{\ell}\sin^2\left[\frac{\pi x}{\ell}\right]\Delta x$$

$$= \frac{2}{(1.00 \times 10^{-10}\,\text{m})}\sin^2\left[\frac{\pi(0.50 \times 10^{-10}\,\text{m})}{(1.00 \times 10^{-10}\,\text{m})}\right](1.0 \times 10^{-12}\,\text{m}) = 0.02.$$

The probability of finding the electron in this region at the center of the well is thus 2%.

NOTE Since $\Delta x = 1.0 \times 10^{-12}\,\text{m}$ is 1% of the well width of $1.00 \times 10^{-10}\,\text{m}$, our result of 2% probability is not what would be expected classically. Classically, the electron would be equally likely to be anywhere in the box, and we would expect the probability to be 1% instead of 2%.

EXAMPLE 38–8 **Probability of e^- in $\frac{1}{4}$ of box.** Determine the probability of finding an electron in the left quarter of a rigid box—i.e., between one wall at $x = 0$ and position $x = \ell/4$. Assume the electron is in the ground state.

APPROACH We cannot make the assumption we did in Example 38–7 that $|\psi|^2 \approx$ constant and Δx is small. Here we need to integrate $|\psi|^2\,dx$ from $x = 0$ to $x = \ell/4$, which is equal to the area under the curve shown colored in Fig. 38–11.

SOLUTION The wave function in the ground state is $\psi_1 = \sqrt{2/\ell}\sin(\pi x/\ell)$. To find the probability of the electron in the left quarter of the box, we integrate just as in Example 38–6 but with different limits on the integral (and now we know that $A = \sqrt{2/\ell}$). That is, we set $\theta = \pi x/\ell$ (then $x = \ell/4$ corresponds to $\theta = \pi/4$) and use the identity $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$. Thus, with $dx = (\ell/\pi)\,d\theta$,

$$\int_0^{\ell/4} |\psi|^2\,dx = \frac{2}{\ell}\int_0^{\ell/4}\sin^2\left(\frac{\pi}{\ell}x\right)dx$$

$$= \frac{1}{\ell}\int_0^{\pi/4}(1 - \cos 2\theta)\left(\frac{\ell}{\pi}\right)d\theta$$

$$= \frac{1}{\pi}\left(\theta - \frac{1}{2}\sin 2\theta\right)\Big|_0^{\pi/4}$$

$$= \frac{1}{4} - \frac{1}{2\pi} = 0.091.$$

NOTE The electron spends only 9.1% of its time in the left quarter of the box. Classically it would spend 25%.

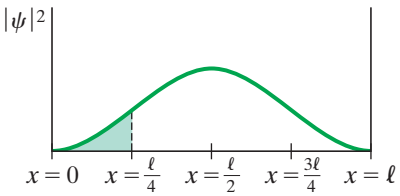


FIGURE 38–11 Ground-state probability distribution $|\psi|^2$ for an electron in a rigid box. Same as $n = 1$ graph of Fig. 38–10; but here we show the area under the curve from $x = 0$ to $x = \ell/4$ which represents the probability of finding the electron in that region.

EXERCISE D What is the probability of finding the electron between $x = \ell/4$ and $x = \ell/2$? (Do you need to integrate?) (*a*) 9.1%; (*b*) 18.2%; (*c*) 25%; (*d*) 33%; (*e*) 40.9%.

EXAMPLE 38–9 **Most likely and average positions.** Two interesting quantities are the most likely position of a particle and the average position of the particle. Consider an electron in a rigid box of width $1.00 \times 10^{-10}\,\text{m}$ in the first excited state $n = 2$. (*a*) What is its most likely position? (*b*) What is its average position?

APPROACH To find (*a*) the most likely position (or positions), we find the maximum value(s) of the probability distribution $|\psi|^2$ by taking its derivative and setting it equal to zero. For (*b*), the average position, we integrate $\bar{x} = \int_0^\ell x\,|\psi|^2\,dx$.

SOLUTION (*a*) The wave function for $n = 2$ is $\psi(x) = \sqrt{\frac{2}{\ell}}\sin\left(\frac{2\pi}{\ell}x\right)$, so $|\psi(x)|^2 = \frac{2}{\ell}\sin^2\left(\frac{2\pi}{\ell}x\right)$. To find maxima and minima, we set $d|\psi|^2/dx = 0$:

$$\frac{d}{dx}|\psi|^2 = \frac{2}{\ell}(2)\frac{2\pi}{\ell}\sin\left(\frac{2\pi}{\ell}x\right)\cos\left(\frac{2\pi}{\ell}x\right).$$

This quantity is zero when either the sine is zero $(2\pi x/\ell = 0, \pi, 2\pi, \cdots)$, or the cosine is zero $(2\pi x/\ell = \pi/2, 3\pi/2, \cdots)$. The maxima and minima occur at $x = 0,\ \ell/2,\ \ell$, and $x = \ell/4,\ 3\ell/4$. The latter $(\ell/4, 3\ell/4)$ are the maxima—see the $n = 2$ curve of Fig. 38–10; the others are minima. [To confirm, you can take the second derivative, $d^2|\psi|^2/dx^2$, which is > 0 for minima and < 0 for maxima.] (*b*) The average position is (again we use $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$):

$$\bar{x} = \int_0^\ell x|\psi|^2\,dx = \int_0^\ell \frac{2}{\ell}x\sin^2\left(\frac{2\pi}{\ell}x\right)dx = \frac{1}{\ell}\int_0^\ell x\left[1 - \cos\left(\frac{4\pi}{\ell}x\right)\right]dx,$$

which gives (integrating by parts, Appendix B: $u = x,\ dv = \cos(4\pi x/\ell)\,dx$):

$$\bar{x} = \frac{1}{\ell}\left[\frac{x^2}{2} - \frac{x\ell}{4\pi}\sin\left(\frac{4\pi}{\ell}x\right) - \frac{\ell^2}{16\pi^2}\cos\left(\frac{4\pi}{\ell}x\right)\right]_0^\ell = \frac{\ell}{2}.$$

Since the curves for $|\psi|^2$ are symmetric about the center of the box, we expect this answer. But note that for $n = 2$, the probability of finding the particle at the point $x = \ell/2$ is actually zero (Fig. 38–10).

EXAMPLE 38–10 **ESTIMATE** **Confined bacterium.** A tiny bacterium with a mass of about $10^{-14}\,\text{kg}$ is confined between two rigid walls 0.1 mm apart. (*a*) Estimate its minimum speed. (*b*) If, instead, its speed is about 1 mm in 100 s, estimate the quantum number of its state.

APPROACH We assume $U = 0$ inside the potential well, so $E = \frac{1}{2}mv^2$. In (*a*) the minimum speed occurs in the ground state, $n = 1$, so $v = \sqrt{2E/m}$ where E is the ground-state energy. In (*b*) we solve Eq. 38–13 for n.

SOLUTION (*a*) With $n = 1$, Eq. 38–13 gives $E = h^2/8m\ell^2$ so

$$v = \sqrt{\frac{2E}{m}} = \sqrt{\frac{h^2}{4m^2\ell^2}} = \frac{h}{2m\ell} = \frac{6.6 \times 10^{-34}\,\text{J}\cdot\text{s}}{2(10^{-14}\,\text{kg})(10^{-4}\,\text{m})} \approx 3 \times 10^{-16}\,\text{m/s}.$$

This is a speed so small that we could not measure it and the object would seem at rest, consistent with classical physics.

(*b*) Given $v = 10^{-3}\,\text{m}/100\,\text{s} = 10^{-5}\,\text{m/s}$, the kinetic energy of the bacterium is

$$E = \tfrac{1}{2}mv^2 = \tfrac{1}{2}(10^{-14}\,\text{kg})(10^{-5}\,\text{m/s})^2 = 0.5 \times 10^{-24}\,\text{J}.$$

From Eq. 38–13, the quantum number of this state is

$$n = \sqrt{E\left(\frac{8m\ell^2}{h^2}\right)} = \sqrt{\frac{(0.5 \times 10^{-24}\,\text{J})(8)(10^{-14}\,\text{kg})(10^{-4}\,\text{m})^2}{(6.6 \times 10^{-34}\,\text{J}\cdot\text{s})^2}} \approx 3 \times 10^{10}.$$

NOTE This number is so large that we could never distinguish between adjacent energy states (between $n = 3 \times 10^{10}$ and $3 \times 10^{10} + 1$). The energy states would appear to form a continuum. Thus, even though the energies involved here are small $(\ll 1\,\text{eV})$, we are still dealing with a macroscopic object (though visible only under a microscope) and the quantum result is not distinguishable from a classical one. This is in accordance with the correspondence principle.

38–9 Finite Potential Well

Let us now look at a particle in a box whose walls are not perfectly rigid. That is, the potential energy outside the box or well is not infinite, but rises to some level U_0, as shown in Fig. 38–12. This is called a **finite potential well**. It can serve as an approximation for, say, a neutron in a nucleus. There are some significant new features that arise for the finite well as compared to the infinite well. We divide the well into three regions as shown in Fig. 38–12. In region II, inside the well, the Schrödinger equation is the same as before $(U = 0)$, although the boundary conditions will be different. So we write the solution for region II as

$$\psi_{II} = A \sin kx + B \cos kx \qquad (0 < x < \ell)$$

but we don't immediately set $B = 0$ or assume that k is given by Eq. 38–12.

In regions I and III, the Schrödinger equation, now with $U(x) = U_0$, is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U_0\psi = E\psi.$$

We rewrite this as

$$\frac{d^2\psi}{dx^2} - \left[\frac{2m(U_0 - E)}{\hbar^2}\right]\psi = 0.$$

Let us assume that E is less than U_0, so the particle is "trapped" in the well (at least classically). There might be only one such **bound state**, or several, or even none, as we shall discuss later. We define the constant G by

$$G^2 = \frac{2m(U_0 - E)}{\hbar^2} \qquad \textbf{(38–16)}$$

and rewrite the Schrödinger equation as

$$\frac{d^2\psi}{dx^2} - G^2\psi = 0.$$

This equation has the general solution

$$\psi_{I,\,III} = Ce^{Gx} + De^{-Gx},$$

which can be confirmed by direct substitution, since

$$\frac{d^2}{dx^2}\left(e^{\pm Gx}\right) = G^2 e^{\pm Gx}.$$

In region I, x is always negative, so D must be zero (otherwise, $\psi \to \infty$ as $x \to -\infty$, giving the unacceptable result that $\int |\psi|^2\, dx \neq 1$). Similarly in region III, where x is always positive, C must be zero. Hence

$$\psi_I = Ce^{Gx} \qquad (x < 0)$$
$$\psi_{III} = De^{-Gx} \qquad (x > \ell).$$

In regions I and III, the wave function decreases exponentially with distance from the well boundaries. The mathematical forms of the wave function inside and outside the well are different, but we insist that the wave function be continuous even at the two walls. Schrödinger also imposed a second condition: that the slope of ψ, which is its first derivative, must be continuous at the walls. Hence we have the boundary conditions:

$$\psi_I = \psi_{II} \quad \text{and} \quad \frac{d\psi_I}{dx} = \frac{d\psi_{II}}{dx} \quad \text{at} \quad x = 0$$
$$\psi_{II} = \psi_{III} \quad \text{and} \quad \frac{d\psi_{II}}{dx} = \frac{d\psi_{III}}{dx} \quad \text{at} \quad x = \ell.$$

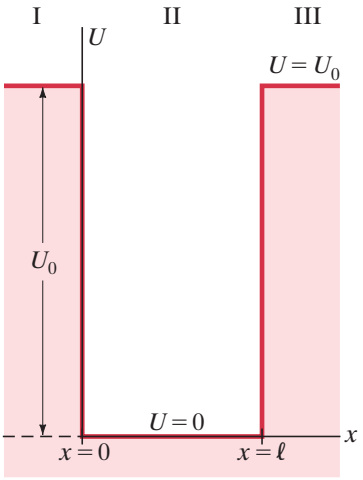


FIGURE 38–12 Potential energy U vs. x for a finite one-dimensional square well.

At the left-hand wall $(x = 0)$ these boundary conditions become

$$Ce^0 = A\sin 0 + B\cos 0 \qquad \text{or} \qquad C = B$$

and

$$GCe^0 = kA\cos 0 - kB\sin 0 \qquad \text{or} \qquad GC = kA.$$

These are two of the relations that link the constants A, B, C, D and the energy E. We get two more relations from the boundary conditions at $x = \ell$, and a fifth relation from normalizing the wave functions over all space, $\int_{-\infty}^{\infty} |\psi|^2\, dx = 1$. These five relations allow us to solve for the five unknowns, including the energy E. We will not go through the detailed mathematics, but we will discuss some of the results.

Figure 38–13a shows the wave function ψ for the three lowest possible states, and Fig. 38–13b shows the probability distributions $|\psi|^2$. We see that the wave functions are smooth at the walls of the well. Within the well ψ has the form of a sinusoidal wave; for the ground state, there is less than a half wavelength. Compare this to the infinite well (Fig. 38–9), where the ground-state wave function is exactly a half wavelength: $\lambda = 2\ell$. For our finite well, $\lambda > 2\ell$. Thus for a finite well the momentum of a particle $(p = h/\lambda)$, and hence its ground-state energy, will be less than for an infinite well of the same width ℓ.

Outside the finite well we see that the wave function drops off exponentially on either side of the walls. That ψ is not zero beyond the walls means that the particle can sometimes be found outside the well. This completely contradicts classical ideas. Outside the well, the potential energy of the particle is greater than its total energy: $U_0 > E$. This violates conservation of energy. But we clearly see in Fig. 38–13b that the particle can spend some time outside the well, where $U_0 > E$ (although the penetration into this classically forbidden region is generally not far since $|\psi|^2$ decreases exponentially with distance from either wall). The penetration of a particle into a classically forbidden region is a very important result of quantum mechanics. But how can it be? How can we accept this nonconservation of energy? We can look to the uncertainty principle, in the form

$$\Delta E\, \Delta t \gtrsim \hbar.$$

It tells us that the energy can be uncertain, and can even be nonconserved, by an amount ΔE for very short times $\Delta t \sim \hbar/\Delta E$.

Now let us consider the situation when the total energy E of the particle is greater than U_0. In this case the particle is a free particle and everywhere its wave function is sinusoidal, Fig. 38–14. Its wavelength is different outside the well than inside, as shown. Since $K = \frac{1}{2}mv^2 = p^2/2m$, the wavelength in region II $(U = 0)$ is

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE}} \qquad 0 < x < \ell,$$

whereas in regions I and III, where $p^2/2m = K = E - U_0$, the wavelength is longer and is given by

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m(E - U_0)}} \qquad x < 0 \quad \text{and} \quad x > \ell.$$

For $E > U_0$, any energy E is possible. But for $E < U_0$, as we saw above, the energy is quantized and only certain states are possible.

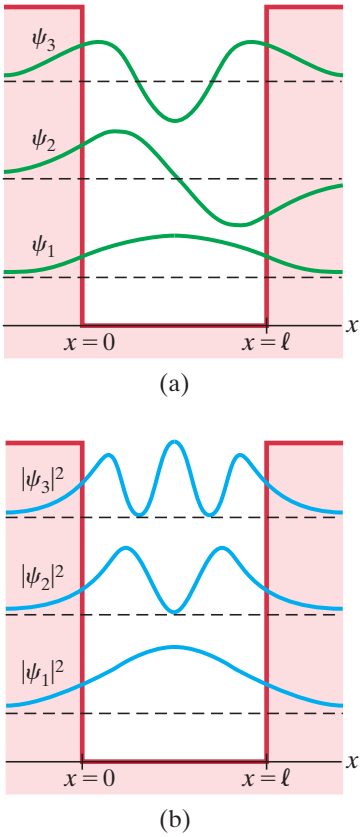


FIGURE 38–13 The wave functions (a), and probability distributions (b), for the three lowest possible states of a particle in a finite potential well. Each of the ψ and $|\psi|^2$ curves has been superposed on its energy level (dashed lines) for convenience.

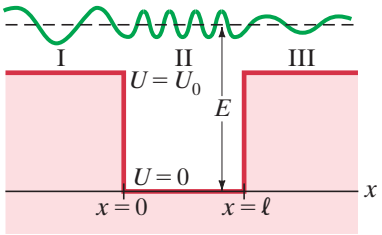


FIGURE 38–14 The wave function of a particle of energy E traveling over a potential well whose depth U_0 is less than E (measured in the well).

PSE5_CH38_v1.7_19-Jun-18

EXERCISE E An electron with energy $E = 6.0\,\text{eV}$ is near a potential well of depth $U_0 = 4.5\,\text{eV}$ and width $\ell = 10.0\,\text{nm}$. What is the wavelength of the electron when it is inside the well? (*a*) 0.50 nm; (*b*) 0.58 nm; (*c*) 1.0 nm; (*d*) 10 nm; (*e*) 20 nm.

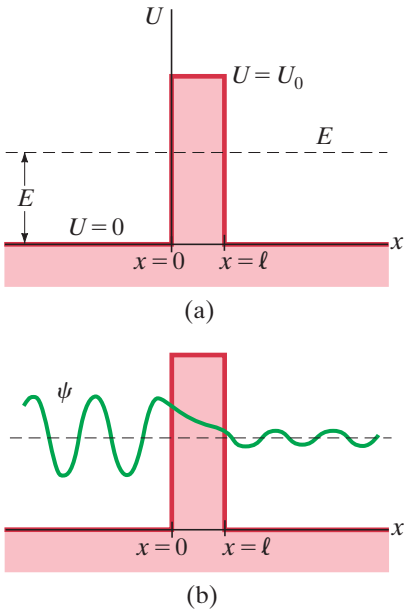
[Another very interesting but more complicated well is the **simple harmonic oscillator** which has $U(x) = \frac{1}{2}Cx^2$ as we discussed in Chapter 14. Some quantum mechanical results, such as a zero-point energy of $\frac{1}{2}\hbar\omega$ (E cannot be zero), are treated briefly in Problem 57.]

38–10 Tunneling through a Barrier

We saw in Section 38–9 that, according to quantum mechanics, a particle such as an electron can penetrate a barrier into a region forbidden by classical mechanics. There are a number of important applications of this phenomenon, particularly as applied to penetration of a thin barrier.

We consider a particle of mass m and energy E traveling to the right along the x axis in free space where the potential energy $U = 0$ so the energy is all kinetic energy ($E = K$). The particle encounters a narrow potential barrier whose height U_0 (in energy units) is greater than E, and whose thickness is ℓ (distance units); see Fig. 38–15a. Since $E < U_0$, we would expect from classical physics that the particle could not penetrate the barrier but would simply be reflected and would return in the opposite direction. Indeed, this is what happens for macroscopic objects. But quantum mechanics predicts a nonzero probability for finding the particle on the other side of the barrier. We can see how this can happen in part (b) of Fig. 38–15, which shows the wave function. The approaching particle has a sinusoidal wave function. Within the barrier the solution to the Schrödinger equation is a decaying exponential just as for the finite well of Section 38–9. However, before the exponential dies away to zero, the barrier ends (at $x = \ell$), and for $x > \ell$ there is again a sinusoidal wave function, since $U = 0$ and $E = K > 0$. But it is a sine wave of greatly reduced amplitude. Nonetheless, because $|\psi|^2$ is nonzero beyond the barrier, we see that there is a nonzero probability that the particle penetrates the barrier. This process is called **tunneling** through the barrier, or **barrier penetration**. Although we cannot observe the particle within the barrier (it would violate conservation of energy), we can detect it after it has penetrated the barrier.

FIGURE 38–15 (a) A potential barrier of height U_0 and thickness ℓ. (b) The wave function for a particle of energy $E\ (< U_0)$ that approaches from the left. The curve for ψ is superposed, for convenience, on the energy level line (dashed).



Quantitatively, we can describe the tunneling probability with a *transmission coefficient*, T, and a *reflection coefficient*, R. Suppose, for example, that $T = 0.03$ and $R = 0.97$; then if 100 particles struck the barrier, on the average 3 would tunnel through and 97 would be reflected. Note that $T + R = 1$, since an incident particle must either reflect or tunnel through. The transmission coefficient can be determined by writing the wave function for each of the three regions, just as we did for the finite well, and then applying the boundary conditions that ψ and $d\psi/dx$ must be continuous at the edges of the barrier ($x = 0$ and $x = \ell$). The calculation shows (see Problem 46) that if T is small ($\ll 1$), then

$$T \approx e^{-2G\ell}, \tag{38–17a}$$

where

$$G = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}}. \tag{38–17b}$$

(This is the same G as in Section 38–9, Eq. 38–16.) We note that increasing the height of the barrier, U_0, or increasing its thickness, ℓ, will drastically reduce T. Indeed, for macroscopic situations, T is extremely small, in accord with classical physics, which predicts no tunneling (again the correspondence principle).

EXAMPLE 38–11 **Barrier penetration.** A 50-eV electron approaches a square barrier 70 eV high and (*a*) 1.0 nm thick, (*b*) 0.10 nm thick. What is the probability that the electron will tunnel through?

APPROACH We convert eV to joules and use Eqs. 38–17.

SOLUTION (*a*) Inside the barrier $U_0 - E = (70\,\text{eV} - 50\,\text{eV})(1.6 \times 10^{-19}\,\text{J/eV}) = 3.2 \times 10^{-18}\,J$. Then, using Eqs. 38–17, we have

$$2G\ell = 2\sqrt{\frac{2(9.11 \times 10^{-31}\,\text{kg})(3.2 \times 10^{-18}\,\text{J})}{(1.055 \times 10^{-34}\,\text{J}\cdot\text{s})^2}}\,(1.0 \times 10^{-9}\,\text{m}) = 46$$

and

$$T = e^{-2G\ell} = e^{-46} \approx 1 \times 10^{-20},$$

which is extremely small.

(*b*) or $\ell = 0.10\,\text{nm}$, $2G\ell = 4.6$ and

$$T = e^{-4.6} = 0.010.$$

Thus the electron has a 1% chance of penetrating a 0.1-nm-thick barrier, but only 1 chance in 10^{20} to penetrate a 1-nm barrier. By reducing the barrier thickness by a factor of 10, the probability of tunneling through increases 10^{18} times! Clearly the transmission coefficient is extremely sensitive to the values of ℓ, $U_0 - E$, and m.

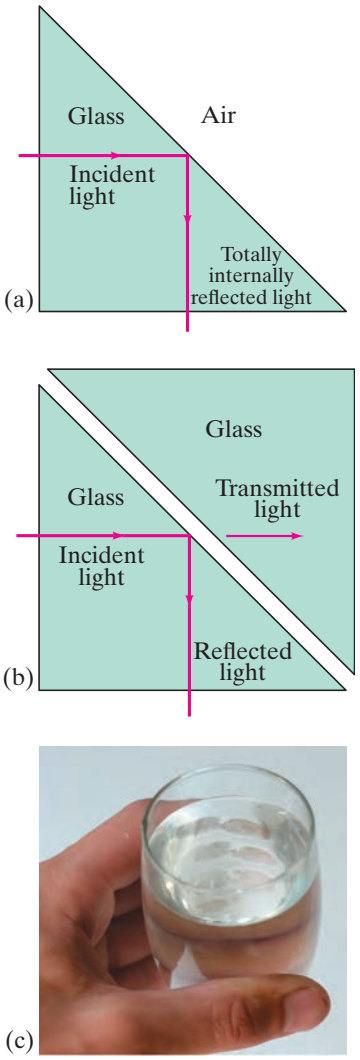


FIGURE 38–16 (a) Light traveling in glass strikes the interface with air at an angle greater than the critical angle, and is totally internally reflected. (b) A small amount of light tunnels through a narrow air gap between two pieces of glass. (c) caption to come.

Tunneling of Light Wave

Tunneling is a result of the wave properties of material particles, and also occurs for classical waves. For example, we saw in Section 32–7 that, when light traveling in glass strikes a glass–air boundary at an angle greater than the critical angle, the light is 100% totally reflected. We studied this phenomenon of total internal reflection from the point of view of ray optics, and we show it here in Fig. 38–16a. The wave theory, however, predicts that waves actually penetrate the air for a few wavelengths—almost as if they "needed" to pass the interface to find out there is air beyond and hence need to be totally reflected. Indeed, if a second piece of glass is brought near the first as shown in Fig. 38–16b, a transmitted wave that has tunneled through the air gap can be experimentally observed. You can actually observe this for yourself by looking down into a glass of water at an angle such that light entering your eye has been totally internally reflected from the (outer) glass surface (it will look silvery). (See Fig. 38–16c.)

If you press a moistened fingertip against the glass, you can see the whorls of the ridges on your fingerprints, because at the ridges you have interfered with the total internal reflection at the outer surface of the glass. So you see light that has penetrated the gap and reflected off the ridges on your finger.

Applications of Tunneling

Tunneling thus occurs even for classical waves. What is new in quantum mechanics is that material particles have wave properties and hence can tunnel. Tunneling has provided the basis for a number of useful devices, as well as helped to explain a number of important phenomena, some of which we mention briefly now.

Some atomic nuclei undergo **radioactive decay** by the emission of an alpha (α) particle, which consists of two protons and two neutrons. Inside a radioactive nucleus, we can imagine that the protons and neutrons are moving about, and sometimes two of each come together and form this stable entity, the alpha particle. We will study alpha decay in more detail in Chapter 41, but for now we note that the potential energy diagram for the alpha particle inside this type of nucleus looks something like Fig. 38–17. The square well represents the attractive nuclear force that holds the nucleus together. To this is added the $1/r$ Coulomb potential energy of repulsion between the positive alpha particle and the remaining positively charged nucleus. The barrier that results is called the **Coulomb barrier**. The wave function for the tunneling particle shown must have energy greater than zero (or the barrier would be infinitely wide and tunneling could not occur), but less than the height of the barrier. If the alpha particle had energy higher than the barrier, it would always be free and the original nucleus wouldn't exist. Thus the barrier keeps the nucleus together, but occasionally a nucleus of this type can decay by the tunneling of an alpha particle. The probability of an α particle escaping, and hence the "lifetime" of a nucleus, depends on the height and width of the barrier, and can take on a very wide range of values for only a limited change in barrier width as we saw in Example 38–11. Lifetimes of α-decaying radioactive nuclei range from less than 1 μs to 10^{10} yr.

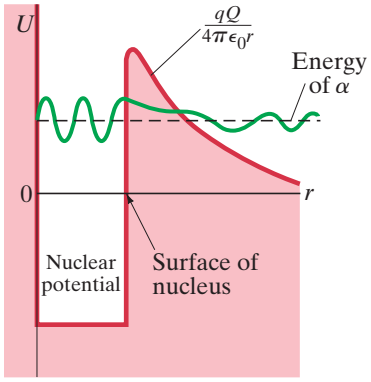
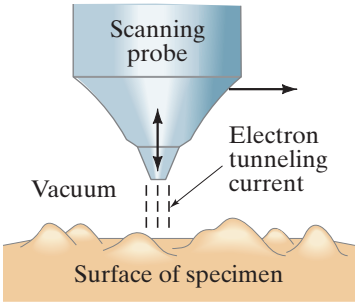


FIGURE 38–17 Potential energy seen by an alpha particle (charge q) in presence of nucleus (charge Q), showing the wave function for tunneling out.

A so-called **tunnel diode** is an electronic device made of two types of semiconductor carrying opposite-sign charge carriers, separated by a very thin neutral region. Current can tunnel through this thin barrier and can be controlled by the voltage applied to it, which affects the height of the barrier.

The **scanning tunneling electron microscope** (STM), developed in the 1980s, makes use of tunneling through a vacuum.[†] A tiny probe, whose tip may be only one (or a few) atoms wide, is moved across the specimen to be examined in a series of linear passes. The tip, as it scans, remains very close to the surface of the specimen, about 1 nm above it, Fig. 38–18. A small voltage applied between the probe and the surface causes electrons to tunnel through the vacuum between them. This "tunneling" current is very sensitive to the gap width (see Example 38–11), so a feedback mechanism can be used to raise and lower the probe so as to maintain a constant electron tunneling current. The probe's vertical motion, following the surface of the specimen, is then plotted as a function of position, scan after scan, producing a three-dimensional image of the surface. Surface features as fine as the size of an atom can be resolved: a resolution better than 50 pm (0.05 nm) laterally and 10^{-2} to 10^{-3} nm vertically. This kind of resolution has given a great impetus to the study of the surface structure of materials. The "topographic" image of a surface actually represents the distribution of electron charge (electron wave probability distributions) on the surface.

FIGURE 38–18 The probe tip of a scanning tunneling microscope, as it is moved horizontally, automatically moves up and down to maintain a constant tunneling current, thus producing an image of the surface.



The new **atomic force microscope** (AFM), developed in the 1980s, is in many ways similar to an STM, but can be used on a wider range of sample materials. Instead of detecting an electric current, the AFM measures the force between a cantilevered tip and the sample, a force which depends strongly on the tip–sample separation at each point. The tip is moved as for the STM.

[†]Other types of electron microscopes were discussed in Chapter 37.

PSE5_CH38_v1.7_19-Jun-18

Summary

In 1925, Schrödinger and Heisenberg separately worked out a new theory, **quantum mechanics**, which is now considered to be the fundamental theory at the atomic level. It is a statistical theory rather than a deterministic one.

An important aspect of quantum mechanics is the Heisenberg **uncertainty principle**. It results from the wave–particle duality and the unavoidable interaction between an observed object and the observer.

One form of the uncertainty principle states that the position x and momentum p_x of an object cannot both be measured precisely at the same time. The products of the uncertainties, $(\Delta x)(\Delta p_x)$, can be no less than $\hbar$ $(=h/2\pi)$:

$$(\Delta x)(\Delta p_x) \gtrsim \hbar. \qquad \textbf{(38–1)}$$

Another form of the uncertainty principle states that the energy can be uncertain (or nonconserved) by an amount ΔE for a time Δt where

$$(\Delta E)(\Delta t) \gtrsim \hbar. \qquad \textbf{(38–2)}$$

A particle such as an electron is represented by a **wave function** ψ. The square of the wave function, $|\psi|^2$, at any point in space represents the **probability** of finding the particle at that point. The wave function must be **normalized**, meaning that $\int |\psi|^2\, dV$ over all space must equal 1, since the particle must be found at one place or another.

In nonrelativistic quantum mechanics, ψ satisfies the **Schrödinger equation**:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U\psi = E\psi, \qquad \textbf{(38–5)}$$

here in its one-dimensional time-independent form, where U is the potential energy as a function of position and E is the total energy of the particle.

A **free particle** subject to no forces has a sinusoidal wave function $\psi = A\sin kx + B\cos kx$ with $k = p/\hbar$ and p is the particle's momentum. Such a wave of fixed momentum is spread out indefinitely in space as a plane wave.

A **wave packet**, localized in space, is a superposition of sinusoidal waves with a range of momenta.

For a particle confined to an **infinitely deep square well potential**, or **rigid box**, the Schrödinger equation gives the wave functions as

$$\psi = A\sin k\ell,$$

where ℓ is the width of the box, $A = \sqrt{2/\ell}$, and $k = n\pi/\ell$ with n an integer, as solutions inside the well. The energy is quantized,

$$E = \frac{\hbar^2 k^2}{2m} = \frac{n^2 h^2}{8m\ell^2}. \qquad \textbf{(38–13)}$$

In a **finite potential well**, the wave function extends into the classically forbidden region where the total energy is less than the potential energy. That this is possible is consistent with the uncertainty principle. The solutions to the Schrödinger equation in these areas are decaying exponentials.

Because quantum-mechanical particles can penetrate such classically forbidden areas, they can **tunnel** through thin barriers even though the potential energy in the barrier is greater than the total energy of the particle.

Questions

1. Compare a matter wave ψ to (a) a wave on a string, (b) an EM wave. Discuss similarities and differences.
2. Explain why Bohr's model of the atom is not compatible with quantum mechanics, particularly the uncertainty principle.
3. Explain why it is that the more massive an object is, the easier it becomes to predict its future position.
4. In view of the uncertainty principle, why does a baseball seem to have a well-defined position and speed, whereas an electron does not?
5. Would it ever be possible to balance a very sharp needle precisely on its point? Explain.
6. When you check the pressure in a tire, doesn't some air inevitably escape? Is it possible to avoid this escape of air altogether? What is the relation to the uncertainty principle?
7. It has been said that the ground-state energy in the hydrogen atom can be precisely known but that the excited states have some uncertainty in their values (an "energy width"). Is this consistent with the uncertainty principle in its energy form? Explain.
8. If Planck's constant were much larger than it is, how would this affect our everyday life?
9. In what ways is Newtonian mechanics contradicted by quantum mechanics? In what ways is Newtonian mechanics still relevant?
10. If you knew the position of an object precisely, with no uncertainty, how well would you know its momentum?
11. A cold thermometer is placed in a hot bowl of soup. Will the temperature reading of the thermometer be the same as the temperature of the hot soup before the measurement was made? Explain.
12. Does the uncertainty principle set a limit to how well you can make any single measurement of position? Explain.
13. Discuss the connection between the zero-point energy for a particle in a rigid box and the uncertainty principle.
14. The wave function for a particle in a rigid box is zero at points within the box (except for $n = 1$). Does this mean that the probability of finding the particle at these points is zero? Does it mean that the particle cannot pass by these points? Explain.
15. What does the probability density look like for a particle in an infinite potential well for large values of n, say $n = 100$ or $n = 1000$? As n becomes very large, do your predictions approach classical predictions in accord with the correspondence principle?
16. For a particle in an infinite potential well the separation between energy states increases as n increases (see Eq. 38–13). But doesn't the correspondence principle require closer spacing between states as n increases so as to approach a classical (nonquantized) situation? Explain.
17. A particle is trapped in an infinite potential well. Describe what happens to the particle's ground-state energy and wave function as the potential walls become finite and get lower and lower until they finally reach zero ($U = 0$ everywhere).

PSE5_CH38_v1.7_19-Jun-18

MisConceptual Questions

1. If a parallel beam of electrons is fired throught a slit,
(*a*) the electrons can be deflected because of their wave properties.
(*b*) only electrons that hit the edge of the slit are deflected.
(*c*) electrons can interact with electromagnetic waves in the slit, forming a diffraction pattern.
(*d*) the probability of an electron making it through the slit depends on the uncertainty principle.

2. According to the uncertainty principle,
(*a*) there is always an uncertainty in a measurement of the position of a particle.
(*b*) there is always an uncertainty in a measurement of the momentum of a particle.
(*c*) there is always an uncertainty in a simultaneous measurement of both the position and momentum of a particle.
(*d*) All of the above.

3. Which of the following is the best paraphrasing of the Heisenberg uncertainty principle?
(*a*) Only if you know the exact position of a particle can you know the exact momentum of a particle.
(*b*) The larger the momentum of a particle, the smaller the position of the particle.
(*c*) The more precisely you know the position of a particle, the less well you can know the momentum of the particle.
(*d*) The better you know the position of a particle, the better you can know the momentum of the particle.
(*e*) How well you can determine the position and momentum of a particle depends on the particle's quantum numbers.

4. If you knew the position of an object precisely, with no uncertainty, how well would you know its momentum?
(*a*) Perfectly, no uncertainty.
(*b*) Not at all—completely uncertain.
(*c*) You would know the direction of the momentum but not its magnitude.
(*d*) You would know the magnitude but not the direction of the momentum.

5. What is uncertain in the Heisenberg uncertainty principle?
(*a*) Position.
(*b*) Momentum.
(*c*) Energy.
(*d*) Time interval.
(*e*) All of the above.

6. A hydrogen atom and a helium atom, each with 4 eV of kinetic energy, approach a thin barrier 6 MeV high. Which has the greater probability of tunneling through?
(*a*) Hydrogen.
(*b*) Helium.
(*c*) Both the same.
(*d*) Cannot evaluate.

7. Why is the wavelike nature of a moving baseball not ever observed?
(*a*) The baseball's frequency is too small.
(*b*) The baseball's wavelength is too small.
(*c*) The baseball's energy is too small.
(*d*) The baseball's velocity is too small.

8. What is a main difference between Newtonian mechanics and quantum mechanics?
(*a*) Newtonian mechanics can be used to predict exactly where a particle will be at any given time, whereas quantum mechanics gives probabilities.
(*b*) Momentum is conserved only in quantum mechanics, not in Newtonian mechanics.
(*c*) Newtonian mechanics is valid only for large objects, whereas quantum mechanics is valid only for very small objects.
(*d*) Energy is conserved only in Newtonian mechanics, not in quantum mechanics.

9. For the wave function Ψ, what is actually waving?
(*a*) The electromagnetic field.
(*b*) The gravitational field.
(*c*) Both (*a*) and (*b*).
(*d*) Neither (*a*) nor (*b*).

10. An electron is in an infinite square well. Which statement(s) are true about the electron?
(*a*) The electron can be at rest or it can be moving.
(*b*) The electron can have only certain energies.
(*c*) The electron can have any energy, but some values are more probable than others.
(*d*) Some locations of the electron are more probable than others.
(*e*) None of these statements is true.

11. A particle trapped in a finite square well has kinetic energy less than the potential energy of the well. The particle
(*a*) will always be found inside the well.
(*b*) may be found occasionally outside the well.
(*c*) may completely escape from the well.
(*d*) Both (*b*) and (*c*).

12. According to quantum mechanics, the energy of a particle will be quantized (only certain discrete values allowed) in all circumstances.
(*a*) True.
(*b*) No, the energy is never quantized.
(*c*) Not all: the energy of a free particle ($U = 0$ for all x) is not quantized.

13. A wave function Ψ must satisfy the normalization condition because
(*a*) Ψ can be infinite.
(*b*) Ψ can be complex.
(*c*) the particle must exist somewhere.
(*d*) the particle may have an uncertain momentum.

Problems

38–2 Wave Function, Double-Slit

1. (II) The neutrons in a parallel beam, each having kinetic energy 0.025 eV, are directed through two slits 0.35 mm apart. How far apart will the interference peaks be on a screen 1.0 m away? [*Hint*: First find the wavelength of the neutron.]

2. (II) Pellets of mass 2.0 g are fired in parallel paths with speeds of 150 m/s through a hole 3.0 mm in diameter. How far from the hole must you be to detect a 1.0-cm-diameter spread in the beam of pellets?

3. (II) A beam of electrons with kinetic energy 45 keV is shot through two narrow slits in a barrier. The slits are a distance 1.6×10^{-6} m apart. If a screen is placed 45.0 cm behind the barrier, calculate the spacing between the "bright" fringes of the interference pattern produced on the screen.

38–3 Uncertainty Principle

4. (I) A proton is traveling with a speed of $(8.940 \pm 0.016) \times 10^5$ m/s. With what maximum precision can its position be ascertained? [*Hint*: $\Delta p = m\,\Delta v$.]

5. (I) If an electron's position can be measured to a precision of 2.1×10^{-8} m, how precisely can its speed be known?

6. (I) An electron remains in an excited state of an atom for typically 10^{-8} s. What is the minimum uncertainty in the energy of the state (in eV)?

7. (II) What is the uncertainty in the mass of a muon $(m = 105.7\ \text{MeV}/c^2)$, specified in eV/c^2, given its lifetime of 2.20 μs?

8. (II) A free neutron $(m = 1.67 \times 10^{-27}\ \text{kg})$ has a mean life of 880 s. What is the uncertainty in its mass (in kg)?

9. (II) An electron and a 140-g baseball are each traveling 120 m/s measured to a precision of 0.075%. Calculate and compare the uncertainty in position of each.

10. (II) A radioactive element undergoes an alpha decay with a lifetime of 12 μs. If alpha particles are emitted with 5.2-MeV kinetic energy, find the percent uncertainty $\Delta E/E$ in the particle energy.

11. (II) If an electron's position can be measured to a precision of 2.6×10^{-8} m, what is the uncertainty in its momentum? Assuming its momentum must be at least equal to its uncertainty, estimate the electron's wavelength.

12. (II) Estimate the lowest possible energy of a neutron contained in a typical nucleus of radius 1.2×10^{-15} m. [*Hint*: Assume a particle can have an energy as large as its uncertainty.]

13. (II) How accurately can the position of a 4.75-keV electron be measured assuming its energy is known to 1.00%?

14. (III) Use the uncertainty principle to show that if an electron were present in the nucleus $(r \approx 10^{-15}\ \text{m})$, its kinetic energy (use relativity) would be hundreds of MeV. (Since such electron energies are not observed, we conclude that electrons are not present in the nucleus.) [*Hint*: Assume a particle can have an energy as large as its uncertainty.]

15. (II) An electron in the $n = 2$ state of hydrogen remains there on average about 10^{-8} s before jumping to the $n = 1$ state. (*a*) Estimate the uncertainty in the energy of the $n = 2$ state. (*b*) What fraction of the transition energy is this? (*c*) What is the wavelength, and width (in nm), of this line in the spectrum of hydrogen?

16. (III) In a double-slit experiment on electrons (or photons), suppose that we use indicators to determine which slit each electron went through (Section 38–2). These indicators must tell us the y coordinate to within $d/2$, where d is the distance between slits. Use the uncertainty principle to show that the interference pattern will be destroyed. [*Note*: First show that the angle θ between maxima and minima of the interference pattern is given by $\frac{1}{2}\lambda/d$, Fig. 38–19.]

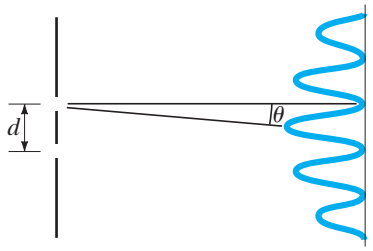


FIGURE 38–19
Problem 16.

*38–6 Time-Dependent Schrödinger Equation

***17.** (II) Show that the superposition principle holds for the time-dependent Schrödinger equation. That is, show that if $\Psi_1(x,t)$ and $\Psi_2(x,t)$ are solutions, then $A\Psi_1(x,t) + B\Psi_2(x,t)$ is also a solution where A and B are arbitrary constants.

***18.** (III) (*a*) Show that $\Psi(x,t) = Ae^{i(kx-\omega t)}$ is a solution to the time-dependent Schrödinger equation for a free particle $[U(x) = U_0 = \text{constant}]$ but that $\Psi(x,t) = A\cos(kx - \omega t)$ and $\Psi(x,t) = A\sin(kx - \omega t)$ are not. (*b*) Show that the valid solution of part (*a*) satisfies conservation of energy if the de Broglie relations hold; $\lambda = h/p$, $\omega = E/\hbar$. That is, show that direct substitution into Eq. 38–7 gives

$$\hbar\omega = \frac{\hbar^2 k^2}{2m} + U_0.$$

38–7 Free Particles; Plane Waves; Wave Packets

19. (I) A free electron has a wave function $\psi(x) = A\sin(2.4 \times 10^{10}\,x)$, where x is given in meters. Determine the electron's (*a*) wavelength, (*b*) momentum, (*c*) speed, and (*d*) kinetic energy.

20. (I) Write the wave function for (*a*) a free electron and (*b*) a free proton, each having a constant velocity $v = 2.7 \times 10^5$ m/s.

21. (III) Show that the uncertainty principle holds for a "wave packet" that is formed by two waves of similar wavelength λ_1 and λ_2. To do so, follow the argument leading up to Eq. 16–8, but use as the two waves $\psi_1 = A\sin k_1 x$ and $\psi_2 = A\sin k_2 x$. Then show that the width of each "wave packet" is $\Delta x = 2\pi/(k_1 - k_2) = 2\pi/\Delta k$ (from $t = 0.05$ s to $t = 0.15$ s in Fig. 16–17). Finally, show that $\Delta x\,\Delta p = h$ for this simple situation.

38–8 Infinite Square Well

22. (II) What is the minimum speed of an electron trapped in a 0.25-nm-wide infinitely deep square well?

23. (II) Show that for a particle in a perfectly rigid box, the wavelength of the wave function for any state is the de Broglie wavelength.

24. (II) An electron trapped in an infinitely deep square well has a ground-state energy $E = 8.4\,\text{eV}$. (*a*) What is the longest wavelength photon that an excited state of this system can emit? (*b*) What is the width of the well?

25. (II) An $n = 4$ to $n = 1$ transition for an electron trapped in a rigid box produces a 370-nm photon. What is the width of the box?

26. (II) For a particle in a box with rigid walls, determine whether our results for the ground state are consistent with the uncertainty principle by calculating the product $\Delta p\,\Delta x$. Take $\Delta x \approx \ell$, since the particle is somewhere within the box. For Δp, note that although p is known ($= \hbar k$), the direction of $\vec{\mathbf{p}}$ is not known, so the x component could vary from $-p$ to $+p$; hence take $\Delta p \approx 2p$.

27. (II) The longest-wavelength line in the spectrum emitted by an electron trapped in an infinitely deep square well is 580 nm. What is the width of the well?

28. (II) Determine the lowest four energy levels and wave functions for an electron trapped in an infinitely deep potential well of width 2.4 nm.

29. (II) Write a formula for the positions of (*a*) the maxima and (*b*) the minima in $|\psi|^2$ for a particle in the nth state in an infinite square well.

30. (II) Consider an atomic nucleus to be a rigid box of width $1.8 \times 10^{-14}\,\text{m}$. What would be the ground-state energy for (*a*) an electron, (*b*) a neutron, and (*c*) a proton in this nucleus?

31. (II) A proton in a nucleus can be roughly modeled as a particle in a box of nuclear dimensions. Calculate the energy released when a proton confined in a nucleus of width $1.2 \times 10^{-14}\,\text{m}$ makes a transition from the first excited state to the ground state.

32. (II) Consider a single oxygen molecule confined in a one-dimensional rigid box of width 3.0 mm. (*a*) Treating this as a particle in a rigid box, determine the ground-state energy. (*b*) If the molecule has an energy equal to the one-dimensional average thermal energy $\frac{1}{2}kT$ at $T = 300\,\text{K}$, what is the quantum number n? (*c*) What is the energy difference between the nth state and the next higher state?

33. (II) An electron is trapped in a 1.50-nm-wide rigid box. Determine the probability of finding the electron within 0.15 nm of the center of the box (on either side of center) for (*a*) $n = 1$, (*b*) $n = 5$, and (*c*) $n = 20$. (*d*) Compare to the classical prediction.

34. (III) If an infinitely deep well of width ℓ is redefined to be located from $x = -\frac{1}{2}\ell$ to $x = \frac{1}{2}\ell$ (as opposed to $x = 0$ to $x = \ell$), speculate how this will change the wave function for a particle in this well. Investigate your speculation(s) by determining the wave functions and energy levels for this newly defined well. [*Hint*: Try $\psi = A\sin(kx + \phi)$.]

38–9 Finite Potential Well

35. (II) An electron with 180 eV of kinetic energy in free space passes over a finite potential well 48 eV deep that stretches from $x = 0$ to $x = 0.50\,\text{nm}$. What is the electron's wavelength (*a*) in free space, (*b*) when over the well? (*c*) Draw a diagram showing the potential energy and total energy as a function of x, and on the diagram sketch a possible wave function.

36. (II) Sketch the wave functions and the probability distributions for the $n = 4$ and $n = 5$ states for a particle trapped in a finite square well.

37. (II) An electron having kinetic energy 3.0 eV is trapped in an 8.0-eV-high potential well of width 8.0 nm. What is the probability of finding the electron outside the confines of the well?

38. (II) Suppose that a particle of mass m is trapped in a finite potential well that has a rigid wall at $x = 0$ ($U = \infty$ for $x < 0$) and a finite wall of height $U = U_0$ at $x = \ell$, Fig. 38–20. (*a*) Sketch the wave functions for the lowest three states. (*b*) What is the form of the wave function in the ground state in the three regions $x < 0$, $0 < x < \ell$, $x < \ell$?

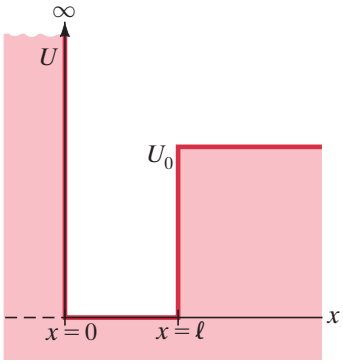


FIGURE 38–20 Problem 38.

39. (II) An electron is trapped in a 0.16-nm-wide finite square well of height $U_0 = 1.8\,\text{keV}$. Estimate at what distance outside the walls of the well the ground state wave function drops to 1.0% of its value at the walls.

38–10 Tunneling

40. (II) A potential barrier has a height $U_0 = 14\,\text{eV}$ and thickness $\ell = 0.85\,\text{nm}$. If the transmission coefficient for an incident electron is 0.00080, what is the electron's energy?

41. (II) An electron approaches a potential barrier 18 eV high and 0.48 nm wide. If the electron has a 1.0% probability of tunneling through the barrier, what is the electron's energy?

42. (II) A proton and a helium nucleus approach a 25-MeV potential energy barrier. If each has a kinetic energy of 7.5 MeV, what is the probability of each to tunnel through the barrier, assuming it is rectangular and 3.6 fm thick?

43. (II) An electron with an energy of 7.2 eV is incident on a potential barrier which is 9.2 eV high and 0.25 nm wide. (*a*) What is the probability that the electron will pass through the barrier? (*b*) What is the probability that the electron will be reflected?

44. (II) A 1.5-mA current of 1.6-MeV protons strikes a 2.6-MeV-high potential barrier $2.8 \times 10^{-13}\,\text{m}$ thick. Estimate the transmitted current.

45. (II) For part (*b*) of Example 38–11, what effect will there be on the transmission coefficient if (*a*) the barrier height is raised 3.0%, (*b*) the barrier thickness is increased by 3.0%?

46. (II) Show that the transmission coefficient is given roughly by Eqs. 38–17 for a high or thick barrier, by calculating $|\psi(x=\ell)|^2/|\psi(0)|^2$. [*Hint*: Assume that ψ is a decaying exponential inside the barrier.]

47. (II) or (III) (*a*) What is the probability that an electron with kinetic energy E can pass through two successive barriers of width ℓ, a distance ℓ apart, both being of height U_0? (*b*) What would be the probability for the electron to pass through one barrier U_0 of thickness 2ℓ? Which is greater?

48. (III) A uranium-238 nucleus $(Q = +92e)$ lasts about 4.5×10^9 years before it decays by emission of an alpha particle $(q = +2e,\ M = 4M_{\text{proton}})$. (*a*) Assuming that the α particle is a point, and the nucleus is roughly 8 fm in radius, estimate the height of the Coulomb barrier (the peak in Fig. 38–17). (*b*) The alpha particle, when free, has kinetic energy ≈ 4 MeV. Estimate the width of the barrier. (*c*) Assuming that the square well has $U = 0$ inside (and $U = 0$ far from the nucleus), calculate the speed of the alpha particle and how often it hits the barrier inside, and from this (and Eqs. 38–17) estimate the uranium lifetime. [*Hint*: Replace the $1/r$ Coulomb barrier with an "averaged" rectangular barrier (as in Fig. 38–15) of width equal to $\frac{1}{3}$ that calculated in (*b*).]

General Problems

49. The Z^0 boson, discovered in 1985, is the mediator of the weak nuclear force, and it typically decays very quickly. Its average rest energy is 91.19 GeV, but its short lifetime shows up as an intrinsic width of 2.5 GeV. What is the lifetime of this particle?

50. Protons are accelerated from rest across 650 V. They are then directed at two slits 0.80 mm apart. How far apart will the interference peaks be on a screen 15 m away?

51. An electron and a proton, each initially at rest, are accelerated across the same voltage. Assuming that the uncertainty in their position is given by their de Broglie wavelength, find the ratio of the uncertainty in their momentum.

52. A neutron is trapped in an infinitely deep potential well 2.5 fm in width. Determine (*a*) the four lowest possible energy states and (*b*) their wave functions. (*c*) What is the wavelength and energy of a photon emitted when the neutron makes a transition between the two lowest states? In what region of the EM spectrum does this photon lie? [*Note*: This is a rough model of an atomic nucleus.]

53. The lifetime of a typical excited state in an atom is about 10 ns. Suppose an atom falls from one such excited state to a lower one, and emits a photon of wavelength about 400 nm. Find the fractional energy uncertainty $\Delta E/E$ and wavelength uncertainty $\Delta\lambda/\lambda$ of this photon.

54. Estimate the kinetic energy and speed of an alpha particle $(q = +2e,\ M = 4M_{\text{proton}})$ trapped in a nucleus 1.8×10^{-14} m wide. Assume an infinitely deep square well potential.

55. Use the uncertainty principle to estimate the position uncertainty for the electron in the ground state of the hydrogen atom. [*Hint*: Determine the momentum using the Bohr model of Section 37–11 and assume the momentum can be anywhere between this value and zero.] How does this result compare to the Bohr radius?

56. Show that the function $\psi(x) = Ae^{ikx}$, where A is a constant and k is given by Eq. 38–11, is a solution of the time-independent Schrödinger equation for the case $U = 0$.

57. Show that the average value of the square of the position x of a particle in state n inside an infinite well of width ℓ is $\overline{x^2} = \int x^2|\psi_n|^2\,dx = \ell^2\left[\frac{1}{3} - \frac{1}{2}(n\pi)^{-2}\right]$. Calculate the values of $\overline{x^2}$ for $n = 1$ to 20 and make a graph of $\overline{x^2}$ versus n. [*Hint*: You may want to consult a detailed Table of integrals.]

58. If your de Broglie wavelength were 0.50 m, how fast would you be moving if your mass is 64.0 kg? Would you notice diffraction effects as you walk through a doorway? Approximately how long would it take you to walk through the doorway?

59. **Simple Harmonic Oscillator**. Suppose that a particle of mass m is trapped not in a square well, but in one whose potential energy is that of a simple harmonic oscillator: $U(x) = \frac{1}{2}Cx^2$. That is, if the particle is displaced from $x = 0$, a restoring force $F = -Cx$ acts on it, where C is constant. (*a*) Sketch this potential energy. (*b*) Show that $\psi = Ae^{-Bx^2}$ is a solution to the Schrödinger equation and that the energy of this state is $E = \frac{1}{2}\hbar\omega$, where $\omega = \sqrt{C/m}$ (as classically, Eq. 14–5) and $B = m\omega/2\hbar$. [*Note*: This is the ground state, and this energy $\frac{1}{2}\hbar\omega$ is the zero-point energy for a harmonic oscillator. The energies of higher states are $E_n = \left(n + \frac{1}{2}\right)\hbar\omega$, where n is an integer.]

60. A small ball of mass 3.0×10^{-6} kg is dropped on a table from a height of 2.0 m. After each bounce the ball rises to 65% of its height before the bounce because of its inelastic collision with the table. Estimate how many bounces occur before the uncertainty principle plays a role in the problem. [*Hint*: Determine when the uncertainty in the ball's speed is comparable to its speed of impact on the table.]

61. An atomic spectrum contains a line with a wavelength centered at 488 nm. Careful measurements show the line is really spread out between 487 and 489 nm. Estimate the lifetime of the excited state that produced this line.

62. A 1500-kg car is traveling with a speed of (22 ± 0.22) m/s. With what maximum precision can its position be determined?

63. By how much does the tunneling current through the tip of an STM change if the tip rises 0.020 nm from some initial height above a sodium surface with a work function $W_0 = 2.28$ eV? [*Hint*: Let the work function (see Section 37–2) equal the energy needed to raise the electron to the top of the barrier.]

64. An electron is trapped in the ground state of an infinite potential well of width $\ell = 0.10$ nm. The probability that the electron will be found in the central 1% of the well was estimated in Example 38–7 by $|\psi|^2\,\Delta x$. Determine how large Δx could be to cause less than a 10% error in such an estimate.

*65. Consider a particle of mass m and energy E traveling to the right where it encounters a narrow potential barrier of height U_0 and width ℓ as shown in Fig. 38–21. It can be shown that:

(i) for $E < U_0$, the transmission probability is

$$T = \left[1 + \frac{(e^{G\ell} - e^{-G\ell})^2}{16(E/U_0)(1 - E/U_0)}\right]^{-1}$$

where

$$G = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}}$$

and the reflection probability is $R = 1 - T$;

(ii) for $E > U_0$, the transmission probability is

$$T = \left[1 + \frac{\sin^2(G'\ell)}{4(E/U_0)(E/U_0 - 1)}\right]^{-1}$$

where

$$G' = \sqrt{\frac{2m(E - U_0)}{\hbar^2}}$$

and $R = 1 - T$. Consider that the particle is an electron and it is incident on a rectangular barrier of height $U_0 = 10\,\text{eV}$ and width $\ell = 1.0 \times 10^{-10}\,\text{m}$. (*a*) Calculate T and R for the electron from $E/U_0 = 0$ to 10, in steps of 0.1. Make a single graph showing the two curves of T and R as a function of E/U_0. (*b*) From the graph determine the energies (E/U_0) at which the electron will have transmission probabilities of 10%, 20%, 50%, and 80%.

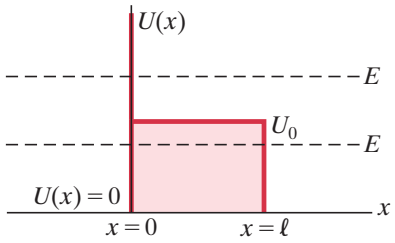


FIGURE 38–21 Problem 63.

66. Show that all three formulations of the uncertainty principle given in this Chapter (position-momentum, energy-time, and angular momentum-angular position) are dimensionally correct.

67. A red ruby laser produces light of wavelength 704.1 nm. If the wavelength is assumed to be known to 0.1 nm, approximately what is the duration of the shortest pulse that a red ruby laser can produce? What is the length of that pulse in nm?

68. A 15.0-gram marble is in a hollow tube that is sealed closed on both ends approximating an infinite square well. If the tube is 7.5 cm long, what is the minimum speed the marble can have?

69. Imagine a line whose length equals the diameter of the smallest orbit in the Bohr model. If we are told only that an electron is located somewhere on this line, then the electron's position can be specified as $x = 0 \pm r_1$, where the origin is at the line's center and r_1 is the Bohr radius. Such an electron can never be observed at rest, but instead at a minimum will have a speed somewhere in the range from $v = 0 - \Delta v$ to $0 + \Delta v$. Determine Δv.

70. Consider a particle that can exist anywhere in space with a wave function given by $\psi(x) = b^{-\frac{1}{2}}|x/b|^{\frac{1}{2}}\, e^{-(x/b)^2/2}$, where $b = 1.0\,\text{nm}$. (*a*) Check that the wave function is normalized. (*b*) What is the most probable position for the particle in the region $x > 0$? (*c*) What is the probability of finding the particle between $x = 0\,\text{nm}$ and $x = 0.25\,\text{nm}$?

71. A 5.0-gram pencil, 18 cm long, is balanced on its point. Classically, this is a configuration of (unstable) equilibrium, so the pencil could remain there forever if it were perfectly placed. A quantum mechanical analysis shows that the pencil must fall. (*a*) Why is this the case? (*b*) Estimate (within a factor of 2) how long it will take the pencil to hit the table if it is initially positioned as well as possible. [*Hint*: Use the uncertainty principle in its angular form to obtain an expression for the initial angle $\phi_0 \approx \Delta\phi$.]

ANSWERS TO EXERCISES

A: (*c*).
B: $2.1 \times 10^{-24}\,\text{kg}\cdot\text{m/s}$, $2.3 \times 10^{6}\,\text{m/s}$.
C: (*e*).
D: (*e*).
E: (*a*).

A neon tube is a thin glass tube, moldable into various shapes, filled with neon (or other) gas that glows with a particular color when a current at high voltage passes through it. Gas atoms, excited to upper energy levels, jump down to lower energy levels and emit light (photons) whose wavelengths (color) are characteristic of the type of gas.

In this Chapter we study what quantum mechanics tells us about atoms, their wave functions and energy levels, including the effect of the exclusion principle for atoms with more than one electron. We also discuss interesting applications such as lasers and holography.



CHAPTER 39

Quantum Mechanics of Atoms

CONTENTS

39 – 1 Quantum-Mechanical View of Atoms
39 – 2 Hydrogen Atom: Schrödinger Equation and Quantum Numbers
39 – 3 Hydrogen Atom Wave Functions
39 – 4 Multielectron Atoms; the Exclusion Principle
39 – 5 Periodic Table of Elements
39 – 6 X-Ray Spectra and Atomic Number
*39 – 7 Magnetic Dipole Moment; Total Angular Momentum
39 – 8 Fluorescence and Phosphorescence
39 – 9 Lasers
*39–10 Holography

CHAPTER-OPENING QUESTIONS—Guess now!

1. Thousands of hydrogen atoms are all in their ground state. Which statement below is true?

(a) All of the atoms have the electron in a circular orbit at the Bohr radius.
(b) All of the atoms have the electron in the same orbit, but it's not the same orbit for the ground state in the Bohr model.
(c) The electron is not in an actual orbit, but the distance between the nucleus and the electron is the same in all of the atoms.
(d) If the distance from the nucleus to the electron could be measured, it would be found at different locations in different atoms. The most probable distance would be the Bohr radius.

2. The state of an electron in an atom can be specified by a set of quantum numbers. Which of the following statements are valid?

(a) No two electrons in the universe can be identical to each other.
(b) All electrons in the universe are identical.
(c) No two electrons in an atom can occupy the same quantum state.
(d) There cannot be more than one electron in a given electron orbit.
(e) Electrons must be excluded from the nucleus because only positive charges are allowed in the nucleus to give rise to stable electronic orbits according to Bohr's model.

At the beginning of Chapter 38 we discussed the limitations of the Bohr model of atomic structure and why a new theory was needed. Although the Bohr model had great success in predicting the wavelengths of light emitted and absorbed by the hydrogen atom, it could not do so for more complex atoms.

PSE5_CH39_v1.9_23-Jun-20

Nor did it explain *fine structure*, the splitting of emission lines into two or more closely spaced lines. And, as a theory, it was an uneasy mixture of classical and quantum ideas.

Quantum mechanics came to the rescue in 1925 and 1926, and in this Chapter we examine the quantum-mechanical theory of atomic structure, which is far more complete than the old Bohr model.

39–1 Quantum-Mechanical View of Atoms

Although the Bohr model has been discarded as an accurate description of nature, nonetheless, quantum mechanics reaffirms certain aspects of the older theory, such as that electrons in an atom exist only in discrete states of definite energy, and that a photon of light is emitted (or absorbed) when an electron makes a transition from one state to another. But quantum mechanics is a much deeper theory, and has provided us with a very different view of the atom. According to quantum mechanics, electrons do not exist in well-defined circular orbits as in the Bohr model. Rather, the electron (because of its wave nature) can be thought of as spread out in space as a "**cloud**." The size and shape of the electron cloud can be calculated for a given state of an atom. For the ground state in the hydrogen atom, the solution of the Schrödinger equation, as we will discuss in more detail in Section 39–3, gives

$$\psi(r) = \frac{1}{\sqrt{\pi r_0^3}} e^{-\frac{r}{r_0}}.$$

Here $\psi(r)$ is the wave function as a function of position, and it depends only on the radial distance r from the center, and not on angular position θ or ϕ (like latitude and longitude on Earth). The constant r_0 happens to be equal to the first Bohr radius. Thus the electron cloud, whose density is $|\psi|^2$, for the ground state of hydrogen is spherically symmetric as shown in Fig. 39–1. The electron cloud at its higher densities roughly indicates the "size" of an atom. But just as a cloud may not have a distinct border, atoms do not have a precise boundary or a well-defined size. Not all electron clouds have a spherical shape, as we shall see later in this Chapter. But note that $\psi(r)$, while becoming extremely small for large r (see the equation above), does not equal zero in any finite region. So quantum mechanics suggests that an atom is not mostly empty space.

The electron cloud can be interpreted from either the particle or the wave viewpoint. Remember that by a particle we mean something that is localized in space—it has a definite position at any given instant. By contrast, a wave is spread out in space. The electron cloud, spread out in space as in Fig. 39–1, is a result of the wave nature of electrons. Electron clouds can also be interpreted as **probability distributions** (or **probability density**) for a particle. As we saw in Section 38–3, we cannot predict the path an electron will follow (thinking of it as a particle). After one measurement of its position we cannot predict exactly where it will be at a later time. We can only calculate the probability that it will be found at different points. If you were to make 500 different measurements of the position of an electron in a hydrogen atom, the majority of the results would show the electron at points where the probability is high (darker area in Fig. 39–1). Only occasionally would the electron be found where the probability is low. The electron cloud on probability distribution becomes small (or thin) at places, especially far away, but never becomes zero. So quantum mechanics suggests that an atom is *not* mostly empty space, and that there is no truly empty space in the universe.

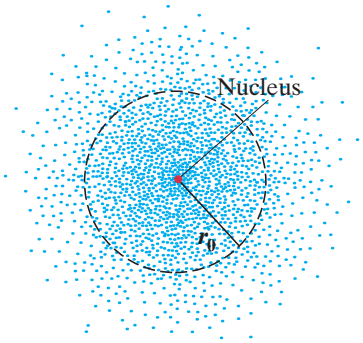


FIGURE 39–1 Electron cloud or "probability distribution" $|\psi|^2$ for the ground state of the hydrogen atom, as seen from afar. The dots represent a hypothetical detection of an electron at each point: dots closer together represent more probable presence of an electron (denser cloud). The (dashed) circle represents the Bohr radius r_0.

39–2 Hydrogen Atom: Schrödinger Equation and Quantum Numbers

The hydrogen atom is the simplest of all atoms, consisting of a single electron of charge $-e$ moving around a central nucleus (a single proton) of charge $+e$. It is with hydrogen that a study of atomic structure must begin.

The Schrödinger equation (see Eq. 38–5) includes a term containing the potential energy. For the hydrogen (H) atom, the potential energy is due to the Coulomb force between electron and proton:

$$U = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r}$$

where r is the radial distance from the proton (situated at $r = 0$) to the electron. See Fig. 39–2. The (time-independent) Schrödinger equation, which must now be written in three dimensions, is then

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right) - \frac{1}{4\pi\epsilon_0}\frac{e^2}{r}\psi = E\psi, \qquad \textbf{(39–1)}$$

where $\partial^2\psi/\partial x^2$, $\partial^2\psi/\partial y^2$, and $\partial^2\psi/\partial z^2$ are partial derivatives with respect to x, y, and z. (As always, $\hbar = h/2\pi$.) To solve the Schrödinger equation for the H atom, it is usual to write it in terms of spherical coordinates (r, θ, ϕ). We will not, however, actually go through the process of solving it. Instead, we look at the properties of the solutions, and (in the next Section) at the wave functions themselves.

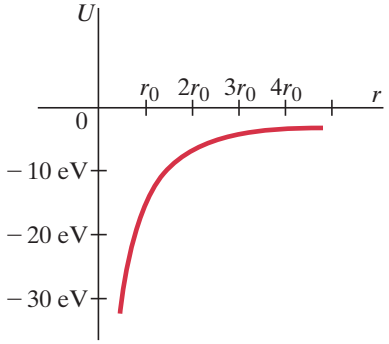


FIGURE 39–2 Potential energy $U(r)$ for the hydrogen atom. The radial distance r of the electron from the nucleus is given in terms of the Bohr radius r_0.

Recall from Chapter 38 that the solutions of the Schrödinger equation in one dimension for the infinite square well were characterized by a single quantum number, which we called n, which arises from applying the boundary conditions. In the three-dimensional H atom, the solutions of the Schrödinger equation are characterized by three quantum numbers corresponding to boundary conditions applied in the three dimensions. However, four different quantum numbers are actually needed to specify each state in the H atom, the fourth coming from a relativistic treatment. We now discuss each of these quantum numbers. Much of our analysis here will also apply to more complex atoms, which we discuss starting in Section 39–4.

Quantum mechanics predicts the same energy levels (Fig. 37–26) for the hydrogen atom as does the Bohr model. That is,

$$E_n = -\frac{13.6\,\text{eV}}{n^2} \qquad n = 1, 2, 3, \cdots, \qquad \textbf{(39–2)}$$

where n is an integer. In the simple Bohr model, there was only one quantum number, n. In quantum mechanics, four different quantum numbers are needed to specify each state in the atom:

(1) The quantum number, n, from the Bohr model is found also in quantum mechanics and is called the **principal quantum number**. It can have any integer value from 1 to ∞. The total energy of a state in the hydrogen atom depends on n, as we saw above.

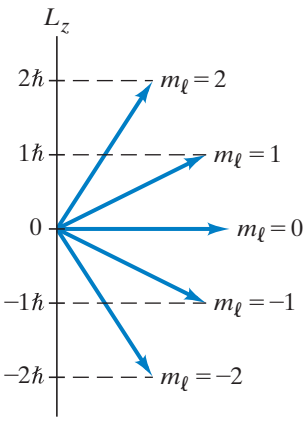
(2) The **orbital quantum number**, ℓ, is related to the magnitude of the angular momentum of the electron; ℓ can take on integer values from 0 to $(n - 1)$. For the ground state, $n = 1$, ℓ can only be zero.† For $n = 3$, ℓ can be 0, 1, or 2. The actual magnitude of the angular momentum L is related to the quantum number ℓ by

$$L = \sqrt{\ell(\ell + 1)}\,\hbar. \qquad \textbf{(39–3)}$$

The value of ℓ has almost no effect on the total energy in the hydrogen atom; only n does to any appreciable extent (but see *fine structure* below). In atoms with two or more electrons, the energy does depend on ℓ as well as n, as we shall see.

(3) The **magnetic quantum number**, m_ℓ, is related to the *direction* of the electron's angular momentum, and it can take on integer values ranging from $-\ell$ to $+\ell$. For example, if $\ell = 2$, then m_ℓ can be $-2, -1, 0, +1$, or $+2$. Since angular momentum is a vector, it is not surprising that both its magnitude and its direction would be quantized. For $\ell = 2$, the five different directions allowed can be represented by the diagram of Fig. 39–3.

FIGURE 39–3 Quantization of angular momentum direction for $\ell = 2$. (Magnitude of $\vec{\mathbf{L}}$ is $L = \sqrt{6}\hbar$.)



†This replaces Bohr model, which assigned $\ell = 1$ to the ground state (Eq. 37–10).

PSE5_CH39_v1.9_23-Jun-20

This limitation on the direction of $\vec{\mathbf{L}}$ is often called **space quantization**. In quantum mechanics, the direction of the angular momentum is usually specified by giving its component along the z axis (this choice is arbitrary). Then L_z is related to m_ℓ by the equation

$$L_z = m_\ell \hbar. \tag{39–4}$$

The values of L_x and L_y are not definite, however. The name for m_ℓ derives not from theory (which relates it to L_z), but from experiment. It was found that when a gas-discharge tube was placed in a magnetic field, the spectral lines were split into several very closely spaced lines. This splitting, known as the **Zeeman effect**, implies that the energy levels must be split (Fig. 39–4), and thus that the energy of a state depends not only on n but also on m_ℓ when a magnetic field is applied—hence the name "magnetic quantum number."

(4) Finally, there is the **spin quantum number**, m_s, which for an electron can have only two values, $m_s = +\frac{1}{2}$ and $m_s = -\frac{1}{2}$. The existence of this quantum number did not come out of Schrödinger's original theory, as did n, ℓ, and m_ℓ. Instead, a subsequent modification by P. A. M. Dirac (1902–1984) explained its presence as a relativistic effect. The first hint that m_s was needed, however, came from experiment. A careful study of the spectral lines of hydrogen showed that each actually consisted of two (or more) very closely spaced lines even in the absence of an external magnetic field. It was at first hypothesized that this tiny splitting of energy levels, called **fine structure**, was due to angular momentum associated with a spinning of the electron. That is, the electron might spin on its axis as well as orbit the nucleus, just as the Earth spins on its axis as it orbits the Sun. The interaction between the tiny current of the spinning electron could then interact with the magnetic field due to the orbiting charge and cause the small observed splitting of energy levels. (The energy thus depends slightly on m_ℓ and m_s. Fine structure is said to be due to a **spin-orbit interaction**.) Today we consider the picture of a spinning electron as not legitimate. We cannot even view an electron as a localized object, much less a spinning one. What is important is that the electron can have two different states due to some intrinsic property that behaves like an angular momentum, and we still call this property "spin." The electron is said to have a spin quantum number $s = \frac{1}{2}$, which produces a spin angular momentum

$$S = \sqrt{s(s+1)}\,\hbar = \frac{\sqrt{3}}{2}\hbar.$$

The z component is

$$S_z = m_s \hbar$$

where the two possible values of $m_s\left(+\frac{1}{2} \text{ and } -\frac{1}{2}\right)$ are often said to be "spin up" and "spin down," referring to the two possible directions of the spin angular momentum. See Fig. 39–5. A state with spin down $\left(m_s = -\frac{1}{2}\right)$ has slightly lower energy than a state with spin up. (Note that we include m_s, but not s, in our list of quantum numbers since s is the same for all electrons.)

The possible values of the four quantum numbers for an electron in the hydrogen atom are summarized in Table 39–1.

TABLE 39–1 Quantum Numbers for an Electron

Name	Symbol	Possible Values
Principal	n	$1, 2, 3, \cdots, \infty$.
Orbital	ℓ	For a given n: ℓ can be $0, 1, 2, \cdots, n-1$.
Magnetic	m_ℓ	For given n and ℓ: m_ℓ can be $\ell, \ell-1, \cdots, 0, \cdots, -\ell$.
Spin	m_s	For each set of n, ℓ, and m_ℓ: m_s can be $+\frac{1}{2}$ or $-\frac{1}{2}$.

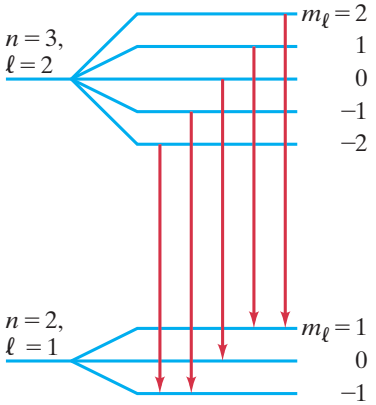
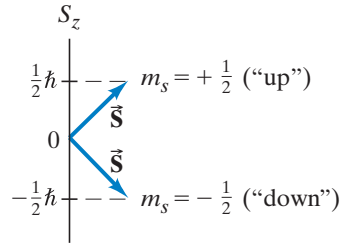


FIGURE 39–4 Energy levels (not to scale). When a magnetic field is applied, an $n = 3$, $\ell = 2$ energy level is split into five separate levels, corresponding to the five values of $m_\ell (2, 1, 0, -1, -2)$. An $n = 2$, $\ell = 1$ level is split into three levels $(m_\ell = 1, 0, -1)$. Transitions can occur between levels (not all transitions are shown), with photons of several slightly different frequencies being given off (the Zeeman effect).

FIGURE 39–5 The spin angular momentum S can take on only two directions, $m_s = +\frac{1}{2}$ or $-\frac{1}{2}$, called "spin up" and "spin down."



PSE5_CH39_v1.9_23-Jun-20

CONCEPTUAL EXAMPLE 39–1 **Possible states for $n = 3$.** How many different states are possible for an electron with principal quantum number $n = 3$?

RESPONSE For $n = 3$, ℓ can have the values $\ell = 2, 1, 0$. For $\ell = 2$, m_ℓ can be $2, 1, 0, -1, -2$, which is five different possibilities. For each of these, m_s can be either up or down $\left(+\frac{1}{2} \text{ or } -\frac{1}{2}\right)$; so for $\ell = 2$, there are $2 \times 5 = 10$ states. For $\ell = 1$, m_ℓ can be $1, 0, -1$, and since m_s can be $+\frac{1}{2}$ or $-\frac{1}{2}$ for each of these, we have 6 more possible states. Finally, for $\ell = 0$, m_ℓ can only be 0, and there are only 2 states corresponding to $m_s = +\frac{1}{2}$ and $-\frac{1}{2}$. The total number of states is $10 + 6 + 2 = 18$, as detailed in the following Table:

n	ℓ	m_ℓ	m_s	n	ℓ	m_ℓ	m_s
3	2	2	$\frac{1}{2}$	3	1	1	$\frac{1}{2}$
3	2	2	$-\frac{1}{2}$	3	1	1	$-\frac{1}{2}$
3	2	1	$\frac{1}{2}$	3	1	0	$\frac{1}{2}$
3	2	1	$-\frac{1}{2}$	3	1	0	$-\frac{1}{2}$
3	2	0	$\frac{1}{2}$	3	1	−1	$\frac{1}{2}$
3	2	0	$-\frac{1}{2}$	3	1	−1	$-\frac{1}{2}$
3	2	−1	$\frac{1}{2}$	3	0	0	$\frac{1}{2}$
3	2	−1	$-\frac{1}{2}$	3	0	0	$-\frac{1}{2}$
3	2	−2	$\frac{1}{2}$				
3	2	−2	$-\frac{1}{2}$				

EXERCISE A An electron has $n = 4$, $\ell = 2$. Which of the following values of m_ℓ are possible: 4, 3, 2, 1, 0, −1, −2, −3, −4?

EXAMPLE 39–2 ***E* and *L* for $n = 3$.** Determine (*a*) the energy and (*b*) the orbital angular momentum for an electron in each of the hydrogen atom states with $n = 3$ (see also Example 39–1).

APPROACH The energy of a state depends only on n, except for the very small corrections mentioned above, which we will ignore. Energy is calculated as in the Bohr model, $E_n = -13.6\,\text{eV}/n^2$. For angular momentum we use Eq. 39–3.

SOLUTION (*a*) Since $n = 3$ for all these states, they all have the same energy,

$$E_3 = -\frac{13.6\,\text{eV}}{(3)^2} = -1.51\,\text{eV}.$$

(*b*) For $\ell = 0$, Eq. 39–3 gives

$$L = \sqrt{\ell(\ell + 1)}\,\hbar = 0.$$

For $\ell = 1$,

$$L = \sqrt{1(1 + 1)}\,\hbar = \sqrt{2}\hbar = 1.49 \times 10^{-34}\,\text{J} \cdot \text{s}.$$

For $\ell = 2$, $L = \sqrt{2(2 + 1)}\,\hbar = \sqrt{6}\hbar$.

NOTE Atomic angular momenta are generally given as a multiple of $\hbar$ $\left(\sqrt{2}\hbar$ or $\sqrt{6}\hbar$ in this case$\right)$, rather than in SI units.

EXERCISE B What is the magnitude of the orbital angular momentum for orbital quantum number $\ell = 3$? (*a*) $3\hbar$; (*b*) $3.5\hbar$; (*c*) $4\hbar$; (*d*) $12\hbar$.

EXERCISE C What are the energy and angular momentum of the electron in a hydrogen atom with $n = 6$ and $\ell = 4$?

Selection Rules: Allowed and Forbidden Transitions

Another prediction of quantum mechanics is that when a photon is emitted or absorbed, transitions can occur only between states with values of ℓ that differ by exactly one unit:

$$\Delta\ell = \pm 1.$$

According to this **selection rule**, an electron in an $\ell = 2$ state can jump only to a state with $\ell = 1$ or $\ell = 3$. It cannot jump to a state with $\ell = 2$ or $\ell = 0$. A transition such as $\ell = 2$ to $\ell = 0$ is called a **forbidden transition**. Actually, such a transition is not absolutely forbidden and can occur, but only with very low probability compared to **allowed transitions**—those that satisfy the selection rule $\Delta\ell = \pm 1$. Experimentally, transitions with $\Delta\ell = \pm 1$ show spectral lines that are bright compared to forbidden transitions which are very dim and often hard to detect.

Since the orbital angular momentum of an H atom must change by one unit when it emits a photon, conservation of angular momentum tells us that the photon must carry off angular momentum. Indeed, experimental evidence of many sorts shows that the photon can be assigned a spin angular momentum of $1\,\hbar$.

39–3 Hydrogen Atom Wave Functions

The solution of the Schrödinger equation for the ground state of hydrogen (the state with lowest energy) has an energy $E_1 = -13.6\,\text{eV}$, as we have seen. The wave function for the ground state depends only on r and so is spherically symmetric. As already mentioned in Section 39–1, its form is

$$\psi_{100} = \frac{1}{\sqrt{\pi r_0^3}} e^{-\frac{r}{r_0}} \qquad \textbf{(39–5a)}$$

where $r_0 = h^2\epsilon_0/\pi m e^2 = 0.0529\,\text{nm}$ is the Bohr radius (Section 37–11). The subscript 100 on ψ represents the quantum numbers n, ℓ, m_ℓ:

$$\psi_{n\ell m_\ell}.$$

For the ground state, $n = 1$, $\ell = 0$, $m_\ell = 0$, and there is only one wave function that serves for both $m_s = +\frac{1}{2}$ and $m_s = -\frac{1}{2}$ (the value of m_s does not affect the spatial dependence of the wave function for any state, since spin is an *internal* or *intrinsic* property of the electron). The probability density for the ground state is

$$|\psi_{100}|^2 = \frac{1}{\pi r_0^3} e^{-\frac{2r}{r_0}} \qquad \textbf{(39–5b)}$$

which falls off exponentially with r. Note that ψ_{100}, like all other wave functions we discuss, has been normalized:

$$\int_{\text{all space}} |\psi_{100}|^2\, dV = 1.$$

The quantity $|\psi|^2\, dV$ gives the probability of finding the electron in a volume dV about a given point. It is often more useful to specify the **radial probability distribution**, P_r, which is defined so that $P_\text{r}\, dr$ is the probability of finding the electron at a radial distance between r and $r + dr$ from the nucleus. That is, $P_\text{r}\, dr$ specifies the probability of finding the electron within a thin shell of thickness dr of inner radius r and outer radius $r + dr$, regardless of direction (see Fig. 39–6). The volume of this shell is the product of its surface area, $4\pi r^2$, and its thickness, dr:

$$dV = 4\pi r^2\, dr.$$

Hence

$$|\psi|^2\, dV = |\psi|^2 4\pi r^2\, dr$$

and the radial probability distribution is

$$P_\text{r} = 4\pi r^2 |\psi|^2. \qquad \textbf{(39–6)}$$

For the ground state of hydrogen, P_r becomes

$$P_\text{r} = 4\frac{r^2}{r_0^3} e^{-\frac{2r}{r_0}} \qquad \textbf{(39–7)}$$

and is plotted in Fig. 39–7. The peak of the curve is the "most probable" value of r and occurs for $r = r_0$, the Bohr radius, which we show in the following Example.

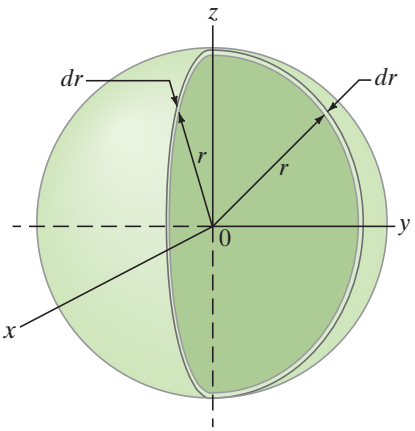
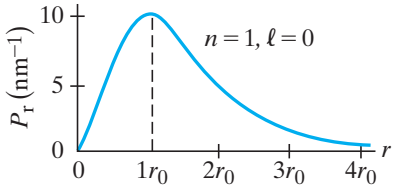


FIGURE 39–6 A spherical shell of thickness dr, inner radius r, and outer radius $r + dr$. Its volume is $dV = 4\pi r^2\, dr$.

FIGURE 39–7 The radial probability distribution P_r for the ground state of hydrogen, $n = 1$, $\ell = 0$. The peak occurs at $r = r_0$, the Bohr radius.



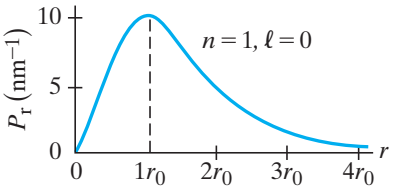


FIGURE 39–7 (Repeated.) The radial probability distribution P_r for the ground state of hydrogen, $n = 1$, $\ell = 0$. The peak occurs at $r = r_0$, the Bohr radius.

EXAMPLE 39–3 **Most probable electron radius in hydrogen.** Determine the most probable distance r from the nucleus at which to find the electron in the ground state of hydrogen.

APPROACH The peak of the curve in Fig. 39–7 corresponds to the most probable value of r. At this point the curve has zero slope, so we take the derivative of Eq. 39–7, set it equal to zero, and solve for r.

SOLUTION We find

$$\frac{d}{dr}\left(4\frac{r^2}{r_0^3}e^{-\frac{2r}{r_0}}\right) = 0$$

$$\left(8\frac{r}{r_0^3} - \frac{8r^2}{r_0^4}\right)e^{-\frac{2r}{r_0}} = 0.$$

Since $e^{-\frac{2r}{r_0}}$ goes to zero only at $r = \infty$, it is the term in parentheses that must be zero:

$$8\frac{r}{r_0^3} - 8\frac{r^2}{r_0^4} = 0.$$

Therefore,

$$\frac{r}{r_0^3} = \frac{r^2}{r_0^4}$$

or

$$r = r_0.$$

The most probable radial distance of the electron from the nucleus according to quantum mechanics is at the Bohr radius, an interesting coincidence.

EXERCISE D Return to the first Chapter-Opening Question, page 1158, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 39–4 **Calculating probability.** Determine the probability of finding the electron in the ground state of hydrogen within two Bohr radii of the nucleus.

APPROACH We need to integrate P_r from $r = 0$ out to $r = 2r_0$.

SOLUTION We want to find

$$P = \int_{r=0}^{2r_0} |\psi|^2\, dV = \int_0^{2r_0} 4\frac{r^2}{r_0^3}e^{-\frac{2r}{r_0}}\, dr.$$

We first make the substitution

$$x = 2\frac{r}{r_0}$$

and then integrate by parts $\left(\int u\, dv = uv - \int v\, du\right)$ letting $u = x^2$ and $dv = e^{-x}\, dx$ (and note that $dx = 2\, dr/r_0$, and the upper limit is $x = 2(2r_0)/r_0 = 4$):

$$P = \tfrac{1}{2}\int_{x=0}^{4} x^2 e^{-x}\, dx = \tfrac{1}{2}\left[-x^2e^{-x} + \int 2xe^{-x}\, dx\right]\Bigg|_0^4.$$

The second term we also integrate by parts with $u = 2x$ and $dv = e^{-x}\, dx$:

$$P = \tfrac{1}{2}\left[-x^2e^{-x} - 2xe^{-x} + 2\int e^{-x}\, dx\right]\Bigg|_0^4$$

$$= \left(-\tfrac{1}{2}x^2 - x - 1\right)e^{-x}\Bigg|_0^4.$$

We evaluate this at $x = 0$ and at $x = 4$:

$$P = (-8 - 4 - 1)e^{-4} + e^0 = 0.76$$

or 76%. Thus the electron would be found 76% of the time within 2 Bohr radii of the nucleus and 24% of the time farther away.

NOTE This result depends on our wave function being properly normalized, which it is, as is readily shown by letting $r \to \infty$ and integrating over all space: $\int_0^\infty |\psi|^2\, dV = 1$; that is, let the upper limit in the equation above be ∞.

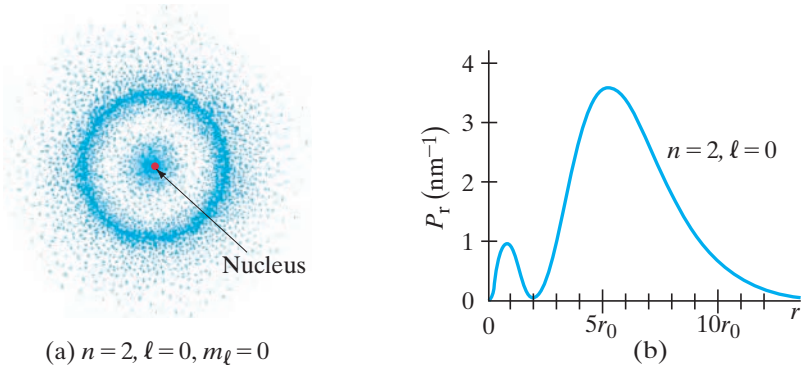


FIGURE 39–8 (a) Electron cloud, or probability distribution, for $n = 2$, $\ell = 0$ state in hydrogen. (b) The radial probability distribution P_r for the $n = 2$, $\ell = 0$ state in hydrogen.

The first excited state in hydrogen has $n = 2$. For $\ell = 0$, the solution of the Schrödinger equation (Eq. 39–1) is a wave function that is again spherically symmetric:

$$\psi_{200} = \frac{1}{\sqrt{32\pi r_0^3}}\left(2 - \frac{r}{r_0}\right)e^{-\frac{r}{2r_0}}. \qquad \textbf{(39–8)}$$

Figure 39–8a shows the probability distribution $|\psi_{200}|^2$ and Fig. 39–8b shows[†] a plot of the radial probability distribution

$$P_r = 4\pi r^2|\psi|^2 = \frac{1}{8}\frac{r^2}{r_0^3}\left(2 - \frac{r}{r_0}\right)^2 e^{-\frac{r}{r_0}}.$$

There are two peaks in this curve; the second, at $r \approx 5r_0$, is higher and corresponds to the most probable value for r in the $n = 2$, $\ell = 0$ state. We see that the electron tends to be somewhat farther from the nucleus in the $n = 2$, $\ell = 0$ state than in the $n = 1$, $\ell = 0$ state. (Compare to the Bohr model that gave $r_2 = 4r_0$.)

For the state with $n = 2$, $\ell = 1$, there are three possible wave functions, corresponding to $m_\ell = +1$, 0, or -1:

$$\begin{aligned}\psi_{210} &= \frac{z}{\sqrt{32\pi r_0^5}}e^{-\frac{r}{2r_0}} \\ \psi_{211} &= \frac{x + iy}{\sqrt{64\pi r_0^5}}e^{-\frac{r}{2r_0}} \\ \psi_{21-1} &= \frac{x - iy}{\sqrt{64\pi r_0^5}}e^{-\frac{r}{2r_0}}\end{aligned} \qquad \textbf{(39–9)}$$

where i is the imaginary number $i = \sqrt{-1}$. These wave functions are *not* spherically symmetric. The probability distributions, $|\psi|^2$, are shown in Fig. 39–9a, where we can see their directional orientation.

You may wonder how such non-spherically symmetric wave functions arise when the potential energy in the Schrödinger equation has spherical symmetry. Indeed, how could an electron select one of these states? In the absence of any external influence, such as a magnetic field in a particular direction, all three of these states are equally likely, and they all have the same energy. Thus an electron can be considered to spend one-third of its time in each of these states. The net effect, then, is the sum of these three wave functions squared:

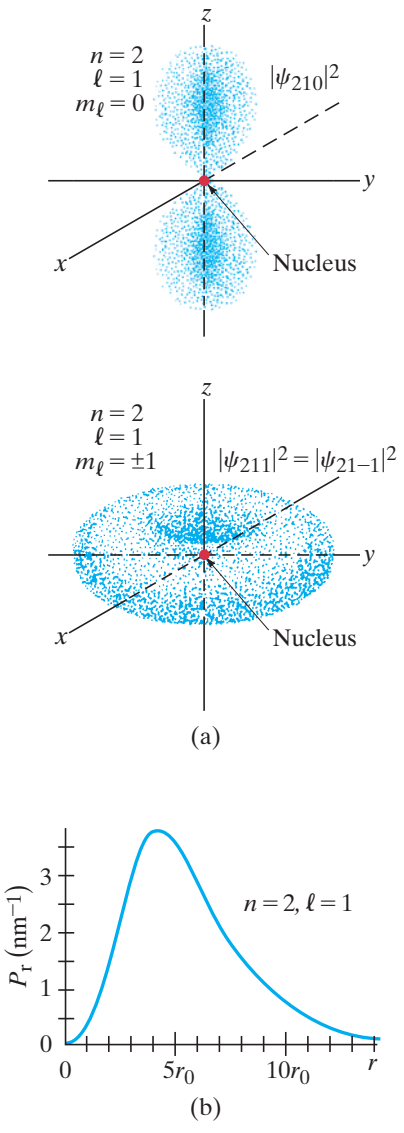
$$|\psi_{210}|^2 + |\psi_{211}|^2 + |\psi_{21-1}|^2,$$

which is spherically symmetric, since $x^2 + y^2 + z^2 = r^2$. The radial probability distribution for this sum is shown in Fig. 39–9b.

Although the spatial distributions of the electron can be calculated for the various states, it is difficult to measure them experimentally. Indeed, most of the experimental information about atoms has come from a careful examination of the emission spectra under various conditions as in Figs. 37–20 and 35–22.

[†]Just as for a particle in a deep square well potential (see Figs. 38–9 and 38–10), the higher the energy, the more nodes there are in ψ and $|\psi|^2$ also for the H atom.

FIGURE 39–9 (a) The probability distribution for the three states with $n = 2$, $\ell = 1$. (b) Radial probability distribution for the sum of the three states with $n = 2$, $\ell = 1$, and $m_\ell = +1$, 0, or -1.



Orbitals

Chemists refer to atomic states, and especially the shape in space of their probability distributions, as **orbitals**. Each atomic orbital is characterized by its quantum numbers n, ℓ, and m_ℓ, and can hold one or two electrons $\left(m_s = +\frac{1}{2} \text{ or } m_s = -\frac{1}{2}\right)$; s-orbitals $(\ell = 0)$ are spherically symmetric, Figs. 39–1 and 39–8a; p-orbitals $(\ell = 1)$ can be dumbbell shaped with lobes, Fig. 39–9a; or donut shaped if combining $m_\ell = +1$ and $m_\ell = -1$, Fig. 39–9b.

39–4 Multielectron Atoms; the Exclusion Principle

We have discussed the hydrogen atom in detail because it is the simplest to deal with. Now we briefly discuss more complex atoms, those that contain more than one electron. Their energy levels can be determined experimentally from an analysis of their emission spectra. The energy levels are *not* the same as in the H atom, because the electrons interact with each other as well as with the nucleus. Each electron in a complex atom still occupies a particular state characterized by the same quantum numbers n, ℓ, m_ℓ, and m_s. For atoms with more than one electron, the energy levels depend on both n and ℓ.

The number of electrons in a neutral atom is called its **atomic number**, Z; Z is also the number of positive charges (protons) in the nucleus, and determines what kind of atom it is. That is, Z determines the fundamental properties that distinguish one type of atom from another.

Quantum mechanics in the years after 1925 proved successful also in dealing with complex atoms. The mathematics becomes very difficult, however, since in multi-electron atoms, each electron is not only attracted to the nucleus but is also repelled by the other electrons.

To understand the possible arrangements of electrons in an atom, a new principle was needed. It was introduced by Wolfgang Pauli (1900–1958; Fig. 38–2) and is called the **Pauli exclusion principle**. It states:

No two electrons in an atom can occupy the same quantum state.

Thus, no two electrons in an atom can have exactly the same set of the quantum numbers n, ℓ, m_ℓ, and m_s. The Pauli exclusion principle forms the basis not only for understanding atoms, but also for understanding molecules and bonding, and other phenomena as well.

Let us now look at the structure of some of the simpler atoms when they are in the ground state. After hydrogen, the next simplest atom is *helium* with two electrons. Both electrons can have $n = 1$, because one can have spin up $\left(m_s = +\frac{1}{2}\right)$ and the other spin down $\left(m_s = -\frac{1}{2}\right)$, thus satisfying the exclusion principle. Since $n = 1$, then ℓ and m_ℓ must be zero (Table 39–1, page 1161). Thus the two electrons have the quantum numbers indicated at the top of Table 39–2.

Lithium has three electrons, two of which can have $n = 1$. But the third cannot have $n = 1$ without violating the exclusion principle. Hence the third electron must have $n = 2$. It happens that the $n = 2$, $\ell = 0$ level has a lower energy than $n = 2$, $\ell = 1$. So the electrons in the ground state have the quantum numbers indicated in Table 39–2. The quantum numbers of the third electron could also be, say, $\left(n, \ell, m_\ell, m_s\right) = \left(3, 1, -1, \frac{1}{2}\right)$. But the atom in this case would be in an excited state because it would have greater energy. It would not be long before it jumped to the ground state with the emission of a photon. At room temperature, unless extra energy is supplied (as in a discharge tube), the vast majority of atoms are in the ground state.

We can continue in this way to describe the quantum numbers of each electron in the ground state of larger and larger atoms. The quantum numbers for sodium, with its eleven electrons, are shown in Table 39–2.

TABLE 39–2 Ground-State Quantum Numbers

n	ℓ	m_ℓ	m_s
Helium, $Z = 2$			
1	0	0	$\frac{1}{2}$
1	0	0	$-\frac{1}{2}$
Lithium, $Z = 3$			
1	0	0	$\frac{1}{2}$
1	0	0	$-\frac{1}{2}$
2	0	0	$\frac{1}{2}$
Sodium, $Z = 11$			
1	0	0	$\frac{1}{2}$
1	0	0	$-\frac{1}{2}$
2	0	0	$\frac{1}{2}$
2	0	0	$-\frac{1}{2}$
2	1	1	$\frac{1}{2}$
2	1	1	$-\frac{1}{2}$
2	1	0	$\frac{1}{2}$
2	1	0	$-\frac{1}{2}$
2	1	−1	$\frac{1}{2}$
2	1	−1	$-\frac{1}{2}$
3	0	0	$\frac{1}{2}$

EXERCISE E Construct a Table of the ground-state quantum numbers for beryllium, $Z = 4$ (like those in Table 39–2).

PSE5_CH39_v1.9_23-Jun-20

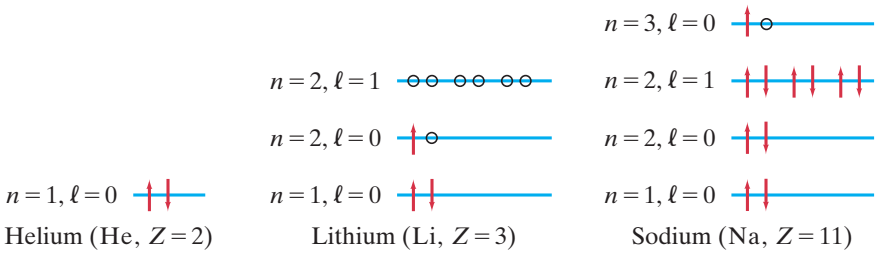


FIGURE 39–10 Energy level diagram (not to scale) showing occupied states (arrows) and unoccupied states (○) for the ground states of He, Li, and Na. Note that we have shown the $n = 2$, $\ell = 1$ level of Li even though it is empty.

Figure 39–10 shows a simple energy level diagram where occupied states are shown as up or down arrows $\left(m_s = +\frac{1}{2} \text{ or } -\frac{1}{2}\right)$, and possible empty states are shown as a small circle.

The ground-state configuration for all atoms is given in the **Periodic Table**, which is displayed inside the back cover of this book, and discussed in the next Section.

The *exclusion principle* applies to identical particles whose spin quantum number is a half-integer $\left(\frac{1}{2}, \frac{3}{2}, \text{ and so on}\right)$, including electrons, protons, and neutrons; such particles are called **fermions**, after Enrico Fermi who derived a statistical theory describing them. A basic assumption is that all electrons are **identical**, indistinguishable one from another. Similarly, all protons are identical, all neutrons are identical, and so on. The exclusion principle does not apply to particles with integer spin (0, 1, 2, and so on), such as the photon and π meson, all of which are referred to as **bosons** (after Satyendranath Bose, who derived a statistical theory for them).

EXERCISE F Return to the second Chapter-Opening Question, page 1158, and answer it again now. Try to explain why you may have answered differently the first time.

39–5 Periodic Table of Elements

More than a century ago, Dmitri Mendeleev (1834–1907) arranged the (then) known elements into what we now call the **Periodic Table** of the elements. The atoms were arranged according to increasing mass, but also so that elements with similar chemical properties would fall in the same column. Today's version is shown inside the back cover of this book. Each square contains the atomic number Z, the symbol for the element, and the atomic mass (in atomic mass units). Finally, the lower left corner shows the configuration of the ground state of the atom. This requires some explanation. Electrons with the same value of n are referred to as being in the same **shell**. Electrons with $n = 1$ are in one shell (the K shell), those with $n = 2$ are in a second shell (the L shell), those with $n = 3$ are in the third (M) shell, and so on. Electrons with the same values of n and ℓ are referred to as being in the same **subshell**. Letters are often used to specify the value of ℓ as shown in Table 39–3. That is, $\ell = 0$ is the s subshell; $\ell = 1$ is the p subshell; $\ell = 2$ is the d subshell; beginning with $\ell = 3$, the letters follow the alphabet, f, g, h, i, and so on. (The first letters s, p, d, and f were originally abbreviations of "sharp," "principal," "diffuse," and "fundamental," terms referring to the experimentally observed spectra.)

The Pauli exclusion principle limits the number of electrons possible in each shell and subshell. For any value of ℓ, there are $2\ell + 1$ possible m_ℓ values (m_ℓ can be any integer from 1 to ℓ, from -1 to $-\ell$, or zero), and two possible m_s values. There can be, therefore, at most $2(2\ell + 1)$ electrons in any ℓ subshell. For example, for $\ell = 2$, five m_ℓ values are possible $(2, 1, 0, -1, -2)$, and for each of these, m_s can be $+\frac{1}{2}$ or $-\frac{1}{2}$ for a total of $2(5) = 10$ states. Table 39–3 lists the maximum number of electrons that can occupy each subshell.

Because the energy levels depend almost entirely on the values of n and ℓ, it is customary to specify the electron configuration simply by giving the n value and the appropriate letter for ℓ, with the number of electrons in each subshell given as a superscript.

TABLE 39–3 Value of ℓ

Value of ℓ	Letter Symbol	Maximum Number of Electrons in Subshell
0	s	2
1	p	6
2	d	10
3	f	14
4	g	18
5	h	22
⋮	⋮	⋮

TABLE 39–4 Electron Configuration of Some Elements

Z (Number of Electrons)	Element[†]	Ground State Configuration (outer electrons)
1	H	$1s^1$
2	He	$1s^2$
3	Li	$2s^1$
4	Be	$2s^2$
5	B	$2s^2 2p^1$
6	C	$2s^2 2p^2$
7	N	$2s^2 2p^3$
8	O	$2s^2 2p^4$
9	F	$2s^2 2p^5$
10	Ne	$2s^2 2p^6$
11	Na	$3s^1$
12	Mg	$3s^2$
13	Al	$3s^2 3p^1$
14	Si	$3s^2 3p^2$
15	P	$3s^2 3p^3$
16	S	$3s^2 3p^4$
17	Cl	$3s^2 3p^5$
18	Ar	$3s^2 3p^6$
19	K	$4s^1$
20	Ca	$4s^2$
21	Sc	$3d^1 4s^2$
22	Ti	$3d^2 4s^2$
23	V	$3d^3 4s^2$
24	Cr	$3d^5 4s^1$
25	Mn	$3d^5 4s^2$
26	Fe	$3d^6 4s^2$

[†]Names of elements can be found in Appendix F.

⚠ CAUTION

Subshells are not always filled in "order"

The ground-state configuration of sodium, for example, is written as $1s^2 2s^2 2p^6 3s^1$. This is simplified in the Periodic Table by specifying the configuration only of the outermost electrons and any other nonfilled subshells (see Table 39–4 here, and the Periodic Table inside the back cover).

CONCEPTUAL EXAMPLE 39–5 Electron configurations. Which of the following electron configurations are possible, and which are not: (*a*) $1s^2 2s^2 2p^6 3s^3$; (*b*) $1s^2 2s^2 2p^6 3s^2 3p^5 4s^2$; (*c*) $1s^2 2s^2 2p^6 2d^1$?

RESPONSE (*a*) This is not allowed, because too many electrons (three) are shown in the *s* subshell of the M $(n = 3)$ shell. The *s* subshell has $m_\ell = 0$, with two slots only, for "spin up" and "spin down" electrons.

(*b*) This is allowed, but it is an excited state. One of the electrons from the 3*p* subshell has jumped up to the 4*s* subshell. Since there are 19 electrons, the element is potassium.

(*c*) This is not allowed, because there is no *d* $(\ell = 2)$ subshell in the $n = 2$ shell (Table 39–1). The outermost electron will have to be (at least) in the $n = 3$ shell.

EXERCISE G Write the complete ground-state configuration for gallium, with its 31 electrons.

The grouping of atoms in the Periodic Table is according to increasing atomic number, Z. There is also a strong regularity according to chemical properties. Although this is treated in chemistry textbooks, we discuss it here briefly because it is a result of quantum mechanics. See the Periodic Table inside the back cover.

All the **noble gases** (in column VIII of the Periodic Table) have completely filled shells or subshells. That is, their outermost subshell is completely full, and the electron distribution is spherically symmetric. With such full spherical symmetry, other electrons are not attracted nor are electrons readily lost (ionization energy is high). This is why the noble gases are chemically inert (more on this when we discuss molecules and bonding in Chapter 40). Column VII contains the **halogens**, which lack one electron from a filled shell. Because of the shapes of the orbits (see Section 40–1), an additional electron can be accepted from another atom, and hence these elements are quite reactive. They have a *valence* of -1, meaning that when an extra electron is acquired, the resulting ion has a net charge of $-1e$. Column I of the Periodic Table contains the **alkali metals**, all of which have a single outer *s* electron. This electron spends most of its time outside the inner filled shells and subshells which shield it from most of the nuclear charge. Indeed, it is relatively far from the nucleus and is attracted to it by a net charge of only about $+1e$, because of the shielding effect of the other electrons. Hence this outer electron is easily removed and can spend much of its time around another atom, forming a molecule. This is why the alkali metals are chemically very reactive and have a valence of $+1$. The other columns of the Periodic Table can be treated similarly.

The presence of the **transition elements** in the center of the Periodic Table, as well as the lanthanides (rare earths) and actinides below, is a result of incomplete inner shells. For the lowest Z elements, the subshells are filled in a simple order: first 1*s*, then 2*s*, followed by 2*p*, 3*s*, and 3*p*. You might expect that 3*d* $(n = 3, \ell = 2)$ would be filled next, but it isn't. Instead, the 4*s* level actually has a slightly lower energy than the 3*d* (due to electrons interacting with each other), so it fills first (K and Ca). Only then does the 3*d* shell start to fill up, beginning with Sc, as can be seen in Table 39–4. (The 4*s* and 3*d* levels are close, so some elements have only one 4*s* electron, such as Cr.) Most of the chemical properties of these transition elements are governed by the relatively loosely held 4*s* electrons, and hence they usually have valences of $+1$ or $+2$. A similar effect is responsible for the *lanthanides* and *actinides*, which are shown at the bottom of the Periodic Table for convenience. All have very similar chemical properties, which are determined by their two outer 6*s* or 7*s* electrons, whereas the different numbers of electrons in the unfilled inner shells have little effect.

PSE5_CH39_v1.9_23-Jun-20

39–6 X-Ray Spectra and Atomic Number

The line spectra of atoms in the visible, UV, and IR regions of the EM spectrum are mainly due to transitions between states of the outer electrons. Much of the positive charge of the nucleus is shielded from these electrons by the negative charge on the inner electrons. But the innermost electrons in the $n = 1$ shell "see" the full charge of the nucleus. Since the energy of a level is proportional to Z^2 (see Eq. 37–14), for an atom with $Z = 50$, we would expect wavelengths about $50^2 = 2500$ times shorter than those found in the Lyman series of hydrogen (around 100 nm), or $(100\,\text{nm})/(2500) \approx 10^{-2}$ to 10^{-1} nm. Such short wavelengths lie in the X-ray region of the spectrum.

X-rays are produced when electrons accelerated by a high voltage strike the metal target inside an X-ray tube (Section 35–10). If we look at the spectrum of wavelengths emitted by an X-ray tube, we see that the spectrum consists of two parts: a continuous spectrum with a cutoff at some λ_0 which depends only on the voltage across the tube, and a series of peaks superimposed. A typical example is shown in Fig. 39–11. The smooth curve and the cutoff wavelength λ_0 move to the left as the voltage across the tube increases. The sharp lines or peaks (labeled K_α and K_β in Fig. 39–11), however, remain at the same wavelength when the voltage is changed, although they are located at different wavelengths when different target materials are used. This observation suggests that the peaks are characteristic of the target material used. Indeed, we can explain the peaks by imagining that the electrons accelerated by the high voltage of the tube can reach sufficient energies that when they collide with the atoms of the target, they can knock out one of the very tightly held inner electrons. Then we explain these **characteristic X-rays** (the peaks in Fig. 39–11) as photons emitted when an electron in an upper state drops down to fill the vacated lower state. The K lines result from transitions *into* the K shell $(n = 1)$. The K_α line consists of photons emitted in a transition that originates from the $n = 2$ (L) shell and drops to the $n = 1$ (K) shell. On the other hand, the K_β line reflects a transition from the $n = 3$ (M) shell down to the K shell. An L line is due to a transition into the L shell, and so on.

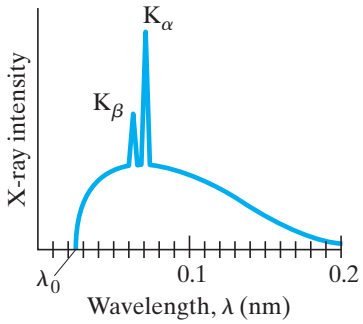
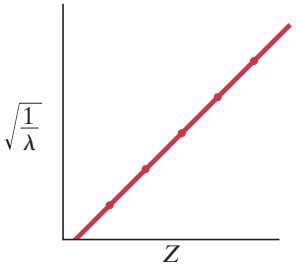


FIGURE 39–11 Spectrum of X-rays emitted from a molybdenum target in an X-ray tube operated at 50 kV.

Measurement of the characteristic X-ray spectra has allowed a determination of the inner energy levels of atoms. It has also allowed the determination of Z values for many atoms, because (as we have seen) the wavelength of the shortest characteristic X-rays emitted will be inversely proportional to Z^2. Actually, for an electron jumping from, say, the $n = 2$ to the $n = 1$ level (K_α line), the wavelength is inversely proportional to $(Z - 1)^2$ because the nucleus is shielded by the one electron that still remains in the 1s level. In 1914, H. G. J. Moseley (1887–1915) found that a plot of $\sqrt{1/\lambda}$ vs. Z produced a straight line, Fig. 39–12 where λ is the wavelength of the K_α line. The Z values of a number of elements were determined by fitting them to such a **Moseley plot**. The work of Moseley put the concept of atomic number on a firm experimental basis.

FIGURE 39–12 Plot of $\sqrt{1/\lambda}$ vs. Z for K_α X-ray lines.



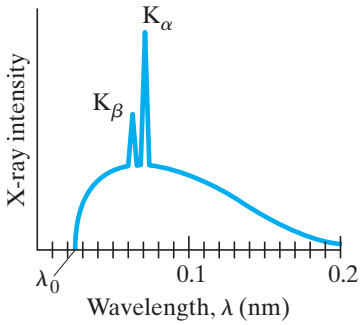


FIGURE 39–11 (Repeated.) Spectrum of X-rays emitted from a molybdenum target in an X-ray tube operated at 50 kV.

EXAMPLE 39–6 **X-ray wavelength.** Estimate the wavelength for an $n = 2$ to $n = 1$ transition in molybdenum $(Z = 42)$. What is the energy of such a photon?

APPROACH We use the Bohr formula, Eq. 37–15 for $1/\lambda$, with Z^2 replaced by $(Z-1)^2 = (41)^2$.

SOLUTION Equation 37–15 gives

$$\frac{1}{\lambda} = \left(\frac{e^4 m}{8\epsilon_0^2 h^3 c}\right)(Z-1)^2\left(\frac{1}{n'^2} - \frac{1}{n^2}\right)$$

where $n = 2$ and $n' = 1$. We substitute in values:

$$\frac{1}{\lambda} = (1.097 \times 10^7\,\text{m}^{-1})(41)^2\left(\frac{1}{1} - \frac{1}{4}\right) = 1.38 \times 10^{10}\,\text{m}^{-1}.$$

So

$$\lambda = \frac{1}{1.38 \times 10^{10}\,\text{m}^{-1}} = 0.072\,\text{nm}.$$

This is close to the measured value (Fig. 39–11) of 0.071 nm. Each of these photons would have energy (in eV) of:

$$E = hf = \frac{hc}{\lambda} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.00 \times 10^8\,\text{m/s})}{(7.2 \times 10^{-11}\,\text{m})(1.60 \times 10^{-19}\,\text{J/eV})} = 17\,\text{keV}.$$

The denominator includes the conversion factor from joules to eV.

EXAMPLE 39–7 **Determining atomic number.** High-energy electrons are used to bombard an unknown material. The strongest peak is found for X-rays emitted with an energy of 7.5 keV. Guess what the material is.

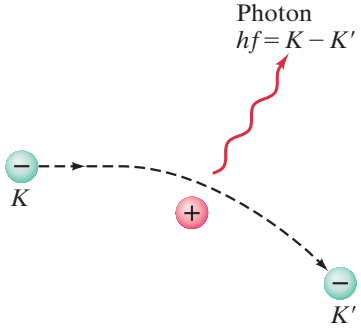
APPROACH The highest intensity X-rays are generally for the K_α line (see Fig. 39–11) which occurs when the high-energy electrons knock out K shell electrons (the innermost orbit, $n = 1$) and their place is taken by electrons from the L shell $(n = 2)$. We use the Bohr model, and assume the electrons of the unknown atoms (Z) "see" a nuclear charge of $Z - 1$ (screened by one electron).

SOLUTION The hydrogen transition $n = 2$ to $n = 1$ would yield $E_H = 13.6\,\text{eV} - 3.4\,\text{eV} = 10.2\,\text{eV}$ (see Fig. 37–26 or Example 37–13). Energy of our unknown E_Z is proportional to Z^2 (Eq. 37–14), or rather $(Z-1)^2$ because the nucleus is shielded by the one electron in a 1*s* state (see above), so we can use ratios:

$$\frac{E_Z}{E_H} = \frac{(Z-1)^2}{1^2} = \frac{7500\,\text{eV}}{10.2\,\text{eV}} = 735,$$

so $Z - 1 = \sqrt{735} = 27$, and $Z = 28$, which makes it cobalt.

FIGURE 39–13 Bremsstrahlung photon produced by an electron decelerated by interaction with a target atom.



Now we briefly analyze the continuous part of an X-ray spectrum (Fig. 39–11) based on the photon theory of light. When electrons strike the target, they collide with atoms of the material and give up most of their energy as heat (about 99%, so X-ray tubes must be cooled). Electrons can also give up energy by emitting a photon of light: an electron decelerated by interaction with atoms of the target (Fig. 39–13) emits radiation because of its deceleration (Chapter 31), and in this case it is called **bremsstrahlung** (German for "braking radiation"). Because energy is conserved, the energy of the emitted photon, hf, equals the loss of kinetic energy of the electron, $\Delta K = K - K'$, so

$$hf = \Delta K.$$

An electron may lose all or a part of its energy in such a collision. The continuous X-ray spectrum (Fig. 39–11) is explained as being due to such bremsstrahlung collisions in which varying amounts of energy are lost by the electrons. The shortest-wavelength X-ray (the highest frequency) must be due to an electron that gives up *all* its kinetic energy to produce one photon in a single collision.

PSE5_CH39_v1.9_23-Jun-20

Since the initial kinetic energy of an electron is equal to the energy given it by the accelerating voltage, V, then $K = eV$. In a single collision in which the electron is brought to rest ($\text{KE}' = 0$), then $\Delta K = eV$ and

$$hf_0 = eV.$$

We set $f_0 = c/\lambda_0$ where λ_0 is the cutoff wavelength (Fig. 39–11) and find

$$\lambda_0 = \frac{hc}{eV}. \qquad \textbf{(39–10)}$$

This prediction for λ_0 corresponds precisely with that observed experimentally. This result is further evidence that X-rays are a form of electromagnetic radiation (light)[†] and that the photon theory of light is valid.

EXAMPLE 39–8 **Cutoff wavelength.** What is the shortest-wavelength X-ray photon emitted in an X-ray tube subjected to 50 kV?

APPROACH The electrons striking the target will have a kinetic energy of 50 keV. The shortest-wavelength photons are due to collisions in which all of an electron's kinetic energy is given to the photon so $K = eV = hf_0$.

SOLUTION From Eq. 39–10,

$$\lambda_0 = \frac{hc}{eV} = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.0 \times 10^8\,\text{m/s})}{(1.6 \times 10^{-19}\,\text{C})(5.0 \times 10^4\,\text{V})} = 2.5 \times 10^{-11}\,\text{m},$$

or 0.025 nm.

NOTE This result agrees well with experiment, Fig. 39–11.

*39–7 Magnetic Dipole Moment; Total Angular Momentum

* Magnetic Dipole Moment and the Bohr Magneton

An electron orbiting the nucleus of an atom can be considered as a current loop, classically, and thus might be expected to have a **magnetic dipole moment** as discussed in Chapter 27. Indeed, in Example 27–12 we did a classical calculation of the magnetic dipole moment of the electron in the ground state of hydrogen based, essentially, on the Bohr model, and found it to give

$$\mu = IA = \tfrac{1}{2}evr.$$

Here v is the orbital velocity of the electron, and for a particle moving in a circle of radius r, its angular momentum is

$$L = mvr.$$

So we can write

$$\mu = \frac{1}{2}\frac{e}{m}L.$$

The direction of the angular momentum $\vec{\mathbf{L}}$ is perpendicular to the plane of the current loop. So is the direction of the magnetic dipole moment vector $\vec{\boldsymbol{\mu}}$, although in the opposite direction since the electron's charge is negative. Hence we can write the vector equation

$$\vec{\boldsymbol{\mu}} = -\frac{1}{2}\frac{e}{m}\vec{\mathbf{L}}. \qquad \textbf{(39–11)}$$

This rough semiclassical derivation was based on the Bohr model. The same result (Eq. 39–11) is obtained using quantum mechanics. Since $\vec{\mathbf{L}}$ is quantized in quantum mechanics, the magnetic dipole moment, too, must be quantized.

[†]If X-rays were not photons but rather neutral particles with mass m, Eq. 39–10 would not be valid.

The magnitude of the dipole moment is given by (see Eq. 39–3)

$$\mu = \frac{e\hbar}{2m}\sqrt{\ell(\ell+1)}.$$

When a magnetic dipole moment is in a magnetic field $\vec{\mathbf{B}}$, it experiences a torque as we saw in Section 27–5, and the potential energy U of such a system depends on $\vec{\mathbf{B}}$ and the orientation of $\vec{\boldsymbol{\mu}}$ relative to $\vec{\mathbf{B}}$ (Eq. 27–12):

$$U = -\vec{\boldsymbol{\mu}} \cdot \vec{\mathbf{B}}.$$

If the magnetic field $\vec{\mathbf{B}}$ is in the z direction, then

$$U = -\mu_z B_z$$

and from Eq. 39–4 $(L_z = m_\ell \hbar)$ and Eq. 39–11, we have

$$\mu_z = -\frac{e\hbar}{2m} m_\ell.$$

(Be careful here not to confuse the electron mass m with the magnetic quantum number, m_ℓ.) It is useful to define the quantity

$$\mu_B = \frac{e\hbar}{2m} \qquad \textbf{(39–12)}$$

which is called the **Bohr magneton** and has the value $\mu_B = 9.27 \times 10^{-24}\,\text{J/T}$ (joule/tesla). Then we can write

$$\mu_z = -\mu_B m_\ell, \qquad \textbf{(39–13)}$$

where m_ℓ has integer values from 0 to $\pm\ell$ (see Table 39–1). An atom placed in a magnetic field would have its energy split into levels that differ by $\Delta U = \mu_B B$; this is the *Zeeman effect*, and was shown in Fig. 39–4.

*Stern-Gerlach Experiment and the *g*-Factor for Electron Spin

The first evidence for the splitting of atomic energy levels, known as *space quantization* (Section 39–2), came in 1922 in a famous experiment known as the **Stern-Gerlach experiment**. Silver atoms (and later others) were heated in an oven from which they escaped as shown in Fig. 39–14. The atoms were made to pass

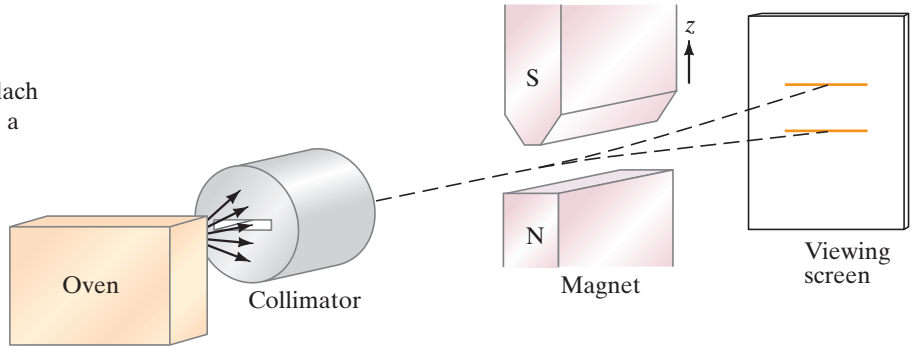


FIGURE 39–14 The Stern-Gerlach experiment, which is done inside a vacuum chamber.

through a collimator, which eliminated all but a narrow beam. The beam then passed into a *nonhomogeneous* magnetic field. The field was deliberately made nonhomogeneous so that it would exert a force on atomic magnetic moments: remember that the potential energy (in this case $-\vec{\boldsymbol{\mu}} \cdot \vec{\mathbf{B}}$) must change in space if there is to be a force $(F_x = -dU/dx,$ etc., Section 8–2). If $\vec{\mathbf{B}}$ has a gradient along the z axis, as in Fig. 39–14, then the force is along z:

$$F_z = -\frac{dU}{dz} = \mu_z \frac{dB_z}{dz}.$$

Thus the silver atoms would be deflected up or down depending on the value of μ for each atom. Classically, we would expect to see a continuous distribution on the viewing screen, since we would expect the atoms to have randomly oriented magnetic moments. But Stern and Gerlach saw instead two distinct lines for silver (and for other atoms sometimes more than two lines). These observations were the first evidence for space quantization, though not fully explained until a few years later.

PSE5_CH39_v1.9_23-Jun-20

If the lines were due to orbital angular momentum, there should have been an odd number of them, corresponding to the possible values of m_ℓ (since $\mu_z = -\mu_B m_\ell$). For $\ell = 0$, there is only one possibility, $m_\ell = 0$. For $\ell = 1$, m_ℓ can be 1, 0, or -1, and we would expect three lines, and so on. Why there are only two lines was eventually explained by the concept of electron spin. With a spin of $\frac{1}{2}$, the electron spin can have only two orientations in space, as we saw in Fig. 39–5. Hence a magnetic dipole moment associated with spin would have only two positions. Thus, the two states for silver seen in the Stern-Gerlach experiment must be due to the spin of its one valence electron. Silver atoms must thus have zero orbital angular momentum but a total spin of $\frac{1}{2}$ due to this one valence electron. (Of silver's 47 electrons, the spins of the first 46 cancel.) For the H atom in its ground state, again only two lines were seen on the screen of Fig. 39–14: due to the spin $\frac{1}{2}$ of its electron since the orbital angular momentum is zero.

The Stern-Gerlach deflection is proportional to the magnetic dipole moment, μ_z, and for a spin $\frac{1}{2}$ particle we expect $\mu_z = -\mu_B m_s = -\left(\frac{1}{2}\right)(e\hbar/2m)$ as for the case of orbital angular momentum, Eq. 39–13. Instead, μ_z for spin was found to be about twice as large:

$$\mu_z = -g\mu_B m_s, \qquad \text{[electron spin]} \quad \textbf{(39–14)}$$

where g, called the ***g*-factor** or **gyromagnetic ratio**, has been measured to be slightly larger than 2: $g = 2.0023\cdots$ for a free electron. This unexpected factor of (about) 2 clearly indicates that spin cannot be viewed as a classical angular momentum. It is a purely quantum-mechanical effect. Equation 39–14 is the same as Eq. 39–13 for orbital angular momentum with m_s replacing m_ℓ. But for the orbital case, $g = 1$.

*Total Angular Momentum $\vec{\mathbf{J}}$

An atom can have both orbital and spin angular momenta. For example, in the $2p$ state of hydrogen $\ell = 1$ and $s = \frac{1}{2}$. In the $4d$ state, $\ell = 2$ and $s = \frac{1}{2}$. The **total angular momentum** is the vector sum of the orbital angular momentum $\vec{\mathbf{L}}$ and the spin $\vec{\mathbf{S}}$:

$$\vec{\mathbf{J}} = \vec{\mathbf{L}} + \vec{\mathbf{S}}.$$

According to quantum mechanics, the magnitude of the total angular momentum $\vec{\mathbf{J}}$ is quantized:

$$J = \sqrt{j(j+1)}\,\hbar. \qquad \textbf{(39–15)}$$

For the single electron in the H atom, quantum mechanics gives the result that j can be

$$j = \ell + s = \ell + \tfrac{1}{2}$$

or

$$j = \ell - s = \ell - \tfrac{1}{2}$$

but never less than zero, just as for ℓ and s. For the $1s$ state, $\ell = 0$ and $j = \frac{1}{2}$ is the only possibility. For p states, say the $2p$ state, $\ell = 1$ and j can be either $\frac{3}{2}$ or $\frac{1}{2}$. The z component for j is quantized in the usual way:

$$m_j = j, j-1, \cdots, -j.$$

For a $2p$ state with $j = \frac{1}{2}$, m_j can be $\frac{1}{2}$ or $-\frac{1}{2}$; for $j = \frac{3}{2}$, m_j can be $\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$, for a total of four states. Note that the state of a single electron can be specified by giving n, ℓ, m_ℓ, m_s, or by giving n, j, ℓ, m_j (only one of these descriptions at a time).

EXERCISE H What are the possibilities for j in the $3d$ state of hydrogen? (*a*) $\frac{3}{2}, \frac{1}{2}$; (*b*) $\frac{5}{2}, \frac{3}{2}, \frac{1}{2}$; (*c*) $\frac{7}{2}, \frac{5}{2}, \frac{3}{2}, \frac{1}{2}$; (*d*) $\frac{5}{2}, \frac{3}{2}$; (*e*) $\frac{7}{2}, \frac{5}{2}$.

The interaction of magnetic fields with atoms, as in the Zeeman effect and the Stern-Gerlach experiment, involves the *total* angular momentum. Thus the Stern-Gerlach experiment on H atoms in the ground state shows two lines (for $m_j = +\frac{1}{2}$ and $-\frac{1}{2}$), but for the first excited state it shows four lines corresponding to the four possible m_j values $\left(\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}\right)$.

*Spectroscopic Notation

We can specify the state of an atom, including the total angular momentum quantum number j, using the following **spectroscopic notation**. For a single electron state we can write

$$nL_j,$$

where the value of L (the orbital quantum number) is specified using the same letters as in Table 39–3, but in upper case:

$$\begin{aligned} L &= 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad \cdots \\ \text{letter} &= \text{S} \quad \text{P} \quad \text{D} \quad \text{F} \quad \text{G} \quad \cdots . \end{aligned}$$

So the $2P_{3/2}$ state has $n = 2$, $\ell = 1$, $j = \frac{3}{2}$, whereas $1S_{1/2}$ specifies the ground state in hydrogen.

*Fine Structure; Spin-Orbit Interaction

A magnetic effect also produces the *fine structure* splitting mentioned in Section 39–2, but it occurs in the absence of any external field. Instead, it is due to a magnetic field produced by the atom itself. We can see how it occurs by putting ourselves in the reference frame of the electron, in which case we see the nucleus revolving about us as a moving charge or electric current that produces a magnetic field, B_n. The electron has an intrinsic magnetic dipole moment μ_s (Eq. 39–14) and hence its energy will be altered by an amount (Eq. 27–12)

$$\Delta U = -\vec{\boldsymbol{\mu}}_s \cdot \vec{\mathbf{B}}_n.$$

Since μ_s takes on quantized values according to the values of m_s, the energy of a single electron state will split into two closely spaced energy levels (for $m_s = +\frac{1}{2}$ and $-\frac{1}{2}$). This tiny splitting of energy levels produces a tiny splitting in spectral lines. For example, in the H atom, the $2P \rightarrow 1S$ transition is split into two lines corresponding to $2P_{1/2} \rightarrow 1S_{1/2}$ and $2P_{3/2} \rightarrow 1S_{1/2}$. The difference in energy between these two is only about $5 \times 10^{-5}\,\text{eV}$, which is very small compared to the $2P \rightarrow 1S$ transition energy of $13.6\,\text{eV} - 3.4\,\text{eV} = 10.2\,\text{eV}$.

The magnetic field $\vec{\mathbf{B}}_n$ produced by the orbital motion is proportional to the orbital angular momentum $\vec{\mathbf{L}}$, and since $\vec{\boldsymbol{\mu}}_s$ is proportional to the spin $\vec{\mathbf{S}}$, then $\Delta U = -\vec{\boldsymbol{\mu}}_s \cdot \vec{\mathbf{B}}_n$ can be written

$$\Delta U \propto \vec{\mathbf{L}} \cdot \vec{\mathbf{S}}.$$

This interaction, which produces the fine structure, is thus called the **spin-orbit interaction**. Its magnitude is related to a dimensionless constant known as the **fine structure constant**,

$$\alpha = \frac{e^2}{2\epsilon_0 hc} \approx \frac{1}{137},$$

which also appears elsewhere in atomic physics

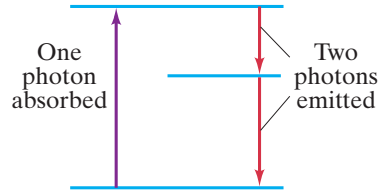
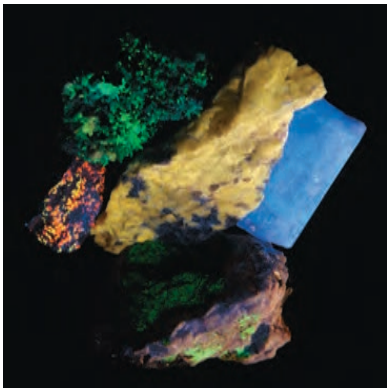


FIGURE 39–15 Fluorescence.

FIGURE 39–16 When UV light (a range of wavelengths) illuminates these various "fluorescent" rocks, they fluoresce in the visible region of the spectrum.



39–8 Fluorescence and Phosphorescence

When an atom is excited from one energy state to a higher one by the absorption of a photon, it may return to the lower level in a series of two (or more) transitions if there is at least one energy level in between (Fig. 39–15). The photons emitted will consequently have lower energy and frequency than the absorbed photon. When the absorbed photon is in the UV and the emitted photons are in the visible region of the spectrum, this phenomenon is called **fluorescence** (Fig. 39–16).

The wavelength for which fluorescence will occur depends on the energy levels of the particular atoms. Because the frequencies are different for different substances, and because many substances fluoresce readily, fluorescence is a powerful tool for identification of compounds. It is also used for assaying—determining how much of a substance is present—and for following substances along a natural *metabolic pathway* in biological organisms. For detection of a given compound, the stimulating light must be monochromatic,

and solvents or other materials present must not fluoresce in the same region of the spectrum. Sometimes the observation of fluorescent light being emitted is sufficient to detect a compound. In other cases, spectrometers are used to measure the wavelengths and intensities of the emitted light.

Fluorescent lightbulbs work in a two-step process. The applied voltage accelerates electrons that strike atoms of the gas in the tube and cause them to be excited. When the excited atoms jump down to their normal levels, they emit UV photons which strike a fluorescent coating on the inside of the tube. The light we see is a result of this material fluorescing in response to the UV light striking it.

Materials such as those used for luminous watch dials, and other glow-in-the-dark products, are said to be **phosphorescent**. When an atom is raised to a normal excited state, it drops back down within about 10^{-8} s. In phosphorescent substances, atoms can be excited by photon absorption to energy levels, called **metastable**, which are states that last much longer because to jump down is a "forbidden" transition as discussed in Section 39–2. Metastable states can last even a few seconds or longer. In a collection of such atoms, many of the atoms will descend to the lower state fairly soon, but many will remain in the excited state for over an hour. Hence light will be emitted even after long periods. When you put a luminous watch dial close to a bright lamp, many atoms are excited to metastable states, and you can see the glow for a long time afterward.

39–9 Lasers

A **laser** is a device that can produce a very narrow intense beam of monochromatic coherent light. (By *coherent*, we mean that across any cross section of the beam, all parts have the same phase.) The emitted beam is a nearly perfect plane wave. An ordinary light source, on the other hand, emits light in all directions (so the intensity decreases rapidly with distance), and the emitted light is incoherent (the different parts of a beam are not in phase with each other). The excited atoms that emit the light in an ordinary lightbulb act independently, so each photon emitted can be considered as a short wave train lasting about 10^{-8} s. Different wave trains bear no phase relation to one another. Just the opposite is true of lasers.

The action of a laser is based on quantum theory. We have seen that a photon can be absorbed by an atom if (and only if) the photon energy hf corresponds to the energy difference between an occupied energy level of the atom and an available excited state, Fig. 39–17a. If the atom is already in the excited state, it may jump down spontaneously (i.e., no stimulus) to the lower state with the emission of a photon. However, if a photon with this same energy strikes the excited atom, it can stimulate the atom to make the transition sooner to the lower state, Fig. 39–17b. This phenomenon is called **stimulated emission**: not only do we still have the original photon, but also a second one of the same frequency as a result of the atom's transition. These two photons are exactly *in phase*, and they are moving in the same direction. This is how coherent light is produced in a laser. The name "laser" is an acronym for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation.

The natural population in energy states of atoms in thermal equilibrium at any temperature T (in K) is given by the **Boltzmann distribution** (or **Boltzmann factor**):

$$N_n = Ce^{-\frac{E_n}{kT}}, \qquad \textbf{(39–16a)}$$

where N_n is the number of atoms in the state with energy E_n. For two states n and n', the ratio of the number of atoms in the two states is

$$\frac{N_n}{N_{n'}} = e^{-\left(\frac{E_n - E_{n'}}{kT}\right)}. \qquad \textbf{(39–16b)}$$

Thus most atoms are in the ground state unless the temperature is very high. In the two-level system of Fig. 39–17, most atoms are normally in the lower state, so the majority of incident photons will be absorbed. In order to obtain the coherent light from stimulated emission, two conditions must be satisfied.

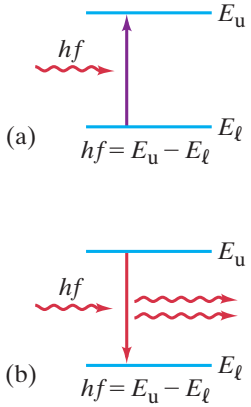


FIGURE 39–17 (a) Absorption of a photon. (b) Stimulated emission. E_u and E_ℓ refer to "upper" and "lower" energy states.



FIGURE 39–18 Two energy levels for a collection of atoms. Each dot represents the energy state of one atom. (a) A normal situation; (b) an inverted population.

⚠ CAUTION

Laser: photons have same frequency and direction, and are in phase

First, the atoms must be excited to the higher state, so that an **inverted population** is produced, one in which more atoms are in the upper state than in the lower one (Fig. 39–18). Then *emission* of photons will dominate over absorption. Hence the system will not be in thermal equilibrium. And second, the higher state must be a **metastable state**—a state in which the electrons remain longer than usual[†] so that the transition to the lower state occurs by stimulated emission rather than spontaneously. (How inverted populations are created will be discussed shortly.)

Figure 39–19 is a schematic diagram of a laser: the "lasing" material is placed in a long narrow tube at the ends of which are two mirrors, one of which is partially transparent (transmitting perhaps 1 or 2%). Some of the excited atoms drop down fairly soon after being excited. One of these is the blue atom shown on the far left in Fig. 39–19. If the emitted photon strikes another atom in the excited state, it stimulates this atom to emit a photon of the *same* frequency, moving in the *same* direction, and *in phase* with it. These two photons then move on to strike other atoms causing more stimulated emission. As the process continues, the number of photons multiplies. When the photons strike the end mirrors, most are reflected back, and as they move in the opposite direction, they continue to stimulate more atoms to emit photons. As the photons move back and forth between the mirrors, a small percentage passes through the partially transparent mirror at one end. These photons make up the narrow coherent external laser beam.

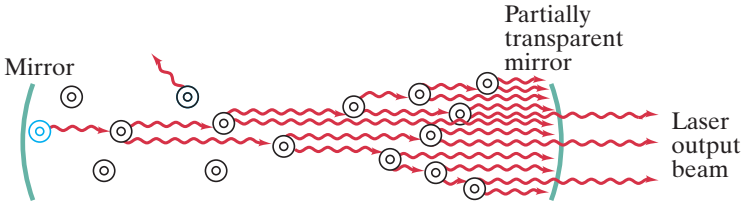


FIGURE 39–19 Laser diagram, showing excited atoms stimulated to emit light.

Inside the tube, some spontaneously emitted photons will be emitted at an angle to the axis, and these will merely go out the side of the tube and not affect the narrow width of the main beam.

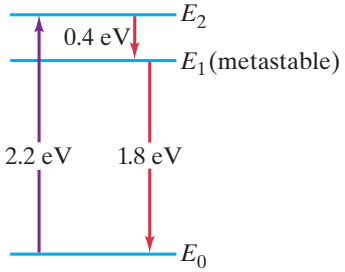
In a well-designed laser, the spreading of the beam is limited only by diffraction, so the angular spread is $\approx \lambda/D$ (see Eq. 35–1 or 35–10), where D is the diameter of the end mirror. The diffraction spreading can be incredibly small. The light energy, instead of spreading out in space as it does for an ordinary light source, can be a pencil-thin beam.

Creating an Inverted Population

The excitation of the atoms in a laser can be done in several ways to produce the necessary inverted population. In a **ruby laser**, the lasing material is a ruby rod consisting of Al_2O_3 with a small percentage of aluminum (Al) atoms replaced by chromium (Cr) atoms. The Cr atoms are the ones involved in lasing. In a process called **optical pumping**, the atoms are excited by strong flashes of light of wavelength 550 nm, which corresponds to a photon energy of 2.2 eV. As shown in Fig. 39–20, the atoms are excited from state E_0 to state E_2. The atoms quickly decay either back to E_0 or to the intermediate state E_1, which is metastable with a lifetime of about 3×10^{-3} s (compared to 10^{-8} s for ordinary levels). With strong pumping action, more atoms can be found in the E_1 state than are in the E_0 state. Thus we have the inverted population needed for lasing. As soon as a few atoms in the E_1 state jump down to E_0, they emit photons that produce stimulated emission of the other atoms, and the lasing action begins. A ruby laser thus emits a beam whose photons have energy 1.8 eV and a wavelength of 694.3 nm (or "ruby-red" light).

In a **helium–neon laser** (He–Ne), the lasing material is a gas, a mixture of about 85% He and 15% Ne. The atoms are excited by applying a high voltage to the tube so that an electric discharge takes place within the gas. In the process,

FIGURE 39–20 Energy levels of chromium in a ruby crystal. Photons of energy 2.2 eV "pump" atoms from E_0 to E_2, which then decay to metastable state E_1. Lasing action occurs by stimulated emission of photons in transition from E_1 to E_0.



[†]An excited atom may land in such a state and can jump to a lower state only by a so-called forbidden transition (Section 39–2), which is why its lifetime is longer than normal.

some of the He atoms are raised to the metastable state E_1 shown in Fig. 39–21, which corresponds to a jump of 20.61 eV, almost exactly equal to an excited state in neon, 20.66 eV. The He atoms do not quickly return to the ground state by spontaneous emission, but instead often give their excess energy to a Ne atom when they collide—see Fig. 39–21. In such a collision, the He drops to the ground state and the Ne atom is excited to the state E'_3 (the prime refers to neon states). The slight difference in energy (0.05 eV) is supplied by the kinetic energy of the moving atoms. In this manner, the E'_3 state in Ne—which is metastable—becomes more populated than the E'_2 level. This inverted population between E'_3 and E'_2 is what is needed for lasing.

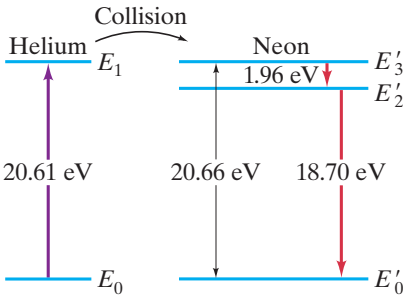


FIGURE 39–21 Energy levels for He and Ne. He is excited in the electric discharge to the E_1 state. This energy is transferred to the E'_3 level of the Ne by collision. E'_3 is metastable and decays to E'_2 by stimulated emission.

Very common now are **semiconductor diode lasers**, also called **pn junction lasers**, which utilize an inverted population of electrons between the conduction band and the lower-energy valence band (Sections 40–7 to 40–9). When an electron jumps down, a photon can be emitted, which in turn can stimulate another electron to make the transition and emit another photon, in phase. The needed mirrors (as in Fig. 39–19) are made by the polished ends of the *pn* crystal. Semiconductor lasers are used in CD and DVD players (see below), and in many other applications.

Other types of laser include: *chemical lasers*, in which the energy input comes from the chemical reaction of highly reactive gases; *dye lasers*, whose frequency is tunable; CO_2 *gas lasers*, capable of high-power output in the infrared; and *rare-earth solid-state lasers* such as the high-power Nd:YAG laser.

The excitation of the atoms in a laser can be done continuously or in pulses. In a **pulsed laser**, the atoms are excited by periodic inputs of energy. In a **continuous laser**, the energy input is continuous: as atoms are stimulated to jump down to the lower level, they are soon excited back up to the upper level so the output is a continuous laser beam.

No laser is a source of energy. Energy must be put in, and the laser converts a part of it into an intense narrow beam output.

CAUTION
Laser not an energy source

*Applications

The unique feature of light from a laser, that it is a coherent narrow beam, has found many applications. In everyday life, lasers are used as bar-code readers (at store checkout stands) and in compact disc (CD) and digital video disc (DVD) players. The laser beam reflects off the stripes and spaces of a bar code, or off the tiny pits of a CD or DVD as shown in Fig. 39–22a. The recorded information on a CD or DVD is a series of pits and spaces representing 0s and 1s (or "off" and "on") of a binary code (Section 23–9) that is decoded electronically before being sent to the audio or video system. The laser of a CD player starts reading at the inside of the disc which rotates at about 500 rpm at the start. As the disc rotates, the laser follows the spiral track (Fig. 39–22b), and as the laser moves outward the disc must slow down because each successive circumference $(C = 2\pi r)$ is slightly longer as r increases; when the laser reaches the outer edge, the disc is rotating about 200 rpm. A 1-hour CD has a track roughly 5 km long; the track width is about 1600 nm $(= 1.6\ \mu\text{m})$ and the distance between pits is about 800 nm. DVDs contain much more information. Standard DVDs use a thinner track $(0.7\ \mu\text{m})$ and shorter pit length (400 nm).

PHYSICS APPLIED
DVD and CD players, bar codes

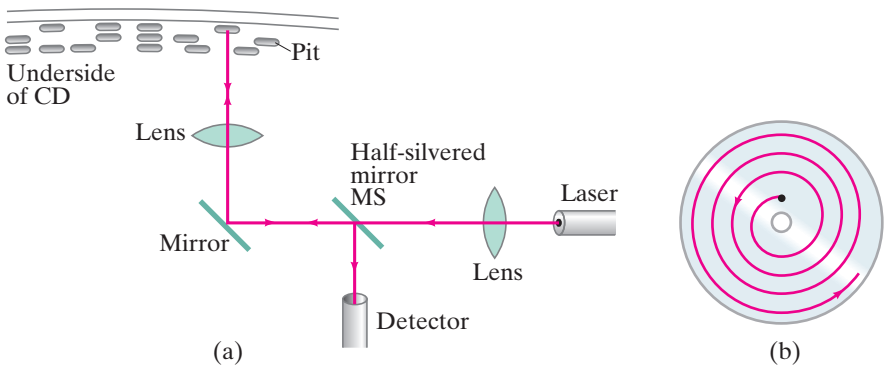


FIGURE 39–22 (a) Reading a CD (or DVD). The fine beam of a laser, focused even more finely with lenses, is directed at the undersurface of a rotating compact disc. The beam is reflected back from the areas between pits but reflects much less from pits. The reflected light is detected as shown, reflected by a half-reflecting mirror MS. The strong and weak reflections correspond to the 0s and 1s of the binary code representing the audio or video signal. (b) A laser follows the CD track which starts near the center and spirals outward.

Blu-ray discs use a "blue" laser with a short wavelength (405 nm) and narrower beam, allowing a narrower track ($0.3\ \mu m$) that can store much more data for high definition. DVDs can also have two layers, one below the other. When the laser focuses on the second layer, the light passes through the semitransparent surface layer. The second layer may start reading at the outer edge instead of inside. DVDs can also have a single or double layer on *both* surfaces of the disc.

PHYSICS APPLIED
Medical and other uses of lasers

Lasers are a useful surgical tool. The narrow intense beam can be used to destroy tissue in a localized area, or to break up gallstones and kidney stones. Because of the heat produced, a laser beam can be used to "weld" broken tissue, such as a detached retina, Fig. 39–23, or to mold the cornea of the eye (by vaporizing tiny bits of material) to correct myopia and other eye defects (LASIK surgery). The laser beam can be carried by an optical fiber (Section 32–7) to the surgical point, sometimes as an additional fiber-optic path on an endoscope (again Section 32–7). An example is the removal of plaque clogging human arteries. Lasers have been used to destroy tiny organelles within a living cell by researchers studying how the absence of that organelle affects the behavior of the cell. Laser beams are used to destroy cancerous and precancerous cells; and the heat seals off capillaries and lymph vessels, thus "cauterizing" the wound to prevent spread of the disease.



FIGURE 39–23 Laser being used in eye surgery.

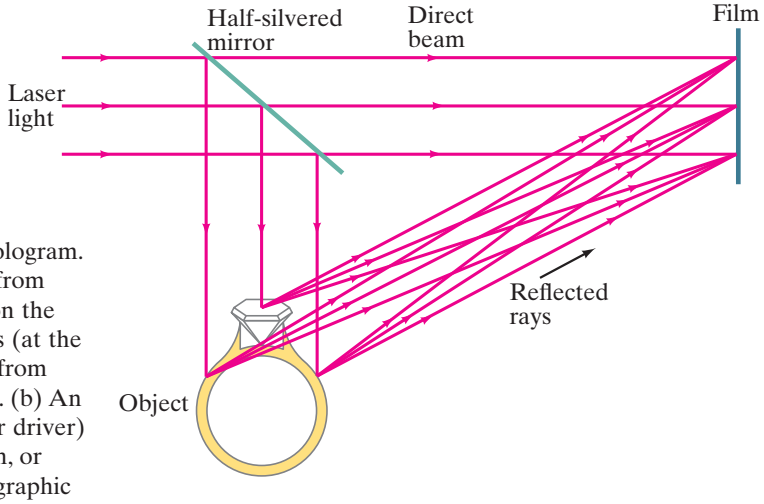
The intense heat produced in a small area by a laser beam is used for welding and machining metals and for drilling tiny holes in hard materials. Because a laser beam is coherent, monochromatic, narrow, and essentially parallel, lenses can be used to focus the light into even smaller areas. The precise straightness of a laser beam is also useful to surveyors for lining up equipment accurately, especially in inaccessible places.

The U.S. Navy is said to have a laser that is strong enough to shoot down small unmanned aircraft such as drones, and to disable small attack boats.

*39–10 Holography

PHYSICS APPLIED
Holography

One of the most interesting applications of laser light is the production of three-dimensional images called **holograms** (see Fig. 39–24). In an ordinary photograph, the film simply records the intensity of light reaching it at each point. When the photograph or transparency is viewed, light reflecting from it or passing through it gives us a two-dimensional picture. In holography, the images are formed by interference, without lenses. A laser hologram is typically made on photographic emulsion (film). A broadened laser beam is split into two parts by a half-silvered mirror, Fig. 39–24a. One part goes directly to the film; the rest passes to the object to be photographed, from which it is reflected to the film. Light from every point on the surface of the object reaches each point on the film, and the interference of the two beams allows the film to record both the intensity and relative phase of the light at each point.



(a)



(b)

FIGURE 39–24
(a) Making a hologram. Light reflected from various points on the object interferes (at the film) with light from the direct beam. (b) An athlete (race-car driver) puts his hand on, or through, a holographic image of himself.

It is crucial that the incident light be coherent—that is, in phase at all points—which is why a laser is used. After the film is developed, it is placed again in a laser beam and a three-dimensional image of the object is created. You can walk around such an image and see it from different sides as if it were the original object. Yet, if you try to touch it with your hand (Fig. 39–24b), there will be nothing material there.

The details of how the image is formed are quite complicated. But we can get the basic idea by considering one single point on the object. In Fig. 39–25a the rays OA and OB have reflected from one point on our object. The rays CA and DB come directly from the source and interfere with OA and OB at points A and B on the film. A set of interference fringes is produced as shown in Fig. 39–25b. The spacing between the fringes changes from top to bottom as shown. Why this happens is explained in Fig. 39–26. Thus the hologram of a single point object would have the fringe pattern shown in Fig. 39–25b. The film in this case looks like a diffraction grating with variable spacing. Hence, when coherent laser light is passed back through the developed film to reconstruct the image, the diffracted rays in the first order maxima occur at slightly different angles because the spacing changes. (Remember Eq. 35–13, $\sin\theta = \lambda/d$: where the spacing d is greater, the angle θ is less.) Hence, the rays diffracted upward (in first order) seem to diverge from a single point, Fig. 39–27. This is a virtual image of the original object, which can be seen with the eye. Rays diffracted in first order *downward* converge to make a real image, which can be seen and also photographed. (Note that the straight-through undiffracted rays are of no interest.) Of course real objects consist of many points, so a hologram will be a complex interference pattern which, when laser light is incident on it, will reproduce an image of the object. Each image point will be at the correct (three-dimensional) position with respect to other points, so the image accurately represents the original object. And it can be viewed from different angles as if viewing the original object. Holograms can be made in which a viewer can walk entirely around the image (360°) and see all sides of it.

Volume or **white-light holograms** do not require a laser to see the image, but can be viewed with ordinary white light (preferably a nearly point source, such as the Sun or a clear bulb with a small bright filament). Such holograms must be made, however, with a laser. They are made not on thin film, but on a *thick* emulsion. The interference pattern in the film emulsion can be thought of as an array of bands or ribbons where constructive interference occurred. This array, and the reconstruction of the image, can be compared to Bragg scattering of X-rays from the atoms in a crystal (see Section 35–10). White light can reconstruct the image because the Bragg condition $(m\lambda = 2d\sin\theta)$ selects out the appropriate single wavelength. If the hologram is originally produced by lasers emitting the three additive primary colors (red, green, and blue), the three-dimensional image can be seen in full color when viewed with white light.

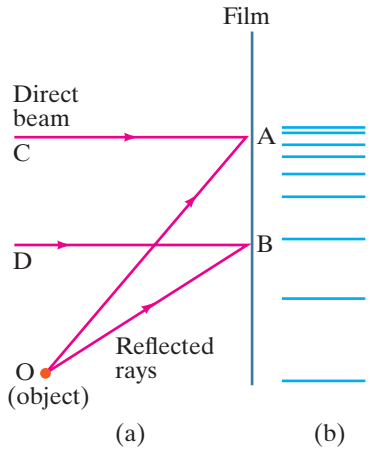
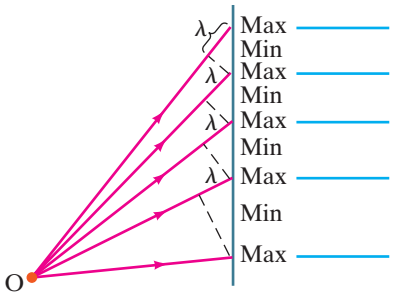


FIGURE 39–25 (a) Light from point O on the object interferes with light of the direct beam (rays CA and DB). (b) Interference fringes produced.

FIGURE 39–26 Each of the rays shown leaving point O is one wavelength shorter than the one above it. If the top ray is in phase with the direct beam (not shown), which has the same phase at all points on the screen, each of the rays shown produces a constructive interference fringe. From this diagram it can be seen that the fringe spacing increases toward the bottom.



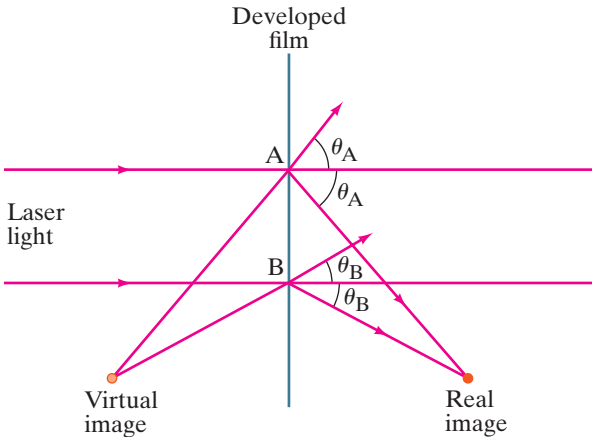


FIGURE 39–27 Reconstructing the image of one point on the object. Laser beam strikes the developed film, which is like a diffraction grating of variable spacing. Rays corresponding to the first diffraction maxima are shown emerging. The angle $\theta_A > \theta_B$ because the spacing at B is greater than at A $(\sin\theta = \lambda/d)$. Hence real and virtual images of the point are reproduced as shown.

Summary

In the quantum-mechanical view of the atom, the electrons do not have well-defined orbits, but instead exist as a "cloud." Electron clouds can be interpreted as an electron wave spread out in space, or as a **probability distribution** for electrons considered as particles.

For the simplest atom, hydrogen, the Schrödinger equation contains the potential energy

$$U = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r}.$$

The solutions give the same values of energy as the old Bohr model.

According to quantum mechanics, the state of an electron in an atom is specified by four **quantum numbers**: n, ℓ, m_ℓ, and m_s:

1. n, the **principal quantum number**, can take on any integer value $(1, 2, 3, \cdots)$ and corresponds to the quantum number of the old Bohr model;
2. ℓ, the **orbital quantum number**, can take on integer values from 0 up to $n - 1$;
3. m_ℓ, the **magnetic quantum number**, can take on integer values from $-\ell$ to $+\ell$;
4. m_s, the **spin quantum number**, can be $+\frac{1}{2}$ or $-\frac{1}{2}$.

The energy levels in the hydrogen atom depend on n, whereas in other atoms they depend on n and ℓ.

The orbital angular momentum of an atom has magnitude $L = \sqrt{\ell(\ell+1)}\,\hbar$ and z component $L_z = m_\ell\hbar$. Spin angular momentum has magnitude $S = \sqrt{s(s+1)}\,\hbar$ and z component $S_z = m_s\hbar$ where $s = \frac{1}{2}$ and $m_s = \pm\frac{1}{2}$.

When an external magnetic field is applied, the spectral lines are split (the **Zeeman effect**), indicating that the energy depends also on m_ℓ in this case.

Even in the absence of a magnetic field, precise measurements of spectral lines show a tiny splitting of the lines called **fine structure**, whose explanation is that the energy depends very slightly on m_ℓ and m_s.

Transitions between states that obey the **selection rule** $\Delta\ell = \pm 1$ are far more probable than other so-called **forbidden transitions**.

The ground-state wave function in hydrogen has spherical symmetry, as do other $\ell = 0$ states. States with $\ell > 0$ have some directionality in space.

The **probability density** (or probability distribution), $|\psi|^2$, and the **radial probability density**, $P_r = 4\pi r^2|\psi|^2$, are both useful to illustrate the spatial extent of the electron cloud.

The arrangement of electrons in multi-electron atoms is governed by the Pauli **exclusion principle**, which states that no two electrons can occupy the same quantum state—that is, they cannot have the same set of quantum numbers n, ℓ, m_ℓ, and m_s.

As a result, electrons in multi-electron atoms are grouped into **shells** (according to the value of n) and **subshells** (according to ℓ).

Electron configurations are specified using the numerical values of n, and using letters for ℓ: s, p, d, f, etc., for $\ell = 0, 1, 2, 3,$ and so on, plus a superscript for the number of electrons in that subshell. Thus, the ground state of hydrogen is $1s^1$, whereas that for oxygen is $1s^2 2s^2 2p^4$.

In the **Periodic Table**, the elements are arranged in horizontal rows according to increasing **atomic number** (= number of electrons in the neutral atom). The shell structure gives rise to a periodicity in the properties of the elements, so that each vertical column can contain elements with similar chemical properties.

X-rays, which are a form of electromagnetic radiation of very short wavelength, are produced when high-speed electrons strike a target. The spectrum of X-rays so produced consists of two parts, a *continuous* spectrum produced when the electrons are decelerated by atoms of the target, and *peaks* representing photons emitted by atoms of the target after being excited by collision with the high-speed electrons. Measurement of these peaks allows determination of inner energy levels of atoms and determination of atomic number Z.

[*An atom has a **magnetic dipole moment** $\vec{\boldsymbol{\mu}}$ related to its orbital angular momentum $\vec{\mathbf{L}}$, which produces a potential energy when in a magnetic field, $U = -\boldsymbol{\mu}\cdot\vec{\mathbf{B}}$. Electron spin also yields a magnetic moment, but the energy in a magnetic field is a factor $g = 2.0023\cdots$ times larger than expected, as determined in Stern-Gerlach experiments.]

[*Atoms have a total angular momentum $\vec{\mathbf{J}} = \vec{\mathbf{L}} + \vec{\mathbf{S}}$ which is quantized as for L and S, namely $J = \sqrt{j(j+1)}\,\hbar$, where j is a half-integer equal to $\ell \pm \frac{1}{2}$ in hydrogen.]

Fluorescence occurs when absorbed UV photons are followed by emission of visible light, due to the special arrangement of energy levels of atoms of the material. **Phosphorescent** materials have **metastable** states (long-lived) that emit light seconds or minutes after absorption of light.

Lasers produce a narrow beam of monochromatic coherent light (light waves *in phase*).

[***Holograms** are images with a 3-dimensional quality, formed by interference of laser light.]

Questions

1. (*a*) List two aspects of the Bohr model that the quantum-mechanical theory of the atom retained. (*b*) Give one major difference between the Bohr model and the quantum-mechanical theory of the atom.
2. The probability density $|\psi|^2$ is a maximum at the center of the H atom $(r = 0)$ for the ground state, whereas the radial probability density $P_r = 4\pi r^2|\psi|^2$ is zero at this point. Explain why.
3. Which model of the hydrogen atom, the Bohr model or the quantum-mechanical model, predicts that the electron spends more time near the nucleus? Explain.
4. The size of atoms varies by only a factor of three or so, from largest to smallest, yet the number of electrons varies from one to over 100. Explain.
5. Excited hydrogen and excited helium atoms both radiate light as they jump down to the $n = 1$, $\ell = 0$, $m_\ell = 0$ state. Why do the two elements have very different emission spectra?
6. In Fig. 39–4, why do the upper and lower levels have different energy splittings in a magnetic field?
7. The 589-nm yellow line in sodium is actually two very closely spaced lines. This splitting is due to an "internal" Zeeman effect. Can you explain this? [*Hint*: Put yourself in the reference frame of the electron.]

8. Which of the following electron configurations are not allowed: (*a*) $1s^2 2s^2 2p^8 3s^1$; (*b*) $1s^2 2s^2 2p^4 3s^2 4p^2$; (*c*) $1s^2 2s^2 2p^6 3s^2 3p^5 4s^2 4d^5 4f^1$? If not allowed, explain why.

9. Give the complete electron configuration for a uranium atom (careful scrutiny across the Periodic Table on the inside back cover will provide useful hints).

10. For noble gases, the halogens, and the alkali metals, explain the atomic structure that is common to each group and how that structure explains a common property of the group.

11. In what column of the Periodic Table would you expect to find the atom with each of the following configurations?

(*a*) $1s^2 2s^2 2p^6 3s^2$; (*b*) $1s^2 2s^2 2p^5$; (*c*) $1s^2 2s^2 2p^6 3s^2 3p^6$; (*d*) $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1$.

12. On what factors does the periodicity of the Periodic Table depend? Consider the exclusion principle, quantization of angular momentum, spin, and any others you can think of.

13. How would the Periodic Table look if there were no electron spin but otherwise quantum mechanics were valid? Consider the first 20 elements or so.

14. The ionization energy for neon $(Z = 10)$ is 21.6 eV and that for sodium $(Z = 11)$ is 5.1 eV. Explain the large difference.

15. Why do chlorine and iodine exhibit similar properties?

16. Explain why potassium and sodium exhibit similar properties.

16'. Briefly explain why noble gases are non-reactive and why alkali metals are highly reactive.

17. Why are the chemical properties of the rare earths so similar? [*Hint*: Examine the Periodic Table.]

18. Why do we not expect perfect agreement between measured values of characteristic X-ray line wavelengths and those calculated using the Bohr model, as in Example 39–6?

19. Why does the Bohr model, which does not work at all well for normal transitions involving the outer electrons for He and more complex atoms, nevertheless predict reasonably well the atomic X-ray spectra for transitions deep inside the atom?

20. Why does the cutoff wavelength in Fig. 39–11 imply a photon nature for light?

21. How would you figure out which lines in an X-ray spectrum correspond to K_α, K_β, L, etc., transitions?

22. Why do the characteristic X-ray spectra vary in a systematic way with Z, whereas the visible spectra (Fig. 35–22) do not?

23. Why do we expect electron transitions deep within an atom to produce shorter wavelengths than transitions by outer electrons?

***24.** Why is the direction of the magnetic dipole moment of an electron opposite to that of its orbital angular momentum?

***25.** Why is a nonhomogeneous field used in the Stern-Gerlach experiment?

26. Compare spontaneous emission to stimulated emission.

27. Does the intensity of light from a laser fall off as the inverse square of the distance? Explain.

28. How does laser light differ from ordinary light? How is it the same?

29. Explain how a 0.0005-W laser beam, photographed at a distance, can seem much stronger than a 1000-W street lamp at the same distance.

MisConceptual Questions

1. An atom has the electron configuration $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1$. How many electrons does this atom have?
(*a*) 15. (*b*) 19. (*c*) 30. (*d*) 46.

2. For the electron configuration of MisConceptual Question 1, what orbital quantum numbers do the electrons have?
(*a*) 0.
(*b*) 0 and 1.
(*c*) 0 and 1 and 2.
(*d*) 0 and 1 and 2 and 3.
(*e*) 0 and 1 and 2 and 3 and 4.

3. What is meant by the ground state of an atom?
(*a*) All of the quantum numbers have their lowest values $(n = 1,\ \ell = m_\ell = 0)$.
(*b*) The principal quantum number of the electrons in the outer shell is 1.
(*c*) All of the electrons are in the lowest energy state, consistent with the exclusion principle.
(*d*) The electrons are in the lowest state allowed by the uncertainty principle.

4. The Pauli exclusion principle applies to all electrons
(*a*) in the same shell, but not electrons in different shells.
(*b*) in the same container of atoms.
(*c*) in the same column of the Periodic Table.
(*d*) in incomplete shells.
(*e*) in the same atom.

5. Which of the following is required by the Pauli exclusion principle?
(*a*) No electron in an atom can have the same set of quantum numbers as any other electron in that atom.
(*b*) Each electron in an atom must have the same n value.
(*c*) Each electron in an atom must have different m_ℓ values.
(*d*) Only two electrons can be in any particular shell of an atom.
(*e*) No two electrons in a collection of atoms can have the exact same set of quantum numbers.

6. Under what condition(s) can the exact location and velocity of an electron be measured at the same time?
(*a*) The electron is in the ground state of the atom.
(*b*) The electron is in an excited state of the atom.
(*c*) The electron is free (not bound to an atom).
(*d*) Both (*a*) and (*b*).
(*e*) Never.

7. Which of the following is *not* always a property of lasers?
(*a*) All of the photons in laser light have the same phase.
(*b*) All laser photons have nearly identical frequencies.
(*c*) Laser light moves as a beam, spreading out very slowly.
(*d*) Laser light is always brighter than other sources of light.
(*e*) Lasers depend on an inverted population of atoms where more atoms occupy a higher energy state than some lower energy state.

8. The letters $s, p, d, f, g, \ldots$
(*a*) summarize all the quantum numbers for electrons in atoms.
(*b*) are an alternative notation for the principal quantum number.
(*c*) apply only to the electron in hydrogen.
(*d*) denote orbitals characterized by the ℓ value of an electron.

9. Which statement below is correct?
(*a*) Forbidden transitions do not occur.
(*b*) Forbidden transitions only occur in lasers.
(*c*) Forbidden transitions occur but with much lower probability than allowed transitions.
(*d*) Forbidden transitions occur in multi-electron atoms but not in hydrogen.

10. How does the energy difference between adjacent states in a hydrogen atom change as the principal quantum number increases?
(*a*) The energy difference increases with increasing values of the principal quantum number.
(*b*) The energy difference remains constant with increasing values of the principal quantum number.
(*c*) The energy difference decreases with increasing values of the principal quantum number.
(*d*) The energy difference varies with increasing values of the principal quantum number.

Problems

39–2 Hydrogen Atom Quantum Numbers

1. (I) For $n = 6$, what values can ℓ have?

2. (I) For $n = 7$, $\ell = 3$, what are the possible values of m_ℓ and m_s?

3. (I) How many different states are possible for an electron whose principal quantum number is $n = 4$? Write down the quantum numbers for each state.

4. (I) If a hydrogen atom has $m_\ell = -5$, what are the possible values of n, ℓ, and m_s?

5. (I) A hydrogen atom has $\ell = 4$. What are the possible values for n, m_ℓ, and m_s?

6. (I) Calculate the magnitude of the angular momentum of an electron in the $n = 5$, $\ell = 4$ state of hydrogen.

7. (II) A hydrogen atom is in the 8*g* state. Determine (*a*) the principal quantum number, (*b*) the energy of the state, (*c*) the orbital angular momentum and its quantum number ℓ, and (*d*) the possible values for the magnetic quantum number.

8. (II) (*a*) Show that the number of different states possible for a given value of ℓ is equal to $2(2\ell + 1)$. (*b*) What is this number for $\ell = 0, 1, 2, 3, 4, 5,$ and 6?

9. (II) Show that the number of different electron states possible for a given value of n is $2n^2$. (See Problem 8.)

10. (II) An excited H atom is in a 5*d* state. (*a*) Name all the states (n, ℓ) to which the atom is "allowed" to make a transition with the emission of a photon. (*b*) How many different wavelengths are there (ignoring fine structure)?

10'. (II) Show that there can be 18 electrons in a "*g*" shell.

11. (II) The magnitude of the orbital angular momentum in an excited state of hydrogen is $6.84 \times 10^{-34}\ \text{J}\cdot\text{s}$ and the z component is $2.11 \times 10^{-34}\ \text{J}\cdot\text{s}$. What are all the possible values of n, ℓ, and m_ℓ for this state?

39–3 Hydrogen Atom Wave Functions

12. (II) Show that the ground-state wave function, Eq. 39–5, is normalized. [*Hint*: See Example 39–4.]

13. (II) For the ground state of hydrogen, what is the value of (*a*) ψ, (*b*) $|\psi|^2$, and (*c*) P_r, at $r = 1.5\, r_0$?

14. (II) For the $n = 2$, $\ell = 0$ state of hydrogen, what is the value of (*a*) ψ, (*b*) $|\psi|^2$, and (*c*) P_r, at $r = 4r_0$?

15. (II) By what factor is it more likely to find the electron in the ground state of hydrogen at the Bohr radius (r_0) than at twice the Bohr radius $(2r_0)$?

16. (II) (*a*) Show that the probability of finding the electron in the ground state of hydrogen at less than one Bohr radius from the nucleus is 32%. (*b*) What is the probability of finding a 1*s* electron between $r = r_0$ and $r = 2r_0$?

17. (III) Determine the radius r of a sphere centered on the nucleus within which the probability of finding the electron for the ground state of hydrogen is (*a*) 50%, (*b*) 90%, (*c*) 99%.

18. (II) (*a*) Estimate the probability of finding an electron, in the ground state of hydrogen, within the nucleus assuming it to be a sphere of radius $r = 1.1$ fm. (*b*) Repeat the estimate assuming the electron is replaced with a muon, which is very similar to an electron (Chapter 43) except that its mass is 207 times greater.

19. (II) Show that the mean value of r for an electron in the ground state of hydrogen is $\bar{r} = \frac{3}{2} r_0$, by calculating

$$\bar{r} = \int_{\text{all space}} r|\psi_{100}|^2\, dV = \int_0^\infty r|\psi_{100}|^2 4\pi r^2\, dr.$$

20. (II) Show that ψ_{200} as given by Eq. 39–8 is normalized.

21. (II) Determine the average radial probability distribution P_r for the $n = 2$, $\ell = 1$ state in hydrogen by calculating

$$P_r = 4\pi r^2\left[\tfrac{1}{3}|\psi_{210}|^2 + \tfrac{1}{3}|\psi_{211}|^2 + \tfrac{1}{3}|\psi_{21-1}|^2\right].$$

22. (II) Use the result of Problem 21 to show that the most probable distance r from the nucleus for an electron in the 2*p* state of hydrogen is $r = 4r_0$, which is just the second Bohr radius (Eq. 37–11, Fig. 37–25).

23. (II) For the ground state of hydrogen, what is the probability of finding the electron within a spherical shell of inner radius $0.99r_0$ and outer radius $1.01r_0$?

24. (III) For the $n = 2$, $\ell = 0$ state of hydrogen, what is the probability of finding the electron within a spherical shell of inner radius $4.00\, r_0$ and outer radius $5.00\, r_0$? [*Hint*: You might integrate by parts.]

25. (III) Show that ψ_{100} (Eq. 39–5a) satisfies the Schrödinger equation (Eq. 39–1) with the Coulomb potential, for energy $E = -me^4/8\epsilon_0^2 h^2$.

26. (III) Show that the probability of finding the electron within 1 Bohr radius of the nucleus in the hydrogen atom is (*a*) 3.4% for the $n = 2$, $\ell = 0$ state, and (*b*) 0.37% for the $n = 2$, $\ell = 1$ state. (See Problem 21.)

27. (III) The wave function for the $n = 3$, $\ell = 0$ state in hydrogen is

$$\psi_{300} = \frac{1}{\sqrt{27\pi r_0^3}}\left(1 - \frac{2r}{3r_0} + \frac{2r^2}{27r_0^2}\right)e^{-\frac{r}{3r_0}}.$$

(*a*) Determine the radial probability distribution P_r for this state, and (*b*) draw the curve for it on a graph. (*c*) Estimate from the graph the most probable distance from the nucleus for an electron in this state.

39–4 and 39–5 Complex Atoms

28. (I) List the quantum numbers for each electron in the ground state of carbon $(Z = 6)$.

29. (I) List the quantum numbers for each electron in the ground state of (*a*) oxygen $(Z = 8)$, (*b*) aluminum $(Z = 13)$.

30. (I) How many electrons can be in the $n = 6$, $\ell = 4$ subshell?

31. (II) An electron has $m_\ell = 3$ and is in its lowest possible energy state. What are the values of n and ℓ for this electron?

32. (II) If the principal quantum number n were limited to the range from 1 to 6, how many elements would we find in nature?

33. (II) What is the full electron configuration for (*a*) cobalt (Co), (*b*) gold (Au), and (*c*) uranium (U)? [*Hint*: See the Periodic Table inside the back cover.]

34. (II) Estimate the binding energy of the third electron in lithium using the Bohr model. [*Hint*: This electron has $n = 2$ and "sees" a net charge of approximately $+1e$.] The measured value is 5.36 eV.

35. (II) Using the Bohr formula for the radius of an electron orbit, estimate the average distance from the nucleus for an electron in the innermost $(n = 1)$ orbit in uranium $(Z = 92)$. Approximately how much energy would be required to remove this innermost electron?

36. (II) Let us apply the exclusion principle to an infinitely high square well (Section 38–8). Let there be five electrons confined to this rigid box whose width is ℓ. Find the lowest energy state of this system, by placing the electrons in the lowest available levels, consistent with the Pauli exclusion principle.

37. (II) Show that the total angular momentum is zero for a filled subshell.

39–6 X-rays

38. (I) If the shortest-wavelength bremsstrahlung X-rays emitted from an X-ray tube have $\lambda = 0.023$ nm, what is the voltage across the tube?

39. (I) For the spectrum of X-rays emitted from the molybdenum target of Fig. 39–11, determine the maximum and minimum energies of the photons.

40. (I) Show that the cutoff wavelength λ_0 in an X-ray spectrum is given by

$$\lambda_0 = \frac{1240}{V}\ \text{nm},$$

where V is the X-ray tube voltage in volts.

41. (II) Estimate the wavelength for an $n = 2$ to $n = 1$ transition in iron $(Z = 26)$.

42. (II) Use the approach of Example 39–6 to estimate the X-ray wavelength emitted when a cobalt atom $(Z = 27)$ makes a transition from $n = 2$ to $n = 1$.

43. (II) A mixture of iron and an unknown material is bombarded with electrons. The wavelength of the K_α lines are 194 pm for iron and 229 pm for the unknown. What is the unknown material?

44. (II) Use the Bohr model to estimate the wavelength for an $n = 3$ to $n = 1$ transition in molybdenum $(Z = 42)$. The measured value is 0.063 nm. Why do we not expect perfect agreement?

45. (III) Use conservation of energy and momentum to show that a moving electron cannot give off an X-ray photon unless there is a third object present, such as an atom or nucleus.

*39–7 Magnetic Dipole Moment; $\vec{J}$

***46.** (I) Verify that the Bohr magneton has the value $\mu_B = 9.27 \times 10^{-24}$ J/T (see Eq. 39–12).

***47.** (I) If the quantum state of an electron is specified by (n, ℓ, m_ℓ, m_s), estimate the energy difference between the states $(1, 0, 0, -\frac{1}{2})$ and $(1, 0, 0, +\frac{1}{2})$ of an electron in the 1*s* state of helium in an external magnetic field of 3.5 T.

***48.** (II) Silver atoms $(\text{spin} = \frac{1}{2})$ are placed in a 2.0-T magnetic field which splits the ground state into two close levels. (*a*) What is the difference in energy between these two levels, and (*b*) what wavelength photon could cause a transition from the lower level to the upper one? (*c*) How would your answer differ if the atoms were hydrogen?

***49.** (II) In a Stern-Gerlach experiment, Ag atoms exit the oven with an average speed of 780 m/s and pass through a magnetic field gradient $dB/dz = 1.8 \times 10^3$ T/m for a distance of 4.0 cm. (*a*) What is the separation of the two beams as they emerge from the magnet? (*b*) What would the separation be if the *g*-factor were 1 for electron spin?

***50.** (II) For an electron in a 5*g* state, what are all the possible values of j, m_j, J, and J_z?

***51.** (II) What are the possible values of j for an electron in (*a*) the 4*p*, (*b*) the 4*f*, and (*c*) the 3*d* state of hydrogen? (*d*) What is J in each case?

***52.** (II) (*a*) Write down the quantum numbers for each electron in the gallium atom. (*b*) Which subshells are filled? (*c*) The last electron is in the 4*p* state; what are the possible values of the total angular momentum quantum number, j, for this electron? (*d*) Explain why the angular momentum of this last electron also represents the total angular momentum for the entire atom (ignoring any angular momentum of the nucleus). (*e*) How could you use a Stern-Gerlach experiment to determine which value of j the atom has?

***53.** (III) The difference between the $2P_{3/2}$ and $2P_{1/2}$ energy levels in hydrogen is about 5×10^{-5} eV, due to the spin-orbit interaction. (*a*) Taking the electron's (orbital) magnetic moment to be 1 Bohr magneton, estimate the internal magnetic field due to the electron's orbital motion. (*b*) Estimate the internal magnetic field using a simple model of the nucleus revolving in a circle about the electron.

39–9 Lasers

54. (II) A laser used to weld detached retinas puts out 23-ms-long pulses of 640-nm light which average 0.68-W output during a pulse. How much energy can be deposited per pulse and how many photons does each pulse contain? [*Hint*: See Example 37–4.]

55. (II) Estimate the angular spread of a laser beam due to diffraction if the beam emerges through a 3.6-mm-diameter mirror. Assume that $\lambda = 658\,\text{nm}$. What would be the diameter of this beam if it struck (*a*) a satellite 380 km above the Earth, (*b*) the Moon? [*Hint*: See Chapter 35.]

56. (II) A low-power laser used in a physics lab might have a power of 0.50 mW and a beam diameter of 3.0 mm. Calculate (*a*) the average light intensity of the laser beam, and (*b*) compare it to the intensity of a 100-W incandescent lightbulb that emits 2.2 W of visible light, that is viewed from a distance of 2.0 m.

57. (II) Calculate the wavelength of a He–Ne laser. See Fig 39–21.

58. (II) Suppose that the energy level system in Fig. 39–20 is not being pumped and is in thermal equilibrium. Determine the fraction of atoms in levels E_2 and E_1 relative to E_0 at $T = 300\,\text{K}$.

59. (II) To what temperature would the system in Fig. 39–20 have to be raised (see Problem 58) so that in thermal equilibrium the level E_2 would have half as many atoms as E_0? (Note that pumping mechanisms do not maintain thermal equilibrium.)

60. (II) Show that a population inversion for two levels (as in a pumped laser) corresponds to a negative Kelvin temperature in the Boltzmann distribution. Explain why such a situation does not contradict the idea that negative Kelvin temperatures cannot be reached in the normal sense of temperature.

General Problems

61. The ionization (binding) energy of the outermost electron in boron is 8.26 eV. (*a*) Use the Bohr model to estimate the "effective charge," Z_{eff}, seen by this electron. (*b*) Estimate the average orbital radius.

62. How many electrons can there be in an "*h*" subshell?

63. What is the full electron configuration in the ground state for elements with Z equal to (*a*) 24, (*b*) 35, (*c*) 40? [*Hint*: See the Periodic Table inside the back cover.]

64. What are the largest and smallest possible values for the angular momentum L of an electron in the $n = 6$ shell?

65. Estimate (*a*) the quantum number ℓ for the orbital angular momentum of the Earth about the Sun, and (*b*) the number of possible orientations for the plane of Earth's orbit.

66. Use the Bohr model (especially Eq. 37–15) to show that the Moseley plot (Fig. 39–12) can be written

$$\sqrt{\frac{1}{\lambda}} = a(Z - b),$$

where $b \approx 1$, and evaluate a.

67. In the so-called **vector model** of the atom, space quantization of angular momentum (Fig. 39–3) is illustrated as shown in Fig. 39–28. The angular momentum vector of magnitude $L = \sqrt{\ell(\ell+1)}\,\hbar$ is thought of as precessing around the z axis (like a spinning top or gyroscope) in such a way that the z component of angular momentum, $L_z = m_\ell \hbar$, also stays constant. Calculate the possible values for the angle θ between $\vec{\mathbf{L}}$ and the z axis (*a*) for $\ell = 1$, (*b*) $\ell = 2$, and (*c*) $\ell = 3$. (*d*) Determine the minimum value of θ for $\ell = 100$ and $\ell = 10^6$. Is this consistent with the correspondence principle?

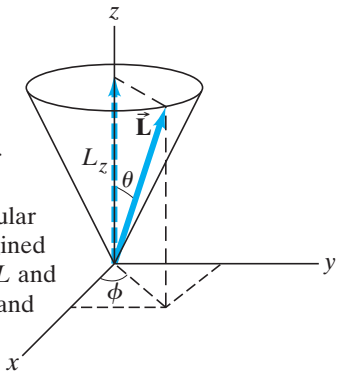


FIGURE 39–28 The vector model for orbital angular momentum. The orbital angular momentum vector $\vec{\mathbf{L}}$ is imagined to precess about the z axis; L and L_z remain constant, but L_x and L_y continually change. Problems 67 and 68.

68. The vector model (Problem 67) gives some insight into the uncertainty principle for angular momentum, which is

$$\Delta L_z \, \Delta\phi \geq \hbar$$

for the z component. Here ϕ is the angular position measured in the plane perpendicular to the z axis. Once m_ℓ for an atom is known, L_z is known precisely, so $\Delta L_z = 0$. (*a*) What does this tell us about ϕ? (*b*) What can you say about L_x and L_y, which are not quantized (only L and L_z are)? (*c*) Show that although L_x and L_y are not quantized, nonetheless $(L_x^2 + L_y^2)^{1/2} = [\ell(\ell+1) - m_\ell^2]^{1/2}\hbar$ is.

69. Determine the most probable distance from the nucleus of an electron in the $n = 3$, $\ell = 0$ state of hydrogen.

70. Show that the diffractive spread of a laser beam, $\approx \lambda/D$ as described in Section 39–9, is precisely what you might expect from the uncertainty principle. [*Hint*: Since the beam's width is constrained by the dimension of the aperture D, the component of the light's momentum perpendicular to the laser axis is uncertain.]

71. (*a*) Show that the mean value for $1/r$ of an electron in the ground state of hydrogen equals $1/r_0$, and from this conclude that the mean value of the potential energy is

$$\overline{U} = -\frac{1}{4\pi\epsilon_0}\frac{e^2}{r_0}.$$

(*b*) Using $E = \overline{U} + \overline{K}$, find a relationship between the average kinetic energy and the average potential energy in the ground state. [*Hint*: For (*a*), see Problem 19 or Example 38–9.]

72. The angular momentum in the hydrogen atom is given both by the Bohr model and by quantum mechanics. Compare the results for $n = 2$.

73. For each of the following atomic transitions, state whether the transition is *allowed* or *forbidden*, and why: (*a*) $4p \rightarrow 3p$; (*b*) $3p \rightarrow 1s$; (*c*) $4d \rightarrow 3d$; (*d*) $4d \rightarrow 3s$; (*e*) $4s \rightarrow 2p$.

74. It is possible for atoms to be excited into states with very high values of the principal quantum number. Electrons in these so-called *Rydberg states* have very small ionization energies and huge orbital radii. This makes them particularly sensitive to external perturbation, as would be the case if the atom were in an electric field. Consider the $n = 45$ state of the hydrogen atom. Determine the binding energy, the radius of the orbit, and the effective cross-sectional area of this Rydberg state.

75. Suppose that the spectrum of an unknown element shows a series of lines with one out of every four matching a line from the Lyman series of hydrogen. Assuming that the unknown element is an ion with Z protons and one electron, determine Z and the element in question.

***76.** Suppose that the splitting of energy levels shown in Fig. 39–4 was produced by a 2.6-T magnetic field. (*a*) What is the separation in energy between adjacent m_ℓ levels for the same ℓ? (*b*) How many different wavelengths will there be for $3d$ to $2p$ transitions, if m_ℓ can change only by ± 1 or 0? (*c*) What is the wavelength for each of these transitions?

77. *Populations in the H atom*. Use the Boltzmann factor (Eq. 39–16) to estimate the fraction of H atoms in the $n = 2$ and $n = 3$ levels (relative to the ground state) for thermal equilibrium at (*a*) $T = 300\,\text{K}$ and (*b*) $T = 6000\,\text{K}$. [Note: Since there are eight states with $n = 2$ and only two with $n = 1$, multiply your result for $n = 2$ by $\frac{8}{2} = 4$; do similarly for $n = 3$.] (*c*) Given 1.0 g of hydrogen, estimate the number of atoms in each state at $T = 6000\,\text{K}$. (*d*) Estimate the number of $n = 3$ to $n = 1$ and $n = 2$ to $n = 1$ photons that will be emitted per second at $T = 6000\,\text{K}$. Assume that the lifetime of each excited state is $10^{-8}\,\text{s}$. [*Hint:* To evaluate a large exponent, you can use base-10 logarithms, Appendix A.]

78. Photons of wavelength 0.154 nm are emitted from the surface of a certain metal when it is bombarded with high-energy radiation. If this photon wavelength corresponds to the K_α line, what is the element?

79. Imagine a line whose length equals the diameter of the smallest orbit in the Bohr model. If we are told only that an electron is located somewhere on this line, then the electron's positions can be specified as $x = 0 \pm r_1$, where the origin is at the line's center and r_1 is the Bohr radius. Such an electron can never be observed at rest, but instead at a minimum will have a speed somewhere in the range from $v = 0 - \Delta v$ to $0 + \Delta v$. Determine Δv.

80. The electron in a hydrogen atom is in a $4s$ state. (*a*) Write down all the lower energy states it can decay to in a single transition. (*b*) How many transitions must it make to get to the ground state? What state(s) could it go through?

Section 39–7 (II) Equation 39–11 can be obtained by a semiclassical derivation but can also be derived quantum mechanically. The semiclassical model has the electron moving in a circle. How fast would a $2p$ electron in H have to be moving to produce a magnetic moment equal to 1 Bohr magneton?

81. Refer to Figure 39–20 and determine the wavelength of a ruby laser.

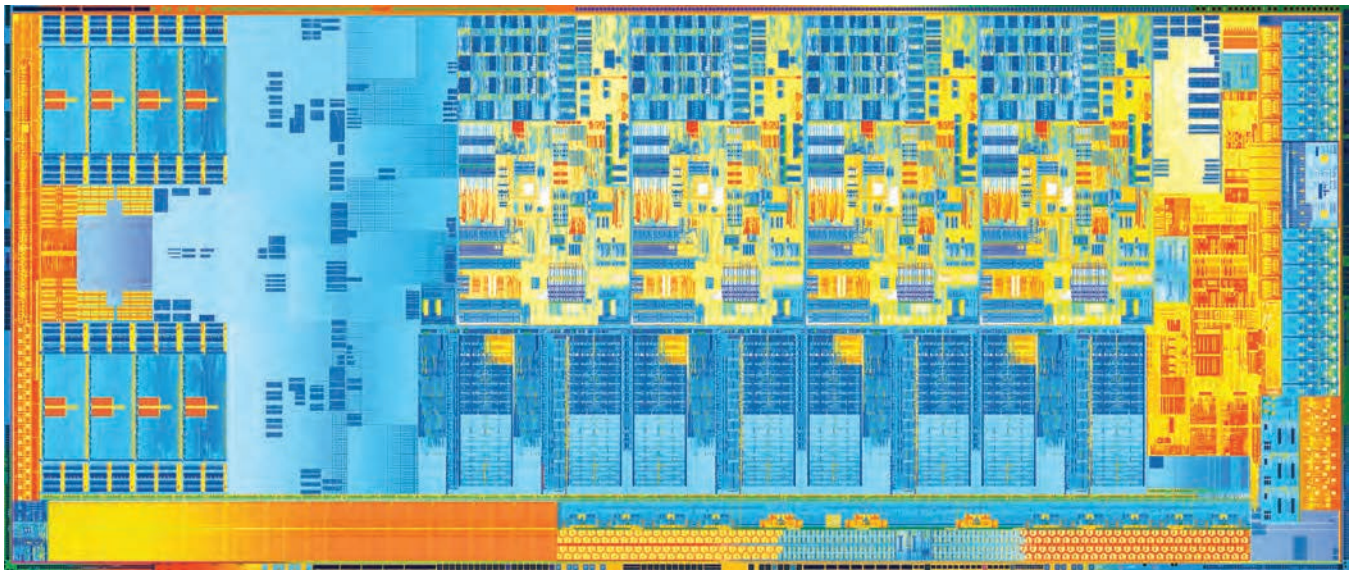
82. Show that the three wave functions of the $n = 2$, $\ell = 1$ are orthogonal to each other by showing that the dot product for each pairing is zero for any point in space.

83. Show that the sum of the probabilities of the electron being in one of the $n = 2$, $\ell = 1$ states is spherically symmetric. Remember to use the complex conjugate in determining the absolute value of a complex number.

84. For the hydrogen atom, show that $|\psi_{210}|^2 + |\psi_{211}|^2 + |\psi_{21-1}|^2$ is spherically symmetric.

ANSWERS TO EXERCISES

A: 2, 1, 0, −1, −2.
B: (*b*).
C: −0.38 eV, $\sqrt{20}\hbar$.
D: (*d*).
E: Add one line to Li in Table 39–2: 2, 0, 0, $-\frac{1}{2}$.
F: (*b*), (*c*).
G: $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^1$.
H: (*d*).



This computer processor chip contains more than 5 billion transistors, plus diodes, resistors, and other electronic elements, all in a space of about 1 cm^2 on a single silicon semiconductor crystal. It uses so-called 14-nm technology, meaning the "wires" (conducting lines) are about 14 nm wide ($\approx$ 30 atoms).

Before discussing semiconductors and their applications, we study the quantum theory description of bonding between atoms to form molecules, and how it explains molecular behavior. We then examine how atoms and molecules form solids, with emphasis on metals as well as on semiconductors and their use in electronics.

CHAPTER 40

Molecules and Solids

CONTENTS

40 – 1 Bonding in Molecules
40 – 2 Potential-Energy Diagrams for Molecules
40 – 3 Weak (van der Waals) Bonds
40 – 4 Molecular Spectra
40 – 5 Bonding in Solids
40 – 6 Free-Electron Theory of Metals; Fermi Energy
40 – 7 Band Theory of Solids
40 – 8 Semiconductors and Doping
40 – 9 Semiconductor Diodes, LEDs, OLEDs
40–10 Transistors: Bipolar and MOSFETs
*40–11 Integrated Circuits, 14-nm Technology

CHAPTER-OPENING QUESTIONS—Guess now!

1. What holds a solid together?
- **(a)** Gravitational forces.
- **(b)** Magnetic forces.
- **(c)** Electric forces.
- **(d)** Glue.
- **(e)** Nuclear forces.

2. As a metal is heated, how does the rms speed (v_{rms}) of the electrons change? Assume a temperature change of about 30C° near room temperature.
- **(a)** v_{rms} increases linearly with temperature.
- **(b)** v_{rms} decreases linearly with temperature.
- **(c)** v_{rms} increases exponentially with temperature.
- **(d)** v_{rms} decreases exponentially with temperature.
- **(e)** v_{rms} changes very little as the temperature is increased.

Since its development in the 1920s, quantum mechanics has had a profound influence on our lives, both intellectually and technologically. Even the way we view the world has changed, as we have seen in the last few Chapters. Now we discuss how quantum mechanics has given us an understanding of the structure of molecules and matter in bulk, as well as important applications including semiconductor devices and applications to molecular biology. Semiconductor devices, such as transistors, now may be only a few atoms thick, which is the realm of quantum mechanics.

PSE5_CH40_v1.7_23-Jun-20

40–1 Bonding in Molecules

One of the great successes of quantum mechanics was to give scientists, at last, an understanding of the nature of chemical bonds. Since it is based in physics, and because this understanding is so important in many fields, we discuss it here.

By a molecule, we mean a group of two or more atoms that are strongly held together so as to function as a single unit. When atoms make such an attachment, we say that a chemical **bond** has been formed. There are two main types of strong chemical bond: covalent and ionic. Many bonds are actually intermediate between these two types.

Covalent Bonds

To understand how *covalent bonds* are formed, we take the simplest case, the bond that holds two hydrogen atoms together to form the hydrogen molecule, H_2. The mechanism is basically the same for other covalent bonds. As two H atoms approach each other, the electron clouds begin to overlap, and the electrons from each atom can "orbit" both nuclei. (This is sometimes called **sharing** electrons.) If both electrons are in the ground state $(n = 1)$ of their respective atoms, there are two possibilities: their spins can be parallel (both up or both down), in which case the total spin is $S = \frac{1}{2} + \frac{1}{2} = 1$; or their spins can be opposite ($m_s = +\frac{1}{2}$ for one, and $m_s = -\frac{1}{2}$ for the other), so that the total spin $S = +\frac{1}{2} + \left(-\frac{1}{2}\right) = 0$. We shall now see that a bond is formed only for the $S = 0$ state, when the spins are opposite.

First we consider the $S = 1$ state, for which the spins are the same. The two electrons cannot both be in the lowest energy state and be attached to the same atom, for then they would have identical quantum numbers in violation of the exclusion principle. The exclusion principle tells us that, because no two electrons can occupy the same quantum state, if two electrons have the same quantum numbers, they must be different in some other way—namely, by being in different places in space (for example, attached to different atoms). Thus, for $S = 1$, when the two atoms approach each other, the electrons will stay away from each other as shown by the probability distribution of Fig. 40–1. The electrons spend very little time between the two nuclei, so the positively charged nuclei repel each other and no bond is formed.

For the $S = 0$ state, on the other hand, the spins are opposite and the two electrons are consequently in different quantum states (m_s is different, $+\frac{1}{2}$ for one, $-\frac{1}{2}$ for the other). Hence the two electrons can come close together, and the probability distribution looks like Fig. 40–2: the electrons can spend much of their time between the two nuclei. The two positively charged nuclei are attracted to the negatively charged electron cloud between them, and this is the attraction that holds the two hydrogen atoms together to form a hydrogen molecule. This is a **covalent bond**.

The probability distributions of Figs. 40–1 and 40–2 can perhaps be better understood on the basis of waves. What the exclusion principle requires is that when the spins are the same, there is destructive interference of the electron wave functions in the region between the two atoms. But when the spins are opposite, constructive interference occurs in the region between the two atoms, resulting in a large amount of negative charge there. Thus a covalent bond can be said to be the result of constructive interference of the electron wave functions in the space between the two atoms, and of the electrostatic attraction of the two positive nuclei for the negative charge concentration between them.

Why a bond is formed can also be understood from the energy point of view. When the two H atoms approach close to one another, if the spins of their electrons are opposite, the electrons can occupy the same space, as discussed above. This means that each electron can now move about in the space of two atoms instead of in the volume of only one. Because each electron now occupies more space, it is less well localized. From the uncertainty principle with Δx larger, we see that Δp and the minimum momentum can be less. With less momentum, each electron has less energy when the two atoms combine than when they are separate. That is, the molecule has less energy than the two separate atoms, and so is more stable.

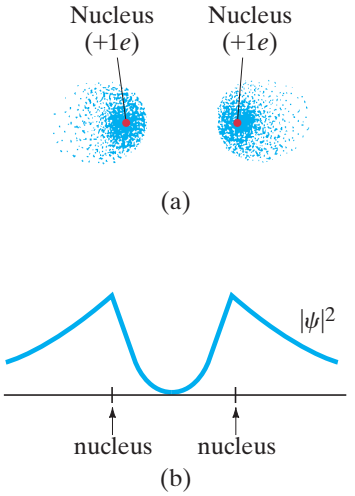
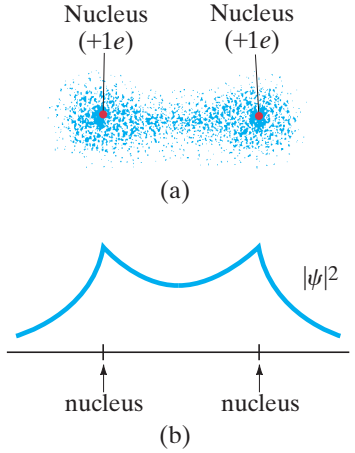


FIGURE 40–1 Electron probability distribution (electron cloud) around two H atoms when the spins are the same $\left(S = \frac{1}{2} + \frac{1}{2} = 1\right)$: (a) electron cloud; (b) projection of $|\psi|^2$ along the line through the centers of the two atoms.

FIGURE 40–2 Electron probability distribution (cloud) around two H atoms when the spins are opposite $(S = 0)$: (a) electron cloud; (b) projection of $|\psi|^2$ along the line through the centers of the atoms. In this case a bond is formed because the positive nuclei are attracted to the concentration of the electron cloud's negative charge between them. This is a hydrogen molecule, H_2.



An energy input is required to break the H_2 molecule into two separate H atoms, so the H_2 molecule is a stable entity. This is what we mean by a *bond*. The energy required to break a bond is called the **bond energy**, the **binding energy**, or the **dissociation energy**. For the hydrogen molecule, H_2, the bond energy is 4.5 eV.

Ionic Bonds

An *ionic bond* is, in a sense, a special case of the covalent bond. Instead of the electrons being shared equally, they are shared unequally. For example, in sodium chloride (NaCl), the outer electron of the sodium spends nearly all its time around the chlorine (Fig. 40–3). The chlorine atom acquires a net negative charge as a result of the extra electron, whereas the sodium atom is left with a net positive charge. The electrostatic attraction between these two charged atoms holds them together. The resulting bond is called an **ionic bond** because it is created by the attraction between the two ions $(Na^+$ and $Cl^-)$. But to understand the ionic bond, we must understand why the extra electron from the sodium spends so much of its time around the chlorine. After all, the chlorine atom is neutral; why should it attract another electron?

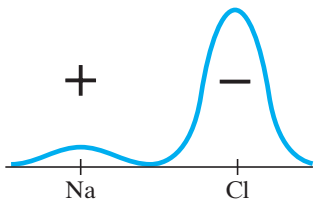
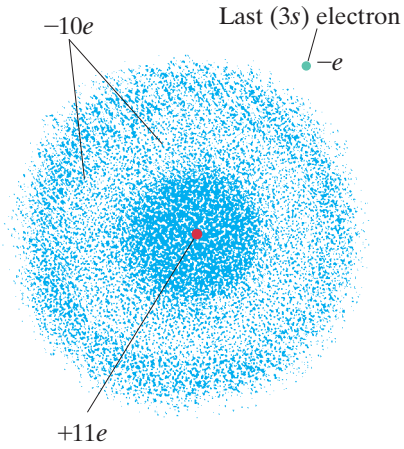


FIGURE 40–3 Probability distribution $|\psi|^2$ for the outermost electron of Na in NaCl.

FIGURE 40–4 In a neutral sodium atom, the 10 inner electrons shield the nucleus, so the single outer electron is attracted by a net charge of $+1e$.



The answer lies in the probability distributions of the electrons in the two neutral atoms. Sodium contains 11 electrons, 10 of which are in spherically symmetric closed shells (Fig. 40–4). The last electron spends most of its time beyond these closed shells. Because the closed shells have a total charge of $-10e$ and the nucleus has charge $+11e$, the outermost electron in sodium "feels" a net attraction due to $+1e$. It is not held very strongly. On the other hand, 12 of chlorine's 17 electrons form closed shells, or subshells (corresponding to $1s^2 2s^2 2p^6 3s^2$). These 12 electrons form a spherically symmetric shield around the nucleus. The other five electrons are in $3p$ states whose probability distributions are not spherically symmetric and have a form similar to those for the $2p$ states in hydrogen shown in Fig. 39–9. Four of these $3p$ electrons can have "doughnut-shaped" distributions symmetric about the z axis, as shown in Fig. 40–5. The fifth can have a "barbell-shaped" distribution (as for $m_\ell = 0$ in Fig. 39–9), which in Fig. 40–5 is shown only in dashed outline because it is half empty. That is, the exclusion principle allows one more electron to be in this state (it will have spin opposite to that of the electron already there). If an extra electron—say from a Na atom—happens to be in the vicinity, it can be in this state, perhaps at point x in Fig. 40–5. It could experience an attraction due to as much as $+5e$ because the $+17e$ of the nucleus is partly shielded at this point by the 12 inner electrons. Thus, the outer electron of a sodium atom will be more strongly attracted by the $+5e$ of the chlorine atom than by the $+1e$ of its own atom. This, combined with the strong attraction between the two ions when the extra electron stays with the Cl^-, produces the charge distribution of Fig. 40–3, and hence the ionic bond.

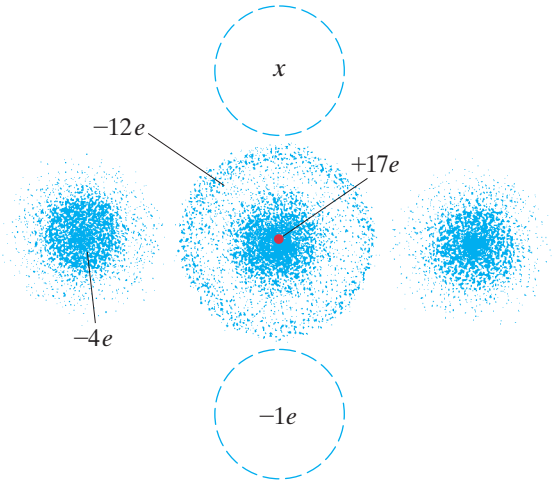


FIGURE 40–5 Neutral chlorine atom. The $+17e$ of the nucleus is shielded by the 12 electrons in the inner shells and subshells. Four of the five $3p$ electrons are shown in doughnut-shaped clouds (seen in cross section at left and right), and the fifth is in the dashed-line cloud concentrated about the z axis (vertical). An extra electron at x will be attracted by a net charge that can be as much as $+5e$.

Partial Ionic Character of Covalent Bonds

A pure covalent bond in which the electrons are shared equally occurs mainly in symmetrical molecules such as H_2, O_2, and Cl_2. When the atoms involved are different from each other, usually the shared electrons are more likely to be

PSE5_CH40_v1.7_23-Jun-20

in the vicinity of one atom than the other. The extreme case is an ionic bond. In intermediate cases the *covalent bond* is said to have a **partial ionic character**. The molecules themselves are **polar**—that is, one part (or parts) of the molecule has a net positive charge and other parts a net negative charge. An example is the water molecule, H_2O (Fig. 40–6). Covalent bonds have shared electrons, which in H_2O are more likely to be found around the oxygen atom than around the two hydrogens. The reason is similar to that discussed above in connection with ionic bonds. Oxygen has eight electrons $(1s^2 2s^2 2p^4)$, of which four form a spherically symmetric core and the other four could have, for example, a doughnut-shaped distribution. The barbell-shaped distribution on the z axis (like that shown dashed in Fig. 40–5) could be empty, so electrons from hydrogen atoms can be attracted by a net charge of $+4e$. They are also attracted by the H nuclei, so they partly orbit the H atoms as well as the O atom. The net effect is that there is a net positive charge on each H atom (less than $+1e$), because the electrons spend only part of their time there. And, there is a net negative charge on the O atom.

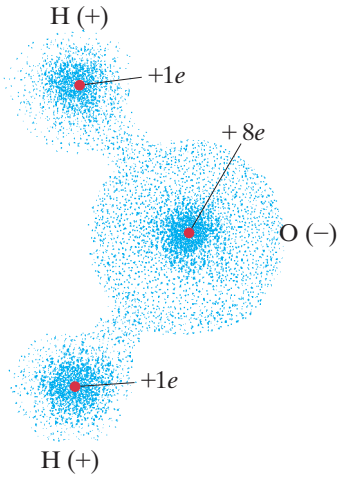


FIGURE 40–6 The water molecule H_2O is polar.

40–2 Potential-Energy Diagrams for Molecules

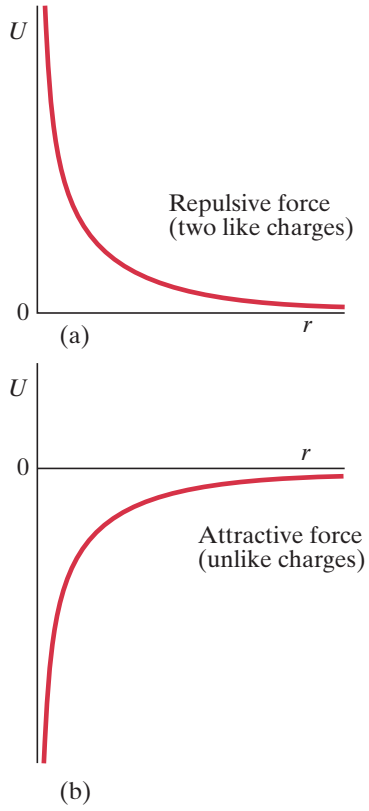
It is useful to analyze the interaction between two objects—say, between two atoms or molecules—with the use of a potential-energy diagram, which is a plot of the potential energy versus the separation distance.

For the simple case of two point charges, q_1 and q_2, the potential energy U is given by (see Chapter 23)

$$U = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r},$$

where r is the distance between the charges, and the constant $(1/4\pi\epsilon_0)$ is equal to $9.0 \times 10^9\ \mathrm{N \cdot m^2/C^2}$. If the two charges have the same sign, the potential energy U is positive for all values of r, and a graph of U versus r in this case is shown in Fig. 40–7a. The force is repulsive (the charges have the *same* sign) and the curve rises as r decreases; this makes sense because work is done to bring the charges together, thereby increasing their potential energy. If, on the other hand, the two charges are of the *opposite* sign, the potential energy is negative because the product $q_1 q_2$ is negative. The force is attractive in this case, and the graph of $U(\propto -1/r)$ versus r looks like Fig. 40–7b. The potential energy becomes more *negative* as r decreases.

FIGURE 40–7 Potential energy U as a function of separation r for two point charges of (a) like sign and (b) opposite sign.



Now let us look at the potential-energy diagram for the formation of a covalent bond, such as for the hydrogen molecule, H_2. The potential energy U of one H atom in the presence of the other is plotted in Fig. 40–8. Starting at large r, U decreases as the atoms approach, because the electrons concentrate between the two nuclei (Fig. 40–2), so attraction occurs. However, at very short distances, the electrons would be "squeezed out"—there is no room for them between the two nuclei. Without the electrons between them, each nucleus would feel a repulsive force due to the other, so the curve rises as r decreases further. There is an optimum separation of the atoms, r_0 in Fig. 40–8, at which the energy is lowest.

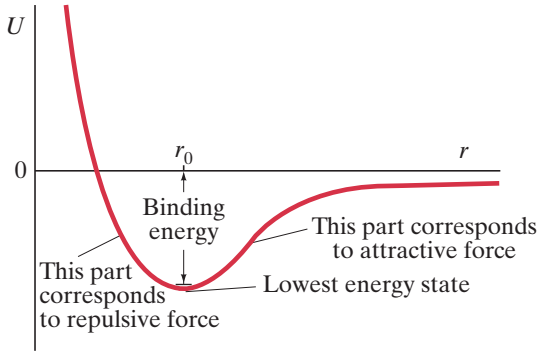


FIGURE 40–8 Potential-energy diagram for H_2 molecule; r is the separation of the two H atoms. The binding energy (the energy difference between $U = 0$ and the lowest energy state near the bottom of the well) is 4.5 eV, and $r_0 = 0.074$ nm.

This is the point of greatest stability for the hydrogen molecule, and r_0 is the average separation of atoms in the H_2 molecule. The depth of this "well" is the *binding energy*,[†] as shown. This is how much energy must be put into the system to separate the two atoms to infinity, where $U = 0$. For the H_2 molecule, the binding energy is about 4.5 eV and $r_0 = 0.074$ nm.

In molecules made of larger atoms, say, oxygen or nitrogen, repulsion also occurs at short distances, because the closed inner electron shells begin to overlap and the exclusion principle forbids their coming too close. The repulsive part of the curve rises even more steeply than $1/r$. A reasonable approximation to the potential energy, at least in the vicinity of r_0, is

$$U = -\frac{A}{r^m} + \frac{B}{r^n}, \tag{40–1}$$

where A and B are constants associated with the attractive and repulsive parts of the potential energy and the exponents m and n are small integers. For ionic and some covalent bonds, the attractive term can often be written with $m = 1$ (Coulomb potential).

For many bonds, the potential-energy curve has the shape shown in Fig. 40–9. There is still an optimum distance r_0 at which the molecule is stable. But when the atoms approach from a large distance, the force is initially repulsive rather than attractive. The atoms thus do not spontaneously form a bond. Some additional energy must be injected into the system to get it over the "hump" (or barrier) in the potential-energy diagram. This required energy is called the **activation energy**.

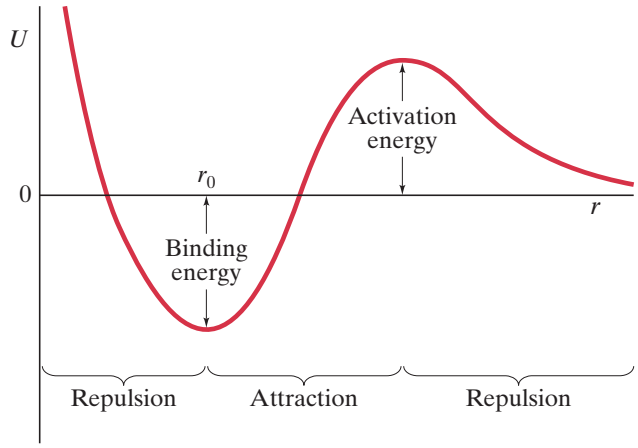


FIGURE 40–9 Potential-energy diagram for a bond requiring an activation energy.

The curve of Fig. 40–9 is much more common than that of Fig. 40–8. The activation energy often reflects a need to break other bonds, before the one under discussion can be made. For example, to make water from O_2 and H_2, the H_2 and O_2 molecules must first be broken into H and O atoms by an input of energy; this is what the activation energy represents. Then the H and O atoms can combine to form H_2O with the release of a great deal more energy than was put in initially. The initial activation energy can be provided by applying an electric spark to a mixture of H_2 and O_2, breaking a few of these molecules into H and O atoms. When the three atoms combine to form H_2O, a lot of energy is released—the ground state is near the bottom of the well (as in Fig. 40–8)—which provides the activation energy needed for further reactions, so additional H_2 and O_2 molecules are broken up and recombined to form H_2O.

The potential-energy diagrams for ionic bonds, such as NaCl, may be more like Fig. 40–8: the Na^+ and Cl^- ions attract each other at distances a bit larger than some r_0, but at shorter distances the overlapping of inner electron shells gives rise to repulsion. The two atoms thus are most stable at some intermediate separation, r_0. For partially ionic bonds, there is usually an activation energy, Fig. 40–9.

[†]The binding energy corresponds not quite to the bottom of the potential energy curve, but to the lowest quantum energy state, slightly above the bottom, as shown in Fig. 40–8.

PSE5_CH40_v1.7_23-Jun-20

EXAMPLE 40–1 **ESTIMATE** **Sodium chloride bond.** A potential-energy diagram for the NaCl ionic bond is shown in Fig. 40–10, where we have set $U = 0$ for free Na and Cl neutral atoms (which are represented on the right in Fig. 40–10). Measurements show that 5.14 eV are required to remove an electron from a neutral Na atom to produce the Na^+ ion; and 3.61 eV of energy is released when an electron is "grabbed" by a Cl atom to form the Cl^- ion. Thus, forming Na^+ and Cl^- ions from neutral Na and Cl atoms requires $5.14\,\text{eV} - 3.61\,\text{eV} = 1.53\,\text{eV}$ of energy, a form of activation energy. This is shown as the "bump" in Fig. 40–10. But note that the potential-energy diagram from here out to the right is not really a function of distance—it is drawn dashed to remind us that it only represents the energy difference between the ions and the neutral atoms (for which we have chosen $U = 0$). (*a*) Calculate the separation distance, r_1, at which the potential of the Na^+ and Cl^- ions drops to zero (measured value is $r_1 = 0.94\,\text{nm}$). (*b*) Estimate the binding energy of the NaCl bond, which occurs at a separation $r_0 = 0.24\,\text{nm}$. Ignore the repulsion of the overlapping electron shells that occurs at this distance (and causes the rise of the potential-energy curve for $r < r_0$, Fig. 40–10).

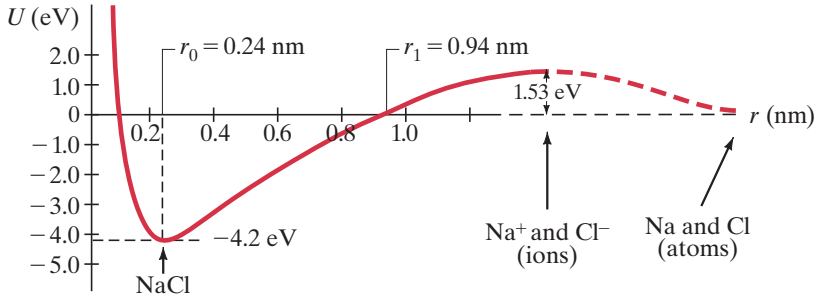


FIGURE 40–10 Example 40–1. Potential-energy diagram for the NaCl bond. Beyond about $r = 1.2\,\text{nm}$, the diagram is schematic only, and represents the energy difference between ions and neutral atoms. $U = 0$ is chosen for the two separated atoms Na and Cl (not for the ions). [For the two ions, Na^+ and Cl^-, the zero of potential energy at $r = \infty$ corresponds to $U \approx 1.53\,\text{eV}$ on this diagram.]

SOLUTION (*a*) The potential energy of two point charges is given by Coulomb's law:

$$U' = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r},$$

where q_1 and q_2 are the charges on the two ions. (This equation for U' differs from the U shown in Fig 40–10 which is "shifted" to show the activation energy and to have $U = 0$ at $r = \infty$.) This formula works for our two ions if we set $U' = 0$ at $r = \infty$, which in the plot of Fig. 40–10 corresponds to $U = +1.53\,\text{eV}$ (Fig. 40–10 is drawn for $U = 0$ for the free *atoms*). The point r_1 in Fig. 40–10 corresponds to $U' = -1.53\,\text{eV}$ relative to the two free ions. We solve the U' equation above for r, setting $q_1 = +e$ and $q_2 = -e$:

$$r_1 = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{U'} = \frac{(9.0\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(-1.60\times10^{-19}\,\text{C})(+1.60\times10^{-19}\,\text{C})}{(-1.53\,\text{eV})(1.60\times10^{-19}\,\text{J/eV})} = 0.94\,\text{nm},$$

which is just the measured value.

(*b*) At $r_0 = 0.24\,\text{nm}$, the potential energy of the two ions (relative to $r = \infty$ for the two ions) is

$$U' = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r} = \frac{(9.0\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(-1.60\times10^{-19}\,\text{C})(+1.60\times10^{-19}\,\text{C})}{(0.24\times10^{-9}\,\text{m})(1.60\times10^{-19}\,\text{J/eV})} = -6.0\,\text{eV}.$$

Thus, we estimate that 6.0 eV of energy is given up when Na^+ and Cl^- ions form a NaCl bond. Put another way, it takes 6.0 eV to break the NaCl bond and form the Na^+ and Cl^- ions. To get the binding energy—the energy to separate the NaCl into Na and Cl *atoms*—we need to subtract out the 1.53 eV (the "bump" in Fig. 40–10) needed to ionize them:

$$\text{binding energy} = 6.0\,\text{eV} - 1.53\,\text{eV} = 4.5\,\text{eV}.$$

The measured value (shown on Fig. 40–10) is 4.2 eV. The difference can be attributed to the energy associated with the repulsion of the electron shells at this distance.

PHYSICS APPLIED
ATP and energy in the cell

Sometimes the potential energy of a bond looks like that of Fig. 40–11. In this case, the energy of the bonded molecule, at a separation r_0, is greater than when there is no bond $(r = \infty)$. That is, an energy *input* is required to make the bond (hence the binding energy is negative), and there is energy release when the bond is broken. Such a bond is stable only because there is the barrier of the activation energy. This type of bond is important in living cells, for it is in such bonds that energy can be stored efficiently in certain molecules, particularly ATP (adenosine triphosphate). The bond that connects the last phosphate group (designated Ⓟ in Fig. 40–11) to the rest of the molecule (ADP, meaning adenosine diphosphate, since it contains only two phosphates) has potential energy of the shape shown in Fig. 40–11. Energy is stored in this bond. When the bond is broken $(\text{ATP} \rightarrow \text{ADP} + \text{Ⓟ})$, energy is released and this energy can be used to make other chemical reactions "go."

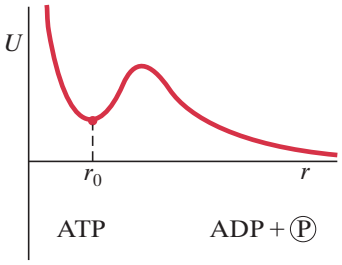


FIGURE 40–11 Potential-energy diagram for the formation of ATP from ADP and phosphate (Ⓟ).

In living cells, many chemical reactions have activation energies that are often on the order of several eV. Such energy barriers are not easy to overcome in the cell. This is where enzymes come in. They act as *catalysts*, which means they act to lower the activation energy so that reactions can occur that otherwise would not. Enzymes act via the electrostatic force to distort the bonding electron clouds, so that the initial bonds are easily broken.

EXAMPLE 40–2 **Bond length.** Suppose a diatomic molecule has a potential energy given by $U = -(1/4\pi\epsilon_0)(e^2/r) + B/r^6$ where $B = 1.0 \times 10^{-78}\,\text{J}\cdot\text{m}^6$ and r is the distance between the centers of the two atoms. Determine the expected equilibrium separation of the two atoms (bond length of the molecule).

APPROACH The force between the two atoms is given by $F = -dU/dr$ (analogous to Eq. 8–7). The classical equilibrium separation is found by setting $F = 0$.

SOLUTION

$$F = -dU/dr = (1/4\pi\epsilon_0)(-e^2/r^2) - B(-6/r^7) = -(1/4\pi\epsilon_0)(e^2/r^2) + 6B/r^7 = 0.$$

Then

$$r^5 = 6B(4\pi\epsilon_0)/e^2 = 2.6 \times 10^{-50}\,\text{m}^5.$$

So $r = 1.2 \times 10^{-10}\,\text{m} = 0.12\,\text{nm}$.

NOTE This can be shown to be a stable equilibrium point by checking the sign of d^2U/dr^2, or by evaluating F at positions slightly larger and smaller than the equilibrium position.

40–3 Weak (van der Waals) Bonds

Once a bond between two atoms or ions is made, energy must normally be supplied to break the bond and separate the atoms. As mentioned in Section 40–1, this energy is called the *bond energy* or *binding energy*. The binding energy for covalent and ionic bonds is typically 2 to 5 eV. These bonds,

PSE5_CH40_v1.7_23-Jun-20

which hold atoms together to *form* molecules, are often called **strong bonds** to distinguish them from so-called "weak bonds." The term **weak bond**, as we use it here, refers to attachments *between* molecules due to simple electrostatic attraction—such as *between* polar molecules (and not *within* a polar molecule, which is a strong bond). The strength of the attachment is much less than for the strong bonds. Binding energies are typically in the range 0.04 to 0.3 eV—hence their name "weak bonds."

Weak bonds are generally the result of attraction between dipoles (Sections 21–11, 23–6). For example, Fig. 40–12 shows two molecules, which have permanent dipole moments, attracting one another. Besides such **dipole–dipole bonds**, there can also be **dipole–induced dipole bonds**, in which a polar molecule with a permanent dipole moment can induce a dipole moment in an otherwise electrically balanced (nonpolar) molecule, just as a single charge can induce a separation of charge in a nearby object (see Fig. 21–7). There can even be an attraction between two nonpolar molecules, because their electrons are moving about: at any instant there may be a transient separation of charge, creating a brief dipole moment and weak attraction. All these weak bonds are referred to as **van der Waals bonds**, and the forces involved **van der Waals forces**. The potential energy has the general shape shown in Fig. 40–8, with the attractive van der Waals potential energy varying as $1/r^6$. The force decreases rapidly with increased distance.

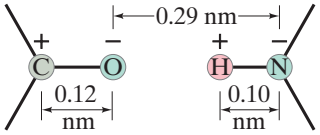
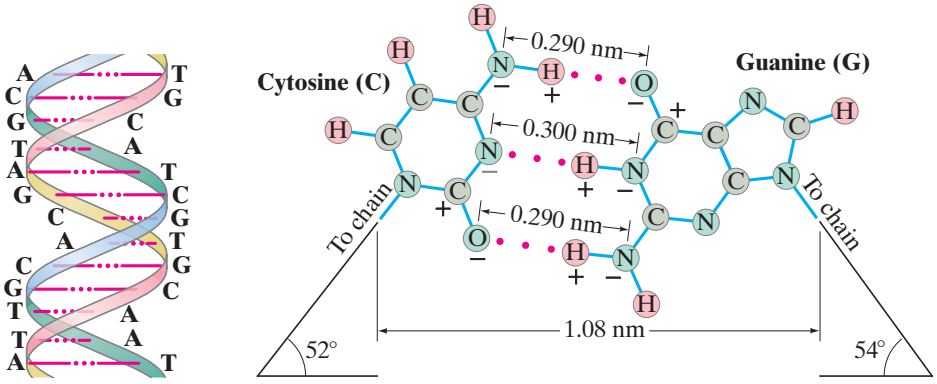


FIGURE 40–12 The $C^+—O^-$ and $H^+—N^-$ dipoles attract each other. (These dipoles may be part of, for example, the nucleotide bases cytosine and guanine in DNA molecules. See Fig. 40–13.) The + and − charges typically have magnitudes of a fraction of e.

When one of the atoms in a dipole–dipole bond is hydrogen, as in Fig. 40–12, it is called a **hydrogen bond**. A hydrogen bond is generally the strongest of the weak bonds, because the hydrogen atom is the smallest atom and can be approached more closely. Hydrogen bonds also have a partial "covalent" character: that is, electrons between the two dipoles may be shared to a small extent, making a stronger, more lasting bond.

Weak bonds are important in liquids and solids when strong bonds are absent (see Section 40–5). They are also very important for understanding the activities of cells, such as the double helix shape of DNA (Fig. 40–13), and DNA replication (see Section 21–12). The average kinetic energy of molecules in a living cell at normal temperatures ($T \approx 300\,\text{K}$) is around $\frac{3}{2}kT \approx 0.04\,\text{eV}$ (kinetic theory, Eq. 18–4), about the magnitude of weak bonds. This means that a weak bond can readily be broken just by a molecular collision. Hence weak bonds are not very permanent—they are, instead, brief attachments. This helps them play particular roles in the cell. On the other hand, strong bonds—those that hold molecules together—are almost never broken simply by molecular collision because their binding energies are much too high (≈ 2 to $5\,\text{eV}$). Thus they are relatively permanent. They can be broken by chemical action (the making of even stronger bonds), and this usually happens in the cell with the aid of an enzyme, which is a protein molecule.

PHYSICS APPLIED
DNA



(a) (b)

FIGURE 40–13 (a) Model of part of a DNA double helix. The red dots represent hydrogen bonds between the two strands. (b) "Close-up" view: cytosine (C) and guanine (G) molecules on separate strands of a DNA double helix are held together by the hydrogen bonds (red dots) involving an H^+ on one molecule attracted to an N^- or $C^+—O^-$ of a molecule on the adjacent chain. See also Section 21–12 and Figs. 21–49 and 21–50.

EXAMPLE 40–3 **Nucleotide energy.** Calculate the potential energy between a $C^+—O^-$ dipole of the nucleotide base cytosine and the nearby $H^+—N^-$ dipole of guanine, assuming that the two dipoles are lined up as shown in Fig. 40–12. Dipole moment $(= q\ell)$ measurements (see Table 23–2 and Fig. 40–12) give

$$q_H = -q_N = \frac{3.0 \times 10^{-30}\,\text{C}\cdot\text{m}}{0.10 \times 10^{-9}\,\text{m}} = 3.0 \times 10^{-20}\,\text{C} = 0.19e,$$

and

$$q_C = -q_O = \frac{8.0 \times 10^{-30}\,\text{C}\cdot\text{m}}{0.12 \times 10^{-9}\,\text{m}} = 6.7 \times 10^{-20}\,\text{C} = 0.42e.$$

APPROACH We want to find the potential energy of the two charges in one dipole due to the two charges in the other, because this will be equal to the work needed to pull the two dipoles infinitely far apart. The potential energy U of a charge q_1 in the presence of a charge q_2 is $U = (1/4\pi\epsilon_0)(q_1 q_2/r_{12})$ where $1/4\pi\epsilon_0 = 9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2$ and r_{12} is the distance between the two charges.

SOLUTION The potential energy U consists of four terms:

$$U = U_{CH} + U_{CN} + U_{OH} + U_{ON},$$

where U_{CH} means the potential energy of C in the presence of H, and similarly for the other terms. We do not have terms corresponding to C and O, or N and H, because the two dipoles are assumed to be stable entities. Then, using the distances shown in Fig. 40–12, we get:

$$\begin{aligned} U &= \frac{1}{4\pi\epsilon_0}\left[\frac{q_C q_H}{r_{CH}} + \frac{q_C q_N}{r_{CN}} + \frac{q_O q_H}{r_{OH}} + \frac{q_O q_N}{r_{ON}}\right] \\ &= \frac{1}{4\pi\epsilon_0} q_C q_H\left(\frac{1}{r_{CH}} - \frac{1}{r_{CN}} - \frac{1}{r_{OH}} + \frac{1}{r_{ON}}\right) \\ &= (9.0 \times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)(6.7)(3.0)\left(\frac{1}{0.31} - \frac{1}{0.41} - \frac{1}{0.19} + \frac{1}{0.29}\right)\frac{(10^{-20}\,\text{C})^2}{(10^{-9}\,\text{m})} \\ &= -1.86 \times 10^{-20}\,\text{J} = -0.12\,\text{eV}. \end{aligned}$$

The potential energy is negative, meaning 0.12 eV of work (or energy input) is required to separate the dipoles. That is, the binding energy of this "weak" or hydrogen bond is 0.12 eV. This is only an estimate, of course, since other charges in the vicinity would have an influence too. Note how small and wide this energy is compared to the 4.5 eV of an H_2 molecule (Fig. 40–8) or the 4.2 eV for the NaCl bond (end of Example 40–1).

*Protein Synthesis

Weak bonds, especially hydrogen bonds, are crucial to the process of protein synthesis. Proteins serve as structural parts of the cell and as enzymes to catalyze chemical reactions needed for the growth and survival of the organism. A protein molecule consists of one or more chains of small molecules known as *amino acids*.[†] There are 20 different amino acids, and a single protein chain may contain hundreds of them in a specific order. The standard model for how amino acids are connected together in the correct order to form a protein molecule is shown schematically in Fig. 40–14.

PHYSICS APPLIED
Protein synthesis

It begins at the DNA double helix: each gene on a chromosome contains the information for producing one protein. The ordering of the four bases, A, C, G, and T, provides the "code," the **genetic code**, for the order of amino acids in the protein. First, the DNA double helix unwinds and a new molecule called *messenger*-RNA (m-RNA) is synthesized using one strand of the DNA as a "template." m-RNA is a chain molecule containing four different bases, like those of DNA (Section 21–12) except that thymine (T) is replaced by the similar uracil molecule (U). Near the top left in Fig. 40–14, a C has just been added to the growing m-RNA chain in much

[†]Typical amino acids have atomic masses of about 100, and are made up mainly of H, C, O, and N.

PSE5_CH40_v1.7_23-Jun-20

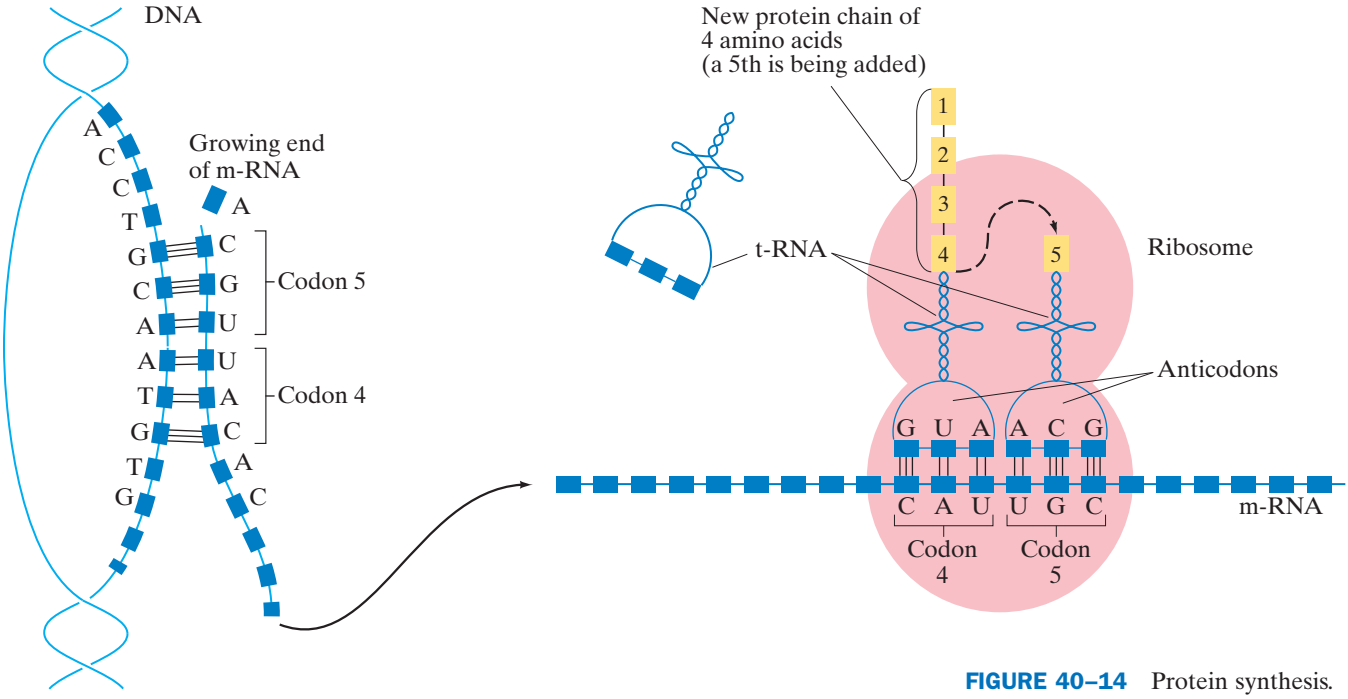


FIGURE 40–14 Protein synthesis. The yellow rectangles represent amino acids. See text for details.

the same way that DNA replicates (Fig. 21–49); and an A, attracted and held close to the T on the DNA chain by the electrostatic force, will soon be attached to the C by an enzyme. The order of the bases, and thus the genetic information, is preserved in the m-RNA because the shapes of the molecules only allow the "proper" one to get close enough so the electrostatic force can act to form weak bonds.

Next, the m-RNA is buffeted about in the cell (kinetic theory, Chapter 18) until it gets close to a tiny organelle known as a *ribosome*, to which it can attach by electrostatic attraction (on the right in Fig. 40–14), because their shapes allow the charged parts to get close enough to form weak bonds (for which the force decreases greatly with increased separation). Also held by the electrostatic force to the ribosome are one or two *transfer*-RNA (t-RNA) molecules. These t-RNA molecules "translate" the genetic code of nucleotide bases into amino acids in the following way. There is a different t-RNA molecule for each amino acid and each combination of three bases. On one end of a t-RNA molecule is an amino acid. On the other end of the t-RNA molecule is the appropriate "anticodon," a set of three nucleotide bases that "code" for that amino acid. If all three bases of an anticodon match the three bases of the "codon" on the m-RNA (in the sense of G to C and A to U), the anticodon is attracted electrostatically to the m-RNA codon and that t-RNA molecule is held there briefly. The ribosome has two particular attachment sites which hold two t-RNA molecules while enzymes bond the two amino acids together to lengthen the amino acid chain (yellow in Fig. 40–14). As each amino acid is connected by an enzyme (four are already connected in Fig. 40–14, top right, and a fifth is about to be connected), the old t-RNA molecule is removed—perhaps by a random collision with some molecule in the cellular fluid. A new one soon becomes attracted as the ribosome moves along the m-RNA.

This process of protein synthesis is often presented as if it occurred in clockwork fashion—as if each molecule knew its role and went to its assigned place. But this is not the case. The forces of attraction between the electric charges of the molecules are rather weak and become significant only when the molecules can come close together and when several weak bonds can be made. Indeed, if the shapes are not just right, the electrostatic attraction is nearly zero, which is why there are few mistakes. The fact that weak bonds are weak is very important. If they were strong, collisions with other molecules would not allow a t-RNA molecule to be released from the ribosome, or the m-RNA to be released from the DNA. If they were not temporary encounters, metabolism would grind to a halt.

As each amino acid is added to the next, the protein molecule grows in length until it is complete. Even as it is being made, this chain is being buffeted about in the cell—we might think of a wiggling worm. But a protein molecule has electrically charged polar groups along its length. And as it takes on various shapes, the electric forces of attraction between different parts of the molecule will eventually lead to a particular shape of the protein which is quite stable. Each type of protein has its own special shape, depending on the location of charged atoms. In the last analysis, the final shape depends on the order of the amino acids.

40–4 Molecular Spectra

When atoms combine to form molecules, the probability distributions of the outer electrons overlap and this interaction alters the energy levels. Nonetheless, molecules can undergo transitions between electron energy levels just as atoms do. For example, the H_2 molecule can absorb a photon of just the right frequency to excite one of its ground-state electrons to an excited state. The excited electron can then return to the ground state, emitting a photon. The energy of photons emitted by molecules can be of the same order of magnitude as for atoms, typically 1 to 10 eV, or less.

Additional energy levels become possible for molecules (but not for atoms) because the molecule as a whole can rotate, and the atoms of the molecule can vibrate relative to each other. The energy levels for both rotational and vibrational levels are quantized, and are generally spaced much more closely $(10^{-3}$ to 10^{-1} eV) than the electronic levels. Each atomic energy level thus becomes a set of closely spaced levels corresponding to the vibrational and rotational motions, Fig. 40–15. Transitions from one level to another appear as many very closely spaced lines. In fact, the lines are not always distinguishable, and these spectra are called **band spectra**. Each type of molecule has its own characteristic spectrum, which can be used for identification and for determination of structure. We now look in more detail at rotational and vibrational states in molecules.

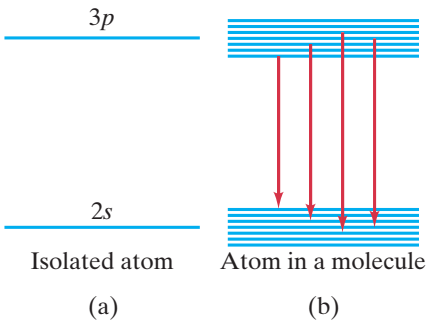


FIGURE 40–15 (a) The individual energy levels of an isolated atom become (b) bands of closely spaced levels in molecules, as well as in solids and liquids.

Rotational Energy Levels in Molecules

We consider only diatomic molecules, although the analysis can be extended to polyatomic molecules. When a diatomic molecule rotates about its center of mass as shown in Fig. 40–16, its kinetic energy of rotation (see Section 10–8) is

$$E_{\text{rot}} = \frac{1}{2} I\omega^2 = \frac{(I\omega)^2}{2I},$$

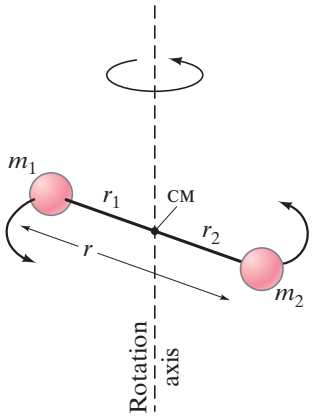
where $I\omega$ is the angular momentum (Section 11–1). Quantum mechanics predicts quantization of angular momentum just as in atoms (see Eq. 39–3):

$$I\omega = \sqrt{\ell(\ell+1)}\,\hbar, \qquad \ell = 0, 1, 2, \cdots,$$

where ℓ is an integer called the **rotational angular momentum quantum number**. Thus the rotational energy is quantized:

$$E_{\text{rot}} = \frac{(I\omega)^2}{2I} = \ell(\ell+1)\frac{\hbar^2}{2I}. \qquad \ell = 0, 1, 2, \cdots. \qquad \textbf{(40–2)}$$

FIGURE 40–16 Diatomic molecule rotating about a vertical axis.



Transitions between rotational energy levels are subject to the **selection rule** (as in Section 39–2):

$$\Delta\ell = \pm 1.$$

The energy of a photon emitted or absorbed for a transition between rotational states with angular momentum quantum number ℓ and $\ell - 1$ will be

$$\Delta E_{\text{rot}} = E_\ell - E_{\ell-1} = \frac{\hbar^2}{2I}\ell(\ell+1) - \frac{\hbar^2}{2I}(\ell-1)(\ell)$$

$$= \frac{\hbar^2}{I}\ell. \qquad \left[\begin{array}{c}\ell \text{ is for upper} \\ \text{energy state}\end{array}\right] \qquad \textbf{(40–3)}$$

We see that the transition energy increases directly with ℓ. Figure 40–17 shows some of the allowed rotational energy levels and transitions. Measured absorption lines fall in the microwave or far-infrared regions of the spectrum (energies $\approx 10^{-3}$ eV), and their frequencies are generally 2, 3, 4, $\cdots$ times higher than the lowest one, as predicted by Eq. 40–3.

EXERCISE A Determine the three lowest rotational energy states (in eV) for a nitrogen molecule which has a moment of inertia $I = 1.39 \times 10^{-46}\,\text{kg}\cdot\text{m}^2$.

The moment of inertia of the molecule in Fig. 40–16 rotating about its center of mass (Section 10–5) is

$$I = m_1 r_1^2 + m_2 r_2^2,$$

where r_1 and r_2 are the distances of each atom from their common center of mass. We can show (in Example 40–4 below) that I can be written

$$I = \frac{m_1 m_2}{m_1 + m_2} r^2 = \mu r^2, \qquad \textbf{(40–4)}$$

where $r = r_1 + r_2$ is the distance between the two atoms of the molecule and $\mu = m_1 m_2/(m_1 + m_2)$ is called the **reduced mass**. If $m_1 = m_2$, then $\mu = \frac{1}{2}m_1 = \frac{1}{2}m_2$.

EXAMPLE 40–4 **Reduced mass.** Show that the moment of inertia of a diatomic molecule rotating about its center of mass can be written

$$I = \mu r^2,$$

where

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

is the reduced mass, Eq. 40–4, and r is the distance between the atoms.

SOLUTION The moment of inertia of a single particle of mass m a distance r from the rotation axis is $I = mr^2$ (Eq. 10–11 or 10–13). For our diatomic molecule (Fig. 40–16)

$$I = m_1 r_1^2 + m_2 r_2^2.$$

Now $r = r_1 + r_2$ and $m_1 r_1 = m_2 r_2$ because the axis of rotation passes through the center of mass. Hence

$$r_1 = r - r_2 = r - \frac{m_1}{m_2} r_1.$$

Solving for r_1 gives

$$r_1 = \frac{r}{1 + \dfrac{m_1}{m_2}} = \frac{m_2 r}{m_1 + m_2}.$$

Similarly,

$$r_2 = \frac{m_1 r}{m_1 + m_2}.$$

Then (see first equation of this Solution)

$$I = m_1\left(\frac{m_2 r}{m_1 + m_2}\right)^2 + m_2\left(\frac{m_1 r}{m_1 + m_2}\right)^2 = \frac{m_1 m_2 (m_1 + m_2) r^2}{(m_1 + m_2)^2}$$

$$= \frac{m_1 m_2}{m_1 + m_2} r^2 = \mu r^2,$$

where

$$\mu = \frac{m_1 m_2}{m_1 + m_2},$$

which is what we wished to show.

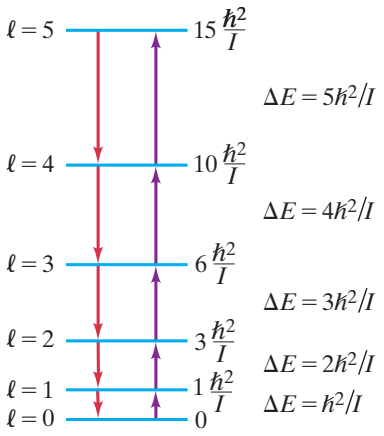


FIGURE 40–17 Rotational energy levels and allowed transitions (emission and absorption) for a diatomic molecule. Upward-pointing arrows represent absorption of a photon, and downward arrows represent emission of a photon.

EXAMPLE 40–5 **Rotational transition.** A rotational transition $\ell = 1$ to $\ell = 0$ for the molecule CO has a measured absorption wavelength $\lambda_1 = 2.60\,\text{mm}$ (microwave region). Use this to calculate (*a*) the moment of inertia of the CO molecule, and (*b*) the CO bond length, *r*. (*c*) Calculate the wavelengths of the next three rotational transitions, and the energies of the photon emitted for each of these four transitions.

APPROACH The absorption wavelength is used to find the energy of the absorbed photon. The moment of inertia I is found from Eq. 40–3, and the bond length r from Eq. 40–4.

SOLUTION (*a*) The photon energy, $E = hf = hc/\lambda$, equals the rotational energy level difference, ΔE_{rot}. From Eq. 40–3, we can write

$$\frac{\hbar^2}{I}\ell = \Delta E_{\text{rot}} = hf = \frac{hc}{\lambda_1}.$$

With $\ell = 1$ (the upper state) in this case, we solve for I:

$$I = \frac{\hbar^2\ell}{hc}\lambda_1 = \frac{h\lambda_1}{4\pi^2 c} = \frac{(6.63\times10^{-34}\,\text{J}\cdot\text{s})(2.60\times10^{-3}\,\text{m})}{4\pi^2(3.00\times10^{8}\,\text{m/s})}$$
$$= 1.46\times10^{-46}\,\text{kg}\cdot\text{m}^2.$$

(*b*) The masses of C and O are 12.0 u and 16.0 u, respectively, where $1\,\text{u} = 1.66\times10^{-27}\,\text{kg}$. Thus the reduced mass is

$$\mu = \frac{m_1 m_2}{m_1 + m_2} = \frac{(12.0)(16.0)}{28.0}(1.66\times10^{-27}\,\text{kg}) = 1.14\times10^{-26}\,\text{kg}$$

or 6.86 u. Then, from Eq. 40–4, the bond length is

$$r = \sqrt{\frac{I}{\mu}} = \sqrt{\frac{1.46\times10^{-46}\,\text{kg}\cdot\text{m}^2}{1.14\times10^{-26}\,\text{kg}}} = 1.13\times10^{-10}\,\text{m} = 0.113\,\text{nm}.$$

(*c*) From Eq. 40–3, $\Delta E \propto \ell$. Hence $\lambda = c/f = hc/\Delta E$ is proportional to $1/\ell$. Thus, for $\ell = 2$ to $\ell = 1$ transitions, $\lambda_2 = \frac{1}{2}\lambda_1 = 1.30\,\text{mm}$. For $\ell = 3$ to $\ell = 2$, $\lambda_3 = \frac{1}{3}\lambda_1 = 0.87\,\text{mm}$. And for $\ell = 4$ to $\ell = 3$, $\lambda_4 = 0.65\,\text{mm}$. All are close to measured values. The energies of the photons, $hf = hc/\lambda$, are respectively $4.8\times10^{-4}\,\text{eV}$, $9.5\times10^{-4}\,\text{eV}$, $1.4\times10^{-3}\,\text{eV}$, and $1.9\times10^{-3}\,\text{eV}$.

Vibrational Energy Levels in Molecules

The potential energy of the two atoms in a typical diatomic molecule has the shape shown in Fig. 40–8 or 40–9, and Fig. 40–18 again shows the potential energy for the H_2 molecule (solid curve). This potential energy curve, at least in the vicinity of the equilibrium separation r_0, closely resembles the potential energy of a simple harmonic oscillator, $U = \frac{1}{2}kx^2$, which is shown superposed in dashed lines. Thus, for small displacements from r_0, each atom experiences a restoring

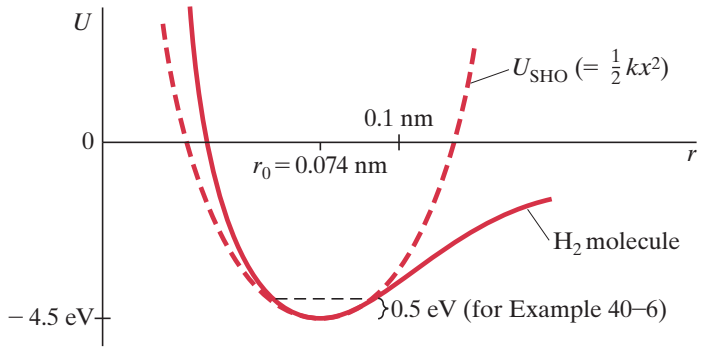


FIGURE 40–18 Potential energy for the H_2 molecule and for a simple harmonic oscillator ($U_{\text{SHO}} = \frac{1}{2}kx^2$, with $|x| = |r - r_0|$) The 0.50-eV energy height marked is for use in Example 40–6 to estimate k. [Note that $U_{\text{SHO}} = 0$ is not the same as $U = 0$ for the molecule.]

PSE5_CH40_v1.7_23-Jun-20

force approximately proportional to the displacement, and the molecule vibrates as a simple harmonic oscillator (SHO). The classical frequency of vibration is

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}, \qquad \textbf{(40–5)}$$

where k is the "stiffness constant" (as for a spring, Chapter 14) and instead of the mass m we must again use the reduced mass $\mu = m_1 m_2/(m_1 + m_2)$. (This is shown in Problem 21.) The Schrödinger equation for the SHO potential energy yields solutions for energy that are quantized according to

$$E_{\text{vib}} = (\nu + \tfrac{1}{2})hf \qquad \nu = 0, 1, 2, \cdots, \qquad \textbf{(40–6)}$$

where the frequency f is given by Eq. 40–5 and ν is an integer called the **vibrational quantum number**. The lowest energy state $(\nu = 0)$ is not zero (as for rotation) but has $E = \frac{1}{2}hf$. This is called the **zero-point energy**.† Higher states have energy $\frac{3}{2}hf$, $\frac{5}{2}hf$, and so on, as shown in Fig. 40–19. Transitions between vibrational energy levels are subject to the **selection rule**

$$\Delta\nu = \pm 1,$$

so allowed transitions occur only between adjacent states‡ and all give off photons of energy

$$\Delta E_{\text{vib}} = hf. \qquad \textbf{(40–7)}$$

This is very close to experimental values for small ν; but for higher energies, the potential-energy curve (Fig. 40–18) begins to deviate from a perfect SHO curve, which affects the wavelengths and frequencies of the transitions. Typical transition energies are on the order of 10^{-1} eV, about 10 to 100 times larger than for rotational transitions, with wavelengths in the infrared region of the spectrum $(\approx 10^{-5}\,\text{m})$.

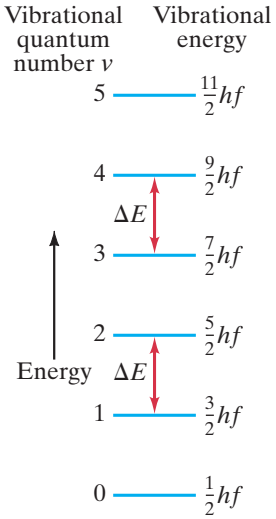


FIGURE 40–19 Allowed vibrational energies of a diatomic molecule, where f is the fundamental frequency of vibration, given by Eq. 40–5. The energy levels are equally spaced. Transitions are allowed only between adjacent levels $(\Delta\nu = \pm 1)$.

EXAMPLE 40–6 ESTIMATE **Wavelength for H_2.** (*a*) Use the curve of Fig. 40–18 to estimate the value of the stiffness constant k for the H_2 molecule, and then (*b*) estimate the fundamental wavelength for vibrational transitions.

APPROACH To find k, we arbitrarily choose an energy height of 0.50 eV which is indicated in Fig. 40–18. By measuring directly on the graph, we find that this energy corresponds to a vibration on either side of $r_0 = 0.074$ nm of about $x = \pm 0.017$ nm.

SOLUTION (*a*) For SHO, $U_{\text{SHO}} = \frac{1}{2}kx^2$ and $U_{\text{SHO}} = 0$ at $x = 0$ $(r = r_0)$; then

$$k = \frac{2U_{\text{SHO}}}{x^2} \approx \frac{2(0.50\,\text{eV})(1.6\times10^{-19}\,\text{J/eV})}{(1.7\times10^{-11}\,\text{m})^2} \approx 550\,\text{N/m}.$$

NOTE This value of k would also be reasonable for a macroscopic spring.
(*b*) The reduced mass is $\mu = m_1 m_2/(m_1 + m_2) = m_1/2 = \frac{1}{2}(1.0\,\text{u})(1.66\times10^{-27}\,\text{kg}) = 0.83\times10^{-27}\,\text{kg}$. Hence, using Eq. 40–5,

$$\lambda = \frac{c}{f} = 2\pi c\sqrt{\frac{\mu}{k}} = 2\pi(3.0\times10^8\,\text{m/s})\sqrt{\frac{0.83\times10^{-27}\,\text{kg}}{550\,\text{N/m}}} = 2300\,\text{nm},$$

which is in the infrared region of the spectrum.

Experimentally, we do the inverse process: The wavelengths of vibrational transitions for a given molecule are measured, and from this the stiffness constant k can be calculated. The values of k calculated in this way are a measure of the strength of the molecular bond.

†Recall this phenomenon for a square well, Fig. 38–8.

‡Forbidden transitions with $\Delta\nu = 2$ occur with much lower probability, but their observation can be important in some cases, such as in astronomy.

EXAMPLE 40–7 **Vibrational energy levels in hydrogen.** Hydrogen molecule vibrations emit infrared radiation of wavelength around 2300 nm. (*a*) What is the separation in energy between adjacent vibrational levels? (*b*) What is the lowest vibrational energy state?

APPROACH The energy separation between adjacent vibrational levels is $\Delta E_{\text{vib}} = hf = hc/\lambda$. The lowest energy (Eq. 40–6) has $\nu = 0$.

SOLUTION (*a*)

$$\Delta E_{\text{vib}} = hf = \frac{hc}{\lambda} = \frac{(6.63\times10^{-34}\,\text{J}\cdot\text{s})(3.00\times10^{8}\,\text{m/s})}{(2300\times10^{-9}\,\text{m})(1.60\times10^{-19}\,\text{J/eV})} = 0.54\,\text{eV},$$

where the denominator includes the conversion factor from joules to eV.

(*b*) The lowest vibrational energy has $\nu = 0$ in Eq. 40–6:

$$E_{\text{vib}} = (\nu + \tfrac{1}{2})hf = \tfrac{1}{2}hf = 0.27\,\text{eV}.$$

EXERCISE B What is the energy of the first vibrational state above the ground state in the hydrogen molecule?

Rotational Plus Vibrational Levels

When energy is imparted to a molecule, both the rotational and vibrational modes can be excited. Because rotational energies are an order of magnitude or so smaller than vibrational energies, which in turn are smaller than the electronic energy levels, we can represent the grouping of levels as shown in Fig. 40–20. Transitions between energy levels, with emission of a photon, are subject to the **selection rules**:

$$\Delta\nu = \pm 1 \qquad \text{and} \qquad \Delta\ell = \pm 1.$$

Some allowed and forbidden (marked ×) transitions are indicated in Fig. 40–20. Not all transitions and levels are shown, and the separation between vibrational levels, and (even more) between rotational levels, has been exaggerated. But we can clearly see the origin of the very closely spaced lines that give rise to the band spectra, as mentioned with reference to Fig. 40–15 earlier in this Section.

The spectra are quite complicated, so we consider briefly only transitions within the same electronic level, such as those at the top of Fig. 40–20. A transition from a state with quantum numbers ν and ℓ, to one with quantum numbers $\nu + 1$ and $\ell \pm 1$ (see the selection rules above), will absorb† a photon of energy:

$$\Delta E = \Delta E_{\text{vib}} + \Delta E_{\text{rot}}$$

$$= hf + (\ell + 1)\frac{\hbar^2}{I} \qquad \begin{bmatrix} \ell \to \ell + 1 \\ (\Delta\ell = +1) \end{bmatrix}, \quad \ell = 0, 1, 2, \cdots \qquad \textbf{(40–8a)}$$

$$= hf - \ell\frac{\hbar^2}{I} \qquad \begin{bmatrix} \ell \to \ell - 1 \\ (\Delta\ell = -1) \end{bmatrix}, \quad \ell = 1, 2, 3, \cdots, \qquad \textbf{(40–8b)}$$

where we have used Eqs. 40–3 and 40–7. Note that for $\ell \to \ell - 1$ transitions, ℓ cannot be zero because there is then no state with $\ell = -1$. Equations 40–8 predict an absorption spectrum like that shown schematically in Fig. 40–21,

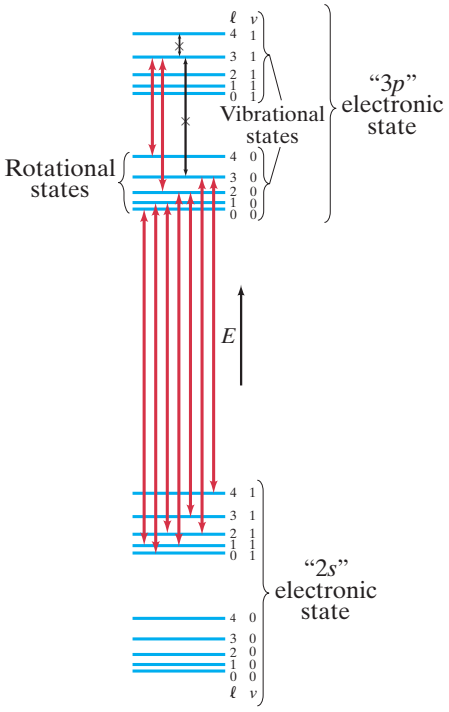


FIGURE 40–20 Combined electronic, vibrational, and rotational energy levels. Transitions marked with an × are not allowed by selection rules.

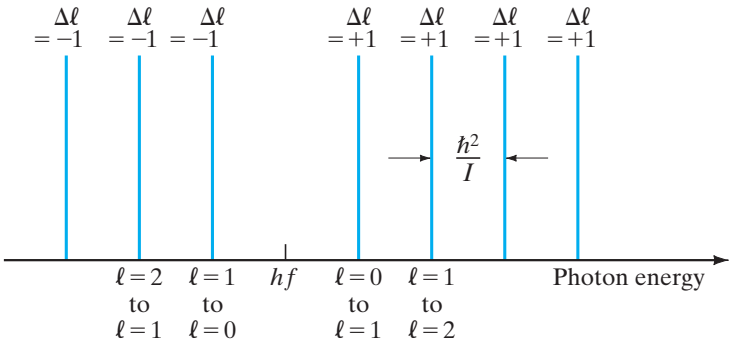


FIGURE 40–21 Expected spectrum for transitions between combined rotational and vibrational states.

†Eqs. 40–8 are for absorption; for emission of a photon, the transition would be $\nu \to \nu - 1$, $\ell \to \ell \pm 1$.

PSE5_CH40_v1.7_23-Jun-20

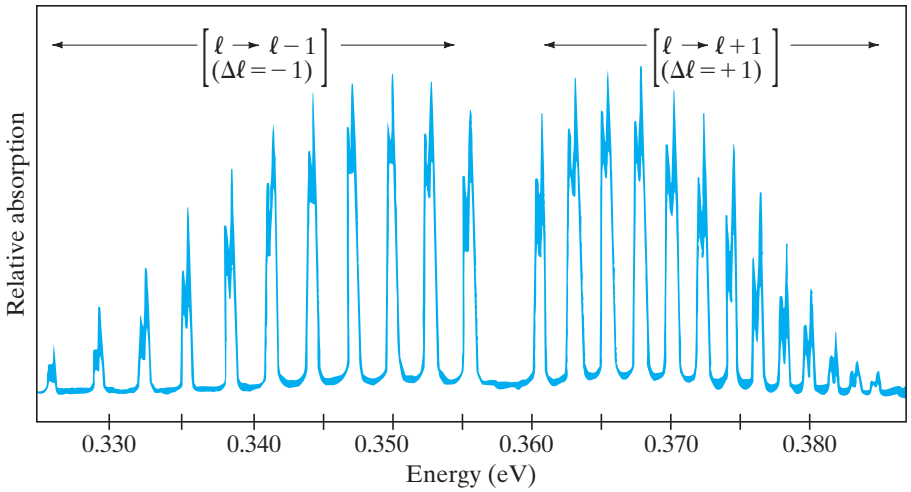


FIGURE 40–22 Absorption spectrum for HCl molecules. Lines on the left correspond to transitions where $\ell \to \ell - 1$; those on the right are for $\ell \to \ell + 1$. Each line has a double peak because chlorine has two isotopes of different mass and different moment of inertia.

with transitions $\ell \to \ell - 1$ on the left and $\ell \to \ell + 1$ on the right. Figure 40–22 shows the molecular absorption spectrum of HCl, which follows this pattern very well. (Each line in the spectrum of Fig. 40–22 is split into two because Cl consists of two isotopes of different mass; hence there are two kinds of HCl molecule with slightly different moments of inertia I.)

EXAMPLE 40–8 **ESTIMATE** **The HCl molecule.** Estimate the moment of inertia of the HCl molecule using the absorption spectrum shown in Fig. 40–22. For the purposes of a rough estimate you can ignore the difference between the two isotopes.

APPROACH The locations of the peaks in Fig. 40–22 should correspond to Eqs. 40–8. We don't know what value of ℓ each peak corresponds to in Fig. 40–22, but we can estimate the energy difference between peaks to be about $\Delta E' = 0.0025\,\text{eV}$.

SOLUTION From Eqs. 40–8, the energy difference between two peaks is given by

$$\Delta E' = \Delta E_{\ell+1} - \Delta E_\ell = \frac{\hbar^2}{I}.$$

Then

$$I = \frac{\hbar^2}{\Delta E'} = \frac{(6.626 \times 10^{-34}\,\text{J}\cdot\text{s}/2\pi)^2}{(0.0025\,\text{eV})(1.6 \times 10^{-19}\,\text{J/eV})} = 2.8 \times 10^{-47}\,\text{kg}\cdot\text{m}^2.$$

NOTE To get an idea of what this number means, we write $I = \mu r^2$ (Eq. 40–4), where μ is the reduced mass (Example 40–4); then we calculate μ:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} = \frac{(1.0\,\text{u})(35\,\text{u})}{36\,\text{u}}(1.66 \times 10^{-27}\,\text{kg/u}) = 1.6 \times 10^{-27}\,\text{kg};$$

the bond length is given by (Eq. 40–4)

$$r = \left(\frac{I}{\mu}\right)^{\frac{1}{2}} = \left(\frac{2.8 \times 10^{-47}\,\text{kg}\cdot\text{m}^2}{1.6 \times 10^{-27}\,\text{kg}}\right)^{\frac{1}{2}} = 1.3 \times 10^{-10}\,\text{m},$$

which is the expected order of magnitude for a bond length.

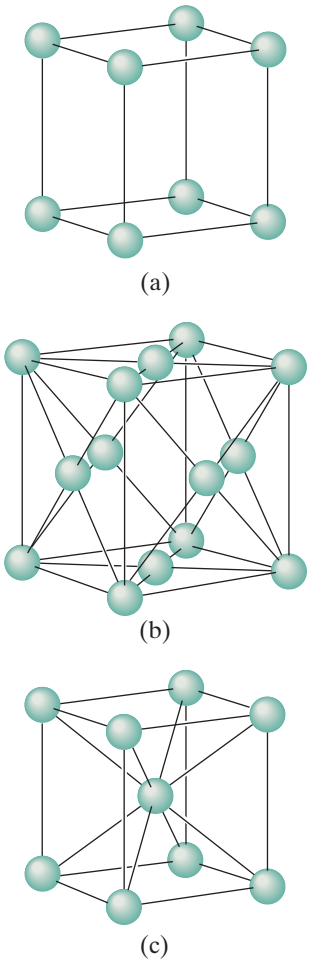


FIGURE 40–23 Arrangement of atoms in (a) a simple cubic crystal, (b) face-centered cubic crystal (note the atom at the center of each face), and (c) body-centered cubic crystal. Each of these "cells" is repeated in three dimensions to the edges of the macroscopic crystal.

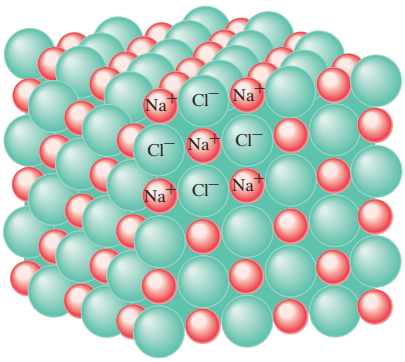


FIGURE 40–24 Diagram of an NaCl crystal, showing the "packing" of atoms.

40–5 Bonding in Solids

Quantum mechanics has been a great tool for understanding the structure of solids. This active field of research today is called **solid-state physics**, or **condensed-matter physics** so as to include liquids as well. The rest of this Chapter is devoted to this subject, and we begin with a brief look at the structure of solids and the bonds that hold them together.

Although some solid materials are *amorphous* in structure (such as glass), in that the atoms and molecules show no long-range order, we are interested here in the large class of *crystalline* substances whose atoms, ions, or molecules are generally accepted to form an orderly array known as a **lattice**. Figure 40–23 shows three of the possible arrangements of atoms in a crystal: simple cubic, face-centered cubic, and body-centered cubic. The NaCl crystal is face-centered cubic (see Fig. 40–24), with one Na^+ ion or one Cl^- ion at each "lattice point" (i.e., considering Na and Cl *separately*).

The molecules of a solid are held together in a number of ways. The most common are by *covalent* bonding (such as between the carbon atoms of the diamond crystal) and by *ionic* bonding (as in a NaCl crystal). Often the bonds are partially covalent and partially ionic. Our discussion of these bonds earlier in this Chapter for molecules applies equally well here to solids.

Let us look for a moment at the NaCl crystal of Fig. 40–24. Each Na^+ ion feels an attractive Coulomb potential due to each of the six "nearest neighbor" Cl^- ions surrounding it.† Note that one Na^+ does not "belong" exclusively to one Cl^-, so we must not think of ionic solids as consisting of individual molecules. Each Na^+ also feels a repulsive Coulomb potential due to other Na^+ ions, although this is weaker since the other Na^+ ions are farther away. Thus we expect a net attractive potential

$$U = -\alpha \frac{1}{4\pi\epsilon_0}\frac{e^2}{r}.$$

The factor α is called the *Madelung constant*. If each Na^+ were surrounded by only the six Cl^- ions, α would be 6, but the influence of all the other ions reduces it to a value $\alpha = 1.75$ for the NaCl crystal. The potential must also include a term representing the repulsive force when the wave functions of the inner shells and subshells overlap, and this has the form $U = B/r^m$, where m is a small integer. The sum of these two terms suggests a potential energy

$$U = -\frac{\alpha}{4\pi\epsilon_0}\frac{e^2}{r} + \frac{B}{r^m}, \qquad \textbf{(40–9)}$$

which has the same form as Eq. 40–1 for molecules (Section 40–2). It can be shown (Problem 28) that, at the equilibrium distance r_0,

$$U = U_0 = -\frac{\alpha}{4\pi\epsilon_0}\frac{e^2}{r_0}\left(1 - \frac{1}{m}\right).$$

This U_0 is known as the *ionic cohesive energy*; it is a sort of "binding energy"—the energy (per ion) needed to take the solid apart into separated ions, one by one.

A different type of bond occurs in metals. Metal atoms have relatively loosely held outer electrons. **Metallic bond** theories propose that in a metallic solid, these outer electrons roam rather freely among all the metal atoms which, without their outer electrons, act like positive ions. According to the theory the electrostatic attraction between the metal ions and this negative electron "gas" is responsible, at least in part, for holding the solid together. The binding energy of metal bonds is typically 1 to 3 eV, somewhat weaker than ionic or covalent bonds (5 to 10 eV in solids). The "free electrons," then, are responsible for the high electrical and

†One to each side in 3 dimensions: to the left, to the right, above, below, in front and behind.

thermal conductivity of metals (see Sections 40–6 and 40–7). This theory also nicely accounts for the shininess of smooth metal surfaces: the free electrons can vibrate at any frequency, so when light of a range of frequencies falls on a metal, the electrons can vibrate in response and re-emit light of those same frequencies. Hence the reflected light will consist largely of the same frequencies as the incident light. Compare this to nonmetallic materials that have a distinct color—the atomic electrons exist only in certain energy states, and when white light falls on them, the atoms absorb at certain frequencies, and reflect other frequencies which make up the color we see.

Here is a brief comparison of important strong bonds:

- ionic: electrons are "grabbed" from one atom by another;
- covalent: electrons are shared by atoms within a single molecule;
- metallic: electrons are shared by all atoms in the metal.

The atoms or molecules of some materials, such as the noble gases, can form only **weak bonds** with each other. As we saw in Section 40–3, weak bonds have very low binding energies and would not be expected to hold atoms together as a liquid or solid at room temperature. The noble gases condense only at very low temperatures, where the (thermal) kinetic energy per atom is small and the weak attraction can then hold the atoms together.

EXERCISE C Return to Chapter-Opening Question 1, page 1186, and answer it again now. Try to explain why you may have answered differently the first time.

40–6 Free-Electron Theory of Metals; Fermi Energy

Let us look more closely at the free-electron theory of metals mentioned in the preceding Section. Let us imagine the electrons trapped within the metal, moving around in a potential well: inside the metal, the potential energy is zero, but at the edges of the metal there are high potential walls. Since very few electrons leave the metal at low temperature, we can imagine the walls as being almost infinitely high (as in Section 38–8). At higher temperatures, electrons do leave the metal (we know that thermionic emission occurs, Section 23–10), so we must recognize that the well is of finite depth. In this simple model, the electrons are trapped within the metal, but are free to move about inside the well whose size is macroscopic—the size of the piece of metal. The energy, which is kinetic energy, will be quantized; but the spacing between energy levels will be very tiny (see Eq. 38–13) because the width of the potential well ℓ is very large. Indeed, for a cube 1 cm on a side, the number of states with energy between, say, 5.0 and 5.5 eV is on the order of 10^{22} (see Example 40–9).

To deal with such vast numbers of states, which are so closely spaced as to seem continuous, we need to use statistical methods. We define a quantity known as the **density of states**, $g(E)$, whose meaning is similar to the Maxwell distribution, Eq. 18–6 (see Section 18–2). That is, the quantity $g(E)\,dE$ represents the number of states per unit volume that have energy between E and $E + dE$. A careful calculation (see Problem 44), which must treat the potential well as three dimensional, shows that

$$g(E) = \frac{8\sqrt{2}\,\pi m^{3/2}}{h^3} E^{1/2} \qquad \textbf{(40–10)}$$

where m is the mass of the electron. This function is plotted in Fig. 40–25.

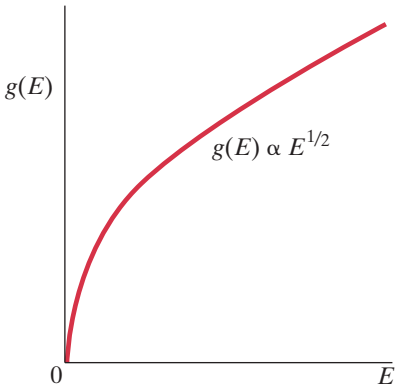


FIGURE 40–25 Density of states $g(E)$ as a function of energy E (Eq. 40–10).

EXERCISE D What are the units for $g(E)$?

EXAMPLE 40–9 | **ESTIMATE** **Electron states in copper.** Estimate the number of states in the range 5.0 to 5.5 eV available to electrons in a 1.0-cm cube of copper metal.

APPROACH Since $g(E)$ is the number of states per unit volume per unit energy interval, the number N of states is approximately (it is approximate because ΔE is not small)

$$N \approx g(E)V\,\Delta E,$$

where the volume $V = 1.0\,\text{cm}^3 = 1.0 \times 10^{-6}\,\text{m}^3$ and $\Delta E = 0.50\,\text{eV}$.

SOLUTION We evaluate $g(E)$ at 5.25 eV, and find (Eq. 40–10):

$$N \approx g(E)V\,\Delta E = \frac{8\sqrt{2}\pi(9.1 \times 10^{-31}\,\text{kg})^{\frac{3}{2}}}{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})^3}\sqrt{(5.25\,\text{eV})(1.6 \times 10^{-19}\,\text{J/eV})}$$
$$\times (1.0 \times 10^{-6}\,\text{m}^3)(0.50\,\text{eV})(1.6 \times 10^{-19}\,\text{J/eV})$$
$$\approx 8 \times 10^{21}$$

states in $1.0\,\text{cm}^3$. Note that the type of metal did not enter the calculation.

Equation 40–10 gives us the density of states. Now we must ask: How are the states available to an electron gas actually populated? Let us first consider the situation at absolute zero, $T = 0\,\text{K}$. For a classical ideal gas, all the particles would be in the lowest state, with zero kinetic energy $(= \frac{3}{2}kT = 0)$—recall Eq. 18–4. But the situation is vastly different for an electron gas because electrons obey the exclusion principle. Electrons do not obey classical statistics but rather a quantum statistics called **Fermi–Dirac statistics**[†] that takes into account the exclusion principle. All particles that have spin $\frac{1}{2}$ (or other half-integral spin: $\frac{3}{2}, \frac{5}{2}$, etc.), such as electrons, protons, and neutrons, obey Fermi–Dirac statistics and are referred to as **fermions**,[‡] as we saw in Section 39–4. The electron gas in a metal is often called a **Fermi gas**. According to the exclusion principle, no two electrons in the metal can have the same set of quantum numbers. Therefore, in each of the energy states available for the electrons in our "gas," there can be at most two electrons: one with spin up $(m_s = +\frac{1}{2})$ and one with spin down $(m_s = -\frac{1}{2})$. (This factor of 2 has already been included in Eq. 40–10.) Thus, at $T = 0\,\text{K}$, the possible energy levels will be filled, two electrons each, up to a maximum level called the **Fermi level**. This is shown in Fig. 40–26, where the vertical axis is labeled $n_o(E)$ for "density of occupied states." The energy of the state at the Fermi level is called the **Fermi energy**, E_F. To determine E_F, we integrate Eq. 40–10 from $E = 0$ to $E = E_F$ (all states up to E_F are filled at $T = 0\,\text{K}$):

$$\frac{N}{V} = \int_0^{E_F} g(E)\,dE, \qquad \textbf{(40–11)}$$

where N/V is the number of conduction electrons per unit volume in the metal. Then, solving for E_F, the result (see Example 40–10 below) is

$$E_F = \frac{h^2}{8m}\left(\frac{3}{\pi}\frac{N}{V}\right)^{\frac{2}{3}}. \qquad \textbf{(40–12)}$$

The average energy in this distribution (see Problem 38) is

$$\overline{E} = \tfrac{3}{5}E_F. \qquad \textbf{(40–13)}$$

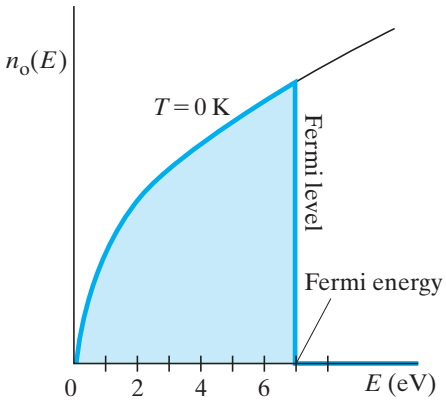


FIGURE 40–26 At $T = 0\,\text{K}$, all states up to energy E_F, called the Fermi energy, are filled. (Shown here for copper.)

[†]Developed independently by Enrico Fermi (Figs. 41–8, 38–2, 37–10) in early 1926 and by P. A. M. Dirac a few months later.

[‡]Particles with integer spin (0, 1, 2, etc.), such as the photon, obey *Bose–Einstein* statistics and are called *bosons*, as mentioned in Section 39–4.

PSE5_CH40_v1.7_23-Jun-20

For copper, $E_F = 7.0\,\text{eV}$ (see Example 40–10) and $\overline{E} = 4.2\,\text{eV}$. This is very much greater than the energy of thermal motion at room temperature $\left(\frac{3}{2}kT \approx 0.04\,\text{eV}\right)$. Clearly, all motion does not stop at absolute zero.

At $T = 0$, all states with energy below E_F are occupied, and all states above E_F are empty. What happens for $T > 0$? We expect that at least some of the electrons will increase in energy due to thermal motion. Classically, the distribution of occupied states would be given by the Boltzmann factor, $e^{-E/kT}$ (see Eqs. 39–16). But for our electron gas, a quantum-mechanical system obeying the exclusion principle, the probability of a given state of energy E being occupied is given by the **Fermi–Dirac probability function** (or **Fermi factor**):

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1}, \qquad \textbf{(40–14)}$$

where E_F is the Fermi energy. This function is plotted in Fig. 40–27 for two temperatures, $T = 0\,\text{K}$ and $T = 1200\,\text{K}$ (just below the melting point of copper).

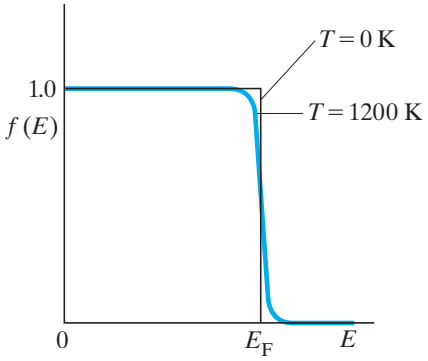


FIGURE 40–27 The Fermi–Dirac probability function for two temperatures, $T = 0\,\text{K}$ (black line) and $T = 1200\,\text{K}$ (blue curve). For $f(E) = 1$, a state with energy E is certainly occupied. For $f(E) = 0.5$, which occurs at $E = E_F$, the state with E_F has a 50% chance of being occupied.

At $T = 0$ (or as T approaches zero) the factor $e^{(E-E_F)/kT}$ in Eq. 40–14 is zero if $E < E_F$ and is ∞ if $E > E_F$. Thus

$$f(E) = \begin{cases} 1 & E < E_F \\ 0 & E > E_F \end{cases} \quad \text{at} \quad T = 0.$$

This is what is plotted in black in Fig. 40–27 and is consistent with Fig. 40–26: all states up to the Fermi level are occupied [probability $f(E) = 1$] and all states above are unoccupied. For $T = 1200\,\text{K}$, the Fermi factor changes only a little, as shown in Fig. 40–27 as the blue curve. Note that at any temperature T, when $E = E_F$, then Eq. 40–14 gives $f(E) = 0.50$, meaning the state at $E = E_F$ has a 50% chance of being occupied by an electron. Alternatively, at $E = E_F$, half the available states are filled and half are empty. To see how $f(E)$ affects the actual distribution of electrons in energy states, we must weight the density of possible states, $g(E)$, by the probability that those states will be occupied, $f(E)$. The product of these two functions gives the **density of occupied states**,

$$n_o(E) = g(E)f(E) = \frac{8\sqrt{2}\,\pi m^{3/2}}{h^3}\frac{E^{1/2}}{e^{(E-E_F)/kT} + 1}. \qquad \textbf{(40–15)}$$

Then $n_o(E)\,dE$ represents the number of electrons per unit volume with energy between E and $E + dE$ in thermal equilibrium at temperature T. This is plotted in Fig. 40–28 for $T = 1200\,\text{K}$, a temperature at which a metal is so hot it would glow. We see that the distribution differs very little from that at $T = 0$. We see also that the changes that do occur are concentrated about the Fermi level.

⚠ CAUTION

At the Fermi level, there are as many filled states as empty ones

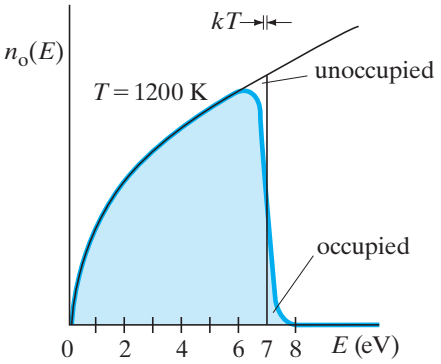


FIGURE 40–28 The density of occupied states for the electron gas in copper. The width kT shown above the graph represents thermal energy at $T = 1200\,\text{K}$.

CAUTION

Average electron energy in metals changes only slightly from $T = 0\,\text{K}$ to $T = 1200\,\text{K}$

A few electrons from slightly below the Fermi level move to energy states slightly above it. The average energy of the electrons increases only very slightly when the temperature is increased from $T = 0\,\text{K}$ to $T = 1200\,\text{K}$. This is very different from the behavior of an ideal gas, for which kinetic energy increases directly with T. Nonetheless, this behavior is readily understood as follows. Energy of thermal motion at $T = 1200\,\text{K}$ is about $\frac{3}{2}kT \approx 0.1\,\text{eV}$. The Fermi level, on the other hand, is on the order of several eV: for copper it is $E_\text{F} \approx 7.0\,\text{eV}$. An electron at $T = 1200\,\text{K}$ may have 7 eV of energy, but it can acquire at most only a few times 0.1 eV of energy by a (thermal) collision with the lattice. Only electrons very near the Fermi level would find vacant states close enough to make such a transition. Essentially none of the electrons could increase in energy by, say, 3 eV, so electrons farther down in the electron gas are unaffected. Only electrons near the top of the energy distribution can be thermally excited to higher states. And their new energy (and speed) is on the average only slightly higher than their old energy. This model of free electrons in a metal as a "gas," though incomplete, provides good explanations for the thermal and electrical conductivity of metals.

EXERCISE E Return to Chapter-Opening Question 2, page 1186, and answer it again now. Try to explain why you may have answered differently the first time.

EXAMPLE 40–10 The Fermi level. For the metal copper, determine (*a*) the Fermi energy, (*b*) the average energy of electrons, and (*c*) the speed of electrons at the Fermi level (this is called the *Fermi speed*).

APPROACH We first derive Eq. 40–12 by combining Eqs. 40–10 and 40–11:

$$\frac{N}{V} = \frac{8\sqrt{2}\,\pi m^{3/2}}{h^3}\int_0^{E_\text{F}} E^{1/2}\,dE = \frac{8\sqrt{2}\,\pi m^{3/2}}{h^3}\frac{2}{3}E_\text{F}^{3/2}$$

$$= \frac{16\sqrt{2}\,\pi m^{3/2}}{3h^3}E_\text{F}^{3/2}.$$

Solving for E_F, we obtain [note: $(16\sqrt{2})^{2/3} = (\sqrt{512})^{2/3} = (512)^{1/3} = 8$]:

$$E_\text{F} = \frac{h^2}{8m}\left(\frac{3}{\pi}\frac{N}{V}\right)^{\frac{2}{3}},$$

and this is Eq. 40–12. We calculated N/V, the number of conduction electrons per unit volume in copper, in Example 25–14 to be $N/V = 8.4 \times 10^{28}\,\text{m}^{-3}$.

SOLUTION (*a*) The Fermi energy for copper is thus

$$E_\text{F} = \frac{(6.63\times10^{-34}\,\text{J}\cdot\text{s})^2}{8(9.1\times10^{-31}\,\text{kg})}\left[\frac{3(8.4\times10^{28}\,\text{m}^{-3})}{\pi}\right]^{\frac{2}{3}}\frac{1}{1.6\times10^{-19}\,\text{J/eV}} = 7.0\,\text{eV}.$$

(*b*) From Eq. 40–13,

$$\overline{E} = \tfrac{3}{5}E_\text{F} = 4.2\,\text{eV}.$$

(*c*) In our model, we have taken $U = 0$ inside the metal (assuming a 3-D infinite potential well, Section 38–8, Eq. 40–10, and Problem 44). Then E is only kinetic energy $= \frac{1}{2}mv^2$. Therefore, at the Fermi level, the Fermi speed is

$$v_\text{F} = \sqrt{\frac{2E_\text{F}}{m}} = \sqrt{\frac{2(7.0\,\text{eV})(1.6\times10^{-19}\,\text{J/eV})}{9.1\times10^{-31}\,\text{kg}}} = 1.6\times10^6\,\text{m/s},$$

a very high speed. The temperature of a classical gas would have to be extremely high to produce an average particle speed this large.

EXAMPLE 40–11 **Incorrect classical calculation.** Let us see what result we get if the electrons are treated as a classical ideal gas. That is, estimate the average kinetic energy of electrons at room temperature using the kinetic theory of gases, Chapter 18.

APPROACH The average kinetic energy of particles in an ideal gas was given in Chapter 18, Eq. 18–4, as

$$\overline{K} = \tfrac{3}{2}kT,$$

where k is Boltzmann's constant and $T \approx 300\,\text{K}$.

SOLUTION The ideal gas model gives

$$\overline{K} = \frac{3}{2}(1.38\times10^{-23}\,\text{J/K})\,(300\,\text{K})\left(\frac{1}{1.6\times10^{-19}\,\text{J/eV}}\right) = 0.039\,\text{eV}.$$

This result is far from the quantum calculation. It is off by a factor of 100: Example 40–10 gave 4.2 eV. The ideal gas model does not work for electrons which obey the exclusion principle. Indeed, we see here how important and powerful the exclusion principle is. Comparing Examples 40–10 and 40–11, we see how profoundly the exclusion principle changes everything.

EXERCISE F Determine the Fermi energy for gold $(\text{density} = 19{,}300\,\text{kg/m}^3)$. (*a*) 5.5 eV, (*b*) 6.2 eV, (*c*) 7.2 eV, (*d*) 8.1 eV, (*e*) 8.4 eV.

The simple model of an electron gas presented here provides good explanations for the electrical and thermal properties of conductors. But it does not explain why some materials are good conductors and others are good insulators. To provide an explanation, our model of electrons inside a metal as moving in a uniform potential well needs to be refined to include the effect of the lattice. Figure 40–29 shows a "periodic" potential that takes into account the attraction of electrons for each atomic ion in the lattice. Here we have taken $U = 0$ for an electron free of the metal; so within the metal, electron energies are less than zero (just as for molecules, or for the H atom in which the ground state has $E = -13.6\,\text{eV}$). The quantity eW_0 represents the minimum energy to remove an electron from the metal, where W_0 is the *work function* (see Section 37–2). The crucial outcome of putting a periodic potential (more easily approximated with narrow square wells) into the Schrödinger equation is that the allowed energy states are divided into *bands*, with energy gaps in between. Only electrons in the highest band, close to the Fermi level, are able to move about freely within the metal crystal. In the next Section we will see physically why there are bands and how they explain the properties of conductors, insulators, and semiconductors.

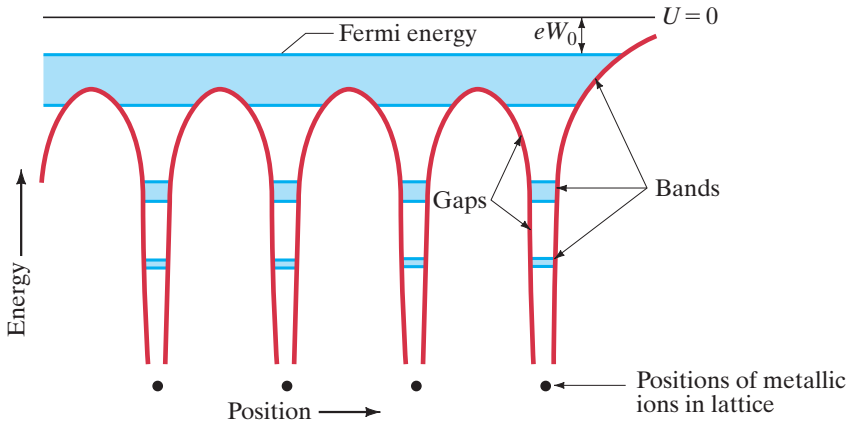


FIGURE 40–29 Potential energy for an electron in a metal crystal, with deep potential wells in the vicinity of each ion in the crystal lattice.

40–7 Band Theory of Solids

We saw in Section 40–1 that when two hydrogen atoms approach each other, the wave functions overlap, and the two 1*s* states (one for each atom) divide into two states of different energy. (As we saw, only one of these states, $S = 0$, has low enough energy to give a bound H_2 molecule.) Figure 40–30a shows this situation for 1*s* and 2*s* states for two atoms: as the two atoms get closer (toward the left in Fig. 40–30a), the 1*s* and 2*s* states split into two levels. If six atoms come together, as in Fig. 40–30b, each of the states splits into six levels. If a large number of atoms come together to form a solid, then each of the original atomic levels becomes a **band** as shown in Fig. 40–30c. The energy levels are so close together in each band that they seem essentially continuous. This is why the spectrum of heated solids (Section 37–1) appears continuous. (See also Fig. 40–15 and its discussion at the start of Section 40–4.)

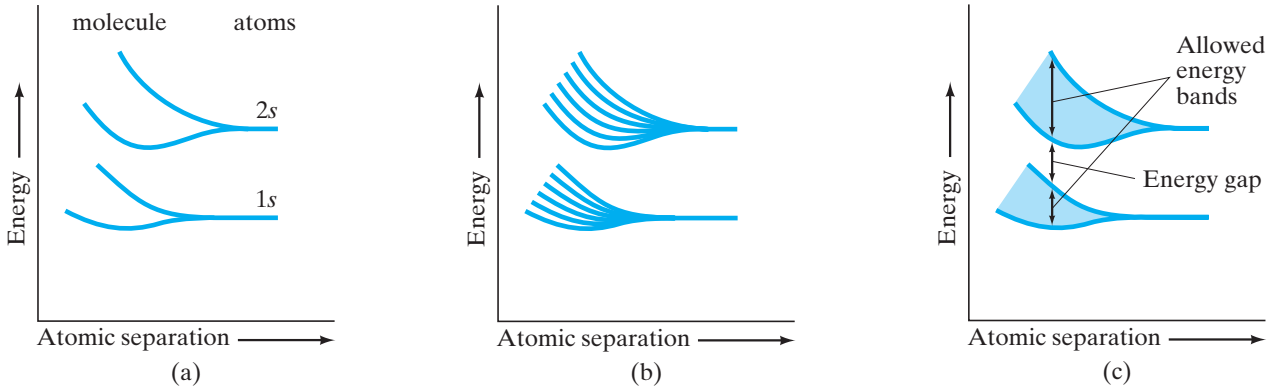
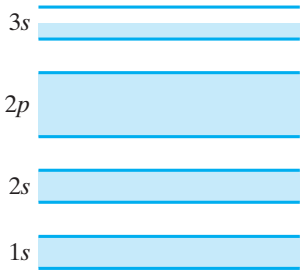


FIGURE 40–30 The splitting of 1*s* and 2*s* atomic energy levels as (a) two atoms approach each other (the atomic separation decreases toward the left on the graph), (b) the same for six atoms, and (c) for many atoms when they come together to form a solid.

FIGURE 40–31 Energy bands for sodium (Na).



The crucial aspect of a good **conductor** is that the highest energy band containing electrons is only partially filled. Consider sodium metal, for example, whose energy bands are shown in Fig. 40–31. The 1*s*, 2*s*, and 2*p* bands are full (just as in a sodium atom) and don't concern us. The 3*s* band, however, is only half full. (Thus, the Fermi level is in the middle of the 3*s* band.) To see why the band is half full, recall that the exclusion principle stipulates that in an atom, only two electrons can be in the 3*s* state, one with spin up and one with spin down. These two spin states have slightly different energy. For a solid consisting of *N* atoms, the 3*s* band will contain 2*N* possible energy states (the 2 because of the 2 spin states). A sodium atom has a single 3*s* electron, so in a sample of sodium metal containing *N* atoms, there are *N* electrons in the 3*s* band, and *N* unoccupied states (total = 2*N*). When a potential difference is applied across the metal, electrons can respond by accelerating and increasing their energy, since there are plenty of unoccupied states of slightly higher energy available. Hence, a current flows readily and sodium is a good conductor. The characteristic of all good conductors is that the highest energy band is only partially filled, or two bands overlap so that unoccupied states are available. An example of the latter is magnesium, which has two 3*s* electrons, so its 3*s* band is filled. But the unfilled 3*p* band overlaps the 3*s* band in energy, so there are lots of available states for the electrons to move into. Thus magnesium, too, is a good conductor.

In a material that is a good **insulator**, on the other hand, the highest band containing electrons, called the **valence band**, is completely filled. The next highest energy band, called the **conduction band**, is separated from the valence band by a "forbidden" **energy gap** (or **band gap**), E_g, of typically 5 to 10 eV. So at room temperature (300 K), where thermal energies (that is, average kinetic energy—see

PSE5_CH40_v1.7_23-Jun-20

Chapter 18) are on the order of $\frac{3}{2}kT \approx 0.04\,\text{eV}$, almost no electrons can acquire the 5 eV needed to reach the conduction band. When a potential difference is applied across the material, no available states are accessible to the electrons, and no current flows. Hence, the material is a good insulator.

Figure 40–32 compares the relevant energy bands (a) for conductors, (b) for insulators, and also (c) for the important class of materials known as **semiconductors**. The bands for a pure (or **intrinsic**) semiconductor, such as silicon or germanium, are like those for an insulator, except that the unfilled conduction band is separated from the filled valence band by a much smaller energy gap, E_g, which for silicon is $E_g = 1.12\,\text{eV}$. At room temperature, electrons are moving about with varying amounts of kinetic energy $\left(K = \frac{3}{2}kT\right)$, according to kinetic theory, Chapter 18. A few electrons can acquire enough thermal energy to reach the conduction band, and so a very small current may flow when a voltage is applied. At higher temperatures, more electrons have enough energy to jump the gap (top end of thermal distribution—see Fig. 18–3). Often this effect can more than offset the effects of more frequent collisions due to increased disorder at higher temperature, so the resistivity of semiconductors can *decrease* with increasing temperature (see Table 25–1). But this is not the whole story of semiconductor conduction. When a potential difference is applied to a semiconductor, the few electrons in the conduction band move toward the positive electrode. Electrons in the valence band try to do the same thing, and a few can because there are a small number of unoccupied states which were left empty by the electrons reaching the conduction band. Such unfilled electron states are called **holes**. Each electron in the valence band that fills a hole in this way as it moves toward the positive electrode leaves behind its own hole, so the holes migrate toward the negative electrode. As the electrons tend to accumulate at one side of the material, the holes tend to accumulate on the opposite side. We will look at this phenomenon in more detail in the next Section.

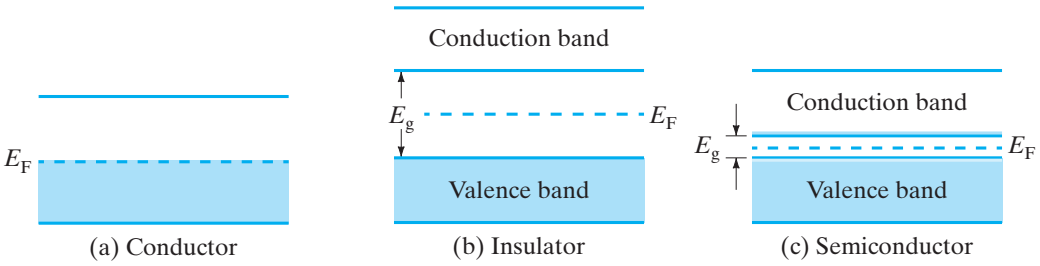


FIGURE 40–32 Energy bands for (a) a conductor, (b) an insulator, which has a large energy gap E_g, and (c) a semiconductor, which has a small energy gap E_g. Shading represents occupied states. Pale shading in (c) represents electrons that can pass from the top of the valence band to the bottom of the conduction band due to thermal agitation at room temperature (exaggerated). The Fermi level E_F is shown for each (see Example 40–13.)

EXAMPLE 40–12 **Calculating the energy gap.** It is found that the conductivity of a certain semiconductor increases when light of wavelength 345 nm or shorter strikes it, suggesting that electrons are being promoted from the valence band to the conduction band. What is the energy gap, E_g, for this semiconductor?

APPROACH The longest wavelength (lowest energy) photon to cause an increase in conductivity has $\lambda = 345\,\text{nm}$, and its energy $(= hf)$ equals the energy gap.

SOLUTION The gap energy equals the energy of a $\lambda = 345\text{-nm}$ photon:

$$E_g = hf = \frac{hc}{\lambda} = \frac{(6.63\times10^{-34}\,\text{J}\cdot\text{s})(3.00\times10^{8}\,\text{m/s})}{(345\times10^{-9}\,\text{m})(1.60\times10^{-19}\,\text{J/eV})} = 3.6\,\text{eV}.$$

We can use the Fermi–Dirac statistics that we discussed in the previous Section also here, to make calculations on semiconductors, as the next Example shows.

EXAMPLE 40–13 **ESTIMATE** **Free electrons in semiconductors and insulators.** Use the Fermi–Dirac probability function, Eq. 40–14, to estimate the order of magnitude of the numbers of free electrons in the conduction band of a solid containing 10^{21} atoms, assuming the solid is at room temperature ($T = 300\,\text{K}$) and is (*a*) a semiconductor with $E_g \approx 1.1\,\text{eV}$, (*b*) an insulator with $E_g \approx 5\,\text{eV}$. Compare to a conductor.

CAUTION
For semiconductors and insulators, the Fermi level is between the valence and conduction bands

APPROACH At $T = 0$, all states above the Fermi energy E_F are empty, and all those below are full. So for semiconductors and insulators we can take E_F to be about midway between the valence and conduction bands, Fig. 40–33, and it does not change significantly as we go to room temperature. We can thus use Eq. 40–14 to find the fraction of electrons in the conduction band at room temperature for the two cases.

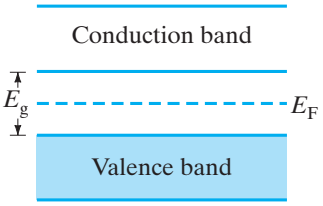


FIGURE 40–33 Example 40–13: The Fermi energy is midway between the valence band and the conduction band.

SOLUTION (*a*) For the semiconductor, the gap $E_g \approx 1.1\,\text{eV}$, so $E - E_F \approx 0.55\,\text{eV}$ for the lowest states in the conduction band. Since at room temperature we have $kT \approx 0.026\,\text{eV}$, then $(E - E_F)/kT \approx 0.55\,\text{eV}/0.026\,\text{eV} \approx 21$ and

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1} \approx \frac{1}{e^{21}} \approx 10^{-9}.$$

Thus about 1 atom in 10^9 can contribute an electron to the conductivity.
(*b*) For the insulator with $E - E_F \approx 5.0\,\text{eV} - \frac{1}{2}(5.0\,\text{eV}) = 2.5\,\text{eV}$, we get

$$f(E) \approx \frac{1}{e^{2.5/0.026} + 1} \approx \frac{1}{e^{96}} \approx 10^{-42}.$$

Thus in an ordinary sample containing 10^{21} atoms, there would be no free electrons in an insulator $(10^{21} \times 10^{-42} = 10^{-21})$, about 10^{12} $(10^{21} \times 10^{-9})$ free electrons in a semiconductor, and about 10^{21} free electrons in a good conductor.

PHYSICS APPLIED
Transparency

CONCEPTUAL EXAMPLE 40–14 **Which is transparent?** The energy gap for silicon is 1.12 eV at room temperature, whereas that of zinc sulfide (ZnS) is 3.6 eV. Which one of these is opaque to visible light, and which is transparent?

RESPONSE Visible light photons span energies from roughly 1.8 eV to 3.1 eV ($E = hf = hc/\lambda$ where $\lambda = 400\,\text{nm}$ to 700 nm and $1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J}$). Light is absorbed by the electrons in a material. Silicon's energy gap is small enough to absorb these photons, thus bumping electrons well up into the conduction band, so silicon is opaque. On the other hand, zinc sulfide's energy gap is so large that no visible light photons would be absorbed; they would pass right through the material which would thus be transparent.

40–8 Semiconductors and Doping

Nearly all electronic devices today use semiconductors—mainly silicon (Si), although the first transistor (1948) was made with germanium (Ge). An atom of silicon has four outer electrons (group IV of the Periodic Table) that act to hold the atoms in the regular lattice structure of the crystal, shown schematically in

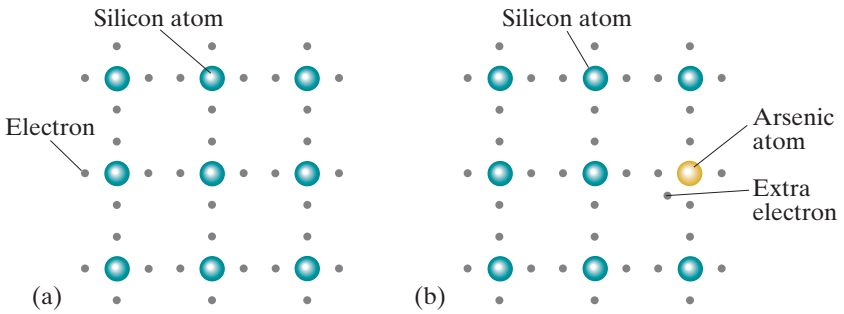


FIGURE 40–34 Two-dimensional representation of a silicon crystal. (a) Four (outer) electrons surround each silicon atom. (b) Silicon crystal doped with a small percentage of arsenic atoms: the extra electron doesn't fit into the crystal lattice and so is free to move about. This is an *n*-type semiconductor.

Fig. 40–34a. Silicon acquires properties useful for electronics when a tiny amount of impurity is introduced into the crystal structure (perhaps 1 part in 10^6 or 10^7). This is called **doping** the semiconductor. Two kinds of doped semiconductor can be made, depending on the type of impurity used. The impurity can be an element whose atoms have five outer electrons (group V in the Periodic Table), such as arsenic. Then we have the situation shown in Fig. 40–34b, with a few arsenic atoms holding positions in the crystal lattice where normally silicon atoms are. Only four of arsenic's electrons fit into the bonding structure. The fifth does not fit in and can move relatively freely, somewhat like the electrons in a conductor. Because of this small number of extra electrons, a doped semiconductor becomes slightly conducting. The density of conduction electrons in an **intrinsic** (= undoped) semiconductor at room temperature is very low, usually less than 1 per 10^9 atoms. With an impurity concentration of perhaps 1 in 10^6 when doped, the conductivity will be much higher and it can be controlled with great precision. An arsenic-doped silicon crystal is an ***n*-type semiconductor** because *negative* charges (electrons) carry the electric current.

CAUTION

Density of free electrons in intrinsic semiconductor is about 1 in 10^9 atoms; if doped, about 1 in 10^6

In a ***p*-type semiconductor**, a small percentage of semiconductor atoms are replaced by atoms with three outer electrons (group III in the Periodic Table), such as boron. As shown in Fig. 40–35a, there is a **hole** in the lattice structure near a boron atom because it has only three outer electrons. An electron from a nearby silicon atom can jump into this hole and fill it. But this leaves a hole where that electron had previously been, Fig. 40–35b. The vast majority of atoms are silicon, so holes are almost always next to a silicon atom. Since silicon atoms require four outer electrons to be neutral, this means there is a net positive charge at the hole. Whenever an electron moves to fill a hole, the positive hole is then at the previous position of that electron. Another electron can then fill this hole, and the hole thus moves to a new location; and so on. This type of semiconductor is called ***p*-type** because it is the positive holes that carry the electric current.[†] Note, however, that both *p-type* and *n-type* semiconductors have *no net charge* on them.

CAUTION

p-type semiconductors act as though + charges move—but electrons actually do the moving

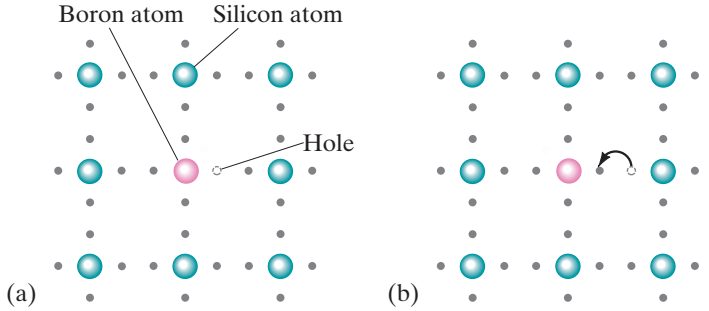


FIGURE 40–35 A *p*-type semiconductor, boron-doped silicon. (a) Boron has only three outer electrons, so there is an empty spot, or *hole* in the structure. (b) An electron from a silicon atom can jump into the hole and fill it. As a result, the hole moves to a new location (to the right in this diagram), to where the electron used to be.

[†]Each electron that fills a hole moves a very short distance (~1 atom < 1 nm) whereas holes move much larger distances and so are the real carriers of the current. We can tell the current is carried by positive charges (holes) by using the Hall effect, Section 27–8.

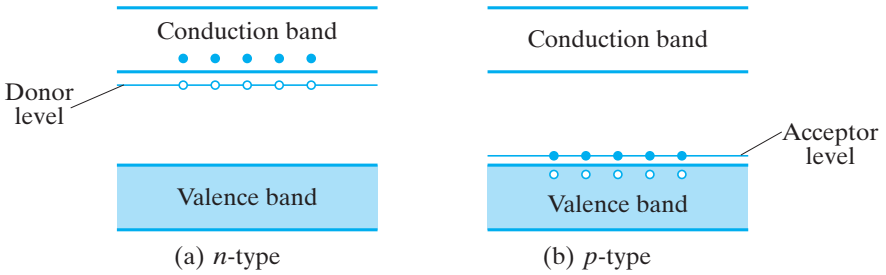
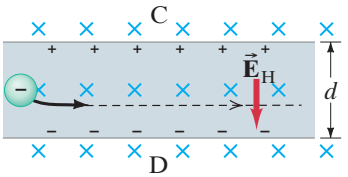


FIGURE 40–36 Impurity energy levels in doped semiconductors.

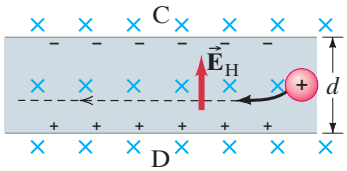
According to band theory (Section 40–7), in a doped semiconductor the impurity provides additional energy states between the bands as shown in Fig. 40–36. In an *n*-type semiconductor, the impurity energy level lies just below the conduction band, Fig. 40–36a. Electrons in this energy level need only about 0.05 eV in Si to reach the conduction band which is on the order of the thermal energy, $\frac{3}{2}kT$ (≈ 0.04 eV at 300 K). At room temperature, the tiny % of electrons in this donor level (roughly ≈ 1 in 10^6) can readily make the transition upward to the conduction band. This energy level can thus supply electrons to the conduction band, so it is called a **donor** level. In *p*-type semiconductors, the impurity energy level is just above the valence band (Fig. 40–36b). It is called an **acceptor** level because electrons from the valence band can jump into it with only average thermal energy. Positive holes are left behind in the valence band, and as other electrons move into these holes, the holes move as discussed earlier.

EXERCISE G Which of the following impurity atoms in silicon would produce a *p*-type semiconductor? (*a*) Ge; (*b*) Ne; (*c*) Al; (*d*) As; (*e*) Ga; (*f*) none of the above.

FIGURE 27–33 (Repeated.) The Hall effect. (a) Negative charges moving to the right as the current. (b) The same current, but as positive charges moving to the left.



(a)



(b)

CONCEPTUAL EXAMPLE 40–15 **Determining charge of conductors.** How can we determine if a *p*-type semiconductor has a current that is really due to the motion of holes? Or, is this just a convenient model?

RESPONSE Recall from Section 27–8 that the Hall effect can be used to distinguish the sign of the charges involved in a current. When placed in a magnetic field, the current in a particular direction can result in a voltage perpendicular to that current due to the magnetic force on the moving charges (Fig. 27–33, repeated here). The direction of this *Hall voltage* depends on the sign of the charges carrying the current. In this way, it has been demonstrated that it really is moving holes that are responsible for the current in a *p*-type semiconductor.

40–9 Semiconductor Diodes, LEDs, OLEDs

Semiconductor diodes and transistors are essential components of modern electronic devices. The miniaturization achieved today allows many billions of diodes, transistors, resistors, etc., to be fabricated (adding doping atoms) on a single *chip* less than a millimeter on a side.

At the interface between an *n*-type and a *p*-type semiconductor, a ***pn* junction diode** is formed. Separately, the two semiconductors are electrically neutral. But near the junction, a few electrons diffuse from the *n*-type into the *p*-type semiconductor, where they fill a few of the holes. The *n*-type is left with a positive charge, and the *p*-type acquires a net negative charge. Thus an "intrinsic" *potential difference* is established, with the *n* side positive relative to the *p* side, and this prevents further diffusion of electrons. The "junction" is actually a very thin layer between the charged *n* and *p* semiconductors where all holes are filled with electrons. This junction region is called the **depletion layer** (depleted of electrons and holes).†

†One way to form the *pn* boundary at the **nanometer** thickness on chips is to implant (or diffuse) *n*-type donor atoms into the surface of a *p*-type semiconductor, converting a layer of the *p*-type semiconductor into *n*-type.

If a battery is connected to a diode with the positive terminal to the *p* side and the negative terminal to the *n* side as in Fig. 40–37a, the externally applied voltage opposes the intrinsic potential difference and the diode is said to be **forward biased**. If the voltage is great enough, about 0.6 V for Si at room temperature, it overcomes that intrinsic potential difference and a large current can flow. The positive holes in the *p*-type semiconductor are repelled by the positive terminal of the battery, and the electrons in the *n*-type are repelled by the negative terminal of the battery. The holes and electrons meet at the junction, and the electrons cross over and fill the holes. A current is flowing. The positive terminal of the battery is continually pulling electrons off the *p* end, forming new holes, and electrons are being supplied by the negative terminal at the *n* end.

When the diode is **reverse biased**, as in Fig. 40–37b, the holes in the *p* end are attracted to the battery's negative terminal and the electrons in the *n* end are attracted to the positive terminal. Almost no current carriers meet near the junction and, ideally, no current flows.

A graph of current versus voltage for a typical diode is shown in Fig. 40–38. A forward bias greater than 0.6 V allows a large current to flow. In reverse bias, a real diode allows a small amount of reverse current to flow; for most practical purposes, it is negligible.†

The symbol for a diode is



[diode]

where the arrow represents the direction conventional (+) current flows readily.

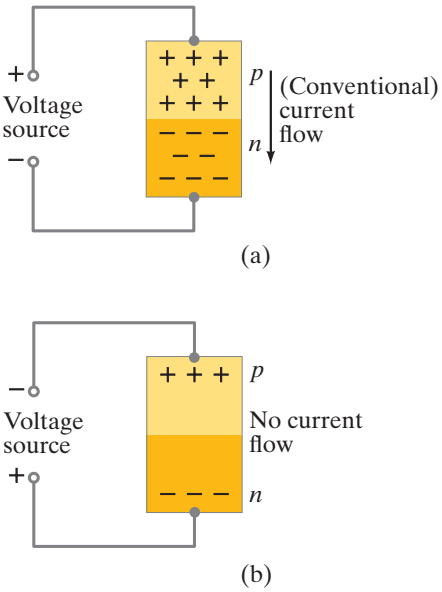


FIGURE 40–37 Schematic diagram showing how a semiconductor diode operates. Current flows when the voltage is connected in forward bias, as in (a), but not when connected in reverse bias, as in (b).

EXAMPLE 40–16 **A diode.** The diode whose current–voltage characteristics are shown in Fig. 40–38 is connected in series with a 4.0-V battery in forward bias and a resistor. If a current of 15 mA is to pass through the diode, what resistance must the resistor have?

APPROACH We use Fig. 40–38, where we see that the voltage drop across the diode is about 0.7 V when the current is 15 mA. Then we use simple circuit analysis and Ohm's law (Chapters 25 and 26).

SOLUTION The voltage drop across the resistor is $4.0\,\text{V} - 0.7\,\text{V} = 3.3\,\text{V}$, so $R = V/I = (3.3\,\text{V})/(1.5 \times 10^{-2}\,\text{A}) = 220\,\Omega$.

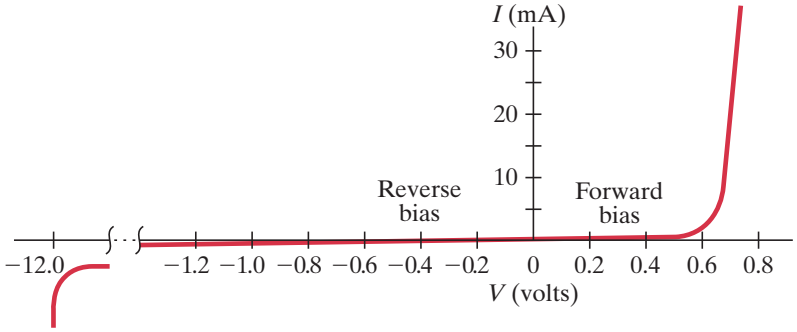


FIGURE 40–38 Current through a silicon *pn* diode as a function of applied voltage.

A diode is called a **nonlinear device** because the current is not proportional to the voltage. That is, a graph of current versus voltage (Fig. 40–38) is not a straight line, as it is for a resistor (which ideally *is* linear).

Zener Diode

If the voltage across a diode connected in reverse bias is increased greatly, breakdown occurs. The electric field across the junction becomes so large that ionization of atoms results. The electrons thus pulled off their atoms contribute to a larger and larger current as breakdown continues. The voltage remains constant over a wide range of currents. This is shown on the far left in Fig. 40–38.

†At room temperature, the reverse current is a few pA in Si; but it increases rapidly with temperature, and may render a diode ineffective above 200°C.

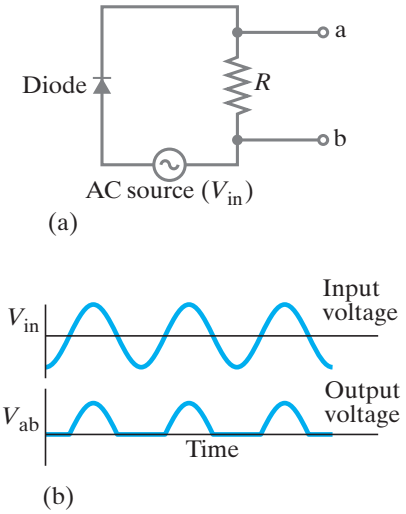


FIGURE 40–39 (a) A simple (half-wave) rectifier circuit using a semiconductor diode. (b) AC source input voltage, and output voltage across R, as functions of time.

This property of diodes can be used to accurately regulate a voltage supply. A diode designed for this purpose is called a **zener diode**. When placed across the output of an unregulated power supply, a zener diode can maintain the voltage at its own breakdown voltage as long as the supply voltage is always above this point. Zener diodes can be obtained corresponding to voltages of a few volts to hundreds of volts.

Rectifiers

Because a *pn* junction diode allows current to flow only in one direction (as long as the voltage is not too high), it can serve as a **rectifier**—to change ac into dc. A simple rectifier circuit is shown in Fig. 40–39a. The ac source applies a voltage across the diode alternately positive and negative. Only during half of each cycle will a current pass through the diode; only then is there a current through the resistor R. Hence, a graph of the voltage V_{ab} across R as a function of time looks like the output voltage shown in Fig. 40–39b. This **half-wave rectification** is not exactly dc, but it is unidirectional. More useful is a **full-wave rectifier** circuit, which uses two diodes (or sometimes four) as shown in Fig. 40–40a. At any given instant, either one diode or the other will conduct current to the right. Therefore, the output across the load resistor R will be as shown in Fig. 40–40b. Actually this is the voltage if the capacitor C were not in the circuit. The capacitor tends to store charge and, if the time constant RC is sufficiently long, helps to smooth out the current as shown in Fig. 40–40c. (The variation in output shown in Fig. 40–40c is called **ripple voltage**.)

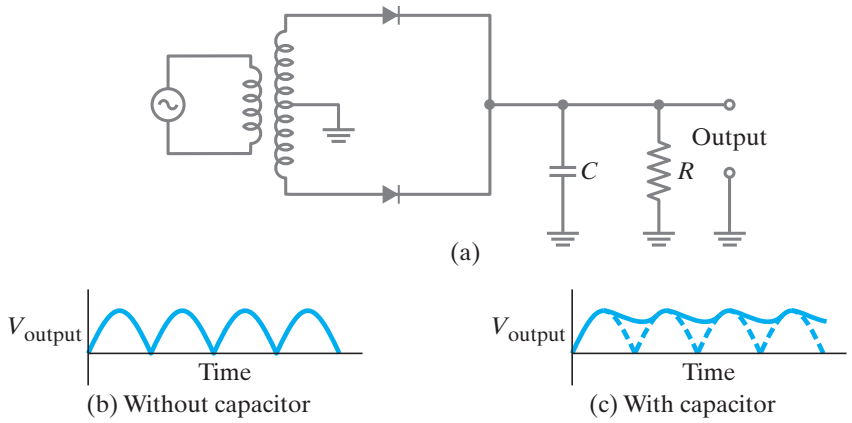


FIGURE 40–40 (a) Full-wave rectifier circuit (including a transformer so the magnitude of the voltage can be changed). (b) Output voltage in the absence of capacitor C. (c) Output voltage with the capacitor in the circuit.

Rectifier circuits are important because most line voltage in buildings is ac, and most electronic devices require a dc voltage for their operation. Hence, diodes are found in nearly all electronic devices including radios, TV sets, computers, and chargers for cell phones and other devices.

Photovoltaic Cells

Solar cells, also called **photovoltaic cells**, are rather heavily doped *pn* junction diodes used to convert sunlight into electric energy. Photons are absorbed, creating electron–hole pairs if the photon energy is greater than the band gap energy, E_g (see Figs. 40–32c and 40–36). That is, the absorbed photon excites an electron from the valence band up to the conduction band, leaving behind a hole in the valence band. The created electrons and holes produce a current that, when connected to an external circuit, becomes a source of emf and power. A typical silicon *pn* junction may produce about 0.6 V. Many are connected in series to produce a higher voltage. Such series strings are connected in parallel within a **photovoltaic panel**. Research includes experimenting with combinations of semiconductors. A good photovoltaic panel can have an output of perhaps 50 W/m^2, averaged over

PSE5_CH40_v1.7_23-Jun-20

day and night, sunny and cloudy. The world's total electricity demand is on the order of 10^{12} W, which could be met with solar cells covering an area of only about 200 km × 200 km of Earth's surface.[†]

Photodiodes (Section 37–2) and **semiconductor particle detectors** (Section 41–11) operate similarly.

LEDs

PHYSICS APPLIED
LEDs and applications
Car safety (brakes)

A **light-emitting diode (LED)** is sort of the reverse of a photovoltaic cell. When a *pn* junction is forward biased, a current begins to flow. Electrons cross from the *n*-region into the *p*-region, recombining with holes, and a photon can be emitted with an energy about equal to the band gap energy, E_g. This does not work well with silicon diodes.[‡] But high light-emission is achieved with **compound semiconductors**, typically involving a group III and a group V element such as gallium and arsenic (= gallium arsenide = GaAs). Remarkably, GaAs has a crystal structure very similar to Si. See Fig. 40–41. For doping of GaAs, group VI atoms (like Se) can serve as donors, and group II atoms (valence +2, such as Zn) as acceptors. The energy gap for GaAs is $E_g = 1.42$ eV, corresponding to near-infrared photons with wavelength 870 nm (almost visible). Such infrared LEDs are suitable for use in remote-control devices for TVs, DVD players, stereos, car door locks, and so on.

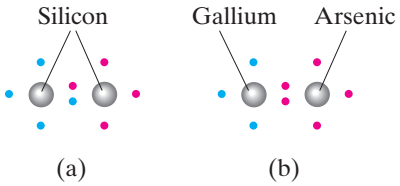


FIGURE 40–41 (a) Two Si atoms forming the covalent bond showing the electrons in different colors for each of the two separate atoms. (In Fig. 40–34a we showed each atom separately to emphasize the four outer electrons in each.) (b) A gallium–arsenic pair, also covalently bonded.

The first visible-light LED, developed in the early 1960s, was made of a semiconductor compound of gallium, arsenic, and phosphorus (= GaAsP) which emitted red light. The red LED soon found use as the familiar indicator lights (on–off) on electronic devices, and as the bright red read-out on calculators and digital clocks (brighter than the dimmer LCD readouts which rely on reflected light and cannot be seen in the dark unless there is built-in backlighting). Further development led to LEDs with higher E_g and shorter wavelengths: first yellow, then finally in 1995, blue (InGaN). A blue LED was important because it gave the possibility of a **white-light LED**. White light can be approximated by LEDs in two ways: (1) using a red, a yellow–green, and a blue LED; (2) using a blue LED with coatings of "powders" or "phosphors" that are fluorescent (Section 39–8). For the latter, the high-energy blue LED photons are themselves emitted, plus they can excite the various phosphors to excited states which decay in two or more steps, emitting light of lower energy and longer wavelengths. Figure 40–42 shows typical spectra of both types. The first type, shown in Fig. 40–42a, can contain 4 or 5 LED types to better reproduce the full visible spectrum. Today the second type with phosphors (Fig. 40–42b) produces such good white color that it is used for household bulbs and even in museums.[§]

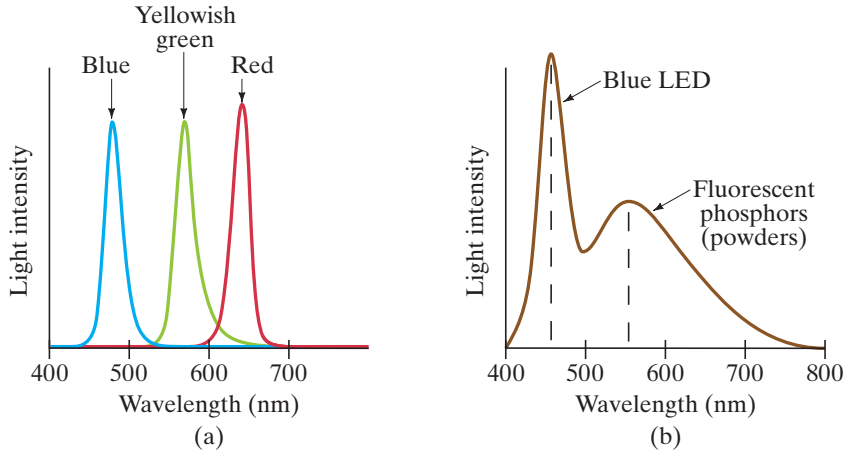


FIGURE 40–42 (a) A combination of three LEDs of three different colors gives a sort of white color, but there are large wavelength gaps, so some colors would not be reflected and would appear black. (b) A blue LED with fluorescent phosphors or powders gives a better approximation of white light. (Thanks to M. Vannoni and G. Molesini for (b).)

[†]Electricity makes up about 5% of total global energy use.

[‡]Electron-hole recombination in silicon results mostly in heat, as lattice vibrations called **phonons**.

[§]The renowned frescoes by Michelangelo in the Sistine Chapel (Rome) are now lit only by LED lighting.



FIGURE 40–43 LED flashlights. Note the tiny LEDs, each maybe $\frac{1}{2}$ cm in diameter.

LED "bulbs" are also used in flashlights (Fig. 40–43), street lighting, traffic signals, car brake lights, billboards, backlighting for LCD screens, and large display screens at stadiums.

LED lights, sometimes called **solid-state** lighting, are longer-lived (20,000 to 50,000 hours vs. 1000–2000 for ordinary bulbs), more efficient (on the order of 10 times better), and rugged. A small town in Italy, Torraca, was the first to have all its street lighting be LED (2007). LEDs can be as small as 1 or 2 mm wide, and are individual units with wires connected directly to them. They can be used for large TV screens in stadiums, but a home TV is not yet available and would require much smaller LED size, meaning fabrication of many on a crystalline semiconductor, and the pixels would be addressed as discussed in Section 23–10 for LCD screens. [When you see a home TV advertised as LED, it is an LCD screen (Section 34–8) using LEDs as the backlight.]

*Pulse Oximeter

A **pulse oximeter** uses two LEDs to measure the % oxygen (O_2) saturation in your blood. One LED is red, 660 nm, and the other IR (900–940 nm). The LED beams pass through a finger (Fig. 40–44) or earlobe and are detected by a photodiode. Oxygenated red blood cells absorb less red and more infrared light than deoxygenated cells. A ratio of absorbed light (red/IR) of 0.5 corresponds to nearly 100% O_2 saturation; a ratio of 1.0 is about 85% and 2.0 corresponds to about 50% (bad). The LED measures during complete pulses, including blood surges, and the device can also count your heartbeat rate.

FIGURE 40–44 A pulse oximeter.



**pn* Diode Lasers

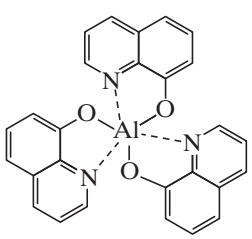
Diode lasers, using a *pn*-junction in forward bias like an LED, are the most compact of lasers and are very common: they read CDs and DVDs and are used as pointers and in laser printers. They emit photons like an LED but, like all lasers (Section 39–9), need to have an *inverted population* of states for the lasing frequency. This is achieved by applying a high forward-bias voltage. The large current brings many electrons into the conduction band at the junction layer, and holes into the valence band, and before the electrons have time to combine with holes, they form an inverted population. When one electron drops down into a hole and emits a photon, that photon stimulates other electrons to drop down as well, *in phase*, creating coherent laser light. Opposite ends of the crystal are made parallel and very smooth so they act as the mirrors needed for lasing, as shown in our laser diagram, Fig. 39–19.

*OLED (Organic LED)

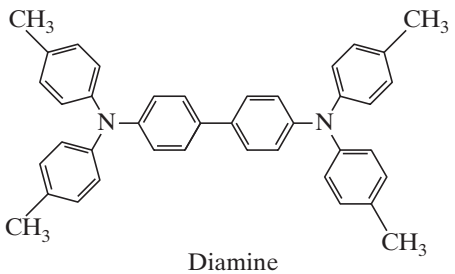
Many organic compounds have semiconductor properties. Useful ones can have mobile electrons and holes. A practical organic **electroluminescent (EL)** device, an **organic light-emitting diode (OLED)** was first described in the late 1980s.

Organic compounds contain carbon (C), hydrogen (H), often nitrogen and oxygen, and sometimes other atoms. We usually think of them as coming from life—plants and animals. They are also found in petroleum, and some can be synthesized in the lab. Organic compounds can be complex, and often contain the familiar hexagonal "benzene ring" with C atoms at all (or most) of the six corners. The two organic compounds shown in Fig. 40–45 were used as *n*-type and *p*-type layers in the earliest useful OLED. **Polymers**, long organic molecules with repeating structural units, can also be used for an OLED.

FIGURE 40–45 These two organic molecules were used in the first OLEDs (1987). The hexagons have carbon at each corner, and an attached hydrogen, unless otherwise noted.



Alq_3



Diamine

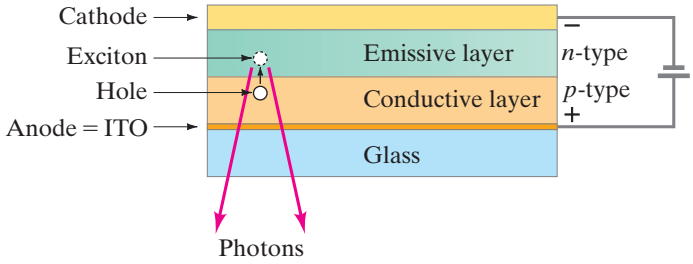


FIGURE 40–46 An OLED with two organic layers. Hole–electron recombination into an exciton (dashed circle) occurs in the emissive layer, followed by photon emission. Photons emitted in the wrong direction (upward in the diagram) reduce efficiency.

The simplest OLED consists of two organic layers, the **emissive layer** and the **conductive layer**, each 20 to 50 nm thick, sandwiched between two electrodes, Fig. 40–46. The anode is typically transparent, to let the light out. It can be made of a very thin layer of indium–tin oxide (**ITO**), which is transparent and conductive, coated on a glass slab. The cathode is often metallic, but could also be made of transparent material.

OLEDs can be smaller and thinner than ordinary inorganic LEDs. They can be more easily constructed as a unit for a screen display than for inorganic LEDs. Their use as screens on cell phones, cameras, and TVs produces greater contrast, and they need less power (important for battery life of portable devices) than LCD screens. Why? They need no backlight (like LCDs) because they emit the light themselves. OLEDs can be fabricated as a matrix, usually active matrix (**AMOLED**), using the same type of addressing described in Section 23–10 for LCDs. OLED displays are much thinner than LCDs and retain brightness at larger viewing angles. They can even be fabricated on curved or flexible substrates—try the windshield of your car (Fig. 40–47). The array may be RGBG (similar to a Bayer mosaic, Fig. 33–18) or RGBW where W = white is meant to give greater brightness. The subpixels can also be stacked, one above the other (similar to the Foveon, Fig. 33–19).

FIGURE 40–47 Head up displays on curved windshields can use curved OLEDs to show, for example, your speed without having to look down at the speedometer.



*OLED Functioning (advanced)

According to band theory, when a voltage is applied (≈ 2 to 5 V), electrons are attracted or "injected" (engineering term) into energy states of the **lowest unoccupied molecular orbitals (LUMO)** of the emissive layer. At the same time, electrons are withdrawn from the **highest occupied molecular orbitals (HOMO)** of the conductive layer—which is equivalent to **holes** being "injected" into the conductive layer. The LUMO and HOMO energy levels are analogous to the conduction and valence bands of inorganic silicon diodes (Fig. 40–36). Holes travel in the HOMO, electrons in the LUMO. ("Orbital" is a chemistry word for the states occupied by the electrons in a molecule.)

When electrons and holes meet near the junction (Fig. 40–46), they can form a sort of bound state (like in the hydrogen atom) known as an **exciton**. An exciton has a small binding energy (0.1 to 1 eV), and a very short lifetime on the order of nanoseconds. When an exciton "decays" (the negative electron and positive hole combine), a photon is emitted. These photons are the useful output.

The energy hf of the photon, and its frequency corresponding to the color, depends on the energy structure of the exciton. The energy gap, LUMO–HOMO, sets an upper limit on hf, but the vibrational energy levels of the molecules reduce that by varying amounts, as does the binding energy of the exciton. The spectrum has a peak, like those in Fig. 40–42a, but is wider, 100–200 nm at half maximum. The organic molecules are chosen so that the photons have frequencies in the color range desired, say for a display subpixel: bluish (B), greenish (G), or red (R).

The conductive layer is also called the **hole transport layer (HTL)**, which name expresses its purpose. The emissive layer, on the other hand (Fig. 40–46), serves two purposes: (1) it serves to transport electrons toward the junction, and (2) it is in this layer (near the junction) that holes meet electrons to form excitons and then combine and emit light. These two functions can be divided in a more sophisticated OLED that has three layers: Adjacent to the cathode is the **electron transport layer (ETL)**, plus there is an **emissive layer (EML)** sandwiched between the ETL and the HTL. The emissive layer can be complex, containing a **host** material plus a **guest** compound in small concentration—a kind of doping—to fine-tune energy levels and efficiency.

40–10 Transistors: Bipolar and MOSFETs

The **bipolar junction transistor** was invented in 1948 by J. Bardeen, W. Shockley, and W. Brattain. It consists of a crystal of one type of doped semiconductor sandwiched between two of the opposite type. Both *npn* and *pnp* transistors can be made, and they are shown schematically in Fig. 40–48a. The three semiconductors are given the names **collector**, **base**, and **emitter**. The symbols for *npn* and *pnp* transistors are shown in Fig. 40–48b. The arrow is always placed on the emitter and indicates the direction of (conventional) current flow in normal operation.

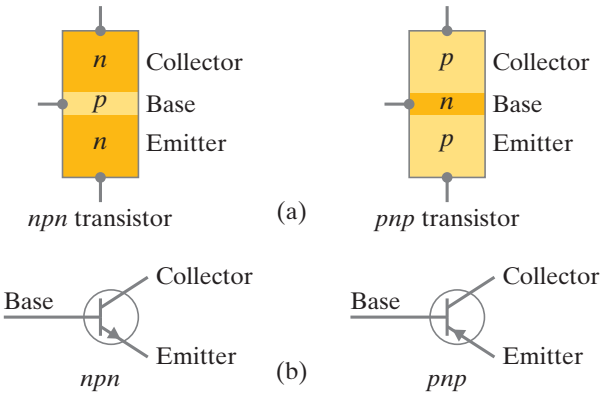


FIGURE 40–48 (a) Schematic diagram of *npn* and *pnp* transistors. (b) Symbols for *npn* and *pnp* transistors.

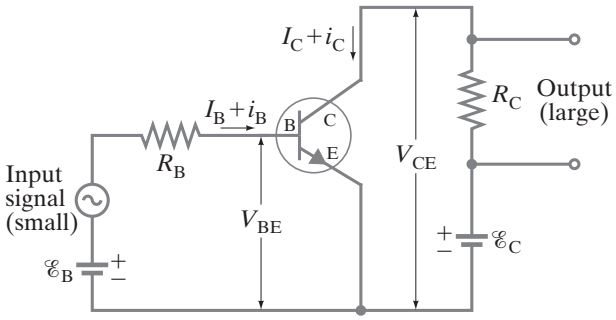


FIGURE 40–49 An *npn* transistor used as an amplifier. I_B is the current produced by $\mathscr{E}_B$ (in the absence of a signal), i_B is the ac signal current (= change in I_B).

The operation of an *npn* transistor as an **amplifier** is shown in Fig. 40–49. A dc voltage V_{CE} is maintained between the collector and emitter by battery $\mathscr{E}_C$. The voltage applied to the base is called the *base bias voltage*, V_{BE}. If V_{BE} is positive, conduction electrons in the emitter are attracted into the base. The base region is very thin, much less than 1 μm, so most of these electrons flow right across into the collector which is maintained at a positive voltage. A large current, I_C, flows between collector and emitter and a much smaller current, I_B, through the base. In the steady state, I_B and I_C can be considered dc. But a small variation (= ac) in the base voltage due to an input signal attracts (or repels) charge that passes through into the collector and thus can cause a large *change* in the collector current and a large change in the voltage drop across the output resistor R_C. Hence a transistor can *amplify* a small signal into a larger one.

Typically a small ac signal (call it i_B) is to be amplified, and when added to the base bias current (and voltage) causes the current and voltage at the collector to vary at the same rate but magnified. Thus, what is important for amplification is the *change* in collector current for a given input *change* in base current. We label these ac signal currents (= changes in I_C and I_B) as i_C and i_B. The **current gain** is defined as the ratio

$$\beta_I = \frac{\text{output (collector) ac current}}{\text{input (base) ac current}} = \frac{i_C}{i_B}.$$

β_I may be on the order of 10 to 100. Similarly, the **voltage gain** is

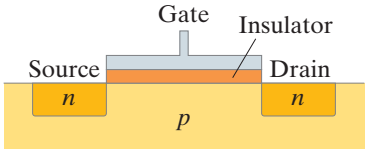
$$\beta_V = \frac{\text{output (collector) ac voltage}}{\text{input (base) ac voltage}}.$$

Transistors are the basic elements in modern electronic amplifiers of all sorts.

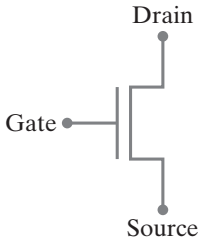
A *pnp* transistor operates like an *npn*, except that holes move instead of electrons. The collector voltage is negative, and so is the base voltage in normal operation.

Another kind of transistor, very important, is the **MOSFET** (metal-oxide semiconductor field-effect transistor) common in **digital circuits** as a type of switch. Its construction is shown in Fig. 40–50a, and its symbol in Fig. 40–50b. What is called the emitter in a bipolar transistor is called the **source** in a MOSFET, and the collector is called the **drain**. The base is called the **gate**. The gate acts to let a current flow, or not, from the source to the drain, depending on the electric field it (the gate)

FIGURE 40–50 (a) Construction of a MOSFET of *n*- and *p*-type semiconductors and a gate of metal or heavily doped silicon (= a good conductor). (b) Symbol for a MOSFET which suggests its function.



(a)



(b)

provides across an insulator that separates it from the *p*-type semiconductor below, Fig. 40–50a. Hence the name "field-effect transistor" (FET).[†] MOSFETs are often used like switches, on or off, which in digital circuits can allow the storage of a binary bit, a "1" or a "0." We discussed uses of MOSFETs relative to digital TV (Section 23–10) and computer memory storage (Section 29–8).

*40–11 Integrated Circuits, 14-nm Technology

Although individual transistors are very small compared to the once-used vacuum tubes, they are huge compared to **integrated circuits** or **chips** (photo at start of this Chapter), invented in 1959 independently by Jack Kilby and Robert Noyce. Tiny amounts of impurities can be inserted or injected at particular locations within a single silicon crystal or wafer. These can be arranged to form diodes, transistors, resistors (undoped semiconductors), and very thin connecting "wires" (= conductors) which are heavily doped thin lines. Capacitors and inductors can also be formed, but also can be connected separately. Integrated circuits are the heart of computers, televisions, calculators, cameras, cell phones, and the electronic instruments that control aircraft, space vehicles, and automobiles.

A tiny chip, a few millimeters on a side, may contain billions of transistors and other circuit elements. The number of elements/mm^2 has been doubling every 2 or 3 years. We often hear of the **technology generation**, which is a number that refers to the minimum width of a conducting line ("wire") or to the gate of a MOSFET (which may be slightly smaller than a wire). Since 2003 we have passed from 90-nm technology to 65-nm, to 45-nm, to 32-nm, to 22-nm, every 2 to 3 years, and now 14-nm technology. The next step to 10-nm, which means less than 20 atoms wide, may involve new structures and quantum-mechanical effects, and may not be easy to achieve. Smaller means more diodes and transistors per mm^2 and therefore greater speed (faster response time) because the distance signals have to travel is less. Smaller also can mean lower power consumption and longer battery life, as well as slimmer and lighter devices. Size, speed, and power have all been improved 10 to 100 million times in the last half century or so.

[†]The "MOS" comes from a version with a **M**etal gate, silicon di**O**xide insulator, and a **S**emiconductor (*p*-type shown in Fig. 40–50a). The gate can also be heavily doped silicon (= good conductor).

Summary

Quantum mechanics explains the bonding together of atoms to form **molecules**. In a **covalent bond**, the atoms share electrons. The electron clouds of two or more atoms overlap because of constructive interference between the electron waves. The positive nuclei are attracted to this concentration of negative charge between them, forming the bond.

An **ionic bond** is an extreme case of a covalent bond in which one or more electrons from one atom spend much more time around the other atom than around their own. The atoms then act as oppositely charged ions that attract each other, forming the bond.

These **strong bonds** hold molecules together, and also hold atoms and molecules together in solids. Also important are **weak bonds** (or **van der Waals bonds**), which are generally dipole attractions between molecules.

When atoms combine to form molecules, the energy levels of the outer electrons are altered because they now interact with each other. Additional energy levels also become possible because the atoms can vibrate with respect to each other, and the molecule as a whole can rotate. The energy levels for both vibrational and rotational motion are quantized, and are very close together (typically, 10^{-1} eV to 10^{-3} eV apart). Each atomic energy level thus becomes a set of closely spaced levels corresponding to the vibrational and rotational motions. Transitions from one level to another appear as many very closely spaced lines. The resulting spectra are called **band spectra**.

The quantized rotational energy levels are given by

$$E_{\text{rot}} = \ell(\ell + 1)\frac{\hbar^2}{2I}, \qquad \ell = 0, 1, 2, \cdots, \qquad \textbf{(40–2)}$$

where I is the moment of inertia of the molecule.

The energy levels for vibrational motion are given by

$$E_{\text{vib}} = \left(\nu + \tfrac{1}{2}\right)hf, \qquad \nu = 0, 1, 2, \cdots, \qquad \textbf{(40–6)}$$

where f is the classical natural frequency of vibration for the molecule. Transitions between energy levels are subject to the selection rules $\Delta\ell = \pm 1$ and $\Delta\nu = \pm 1$.

Some **solids** are bound together by covalent and ionic bonds, just as molecules are. In metals, the electrostatic force between free electrons and positive ions helps form the **metallic bond**.

In the free-electron theory of metals, electrons occupy the possible energy states according to the exclusion principle. At $T = 0\,\text{K}$, all possible states are filled up to a maximum energy level called the **Fermi energy**, E_F, the magnitude of which is typically a few eV. All states above E_F are vacant at $T = 0\,\text{K}$.

At normal temperatures (300 K) the distribution of occupied states is only slightly altered and is given by the **Fermi–Dirac probability function**

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1}. \qquad \textbf{(40–14)}$$

In a crystalline solid, the possible energy states for electrons are arranged in **bands**. Within each band the levels are very close together, but between the bands there may be forbidden **energy gaps**. Good conductors are characterized by the highest occupied band (the **conduction band**) being only partially full, so lots of states are available to electrons to move about and accelerate when a voltage is applied. In a good insulator, the highest occupied energy band (the **valence band**) is completely full, and there is a large energy gap (5 to 10 eV) to the next highest band, the *conduction band*. At room temperature, molecular kinetic energy (thermal energy) available due to collisions is only about 0.04 eV, so almost no electrons can jump from the valence to the conduction band in an insulator. In a **semiconductor**, the gap between valence and conduction bands is much smaller, on the order of 1 eV, so a few electrons can make the transition from the essentially full valence band to the nearly empty conduction band, allowing a small amount of conductivity.

In a **doped** silicon semiconductor, a small percentage of impurity atoms with five or three valence electrons replace a few of the normal silicon atoms with their four valence electrons. A five-electron impurity produces an ***n*-type** semiconductor with negative electrons as carriers of current. A three-electron impurity produces a ***p*-type** semiconductor in which positive **holes** carry the current. The energy level of impurity atoms lies slightly below the conduction band in an *n*-type semiconductor, and acts as a **donor** from which electrons readily pass into the conduction band. The energy level of impurity atoms in a *p*-type semiconductor lies slightly above the valence band and acts as an **acceptor** level, since electrons from the valence band easily reach it, leaving holes behind to act as charge carriers.

A semiconductor **diode** consists of a ***pn* junction** and allows current to flow in one direction only; *pn* junction diodes are used as **rectifiers** to change ac to dc, as photovoltaic cells to produce electricity from sunlight, and as lasers. **Light-emitting diodes (LED)** use compound semiconductors which can emit light when a forward-bias voltage is applied; uses include readouts, infrared remote controls, normal lighting (household light bulbs, flashlights, street lights), and very large TV screens. LEDs using organic molecules or polymers (**OLED**) are used as screens on cell phones and other displays. Common **transistors** consist of three semiconductor sections, either as ***pnp*** or ***npn***. Transistors can amplify electrical signals and in computers serve as switches or **gates** for the 1s and 0s of digital bits. An integrated circuit consists of a tiny semiconductor crystal or **chip** on which many transistors, diodes, resistors, and other circuit elements are constructed by placement of impurities.

Questions

1. What type of bond would you expect for (*a*) the N_2 molecule, (*b*) the HCl molecule, (*c*) Fe atoms in a solid?
2. Describe how the molecule $CaCl_2$ could be formed.
3. Does the H_2 molecule have a permanent dipole moment? Does O_2? Does H_2O? Explain.
4. Although the molecule H_3 is not stable, the ion H_3^+ is. Explain, using the Pauli exclusion principle.
5. Would you expect the molecule H_2^+ to be stable? If so, where would the single electron spend most of its time?
6. Compare the potential energy diagram for an H_2 molecule with the potential energy diagram for ATP formation from ADP and Ⓟ. Explain the significance of the difference in shapes of the two diagrams. (See Section 40–2.)
7. Explain why the carbon atom $(Z = 6)$ usually forms four bonds with hydrogen-like atoms.
8. The energy of a molecule can be divided into four categories. What are they?
9. (*a*) Why are weak bonds important in cells? (*b*) Explain why heating proteins too much may cause them to denature—that is, lose the specific shape they need to function. (See Section 40–3.) (*c*) What is the strongest weak bond, and why? (*d*) If this bond, and the other weak bonds, were stronger (that is, too strong), what would be the consequence for protein synthesis?
10. Explain why metals are shiny.
11. If conduction electrons are free to roam about in a metal, why don't they leave the metal entirely?
12. Explain why the resistivity of metals increases with increasing temperature whereas the resistivity of semiconductors may decrease with increasing temperature.
13. Compare the resistance of a *pn* junction diode connected in forward bias to its resistance when connected in reverse bias.
14. Explain on the basis of energy bands why the sodium chloride crystal is a good insulator. [*Hint*: Consider the shells of Na^+ and Cl^- ions.]
15. Which aspects of Fig. 40–28 are peculiar to copper, and which are valid in general?
16. What is the main difference between *n*-type and *p*-type semiconductors?
17. A silicon semiconductor is doped with phosphorus. Will these atoms be donors or acceptors? What type of semiconductor will this be?
18. Figure 40–51 shows a "bridge-type" full-wave rectifier. Explain how the current is rectified and how current flows during each half cycle.

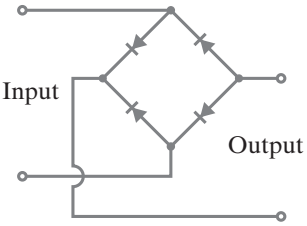


FIGURE 40–51
Question 18.

19. A transistor can amplify an electronic signal, meaning it can increase the power of an input signal. Where does it get the energy to increase the power?

PSE5_CH40_v1.7_23-Jun-20

20. In a transistor, the base–emitter junction and the base–collector junction are essentially diodes. Are these junctions reverse-biased or forward-biased in the application shown in Fig. 40–49?

21. Do diodes and transistors obey Ohm's law? Explain.

22. Can a diode be used to amplify a signal? Explain.

23. Explain how a transistor can be used as a switch.

MisConceptual Questions

1. Which of the following is true for covalently bound diatomic molecules such as H_2?
(*a*) All electrons in the two atoms have identical quantum numbers.
(*b*) The molecule has fewer electrons than the two separate atoms do.
(*c*) The molecule has less energy than two separate atoms.
(*d*) The energy of the molecule is greatest when the atoms are separated by one bond length.

2. A hydrogen atom $(Z = 1)$ is bonded to a lithium atom $(Z = 3)$ in lithium hydride, LiH. Which of the following are possible spin states of the two shared electrons?
(*a*) $+\frac{1}{2}, +\frac{1}{2}$.
(*b*) $-\frac{1}{2}, -\frac{1}{2}$.
(*c*) $+\frac{1}{2}, -\frac{1}{2}$.
(*d*) Both (*a*) and (*b*).
(*e*) Any of the above.

3. Ionic bonding is related to
(*a*) magnetic dipole interactions.
(*b*) the transfer of one or more electrons from one atom to another.
(*c*) the sharing of electrons among atoms.
(*d*) the transfer of electrons to the solid.
(*e*) oscillation dipoles.

4. Consider Fig. 40–11. As the last phosphate group approaches and then bonds to the ADP molecule, which of the following is true? Choose all that apply.
(*a*) The phosphate group is first repelled and then attracted to the ADP molecule.
(*b*) The phosphate group is always attracted to the ADP molecule.
(*c*) The phosphate group is always repelled by the ADP molecule.
(*d*) The system first loses and then stores potential energy.
(*e*) Both binding energy and activation energy are negative.
(*f*) Both binding energy and activation energy are positive.

5. Which type of bond holds the molecules of the DNA double helix together?
(*a*) Covalent bond.
(*b*) Ionic bond.
(*c*) Einstein bond.
(*d*) Van der Waals bond.

6. Which of the following is usually the weakest bond?
(*a*) Hydrogen bond.
(*b*) Covalent bond.
(*c*) Ionic bond.
(*d*) Van der Waals bond.

7. Which of the following apply to metallic bonds?
(*a*) Free electrons are shared by all atoms.
(*b*) Atoms bond to hydrogen molecules.
(*c*) Unequal charge distributions around neutral atoms create strong bonds.
(*d*) Electrons are stolen from one atom by another.

8. Why are metals good conductors?
(*a*) Gaining a tiny bit of energy by applying a voltage allows their electrons to move.
(*b*) They have more electrons than protons, so some of the electrons are extra and free to move.
(*c*) They have more protons than electrons, so some of the protons are extra and free to move.
(*d*) Gaining a tiny bit of energy allows their protons to move.
(*e*) Electrons are tightly bound to their atoms.

9. The speed of the conduction electrons in a metal
(*a*) is nearly the speed of light.
(*b*) depends mostly on the Fermi energy.
(*c*) depends on the absolute temperature according to $v = \sqrt{3KT/m}$.
(*d*) depends mostly on the current flow.

10. In an insulator
(*a*) electrons must gain a large amount of energy to find an empty state in the conduction band.
(*b*) there is a slight deficit of electrons making it difficult for electrons to move.
(*c*) the exclusion principle keeps electrons out of the conduction band.
(*d*) there is no conduction band in an insulator.

11. Doping a semiconductor will
(*a*) increase the number of outer electrons of atoms.
(*b*) decrease the number of outer electrons of atoms.
(*c*) change the size of the band gap.
(*d*) either (*a*) or (*b*).
(*e*) either (*a*), (*b*), or (*c*).

12. In a *p*-type semiconductor, a hole is
(*a*) a region in the molecular structure where an atom is missing.
(*b*) an extra electron from one of the donor atoms.
(*c*) an extra positively charged particle in the molecular structure.
(*d*) a region missing an electron relative to the rest of the molecular structure.

13. Which of the following would *not* be used as an impurity in doping silicon?
(*a*) Germanium.
(*b*) Gallium.
(*c*) Boron.
(*d*) Phosphorus.
(*e*) Arsenic.

14. Which of the following statements about a diode is true?
(*a*) When reverse biased, a diode normally passes very little current.
(*b*) When you double the voltage applied across a forward biased diode, you double the current flowing through it.
(*c*) A diode is either *npn* or *pnp*.
(*d*) A forward biased diode always has the same current flow regardless of the applied voltage.
(*e*) None of these statements are true.

Problems

40–1 to 40–3 Molecular Bonds

1. (I) Estimate the binding energy of a KCl molecule by calculating the electrostatic potential energy when the K^+ and Cl^- ions are at their stable separation of 0.28 nm. Assume each has a charge of magnitude 1.0 e.
2. (II) The measured binding energy of KCl is 4.43 eV. From the result of Problem 1, estimate the contribution to the binding energy of the repelling electron clouds at the equilibrium distance $r_0 = 0.28$ nm.
3. (II) The equilibrium distance r_0 between two atoms in a molecule is called the **bond length**. Using the bond lengths of homogeneous molecules (like H_2, O_2, and N_2), one can estimate the bond length of heterogeneous molecules (like CO, CN, and NO). This is done by summing half of each bond length of the homogenous molecules to estimate that of the heterogeneous molecule. Given the following bond lengths: H_2 (= 74 pm), N_2 (= 145 pm), O_2 (= 121 pm), C_2 (= 154 pm), estimate the bond lengths for: HN, CN, and NO.
4. (II) Binding energies are often measured experimentally in kcal per mole, and then the binding energy in eV per molecule is calculated from that result. What is the conversion factor in going from kcal per mole to eV per molecule? What is the binding energy of KCl (= 4.43 eV) in kcal per mole?
5. (III) Estimate the binding energy of the H_2 molecule, assuming the two H nuclei are 0.074 nm apart and the two electrons spend 33% of their time midway between them.
6. (III) Estimate the energy associated with the repulsion of the electron shells of a lithium fluoride (LiF) molecule. The ionization energy of lithium is 5.39 eV, and it takes 3.41 eV to remove the extra electron from an F^- ion. The bond length is 0.156 nm, and the binding energy of LiF is 5.95 eV.
7. (III) (*a*) Apply reasoning similar to that in the text for the $S = 0$ and $S = 1$ states in the formation of the H_2 molecule to show why the molecule He_2 is *not* formed. (*b*) Explain why the He_2^+ molecular ion *could* form. (Experiment shows it has a binding energy of 3.1 eV at $r_0 = 0.11$ nm.)

40–4 Molecular Spectra

8. (I) Show that the quantity $\hbar^2/I$ has units of energy.
9. (I) What is the reduced mass of the molecules (*a*) KCl; (*b*) O_2; (*c*) HCl?
10. (I) Estimate the moment of inertia of a hydrogen molecule around an axis perpendicular to a line connecting the two atoms, at their CM.
11. (II) Suppose two 1.5-mg masses were made into a small dumbbell. (*a*) Given that the rotational states of this system would be quantized, what would the angular velocity of the dumbbell be in the $\ell = 1$ state? Assume the masses are separated by 1.0 mm. (*b*) If this dumbbell rotated at 50 rpm, what would be its singular momentum state ℓ?
12. (II) (*a*) Calculate the "characteristic rotational energy," $\hbar^2/2I$, for the O_2 molecule whose bond length is 0.121 nm. (*b*) What are the energy and wavelength of photons emitted in an $\ell = 3$ to $\ell = 2$ transition?
13. (II) The "characteristic rotational energy," $\hbar^2/2I$, for N_2 is 2.48×10^{-4} eV. Calculate the N_2 bond length.
14. (II) The equilibrium separation of H atoms in the H_2 molecule is 0.074 nm (Fig. 40–8). Calculate the energies and wavelengths of photons for the rotational transitions (*a*) $\ell = 1$ to $\ell = 0$, (*b*) $\ell = 2$ to $\ell = 1$, and (*c*) $\ell = 3$ to $\ell = 2$.
15. (II) Determine the wavelength of the photon emitted when the CO molecule makes the rotational transition $\ell = 5$ to $\ell = 4$. [*Hint*: See Example 40–5.]
16. (II) Calculate the bond length for the NaCl molecule given that three successive wavelengths for rotational transitions are 23.1 mm, 11.6 mm, and 7.71 mm.
17. (II) Estimate the longest wavelength emitted by a lithium hydride (LiH) molecule for a change in its rotational state if its equilibrium separation is 0.16 nm.
18. (II) Explain why there is no transition for $\Delta E = hf$ in Fig. 40–21 (and Fig. 40–22). See Eqs. 40–8.
19. (II) The fundamental vibration frequency for the CO molecule is 6.42×10^{13} Hz. Determine (*a*) the reduced mass, and (*b*) the effective value of the "stiffness" constant k. Compare to k for the H_2 molecule.
20. (II) Li and Br form a molecule for which the lowest vibrational frequency is 1.7×10^{13} Hz. What is the effective stiffness constant k?
21. (II) (*a*) Use the curve of Fig. 40–18 to estimate the stiffness constant k for the H_2 molecule. (Recall that $U = \frac{1}{2}kx^2$.) (*b*) Then estimate the fundamental wavelength for vibrational transitions using the classical formula (Chapter 14), but use only $\frac{1}{2}$ the mass of an H atom (because both H atoms move).
22. (III) Imagine the two atoms of a diatomic molecule as if they were connected by a spring, Fig. 40–52. Show that the classical frequency of vibration is given by Eq. 40–5. [*Hint*: Let x_1 and x_2 be the displacements of each mass from initial equilibrium positions; then $m_1 d^2x_1/dt^2 = -kx$, and $m_2 d^2x_2/dt^2 = -kx$, where $x = x_1 + x_2$. Find another relationship between x_1 and x_2, assuming that the center of mass of the system stays at rest, and then show that $\mu\, d^2x/dt^2 = -kx$.]

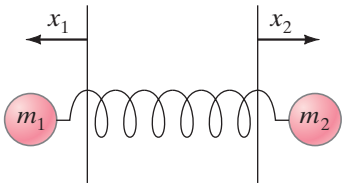


FIGURE 40–52
Problem 22.

40–5 Bonding in Solids

23. (I) Estimate the ionic cohesive energy for NaCl taking $\alpha = 1.75$, $m = 8$, and $r_0 = 0.28$ nm.
24. (II) Common salt, NaCl, has a density of 2.165 g/cm^3. The molecular weight of NaCl is 58.44. Estimate the distance between nearest neighbor Na and Cl ions. [*Hint*: Each ion can be considered to be at the corner of a cube.]
25. (II) Repeat the previous Problem for KCl whose density is 1.99 g/cm^3.
26. (II) The spacing between "nearest neighbor" Na and Cl ions in a NaCl crystal is 0.24 nm. What is the spacing between two nearest neighbor Na ions?
27. (III) For a long one-dimensional chain of alternating positive and negative ions, show that the Madelung constant would be $\alpha = 2 \ln 2$. [*Hint*: Use a series expansion for $\ln(1 + x)$.]

28. (III) (*a*) Starting from Eq. 40–9, show that the ionic cohesive energy is given by $U_0 = -(\alpha e^2/4\pi\epsilon_0 r_0)(1 - 1/m)$. Determine U_0 for (*b*) NaI $(r_0 = 0.33\,\text{nm})$ and (*c*) MgO $(r_0 = 0.21\,\text{nm})$. Assume $m = 10$. (*d*) If you used $m = 8$ instead, how far off would your answers be? Assume $\alpha = 1.75$.

40–6 Free-Electron Theory of Metals

29. (II) Estimate the number of possible electron states in a 1.00-cm^3 cube of silver between $0.975E_F$ and $E_F\,(= 5.48\,\text{eV})$.

30. (II) Estimate the number of states between 7.00 eV and 7.05 eV that are available to electrons in a 1.0-cm^3 cube of copper.

31. (II) What, roughly, is the ratio of the density of molecules in an ideal gas at 285 K and 1 atm (say O_2) to the density of free electrons (assume one per atom) in a metal (copper) also at 285 K?

32. (II) Calculate the energy which has 85.0% occupancy probability for copper at (*a*) $T = 295\,\text{K}$; (*b*) $T = 750\,\text{K}$.

33. (II) Calculate the energy which has 15.0% occupancy probability for copper at (*a*) $T = 295\,\text{K}$; (*b*) $T = 950\,\text{K}$.

34. (II) What is the occupancy probability for a conduction electron in copper at $T = 295\,\text{K}$ for an energy $E = 1.015E_F$?

35. (II) The atoms in zinc metal $(\rho = 7.1 \times 10^3\,\text{kg/m}^3)$ each have two free electrons. Calculate (*a*) the density of conduction electrons, (*b*) their Fermi energy, and (*c*) their Fermi speed.

36. (II) Calculate the Fermi energy and Fermi speed for sodium, which has a density of $0.97 \times 10^3\,\text{kg/m}^3$ and has one conduction electron per atom.

37. (II) Given that the Fermi energy of aluminum is 11.63 eV, (*a*) calculate the density of free electrons using Eq. 40–12, and (*b*) estimate the valence of aluminum using this model and the known density $(2.70 \times 10^3\,\text{kg/m}^3)$ and atomic mass (27.0) of aluminum.

38. (II) Show that the average energy of conduction electrons in a metal at $T = 0\,\text{K}$ is $\overline{E} = \frac{3}{5}E_F$ (Eq. 40–13) by calculating

$$\overline{E} = \frac{\int E\,n_o(E)\,dE}{\int n_o(E)\,dE}.$$

39. (II) The neutrons in a neutron star (Chapter 44) can be treated as a Fermi gas with neutrons in place of the electrons in our model of an electron gas. Determine the Fermi energy for a neutron star of radius 12 km and mass 2.5 times that of our Sun. Assume that the star is made entirely of neutrons and is of uniform density.

40. (II) For a one-dimensional potential well of width ℓ, start with Eq. 38–13 and show that the number of states per unit energy interval for an electron gas is given by

$$g_\ell(E) = \sqrt{\frac{8m\ell^2}{h^2 E}}.$$

Remember that there can be two electrons (spin up and spin down) for each value of n. [*Hint*: Write the quantum number n in terms of E. Then $g_\ell(E) = 2\,dn/dE$ where dn is the number of energy levels between E and $E + dE$.]

41. (II) Show that the probability for the state at the Fermi energy being occupied is exactly $\frac{1}{2}$, independent of temperature.

42. (II) A very simple model of a "one-dimensional" metal consists of N electrons confined to a rigid box of width ℓ. We neglect the Coulomb interaction between the electrons. (*a*) Calculate the Fermi energy for this one-dimensional metal (E_F = the energy of the most energetic electron at $T = 0\,\text{K}$), taking into account the Pauli exclusion principle. You can assume for simplicity that N is even. (*b*) What is the smallest amount of energy, ΔE, that this 1-D metal can absorb? (*c*) Find the limit of $\Delta E/E_F$ for large N. What does this result say about how well metals can conduct?

43. (II) (*a*) For copper at room temperature $(T = 293\,\text{K})$, calculate the Fermi factor, Eq. 40–14, for an electron with energy 0.12 eV above the Fermi energy. This represents the probability that this state is occupied. Is this reasonable? (*b*) What is the probability that a state 0.12 eV below the Fermi energy is occupied? (*c*) What is the probability that the state in part (*b*) is unoccupied?

44. (III) Proceed as follows to derive the density of states, $g(E)$, the number of states per unit volume per unit energy interval, Eq. 40–10. Let the metal be a cube of side ℓ. Extend the discussion of Section 38–8 for an infinite well to three dimensions, giving energy levels

$$E = \frac{h^2}{8m\ell^2}(n_1^2 + n_2^2 + n_3^2).$$

(Explain the meaning of n_1, n_2, n_3.) Each set of values for the quantum numbers n_1, n_2, n_3 corresponds to one state. Imagine a space where n_1, n_2, n_3 are the axes, and each state is represented by a point on a cubic lattice in this space, each separated by 1 unit along an axis. Consider the octant $n_1 > 0$, $n_2 > 0$, $n_3 > 0$. Show that the number of states N within a radius $R = (n_1^2 + n_2^2 + n_3^2)^{\frac{1}{2}}$ is $2\left(\frac{1}{8}\right)\left(\frac{4}{3}\pi R^3\right)$. Then, to get Eq. 40–10, set $g(E) = (1/V)(dN/dE)$, where $V = \ell^3$ is the volume of the metal.

40–7 Band Theory of Solids

45. (I) A semiconductor is struck by light of slowly increasing frequency and begins to conduct when the wavelength of the light is 620 nm; estimate the energy gap E_g.

46. (I) Calculate the longest-wavelength photon that can cause an electron in silicon $(E_g = 1.12\,\text{eV})$ to jump from the valence band to the conduction band.

47. (II) The energy gap between valence and conduction bands in germanium is 0.72 eV. What range of wavelengths can a photon have to excite an electron from the top of the valence band into the conduction band?

48. (II) We saw that there are $2N$ possible electron states in the 3*s* band of Na, where N is the total number of atoms. How many possible electron states are there in the (*a*) 2*s* band, (*b*) 2*p* band, and (*c*) 3*p* band? (*d*) State a general formula for the total number of possible states in any given electron band.

49. (II) The energy gap E_g in germanium is 0.72 eV. When used as a photon detector, roughly how many electrons can be made to jump from the valence to the conduction band by the passage of a 710-keV photon that loses all its energy in this fashion?

40–8 Semiconductors and Doping

50. (II) Suppose that a silicon semiconductor is doped with phosphorus so that one silicon atom in 1.3×10^6 is replaced by a phosphorus atom. Assuming that the "extra" electron in every phosphorus atom is donated to the conduction band, by what factor is the density of conduction electrons increased? The density of silicon is 2330 kg/m^3, and the density of conduction electrons in pure silicon is about 10^{16} m^{-3} at room temperature.

40–9 Diodes

51. (I) At what wavelength will an LED radiate if made from a material with an energy gap $E_g = 1.8$ eV?

52. (I) If an LED emits light of wavelength $\lambda = 630$ nm, what is the energy gap (in eV) between valence and conduction bands?

53. (II) A silicon diode, whose current–voltage characteristics are given in Fig. 40–38, is connected in series with a battery and an 860-Ω resistor. What battery voltage is needed to produce a 16-mA current?

54. (II) Suppose that the diode of Fig. 40–38 is connected in series to a 120-Ω resistor and a 2.0-V battery. What current flows in the circuit? [*Hint*: Draw a line on Fig. 40–38 representing the current in the resistor as a function of the voltage across the diode; the intersection of this line with the characteristic curve will give the answer.]

55. (II) Sketch the resistance as a function of current, for $V > 0$, for the diode shown in Fig. 40–38.

56. (II) An ac voltage of 120 V rms is to be rectified. Estimate very roughly the average current in the output resistor R (42 kΩ) for (*a*) a half-wave rectifier (Fig. 40–39), and (*b*) a full-wave rectifier (Fig. 40–40) without capacitor.

57. (II) A semiconductor diode laser emits 1.1-μm light. Assuming that the light comes from electrons and holes recombining, what is the band gap in this laser material?

58. (III) A silicon diode passes significant current only if the forward-bias voltage exceeds about 0.6 V. Make a rough estimate of the average current in the output resistor R of (*a*) a half-wave rectifier (Fig. 40–39), and (*b*) a full-wave rectifier (Fig. 40–40) without a capacitor. Assume that $R = 150\ \Omega$ in each case and that the ac voltage is 9.0 V rms in each case.

59. (III) A 120-V rms 60-Hz voltage is to be rectified with a full-wave rectifier as in Fig. 40–40, where $R = 24\ \text{k}\Omega$, and $C = 35\ \mu\text{F}$. (*a*) Make a rough estimate of the average current. (*b*) What happens if $C = 0.10\ \mu\text{F}$? [*Hint*: See Section 26–5.]

40–10 Transistors

60. (I) Draw a circuit diagram (like Fig. 40–49) showing how a *pnp* transistor can operate as an amplifier. Show polarities, etc.

61. (II) If the current gain of the transistor amplifier in Fig. 40–49 is $\beta = i_C/i_B = 95$, what value must R_C have if a 1.0-μA ac base current is to produce an ac output voltage of 0.40 V?

62. (II) Suppose that the current gain of the transistor in Fig. 40–49 is $\beta = i_C/i_B = 85$. If $R_C = 3.9\ \text{k}\Omega$, calculate the ac output voltage for an ac input current of 2.0 μA.

63. (II) An amplifier has a voltage gain of 75 and a 25-kΩ load (output) resistance. What is the peak output current through the load resistor if the input voltage is an ac signal with a peak of 0.080 V?

64. (II) A transistor, whose current gain $\beta = i_C/i_B = 65$, is connected as in Fig. 40–49 with $R_B = 3.8\ \text{k}\Omega$ and $R_C = 7.8\ \text{k}\Omega$. Calculate (*a*) the voltage gain, and (*b*) the power amplification.

65. (II) From Fig. 40–49, write an equation for the relationship between the base current (I_B), the collector current (I_C), and the emitter current (I_E, not labeled in Fig. 40–49). Assume the ac current $i_B = i_C = 0$.

General Problems

66. Estimate the number of states from 4.0 eV to 6.2 eV available to electrons in a 10-cm cube of iron.

67. A full-wave rectifier (Fig. 40–40) uses two diodes to rectify a 75-V rms 60-Hz ac voltage. If $R = 7.8\ \text{k}\Omega$ and $C = 36\ \mu\text{F}$, what will be the approximate percent variation in the output voltage? The variation in output voltage (Fig. 40–40c) is called *ripple voltage*. [*Hint*: See Section 26–5 and assume the discharge of the capacitor is approximately linear.]

68. In the ionic salt KF, the separation distance between ions is about 0.27 nm. (*a*) Estimate the electrostatic potential energy between the ions assuming them to be point charges (magnitude 1*e*). (*b*) When F "grabs" an electron, it releases 3.41 eV of energy, whereas 4.34 eV is required to ionize K. Find the binding energy of KF relative to free K and F atoms, neglecting the energy of repulsion.

69. A strip of silicon 1.8 cm wide and 1.0 mm thick is immersed in a magnetic field of strength 1.3 T perpendicular to the strip (Fig. 40–53). When a current of 0.28 mA is run through the strip, there is a resulting Hall effect voltage of 18 mV across the strip (Section 27–8). How many electrons per silicon atom are in the conduction band? The density of silicon is 2330 kg/m^3.

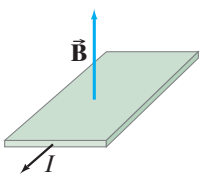


FIGURE 40–53
Problem 69.

70. One possible form for the potential energy (U) of a diatomic molecule (Fig. 40–8) is called the *Morse Potential*:

$$U = U_0\left[1 - e^{-a(r-r_0)}\right]^2.$$

(*a*) Show that r_0 represents the equilibrium distance and U_0 the dissociation energy. (*b*) Graph U from $r = 0$ to $r = 4r_0$, assuming $a = 18\ \text{nm}^{-1}$, $U_0 = 4.6$ eV, and $r_0 = 0.13$ nm.

71. Do we need to consider quantum effects for everyday rotating objects? Estimate the differences between rotational energy levels for a spinning baton compared to the energy of the baton. Assume the baton consists of a uniform 32-cm-long bar with a mass of 260 g and two small end masses, each of mass 350 g, and that it rotates at 1.6 rev/s about the bar's center.

72. For H_2, estimate how many rotational states there are between vibrational states.

73. Explain, using the Boltzmann factor (Eq. 39–16), why the heights of the peaks in Fig. 40–22 are different from one another. Explain also why the lines are not equally spaced. [*Hint*: Does the moment of inertia necessarily remain constant?]

74. The rotational absorption spectrum of a molecule displays peaks about 9.2×10^{11} Hz apart. Determine the moment of inertia of this molecule.

75. A TV remote control emits IR light. If the detector on the TV set is *not* to react to visible light, could it make use of silicon as a "window" with its energy gap $E_g = 1.12\,\text{eV}$? What is the shortest-wavelength light that can strike silicon without causing electrons to jump from the valence band to the conduction band?

76. For O_2 with a bond length 0.121 nm, what is the moment of inertia about an axis through the CM perpendicular to the line joining the atoms?

77. The **Fermi temperature** T_F is defined as that temperature at which the thermal energy kT (without the $\frac{3}{2}$) is equal to the Fermi energy: $kT_F = E_F$. (*a*) Determine the Fermi temperature for copper. (*b*) Show that for $T \gg T_F$, the Fermi factor (Eq. 40–14) approaches the Boltzmann factor. (Note: This last result is not very useful for understanding conductors. Why?)

78. Consider a monatomic solid with a weakly bound cubic lattice, with each atom connected to six neighbors, each bond having a binding energy of $3.6 \times 10^{-3}\,\text{eV}$. When this solid melts, its latent heat of fusion goes directly into breaking the bonds between the atoms. Estimate the latent heat of fusion for this solid, in J/mol. [*Hint*: Show that in a simple cubic lattice (Fig. 40–54), there are *three* times as many bonds as there are atoms, when the number of atoms is large.]

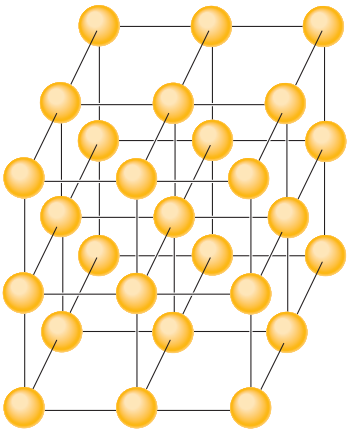


FIGURE 40–54 Problem 78.

79. The energy gap between valence and conduction bands in zinc sulfide is 3.6 eV. What range of wavelengths can a photon have to excite an electron from the top of the valence band into the conduction band?

80. Most of the Sun's radiation has wavelengths shorter than 1100 nm. For a solar cell to absorb all this, what energy gap ought the material have?

81. Green and blue LEDs became available many years after red LEDs were first developed. Approximately what energy gaps would you expect to find in green (525 nm) and in blue (465 nm) LEDs?

82. For an arsenic donor atom in a doped silicon semiconductor, assume that the "extra" electron moves in a Bohr orbit about the arsenic ion. For this electron in the ground state, take into account the dielectric constant $K = 12$ of the Si lattice (which represents the weakening of the Coulomb force due to all the other atoms or ions in the lattice), and estimate (*a*) the binding energy, and (*b*) the orbit radius for this extra electron. [*Hint*: Substitute $\epsilon = K\epsilon_0$ in Coulomb's law; see Section 24–5.]

83. The fundamental vibration frequency for the HCl molecule is $8.66 \times 10^{13}\,\text{Hz}$. Determine (*a*) the reduced mass, and (*b*) the effective value of the stiffness constant k. Compare to k for the H_2 molecule.

84. A zener diode voltage regulator is shown in Fig. 40–55. Suppose that $R = 2.80\,\text{k}\Omega$ and that the diode breaks down at a reverse voltage of 130 V. (The current increases rapidly at this point, as shown on the far left of Fig. 40–38 at a voltage of -12 V on that diagram.) The diode is rated at a maximum current of 120 mA. (*a*) If $R_{\text{load}} = 21.0\,\text{k}\Omega$, over what range of supply voltages will the circuit maintain the output voltage at 130 V? (*b*) If the supply voltage is 275 V, over what range of load resistance will the voltage be regulated?

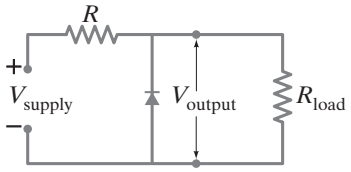


FIGURE 40–55 Problem 84.

85. The average translational kinetic energy of an atom or molecule is about $\overline{K} = \frac{3}{2}kT$ (see Chapter 18), where $k = 1.38 \times 10^{-23}\,\text{J/K}$ is Boltzmann's constant. At what temperature T will $\overline{K}$ be on the order of the bond energy (and hence the bond easily broken by thermal motion) for (*a*) a covalent bond (say H_2) of binding energy 4.0 eV, and (*b*) a "weak" hydrogen bond of binding energy 0.12 eV?

86. When EM radiation is incident on diamond, it is found that light with wavelengths shorter than 226 nm will cause the diamond to conduct. What is the energy gap between the valence band and the conduction band for diamond?

87. A diatomic molecule is found to have an activation energy of 1.4 eV. When the molecule is disassociated, 1.6 eV of energy is released. Draw a potential energy curve for this molecule.

88. Use the uncertainty principle to estimate the binding energy of the H_2 molecule by calculating the difference in kinetic energy of the electrons between (i) when they are in separate atoms and (ii) when they are in the molecule. Take Δx for the electrons in the separated atoms to be the radius of the first Bohr orbit, 0.053 nm, and for the molecule take Δx to be the separation of the nuclei, 0.074 nm. [*Hint*: Let $\Delta p \approx \Delta p_x$.]

89. The band gap of silicon is 1.12 eV. (*a*) For what range of wavelengths will silicon be transparent? (See Example 40–14.) In what region of the electromagnetic spectrum does this transparent range begin? (*b*) If window glass is transparent for all visible wavelengths, what is the minimum possible band gap value for glass (assume $\lambda = 400\,\text{nm}$ to $700\,\text{nm}$)? [*Hint*: If the photon has less energy than the band gap, the photon will pass through the solid without being absorbed.]

90. For a certain semiconductor, the longest wavelength radiation that can be absorbed is 2.04 mm. What is the energy gap in this semiconductor?

ANSWERS TO EXERCISES

A: 0, $5.00 \times 10^{-4}\,\text{eV}$, $1.50 \times 10^{-3}\,\text{eV}$.
B: 0.81 eV.
C: (*c*).
D: (*e*).
E: number/$(\text{J} \cdot \text{m}^3)$.
F: (*a*).
G: (*c*) and (*e*).

91. Assume conduction electrons in a semiconductor behave as an ideal gas. (This is not true for conduction electrons in a metal.) (*a*) Taking mass $m = 9 \times 10^{-31}$ kg and temperature $T = 300$ K, determine the de Broglie wavelength of a semiconductor's conduction electrons. (*b*) Given that the spacing between atoms in a semiconductor's atomic lattice is on the order of 0.3 nm, would you expect room-temperature conduction electrons to travel in straight lines or diffract when traveling through this lattice? Explain.

***92.** A simple picture of an H_2 molecule sharing two electrons is shown in Fig. 40–56. We assume the electrons are symmetrically located between the two protons, which are separated by $r_0 = 0.074$ nm. (*a*) When the electrons are separated by a distance d, write the total potential energy U in terms of d and r_0. (*b*) Make a graph of U in eV as a function of d in nm, and state where U has a minimum on your graph, and for what range of d values U is negative. (*c*) Determine analytically the value of d that gives minimum U (greatest stability).

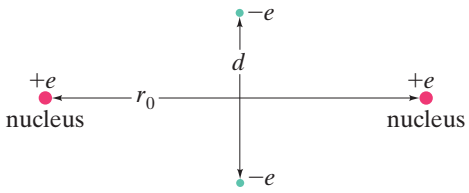
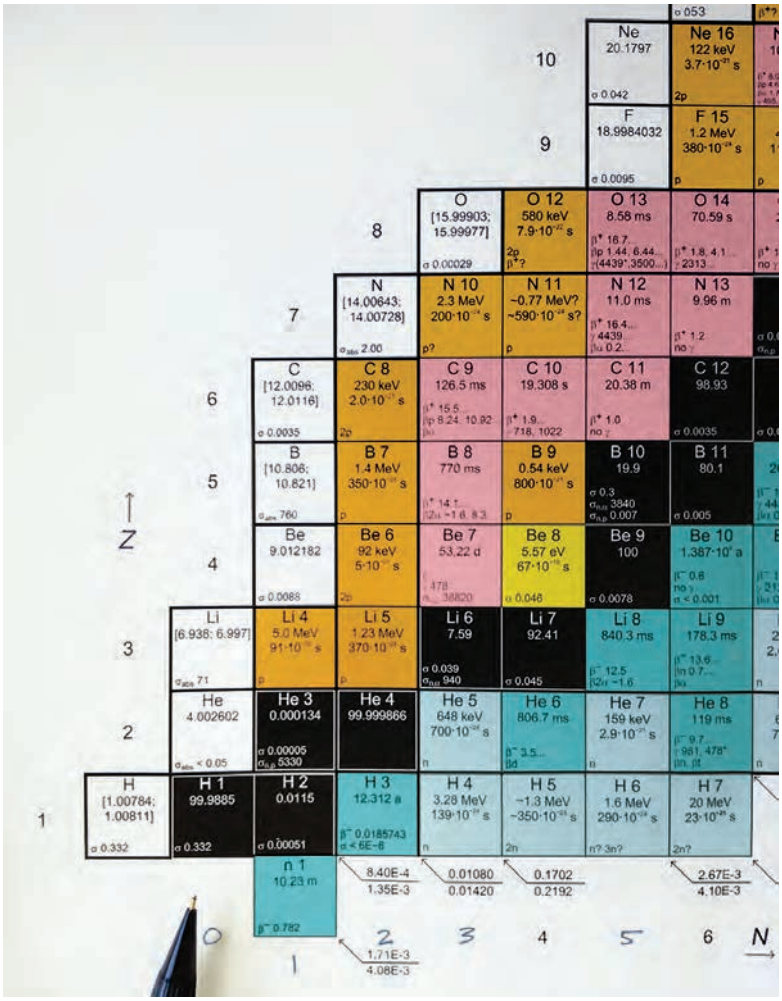


FIGURE 40–56 Problem 92.

Now we begin our discussion of nuclear physics. We study the properties of nuclei, the various forms of radioactivity, and how radioactive decay can be used in a variety of fields to determine the age of old objects, from bones and trees to rocks and other mineral substances, and obtain information on the history of the Earth.

Shown is one version of a **Chart of the Nuclides**. Each horizontal row has a square for each known isotope (nuclide) of one element with a particular Z value (= number of electrons in the neutral atom = number of protons in the nucleus). At the far left is a white box with the average atomic weight (or a range if uncertain) of the naturally occurring isotopes of that element. Each vertical column contains nuclides with the same neutron number N. For $N = 1$ (to right of pencil), starting at the bottom, there is a lone neutron, then above it ${}^{2}_{1}\text{H}$, then ${}^{3}_{2}\text{He}$ and ${}^{4}_{3}\text{Li}$. Each square is color coded: black means a stable nuclide. Radioactive nuclides are blue green for β^- decay, pink for β^+ decay or electron capture (ε) such as ${}^{7}_{4}\text{Be}$, yellow for α decay, and so on. Thus ${}^{1}_{1}\text{H}$ and ${}^{2}_{1}\text{H}$ are stable but ${}^{3}_{1}\text{H}$ (tritium) undergoes β^- decay with half-life = 12.3 years ("a" is for Latin "anno" = year). The squares contain the atomic mass of that isotope, or half-life and energy released if radioactive. Other details may be alternate decay modes and certain cross sections (σ).



CHAPTER 41

Nuclear Physics and Radioactivity

CONTENTS

41 – 1 Structure and Properties of the Nucleus
41 – 2 Binding Energy and Nuclear Forces
41 – 3 Radioactivity
41 – 4 Alpha Decay
41 – 5 Beta Decay
41 – 6 Gamma Decay
41 – 7 Conservation of Nucleon Number and Other Conservation Laws
41 – 8 Half-Life and Rate of Decay
41 – 9 Decay Series
41–10 Radioactive Dating
41–11 Detection of Particles

CHAPTER-OPENING QUESTION—Guess now!

If half of an 80-μg sample of ${}^{60}_{27}\text{Co}$ decays in 5.3 years, how much ${}^{60}_{27}\text{Co}$ is left in 10.6 years?

(a) 10 μg.
(b) 20 μg.
(c) 30 μg.
(d) 40 μg.
(e) 0 μg.

In the early part of the twentieth century, Rutherford's experiments (Section 37–9) led to the idea that at the center of an atom there is a tiny but massive nucleus with a positive charge. At the same time that quantum theory was being developed and scientists were attempting to understand the structure of the atom and its electrons, investigations into the nucleus itself had also begun. In this Chapter and the next, we take a brief look at *nuclear physics*.

41–1 Structure and Properties of the Nucleus

An important question for physicists was whether the nucleus had a structure, and what that structure might be. By the early 1930s, a model of the nucleus had been developed that is still useful. According to this model, a nucleus is made up of two types of particles: protons and neutrons. [These "particles" also have wave properties, but for ease of visualization and language, we usually refer to them simply as "particles."] A **proton** is the nucleus of the simplest atom, hydrogen. The proton has a positive charge $(= +e = +1.60 \times 10^{-19}\,\text{C},$ the same magnitude as for the electron) and its mass is measured to be

$$m_p = 1.67262 \times 10^{-27}\,\text{kg}.$$

The **neutron**, whose existence was ascertained in 1932 by the English physicist James Chadwick (1891–1974), is electrically neutral $(q = 0),$ as its name implies. Its mass is very slightly larger than that of the proton:

$$m_n = 1.67493 \times 10^{-27}\,\text{kg}.$$

These two constituents of a nucleus, neutrons and protons, are referred to collectively as **nucleons**.

Although a normal hydrogen nucleus consists of a single proton alone, the nuclei of all other elements consist of both neutrons and protons. The different nuclei are often referred to as **nuclides**. The number of protons in a nucleus (or nuclide) is called the **atomic number** and is designated by the symbol Z. The total number of nucleons, neutrons plus protons, is designated by the symbol A and is called the **atomic mass number**, or sometimes simply **mass number**. This name is used since the mass of a nucleus is very closely A times the mass of one nucleon. A nuclide with 7 protons and 8 neutrons thus has $Z = 7$ and $A = 15$. The **neutron number** N is $N = A - Z$.

To specify a given nuclide, we need give only A and Z. A special symbol is commonly used which takes the form

$$^{A}_{Z}\text{X},$$

where X is the chemical symbol for the element (see Appendix G, and the Periodic Table inside the back cover), A is the atomic mass number, and Z is the atomic number. For example, $^{15}_{7}\text{N}$ means a nitrogen nucleus containing 7 protons and 8 neutrons for a total of 15 nucleons. In a neutral atom, the number of electrons orbiting the nucleus is equal to the atomic number Z (since the charge on an electron has the same magnitude but opposite sign to that of a proton). The main properties of an atom, and how it interacts with other atoms, are largely determined by the number of electrons. Hence Z determines what kind of atom it is: carbon, oxygen, gold, or whatever. It is redundant to specify both the symbol of a nucleus and its atomic number Z as described above. If the nucleus is nitrogen, for example, we know immediately that $Z = 7$. The subscript Z is thus sometimes dropped and $^{15}_{7}\text{N}$ is then written simply ^{15}N; in words we say "nitrogen fifteen."

For a particular type of atom (say, carbon), nuclei are found to contain different numbers of neutrons, although they all have the same number of protons. For example, carbon nuclei always have 6 protons, but they may have 5, 6, 7, 8, 9, or 10 neutrons. Nuclei that contain the same number of protons but different numbers of neutrons are called **isotopes**. Thus, $^{11}_{6}\text{C}$, $^{12}_{6}\text{C}$, $^{13}_{6}\text{C}$, $^{14}_{6}\text{C}$, $^{15}_{6}\text{C}$, and $^{16}_{6}\text{C}$ are all isotopes of carbon. The isotopes of a given element are not all equally common. For example, 98.9% of naturally occurring carbon (on Earth) is the isotope $^{12}_{6}\text{C}$, and about 1.1% is $^{13}_{6}\text{C}$. These percentages are referred to as the **natural abundances**.[†] Even hydrogen has isotopes: 99.99% of natural hydrogen is $^{1}_{1}\text{H}$, a simple proton, as the nucleus; there are also $^{2}_{1}\text{H}$, called **deuterium**, and $^{3}_{1}\text{H}$, **tritium**, which besides the proton contain 1 or 2 neutrons. (The bare nucleus in each case is called the **deuteron** and **triton**.)

[†]The mass value for each element as given in the Periodic Table (inside back cover) is an average weighted according to the natural abundances of its isotopes.

Many isotopes that do not occur naturally can be produced in the laboratory by means of nuclear reactions (more on this later). Indeed, all elements beyond uranium $(Z > 92)$ do not occur naturally on Earth and are only produced artificially (in the laboratory), as are many nuclides with $Z \leq 92$.

The approximate size of nuclei was determined originally by Rutherford from the scattering of charged particles by thin metal foils (Fig. 37–17). We cannot speak about a definite size for nuclei because of the wave–particle duality (Section 37–6): their spatial extent must remain somewhat fuzzy. Nonetheless a rough "size" can be measured by scattering high-speed electrons off nuclei. It is found that nuclei have a roughly spherical shape with a radius that increases with A according to the approximate formula

$$r \approx (1.2 \times 10^{-15}\,\text{m})(A^{\frac{1}{3}}). \qquad \textbf{(41–1)}$$

Since the volume of a sphere is $V = \frac{4}{3}\pi r^3$, we see that the volume of a nucleus is approximately proportional to the number of nucleons, $V \propto A$ (because $(A^{\frac{1}{3}})^3 = A$). This is what we would expect if nucleons were like impenetrable billiard balls: if you double the number of balls, you double the total volume. Hence, all nuclei have nearly the same density, and it is enormous (see Example 41–2).

The metric abbreviation for $10^{-15}\,\text{m}$ is the fermi (after Enrico Fermi, Fig. 41–8) or the femtometer, fm (Table 1–4 or inside the front cover). Thus $1.2 \times 10^{-15}\,\text{m} = 1.2\,\text{fm}$ or 1.2 fermis.

EXAMPLE 41–1 ESTIMATE **Nuclear sizes.** Estimate the diameter of the smallest and largest naturally occurring nuclei: (*a*) ${}^{1}_{1}\text{H}$, (*b*) ${}^{238}_{92}\text{U}$.

APPROACH The radius r of a nucleus is related to its number of nucleons A by Eq. 41–1. The diameter $d = 2r$.

SOLUTION (*a*) For hydrogen, $A = 1$, Eq. 41–1 gives

$$d = \text{diameter} = 2r \approx 2(1.2 \times 10^{-15}\,\text{m})(A^{\frac{1}{3}}) = 2.4 \times 10^{-15}\,\text{m}$$

since $A^{\frac{1}{3}} = 1^{\frac{1}{3}} = 1$.

(*b*) For uranium $d \approx (2.4 \times 10^{-15}\,\text{m})(238)^{\frac{1}{3}} = 15 \times 10^{-15}\,\text{m}$.

The range of nuclear diameters is only from 2.4 fm to 15 fm.

Because nuclear radii vary as $A^{\frac{1}{3}}$, the largest nuclei, such as uranium with $A = 238$, have a radius only about $\sqrt[3]{238} \approx 6$ times that of the smallest, hydrogen $(A = 1)$.

EXAMPLE 41–2 ESTIMATE **Nuclear and atomic densities.** Compare the density of nuclear matter to the density of normal solids.

APPROACH The density of normal liquids and solids is on the order of 10^3 to $10^4\,\text{kg/m}^3$ (see Table 13–1), and because the atoms are close packed, atoms have about this density too. We therefore compare the density (mass per volume) of a nucleus to that of its atom as a whole.

SOLUTION The mass of a proton is greater than the mass of an electron by a factor

$$\frac{1.67 \times 10^{-27}\,\text{kg}}{9.1 \times 10^{-31}\,\text{kg}} \approx 2000.$$

Thus, over 99.9% of the mass of an atom is in the nucleus, and for our estimate we can say the mass of the atom equals the mass of the nucleus, $m_{\text{nucl}}/m_{\text{atom}} = 1$. Atoms have a radius of about $10^{-10}\,\text{m}$ (Chapter 37) and nuclei on the order of $10^{-15}\,\text{m}$ (Eq. 41–1). Thus the ratio of nuclear density to atomic density is about

$$\frac{\rho_{\text{nucl}}}{\rho_{\text{atom}}} = \frac{(m_{\text{nucl}}/V_{\text{nucl}})}{(m_{\text{atom}}/V_{\text{atom}})} = \left(\frac{m_{\text{nucl}}}{m_{\text{atom}}}\right)\frac{\frac{4}{3}\pi r_{\text{atom}}^3}{\frac{4}{3}\pi r_{\text{nucl}}^3} \approx (1)\,\frac{(10^{-10})^3}{(10^{-15})^3} = 10^{15}.$$

The nucleus is 10^{15} times more dense than ordinary matter.

PSE5_CH41_v1.8_23Jun-20

The masses of nuclei can be determined from the radius of curvature of fast-moving nuclei (as ions) in a known magnetic field using a mass spectrometer, as discussed in Section 27–9. Indeed the existence of different isotopes of the same element (different number of neutrons) was discovered using this device. Nuclear masses can be specified in **unified atomic mass units** (u). On this scale, a neutral ${}^{12}_{6}\text{C}$ atom is given the exact value 12.000000 u. (A ${}^{12}_{6}\text{C}$ nucleus, without electrons, is slightly less than 12.000000.) A neutron then has a measured mass of 1.008665 u, a proton 1.007276 u, and a neutral hydrogen atom ${}^{1}_{1}\text{H}$ (proton plus electron) 1.007825 u. The masses of many nuclides are given in Appendix G. It should be noted that the masses in this Table, as is customary, are for the *neutral atom* (including electrons), and not for a bare nucleus.

CAUTION
Masses are for neutral atom (nucleus plus electrons)

Masses may be specified using the electron-volt energy unit, $1\,\text{eV} = 1.6022 \times 10^{-19}\,\text{J}$ (Section 23–8). This can be done because mass and energy are related, and the precise relationship is given by Einstein's equation $E = mc^2$ (Chapter 36). Since the mass of a proton is $1.67262 \times 10^{-27}\,\text{kg}$, or 1.007276 u, then 1 u is equal to

$$1.0000\,\text{u} = (1.0000\,\text{u})\left(\frac{1.67262 \times 10^{-27}\,\text{kg}}{1.007276\,\text{u}}\right) = 1.66054 \times 10^{-27}\,\text{kg};$$

this is equivalent to an energy (see Table inside front cover) in MeV ($= 10^6\,\text{eV}$) of

$$E = mc^2 = \frac{(1.66054 \times 10^{-27}\,\text{kg})(2.9979 \times 10^8\,\text{m/s})^2}{(1.6022 \times 10^{-19}\,\text{J eV})} = 931.5\,\text{MeV}.$$

The mass of one atomic mass unit is (solving for $m = E/c^2$)

PROBLEM SOLVING
Atomic mass unit in MeV/c^2

$$1\,\text{u} = 1.6605 \times 10^{-27}\,\text{kg} = 931.5\,\text{MeV}/c^2.$$

The rest masses of some of the basic particles are given in Table 41–1. As a rule of thumb, to remember, the masses of neutron and proton are about $1\,\text{GeV}/c^2$ ($= 1000\,\text{MeV}/c^2$) which is about 2000 times the mass of an electron ($\approx \frac{1}{2}\,\text{MeV}/c^2$).

PROBLEM SOLVING
Masses of p and n are $\approx 1\,\text{GeV}/c^2$

TABLE 41–1
Rest Masses in Kilograms, Unified Atomic Mass Units, and MeV/c^2

	Mass		
Object	**kg**	**u**	**MeV/c^2**
Electron	9.1094×10^{-31}	0.00054858	0.51100
Proton	1.67262×10^{-27}	1.007276	938.27
${}^{1}_{1}\text{H}$ atom	1.67353×10^{-27}	1.007825	938.78
Neutron	1.67493×10^{-27}	1.008665	939.57

Just as an electron has intrinsic spin and orbital angular momentum quantum numbers, so too do nuclei and their constituents, the proton and neutron. Both the proton and the neutron are spin-$\frac{1}{2}$ particles, just like the electron. A nucleus, made up of protons and neutrons, has a **nuclear spin** quantum number I that is the vector sum of the spins of all its nucleons (plus any orbital angular momentum), and can be either integer or half integer, depending on whether it is made up of an even or an odd number of nucleons. [Orbital angular momentum is integer and doesn't affect half integer or whole for I.]

PSE5_CH41_v1.8_23Jun-20

The **total nuclear angular momentum** of a nucleus is given, as might be expected (see Section 39–2 and Eq. 39–15), by $\sqrt{I(I+1)}\,\hbar$.

Nuclear magnetic moments are measured in terms of the **nuclear magneton**

$$\mu_N = \frac{e\hbar}{2m_p}, \tag{41–2}$$

which is defined by analogy with the Bohr magneton for electrons ($\mu_B = e\hbar/2m_e$, Section 39–7). Since μ_N contains the proton mass, m_p, instead of the electron mass, it is about 2000 times smaller. The electron spin magnetic moment is about 2 Bohr magnetons. The proton's magnetic moment μ_p has been measured to be

$$\mu_p = 2.7928\,\mu_N.$$

There is no satisfactory explanation for this large factor. The neutron has a magnetic moment

$$\mu_n = -1.9135\,\mu_N,$$

which suggests that, although the neutron carries no net charge, it may have internal structure (quarks, as we discuss later). The minus sign for μ_n indicates that its magnetic moment is opposite to its spin.

Important applications based on nuclear spin are nuclear magnetic resonance (NMR) and magnetic resonance imaging (MRI). They are discussed in the next Chapter (Section 42–10).

41–2 Binding Energy and Nuclear Forces

Binding Energies

The total mass of a stable nucleus is always less than the sum of the masses of its separate protons and neutrons, as the following Example shows.

EXAMPLE 41–3 **$^4_2$He mass compared to its constituents.** Compare the mass of a $^4_2$He atom to the total mass of its constituent particles.

PROBLEM SOLVING
Keep track of electron masses

APPROACH The $^4_2$He nucleus contains 2 protons and 2 neutrons. Tables normally give the masses of neutral atoms—that is, nucleus plus its Z electrons—since this is how masses are measured. We must therefore be sure to balance out the electrons when we compare masses. Thus we use the mass of a $^1_1$H atom rather than that of a proton alone. We look up the mass of the $^4_2$He atom in Appendix G (it includes the mass of 2 electrons), as well as the mass for the 2 neutrons and 2 hydrogen atoms (= 2 protons + 2 electrons).

SOLUTION The mass of a neutral $^4_2$He atom, from Appendix G, is 4.002603 u. The mass of two neutrons and two H atoms (2 protons including the 2 electrons) is

$$\begin{aligned} 2m_n &= 2(1.008665\,\text{u}) = 2.017330\,\text{u} \\ 2m(^1_1\text{H}) &= 2(1.007825\,\text{u}) = 2.015650\,\text{u} \\ \text{sum} &= 4.032980\,\text{u}. \end{aligned}$$

Thus the mass of $^4_2$He is measured to be less than the masses of its constituents by an amount $4.032980\,\text{u} - 4.002603\,\text{u} = 0.030377\,\text{u}$.

Where has this lost mass of 0.030377 u disappeared to? It must be $E = mc^2$.

PSE5_CH41_v1.8_23Jun-20

If the four nucleons suddenly came together to form a $^4_2$He nucleus, the mass "loss" would appear as energy of another kind (such as γ radiation, or kinetic energy). The mass (or energy) difference in the case of $^4_2$He, given in energy units, is $(0.030377\,\text{u})\,(931.5\,\text{MeV/u}) = 28.30\,\text{MeV}$. This difference is referred to as the **total binding energy** of the nucleus. The total binding energy represents the amount of energy that must be put *into* a nucleus in order to break it apart into its constituents. If the mass of, say, a $^4_2$He nucleus were exactly equal to the mass of two neutrons plus two protons, the nucleus could fall apart without any input of energy. To be stable, the mass of a nucleus *must* be less than that of its constituent nucleons, so that energy input *is* needed to break it apart.

CAUTION
Mass of nucleus must be less than mass of constituents

Binding energy is not something a nucleus has—it is energy it "lacks" relative to the total mass of its separate constituents.

As a comparison, we saw in Chapter 37 that the binding energy of the one electron in the hydrogen atom is 13.6 eV; so the mass of a $^1_1$H atom is less than that of a single proton plus a single electron by $13.6\,\text{eV}/c^2$. Compared to the total mass of the hydrogen atom $(939\,\text{MeV}/c^2)$, this is incredibly small, 1 part in 10^8. The binding energies of nuclei are on the order of MeV, so the eV binding energies of electrons can be ignored. Note that nuclear binding energies, compared to nuclear masses, are on the order of $(28\,\text{MeV}/4000\,\text{MeV}) \approx 1 \times 10^{-2}$, where we used helium's binding energy (see above) and mass $\approx 4 \times 940\,\text{MeV} \approx 4000\,\text{MeV}$.

EXERCISE A Determine how much less the mass of the $^7_3$Li nucleus is compared to that of its constituents. See Appendix G.

The difference between the mass of a nucleus, M, and its atomic number A, is called the **mass excess** $\Delta = M - A$. The quantity $A - M$ is called the **mass defect**. We mention this because some references use these quantities instead of M as we do in Appendix G.

The **binding energy per nucleon** is defined as the total binding energy of a nucleus divided by A, the total number of nucleons. We calculated above that the binding energy of $^4_2$He is 28.3 MeV, so its binding energy per nucleon is $28.3\,\text{MeV}/4 = 7.1\,\text{MeV}$. Figure 41–1 shows the measured binding energy per nucleon as a function of A for stable nuclei. The curve rises as A increases and reaches a plateau at about 8.7 MeV per nucleon above $A \approx 40$. Beyond $A \approx 80$, the curve decreases slowly, indicating that larger nuclei are held together a little less tightly than those in the middle of the Periodic Table. We will see later that these characteristics allow the release of nuclear energy in the processes of fission and fusion.

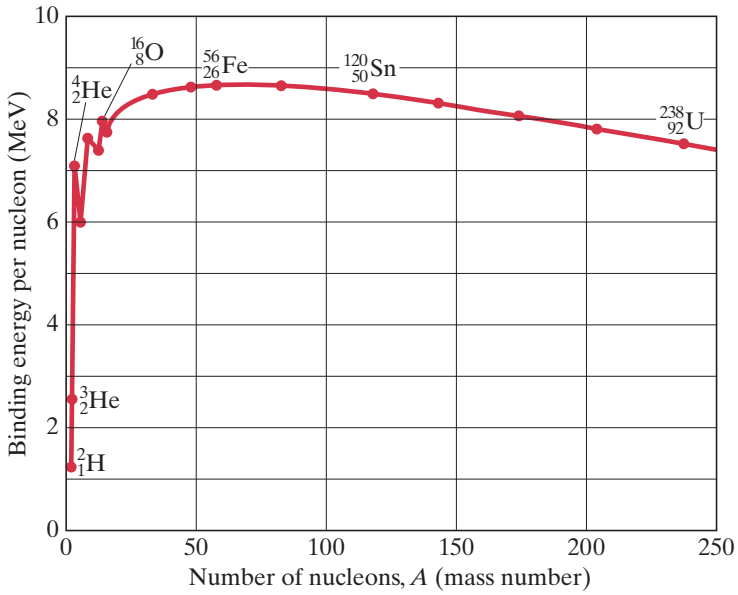


FIGURE 41–1 Binding energy per nucleon for the more stable nuclides as a function of mass number A.

EXAMPLE 41–4 **Binding energy for iron.** Calculate the total binding energy and the binding energy per nucleon for $^{56}_{26}\text{Fe}$, the most common stable isotope of iron.

APPROACH We subtract the mass of an $^{56}_{26}\text{Fe}$ atom from the total mass of 26 hydrogen atoms and 30 neutrons, all found in Appendix G. Then we convert mass units to energy units; finally we divide by $A = 56$, the total number of nucleons.

SOLUTION $^{56}_{26}\text{Fe}$ has 26 protons and 30 neutrons whose separate masses are

$$\begin{aligned} 26m\left(^{1}_{1}\text{H}\right) &= (26)(1.007825\,\text{u}) = 26.20345\,\text{u (includes 26 electrons)} \\ 30m_{\text{n}} &= (30)(1.008665\,\text{u}) = \underline{30.25995\,\text{u}} \\ \text{sum} &= 56.46340\,\text{u}. \\ \text{Subtract mass of } ^{56}_{26}\text{Fe}: &= \underline{-55.93494\,\text{u}}\ \text{(Appendix G)} \\ \Delta m &= 0.52846\,\text{u}. \end{aligned}$$

The total binding energy is thus

$$(0.52846\,\text{u})(931.5\,\text{MeV/u}) = 492.26\,\text{MeV}$$

and the binding energy per nucleon is

$$\frac{492.26\,\text{MeV}}{56\ \text{nucleons}} = 8.79\,\text{MeV}.$$

NOTE The binding energy per nucleon graph (Fig. 41–1) peaks about here, for iron. So the iron nucleus, and its neighbors are the most stable of nuclei.

EXERCISE B Determine the binding energy per nucleon for $^{16}_{8}\text{O}$.

EXAMPLE 41–5 **Binding energy of last neutron.** What is the binding energy of the last neutron in $^{13}_{6}\text{C}$?

APPROACH If $^{13}_{6}\text{C}$ lost one neutron, it would be $^{12}_{6}\text{C}$. We subtract the mass of $^{13}_{6}\text{C}$ from the masses of $^{12}_{6}\text{C}$ and a free neutron.

SOLUTION Obtaining the masses from Appendix G, we have

$$\begin{aligned} \text{Mass } ^{12}_{6}\text{C} &= 12.000000\,\text{u} \\ \text{Mass } ^{1}_{0}\text{n} &= \underline{1.008665\,\text{u}} \\ \text{Total} &= 13.008665\,\text{u}. \\ \text{Subtract mass of } ^{13}_{6}\text{C}: &\quad \underline{-13.003355\,\text{u}} \\ \Delta m &= 0.005310\,\text{u} \end{aligned}$$

which in energy is $(931.5\,\text{MeV/u})(0.005310\,\text{u}) = 4.95\,\text{MeV}$. That is, it would require 4.95 MeV input of energy to remove one neutron from $^{13}_{6}\text{C}$.

Nuclear Forces

We can analyze nuclei not only from the point of view of energy, but also from the point of view of the forces that hold them together. We would not expect a collection of protons and neutrons to come together spontaneously, since protons are all positively charged and thus exert repulsive electric forces on each other. Since stable nuclei *do* stay together, another force must be acting. This new force has to be stronger than the electric force in order to hold the nucleus together, and is called the **strong force** or **strong nuclear force**. The strong nuclear force acts as an attractive force between all nucleons, protons and neutrons alike. Thus protons attract each other via the strong nuclear force at the same time they repel each other via the electric force.

Neutrons, because they are electrically neutral, only attract other neutrons or protons via the strong nuclear force.

The strong nuclear force turns out to be far more complicated than the gravitational and electromagnetic forces. One important aspect of the strong nuclear force is that it is a **short-range** force: it acts only over a very short distance. It is very strong between two nucleons if they are less than about 10^{-15} m apart, but it is essentially zero if they are separated by a distance greater than this. Compare this to electric and gravitational forces, which decrease as $1/r^2$ but continue acting over any distances and are therefore called **long-range** forces.

The strong nuclear force has some strange features. For example, if a nuclide contains too many or too few neutrons relative to the number of protons, the binding of the nucleons is reduced; nuclides that are too unbalanced in this regard are unstable. As shown in Fig. 41–2, stable nuclei tend to have the same number of protons as neutrons $(N = Z)$ up to about $A = 30$. Beyond this, stable nuclei contain more neutrons than protons. This makes sense since, as Z increases, the electrical repulsion increases, so a greater number of neutrons—which exert only the attractive strong nuclear force—are required to maintain stability. For very large Z, no number of neutrons can overcome the greatly increased electric repulsion. Indeed, there are no completely stable nuclides above $Z = 82$.

What we mean by a *stable nucleus* is one that stays together indefinitely. What then is an *unstable nucleus*? It is one that comes apart; and this results in radioactive decay. Before we discuss the important subject of radioactivity (next Section), we note that there is a second type of nuclear force that is much weaker than the strong nuclear force. It is called the **weak nuclear force** (or **weak force**), and we are aware of its existence only because it shows itself in certain types of radioactive decay. These two nuclear forces, the strong and the weak, together with the gravitational and electromagnetic forces, comprise the four fundamental types of force in nature.

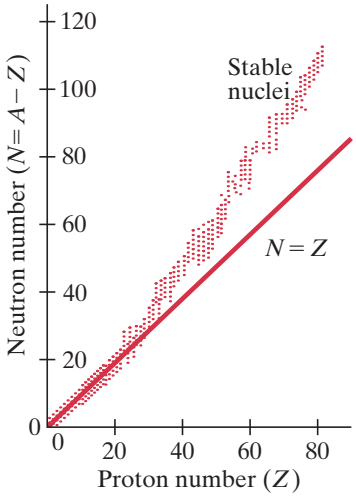


FIGURE 41–2 Number of neutrons versus number of protons for stable nuclides, which are represented by dots. The straight line represents $N = Z$.

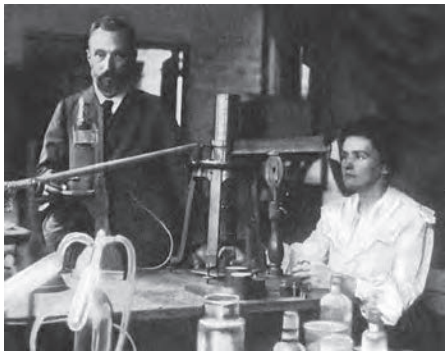
41–3 Radioactivity

Nuclear physics had its beginnings in 1896. In that year, Henri Becquerel (1852–1908) made an important discovery: in his studies of phosphorescence, he found that a certain mineral (which happened to contain uranium) would darken a photographic plate even when the plate was wrapped to exclude light. It was clear that the mineral emitted some new kind of radiation that, unlike X-rays (Section 35–10), occurred without any external stimulus. This new phenomenon eventually came to be called **radioactivity**.

Soon after Becquerel's discovery, Marie Curie (1867–1934) and her husband, Pierre Curie (1859–1906), isolated two previously unknown elements that were very highly radioactive (Fig. 41–3). These were named polonium and radium. Other radioactive elements were soon discovered as well. The radioactivity was found in every case to be unaffected by the strongest physical and chemical treatments, including strong heating or cooling or the action of strong chemicals. It was suspected that the source of radioactivity must be deep within the atom, coming from the nucleus. It became apparent that radioactivity is the result of the **disintegration** or **decay** of an unstable nucleus. Certain isotopes are not stable, and they decay with the emission of some type of radiation or "rays."

Many unstable isotopes occur in nature, and such radioactivity is called "natural radioactivity." Other unstable isotopes can be produced in the laboratory by nuclear reactions (Section 42–1); these are said to be produced "artificially" and to have "artificial radioactivity." Radioactive isotopes are sometimes referred to as **radioisotopes** or **radionuclides**.

FIGURE 41–3 Marie and Pierre Curie in their laboratory (about 1906) where radium was discovered.



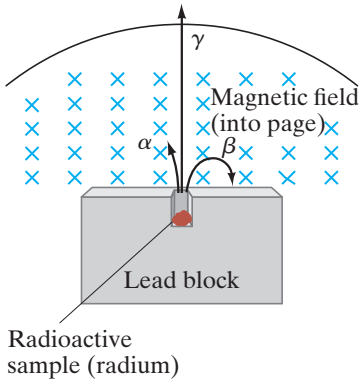


FIGURE 41–4 Alpha and beta rays are bent in opposite directions by a magnetic field, whereas gamma rays are not bent at all.

Rutherford and others began studying the nature of the rays emitted in radioactivity about 1898. They classified the rays into three distinct types according to their penetrating power. One type of radiation could barely penetrate a piece of paper. The second type could pass through as much as 3 mm of aluminum. The third was extremely penetrating: it could pass through several centimeters of lead and still be detected on the other side. They named these three types of radiation alpha (α), beta (β), and gamma (γ), respectively, after the first three letters of the Greek alphabet.

Each type of ray was found to have a different charge and hence is bent differently in a magnetic field, Fig. 41–4; α rays are positively charged, β rays are negatively charged, and γ rays are neutral. It was soon found that all three types of radiation consisted of familiar kinds of particles. Gamma rays are very high-energy *photons* whose energy is even higher than that of X-rays. Beta rays were identified as *electrons*, identical to those that orbit the nucleus, but they are created within the nucleus itself when it decays. Alpha rays (or α particles) are simply the nuclei of *helium* atoms, ${}^4_2\text{He}$; that is, an α ray consists of two protons and two neutrons bound together.

We now discuss each of these three types of radioactivity, or decay, in more detail.

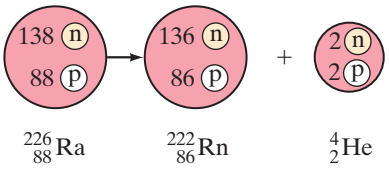
41–4 Alpha Decay

Experiments show that when nuclei decay, the number of nucleons (= mass number A) is conserved, as well as electric charge ($= Ze$). When a nucleus emits an α particle $\left({}^4_2\text{He}\right)$, the remaining nucleus will be different from the original: it has lost two protons and two neutrons. Radium 226 $\left({}^{226}_{88}\text{Ra}\right)$, for example, is an α emitter. It decays to a nucleus with $Z = 88 - 2 = 86$ and $A = 226 - 4 = 222$. The nucleus with $Z = 86$ is radon (Rn)—see Appendix G or the Periodic Table. Thus the radium decays to radon with the emission of an α particle. This is written

$$ {}^{226}_{88}\text{Ra} \rightarrow {}^{222}_{86}\text{Rn} + {}^4_2\text{He}. $$

See Fig. 41–5.

FIGURE 41–5 Radioactive decay of radium to radon with emission of an alpha particle.



When α decay occurs, a different element is formed. The **daughter** nucleus $\left({}^{222}_{86}\text{Rn}\right.$ in this case$\left.\right)$ is different from the **parent** nucleus $\left({}^{226}_{88}\text{Ra}\right.$ in this case$\left.\right)$. This changing of one element into another is called **transmutation** of the elements.

Alpha decay can be written in general as

$$ {}^A_Z\text{N} \rightarrow {}^{A-4}_{Z-2}\text{N}' + {}^4_2\text{He} \qquad [\alpha \text{ decay}] $$

where N is the parent, N' the daughter, and Z and A are the atomic number and atomic mass number, respectively, of the parent.

EXERCISE C ${}^{154}_{66}\text{Dy}$ decays by α emission to what element? (*a*) Pb, (*b*) Gd, (*c*) Sm, (*d*) Er, (*e*) Yb.

Alpha decay occurs because the strong nuclear force is unable to hold very large nuclei together. The nuclear force is a short-range force: it acts only between neighboring nucleons. But the electric force acts all the way across a large nucleus. For very large nuclei, the large Z means the repulsive electric force becomes so large (Coulomb's law) that the strong nuclear force is unable to hold the nucleus together.

We can express the instability of the parent nucleus in terms of energy (or mass): decay cannot occur unless the mass of the parent nucleus is greater than the mass of the daughter nucleus plus the mass of the α particle. The mass difference appears as kinetic energy, which is carried away by the α particle and the recoiling daughter nucleus. The total energy released is called the **disintegration energy**, Q, or the ***Q*-value** of the decay. From conservation of energy,

$$ M_\text{P}c^2 = M_\text{D}c^2 + m_\alpha c^2 + Q, $$

where Q equals the kinetic energy of the daughter and α particle, and M_P, M_D, and m_α are the masses of the parent, daughter, and α particle, respectively.

PSE5_CH41_v1.8_23Jun-20

Thus

$$Q = M_P c^2 - (M_D + m_\alpha)c^2. \quad \textbf{(41–3)}$$

If the parent had *less* mass than the daughter plus the α particle (so $Q < 0$), the decay would violate conservation of energy. Such decays have never been observed, another confirmation of this great conservation law.

EXAMPLE 41–6 **Uranium decay energy release.** Calculate the disintegration energy when $^{232}_{92}\text{U}$ (mass = 232.037156 u) decays to $^{228}_{90}\text{Th}$ (228.028741 u) with the emission of an α particle. (As always, masses given are for neutral atoms.)

APPROACH We use conservation of energy as expressed in Eq. 41–3. $^{232}_{92}\text{U}$ is the parent, $^{228}_{90}\text{Th}$ is the daughter.

SOLUTION Since the mass of the $^{4}_{2}\text{He}$ is 4.002603 u (Appendix G), the total mass in the final state $(m_{Th} + m_{He})$ is

$$228.028741\text{ u} + 4.002603\text{ u} = 232.031344\text{ u}.$$

The mass lost when the $^{232}_{92}\text{U}$ decays $(m_U - m_{Th} - m_{He})$ is

$$232.037156\text{ u} - 232.031344\text{ u} = 0.005812\text{ u}.$$

Because 1 u = 931.5 MeV, the energy Q released is

$$Q = (0.005812\text{ u})(931.5\text{ MeV/u}) = 5.4\text{ MeV}$$

and this energy appears as kinetic energy of the α particle and the daughter nucleus.

EXAMPLE 41–7 **Kinetic energy of the α in $^{232}_{92}\text{U}$ decay.** For the $^{232}_{92}\text{U}$ decay of Example 41–6, how much of the 5.4-MeV disintegration energy will be carried off by the α particle?

APPROACH In any reaction, momentum must be conserved as well as energy. The given 5.4 MeV $\ll$ masses of U and the α, so we can use classical expression for kinetic energy.

SOLUTION Before disintegration, assuming the nucleus to be at rest, the total momentum was zero. After disintegration, the total vector momentum must still be zero so the magnitude of the α particle's momentum must equal the magnitude of the daughter's momentum (Fig. 41–6):

$$m_\alpha v_\alpha = m_D v_D.$$

Thus $v_\alpha = m_D v_D / m_\alpha$ and the α's kinetic energy is

$$K_\alpha = \tfrac{1}{2} m_\alpha v_\alpha^2 = \tfrac{1}{2} m_\alpha \left(\frac{m_D v_D}{m_\alpha}\right)^2 = \tfrac{1}{2} m_D v_D^2 \left(\frac{m_D}{m_\alpha}\right) = \left(\frac{m_D}{m_\alpha}\right) K_D$$

$$= \left(\frac{228.028741\text{ u}}{4.002603\text{ u}}\right) K_D = 57 K_D.$$

The total disintegration energy is $Q = K_\alpha + K_D = 57K_D + K_D = 58K_D$. Hence

$$K_\alpha = 57 K_D = \frac{57}{58} Q = 5.3\text{ MeV}.$$

The lighter α particle carries off (57/58) or 98% of the total kinetic energy. The total energy released is 5.4 MeV, so the daughter nucleus, which recoils in the opposite direction, carries off only 0.1 MeV.

NOTE The 5.4 MeV ***

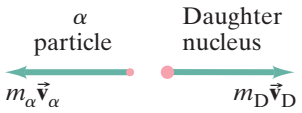


FIGURE 41–6 Momentum conservation in Example 41–7.

α-Decay Theory—Tunneling

If the mass of the daughter nucleus plus the mass of the α particle is less than the mass of the parent nucleus (so the parent is energetically allowed to decay), why are there any parent nuclei at all? That is, why haven't radioactive nuclei all decayed long ago, right after they were formed? Why doesn't an unstable nucleus fall apart immediately? Why do many have such long half-lives? Consider a model of a nucleus that has an alpha particle trapped inside it. The potential energy "seen" by the α particle would have a shape like that shown in Fig. 41–7.

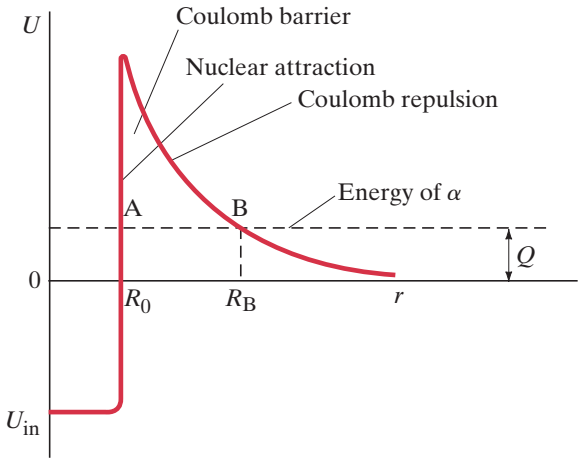


FIGURE 41–7 Potential energy for alpha particle and (daughter) nucleus, showing the Coulomb barrier through which the α particle must tunnel to escape. The Q-value of the reaction is also shown. This plot assumes spherical symmetry, so the central well has diameter $2R_0$.

The potential energy well (approximately square) between $r = 0$ and $r = R_0$ represents the short-range attractive nuclear force. Beyond the nuclear radius, R_0, the Coulomb repulsion dominates (since the nuclear force drops to zero) and we see the characteristic $1/r$ dependence of the Coulomb potential. The α particle, trapped within the nucleus, can be thought of as moving back and forth between the potential walls. Since the potential energy just beyond $r = R_0$ is greater than the energy of the α particle (dashed line), the α particle could not escape the nucleus if it were governed by classical physics. But according to quantum mechanics, there is a nonzero probability that the α particle can **tunnel** through the **Coulomb barrier**, from point A to point B in Fig. 41–7, as we discussed in Section 38–10. The height and width of the barrier affect the rate at which nuclei decay (Section 41–8). Because of this barrier, the lifetimes of α-unstable nuclei can be quite long, from a fraction of a microsecond to over 10^{10} years. Note in Fig. 41–7 that the Q-value represents the total kinetic energy when the α particle is far from the nucleus.

We can look at tunneling via the uncertainty principle which tells us that energy conservation can be violated by an amount ΔE for a length of time Δt given by

$$(\Delta E)(\Delta t) \approx \frac{h}{2\pi}.$$

Thus quantum mechanics allows conservation of energy to be violated for brief periods that may be long enough for an α particle to "tunnel" through the barrier. ΔE would represent the energy difference between the average barrier height and the particle's energy, and Δt the time to pass through the barrier. The height and width of this barrier controls the rate of decay and half-life of an isotope. The higher and wider the barrier, the less time an α particle has to escape (and less likely to do so) so the rate of decay would be lower and the half-life longer.

Why α Particles?

Why, you may wonder, do nuclei emit this combination of four nucleons called an α particle? Why not just four separate nucleons, or even one? The answer

PSE5_CH41_v1.8_23Jun-20

is that the α particle is very strongly bound, so that its mass is significantly less than that of four separate nucleons. That helps the final state in α decay to have less total mass, thus allowing certain nuclides to decay which could not decay to, say, 2 protons plus 2 netrons. For example, $^{232}_{92}\text{U}$ could not decay to $2\text{p} + 2\text{n}$ because the masses of the daughter $^{228}_{90}\text{Th}$ plus four separate nucleons is $228.028741\ \text{u} + 2(1.007825\ \text{u}) + 2(1.008665\ \text{u}) = 232.061721\ \text{u}$, which is greater than the mass of the $^{232}_{92}\text{U}$ parent (232.037156 u). Such a decay would violate the conservation of energy. Indeed, we have never seen $^{232}_{92}\text{U} \rightarrow {}^{228}_{90}\text{Th} + 2\text{p} + 2\text{n}$. Similarly, it is almost always true that the emission of a single nucleon is energetically not possible; see Example 41–5.

Smoke Detectors—An Application

One widespread application of nuclear physics is present in nearly every home in the form of an ordinary **smoke detector**. One type of smoke detector contains about 0.2 mg of the radioactive americium isotope, $^{241}_{95}\text{Am}$, in the form of AmO_2. The α radiation continually ionizes the nitrogen and oxygen molecules in the air space between two oppositely charged plates. The resulting conductivity allows a small steady electric current. If smoke enters, the radiation is absorbed by the smoke particles rather than by the air molecules, thus reducing the current. The current drop is detected by the device's electronics and sets off the alarm. The radiation dose that escapes from an intact americium smoke detector is much less than the natural radioactive background, and so can be considered relatively harmless. There is no question that smoke detectors save lives and reduce property damage.

PHYSICS APPLIED
Smoke detector

41–5 Beta Decay

β^- Decay

Transmutation of elements also occurs when a nucleus decays by β decay—that is, with the emission of an electron or β^- particle. The nucleus $^{14}_{6}\text{C}$, for example, emits an electron when it decays:

$$^{14}_{6}\text{C} \rightarrow {}^{14}_{7}\text{N} + \text{e}^- + \text{neutrino},$$

where e^- is the symbol for the electron. The particle known as the neutrino has charge $q = 0$ and a very small mass, long thought to be zero. It was not initially detected and was only later hypothesized to exist, as we shall discuss later in this Section. No nucleons are lost when an electron is emitted, and the total number of nucleons, A, is the same in the daughter nucleus as in the parent. But because an electron has been emitted from the nucleus itself, the charge on the daughter nucleus is $+1e$ greater than that on the parent. The parent nucleus in the decay written above had $Z = +6$, so from charge conservation the nucleus remaining behind must have a charge of $+7e$. So the daughter nucleus has $Z = 7$, which is nitrogen.

It must be carefully noted that the electron emitted in β decay is *not* an orbital electron. Instead, the electron is created *within the nucleus itself*. What happens is that one of the neutrons changes to a proton and in the process (to conserve charge) emits an electron. Indeed, free neutrons (whose mass is a bit more than a proton's) actually do decay in this fashion:

$$\text{n} \rightarrow \text{p} + \text{e}^- + \text{neutrino}.$$

To remind us of their origin in the nucleus, the electrons emitted in β decay are often referred to as "β particles." They are, nonetheless, indistinguishable from orbital electrons.

CAUTION
β-decay e^- comes from nucleus (it is not an orbital electron)

EXAMPLE 41–8 **Energy release in $^{14}_{6}$C decay.** How much energy is released when $^{14}_{6}$C decays to $^{14}_{7}$N by β emission?

APPROACH We find the mass difference before and after decay, Δm. The energy released is $E = (\Delta m)c^2$. The masses given in Appendix G are those of the neutral atom, and we have to keep track of the electrons involved. Assume the parent nucleus has six orbiting electrons so it is neutral; its mass is 14.003242 u. The daughter in this decay, $^{14}_{7}$N, is not neutral since it has the same six orbital electrons circling it but the nucleus has a charge of $+7e$. However, the mass of this daughter with its six electrons, plus the mass of the emitted electron (which makes a total of seven electrons), is just the mass of a neutral nitrogen atom (including its 7 electrons).

SOLUTION The total mass in the final state is

$$(\text{mass of } ^{14}_{7}\text{N nucleus} + 6 \text{ electrons}) + (\text{mass of 1 electron}),$$

and this is equal to

$$\text{mass of neutral } ^{14}_{7}\text{N (includes 7 electrons)},$$

which from Appendix G is a mass of 14.003074 u. So the mass difference is $14.003242\text{ u} - 14.003074\text{ u} = 0.000168\text{ u}$, which is equivalent to an energy change $\Delta m\,c^2 = (0.000168\text{ u})\,(931.5\text{ MeV/u}) = 0.156\text{ MeV}$ or 156 keV. This energy goes into the kinetic energy of the electron and neutrino.

NOTE The neutrino doesn't contribute to either the mass or charge balance because it has $q = 0$ and $m \approx 0$.

CAUTION
Be careful with atomic and electron masses in β decay

According to Example 41–8, we would expect the emitted electron to have a kinetic energy of 156 keV. (The daughter nucleus, because its mass is very much larger than that of the electron, recoils with very low velocity and hence gets very little of the kinetic energy—see Example 41–7.) Indeed, very careful measurements indicate that a few emitted β particles do have kinetic energy close to this calculated value. But the vast majority of emitted electrons have somewhat less energy. In fact, the energy of the emitted electron can be anywhere from zero up to the maximum value as calculated above. This range of electron kinetic energy was found for any β decay. It was as if the law of conservation of energy was being violated, and Bohr actually considered this possibility. Careful experiments indicated that linear momentum and angular momentum also did not seem to be conserved. Physicists were troubled at the prospect of giving up these laws, which had worked so well in all previous situations.

In 1930, Wolfgang Pauli proposed an alternate solution: perhaps a new particle that was very difficult to detect was emitted during β decay in addition to the electron. This hypothesized particle could be carrying off the energy, momentum, and angular momentum required to maintain the conservation laws. This new particle was named the **neutrino**—meaning "little neutral one"—by the great Italian physicist Enrico Fermi (1901–1954; Fig. 41–8), who in 1934 worked out a detailed theory of β decay. (It was Fermi who, in this theory, postulated the existence of the fourth force in nature which we call the *weak nuclear force*.) The neutrino has zero charge, spin of $\frac{1}{2}\hbar$, and was long thought to have zero mass, although today we are quite sure that it has a very tiny mass $(< 0.14\,\text{eV}/c^2)$. If its mass were zero, it would be much like a photon in that it is neutral and would travel at the speed of light. But the neutrino is very difficult to detect. In 1956, complex experiments produced further evidence for the existence of the neutrino; but by then, most physicists had already accepted its existence.

The symbol for the neutrino is the Greek letter nu (ν). The correct way of writing the decay of $^{14}_{6}$C is then

$$^{14}_{6}\text{C} \rightarrow {}^{14}_{7}\text{N} + \text{e}^- + \bar{\nu}.$$

The bar $(^-)$ over the neutrino symbol is to indicate that it is an "antineutrino."



FIGURE 41–8 Enrico Fermi, as portrayed on a US postage stamp. Fermi contributed significantly to both *theoretical* and *experimental* physics, a feat almost unique in modern times:
theory: statistical theory of identical particles that obey the exclusion principle (= fermions); theory of the weak interaction and β decay;
experiments: neutron physics; induced radioactivity and new elements; first nuclear reactor; first resonance of particle physics;
led and inspired a vast amount of other nuclear research.

(Why this is called an antineutrino rather than simply a neutrino is discussed in Chapter 43.)

In β decay, including β^+ decay and EC, it is the weak nuclear force that plays the crucial role. The neutrino is unique in that it interacts with matter only via the weak force, which is why it is so hard to detect.

β^+ Decay

Many isotopes decay by electron emission. They are always isotopes that have too many neutrons compared to the number of protons. That is, they are isotopes that lie above the stable isotopes plotted in Fig. 41–2. But what about unstable isotopes that have too few neutrons compared to their number of protons—those that fall below the stable isotopes of Fig. 41–2? These, it turns out, decay by emitting a **positron** instead of an electron. A positron (sometimes called an e^+ or β^+ particle) has the same mass as the electron, but it has a positive charge of $+1e$. Because it is so like an electron, except for its charge, the positron is called the **antiparticle**[†] to the electron. An example of a β^+ decay is that of $^{19}_{10}$Ne:

$$^{19}_{10}\text{Ne} \rightarrow {}^{19}_{9}\text{F} + \text{e}^+ + \nu,$$

where e^+ stands for a positron. Note that the ν emitted here is a neutrino, whereas that emitted in β^- decay is called an antineutrino. Thus an antielectron (= positron) is emitted with a neutrino, whereas an antineutrino is emitted with an electron; this gives a certain balance as discussed in Chapter 43.

We can write β^- and β^+ decay, in general, as follows:

$$^{A}_{Z}\text{N} \rightarrow {}^{A}_{Z+1}\text{N}' + \text{e}^- + \bar{\nu} \qquad [\beta^- \text{ decay}]$$

$$^{A}_{Z}\text{N} \rightarrow {}^{A}_{Z-1}\text{N}' + \text{e}^+ + \nu, \qquad [\beta^+ \text{ decay}]$$

where N is the parent nucleus and N′ is the daughter.

Electron Capture

Besides β^- and β^+ emission, there is a third related process. This is **electron capture** (abbreviated EC) and occurs when a nucleus absorbs one of its orbiting electrons. An example is $^{7}_{4}$Be, which as a result becomes $^{7}_{3}$Li. The process is written

$$^{7}_{4}\text{Be} + \text{e}^- \rightarrow {}^{7}_{3}\text{Li} + \nu,$$

or, in general,

$$^{A}_{Z}\text{N} + \text{e}^- \rightarrow {}^{A}_{Z-1}\text{N}' + \nu. \qquad [\text{electron capture}]$$

Usually it is an electron in the innermost (K) shell that is captured, in which case the process is called **K-capture**. The electron disappears in the process, and a proton in the nucleus becomes a neutron; a neutrino is emitted as a result. This process is inferred experimentally by detection of emitted X-rays (due to other electrons jumping down to fill the state of the captured e^-).

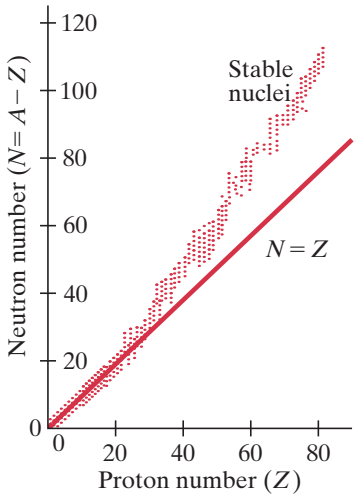


FIGURE 41–2 (Repeated.) Number of neutrons versus number of protons for stable nuclides, which are represented by dots. The straight line represents $N = Z$.

41–6 Gamma Decay

Gamma rays are photons having very high energy. They have their origin in the decay of a nucleus, much like emission of photons by excited atoms. Like an atom, a nucleus itself can be in an excited state. When it jumps down to a lower energy state, or to the ground state, it emits a photon which we call a γ ray. The possible states of a nucleus are much farther apart in energy than those of an atom: on the order of keV or MeV, as compared to a few eV for electrons in an atom. Hence, the emitted photons have energies that can range from a few keV to several MeV. For a given decay, the γ ray always has the same energy. Since a γ ray carries no charge, there is no change in the element as a result of a γ decay.

[†]Discussed in Chapter 43. Briefly, an antiparticle has the same mass as its corresponding particle, but opposite charge. A particle and its antiparticle can quickly annihilate each other, releasing energy in the form of two γ rays: $e^+ + e^- \rightarrow 2\gamma$.

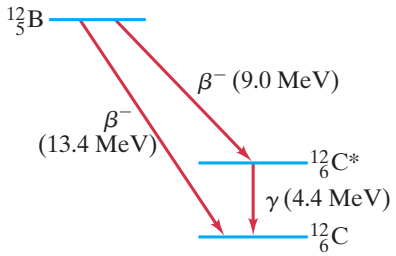


FIGURE 41–9 Energy-level diagram showing how $^{12}_{5}$B can decay to the ground state of $^{12}_{6}$C by β decay (total energy released = 13.4 MeV), or can instead β decay to an excited state of $^{12}_{6}$C (indicated by *), which subsequently decays to its ground state by emitting a 4.4-MeV γ ray.

How does a nucleus get into an excited state? It may occur because of a violent collision with another particle. More commonly, the nucleus remaining after a previous radioactive decay may be in an excited state. A typical example is shown in the energy-level diagram of Fig. 41–9. $^{12}_{5}$B can decay by β decay directly to the ground state of $^{12}_{6}$C; or it can go by β decay to an excited state of $^{12}_{6}$C, written $^{12}_{6}$C*, which itself decays by emission of a 4.4-MeV γ ray to the ground state of $^{12}_{6}$C.

We can write γ decay as

$$^{A}_{Z}\mathrm{N}^* \longrightarrow {}^{A}_{Z}\mathrm{N} + \gamma, \qquad [\gamma \text{ decay}]$$

where the asterisk means "excited state" of that nucleus.

What, you may wonder, is the difference between a γ ray and an X-ray? They both are electromagnetic radiation (photons) and, though γ rays usually have higher energy than X-rays, their range of energies overlap to some extent. The difference is not intrinsic. We use the term X-ray if the photon is produced by an electron–atom interaction, and γ ray if the photon is produced in a nuclear process.

Isomers; Internal Conversion

In some cases, a nucleus may remain in an excited state for some time before it emits a γ ray. The nucleus is then said to be in a **metastable state** and is called an **isomer**.

An excited nucleus can sometimes return to the ground state by another process known as **internal conversion** with no γ ray emitted. In this process, the excited nucleus interacts with one of the orbital electrons and ejects this electron from the atom with the same kinetic energy (minus the binding energy of the electron) that an emitted γ ray would have had.

TABLE 41–2 The Three Types of Radioactive Decay

α decay:

$$^{A}_{Z}\mathrm{N} \longrightarrow {}^{A-4}_{Z-2}\mathrm{N}' + {}^{4}_{2}\mathrm{He}$$

β decay:

$$^{A}_{Z}\mathrm{N} \longrightarrow {}^{A}_{Z+1}\mathrm{N}' + \mathrm{e}^- + \bar{\nu}$$

$$^{A}_{Z}\mathrm{N} \longrightarrow {}^{A}_{Z-1}\mathrm{N}' + \mathrm{e}^+ + \nu$$

$$^{A}_{Z}\mathrm{N} + \mathrm{e}^- \longrightarrow {}^{A}_{Z-1}\mathrm{N}' + \nu \; [\mathrm{EC}]^{\dagger}$$

γ decay:

$$^{A}_{Z}\mathrm{N}^* \longrightarrow {}^{A}_{Z}\mathrm{N} + \gamma$$

†Electron capture.
*Indicates the excited state of a nucleus.

41–7 Conservation of Nucleon Number and Other Conservation Laws

In all three types of radioactive decay, the classical conservation laws hold. Energy, linear momentum, angular momentum, and electric charge are all conserved. These quantities are the same before the decay as after. But a new conservation law is also revealed, the **law of conservation of nucleon number**. According to this law, the total number of nucleons (A) remains constant in any process, although one type can change into the other type (protons into neutrons or vice versa). This law holds in all three types of decay. [In Chapter 43 we will generalize this and call it conservation of baryon number.]

Table 41–2 gives a summary of α, β, and γ decay.

41–8 Half-Life and Rate of Decay

A macroscopic sample of any radioactive isotope consists of a vast number of radioactive nuclei. These nuclei do not all decay at one time. Rather, they decay one by one over a period of time. This is a random process: we can not predict exactly when a given nucleus will decay. But we can determine, on a probabilistic basis, approximately how many nuclei in a sample will decay over a given time period, by assuming that each nucleus has the same probability of decaying in each second that it exists.

The number of decays ΔN that occur in a very short time interval Δt is then proportional to Δt and to the total number N of radioactive nuclei present:

$$\Delta N = -\lambda N \, \Delta t \qquad \textbf{(41–4a)}$$

where the minus sign means N is decreasing. We rewrite this to get the **rate of decay** (number of decays per second):

$$\frac{\Delta N}{\Delta t} = -\lambda N. \qquad \textbf{(41–4b)}$$

In these equations, λ is a measurable constant called the **decay constant**, which is different for different isotopes. The greater λ is, the greater the rate of decay, $\Delta N/\Delta t$,

PSE5_CH41_v1.8_23Jun-20

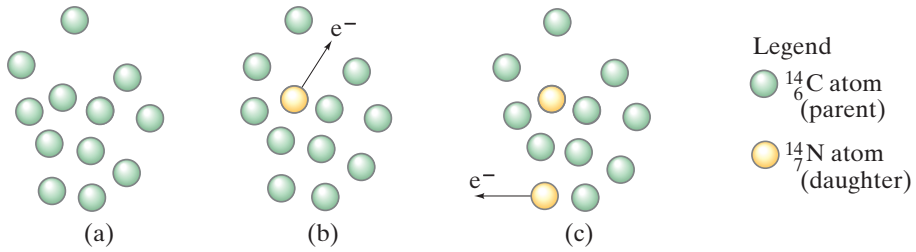


FIGURE 41–10 Radioactive nuclei decay one by one. Hence, the number of parent nuclei in a sample is continually decreasing. When a $^{14}_{6}$C nucleus emits an electron, the nucleus becomes a $^{14}_{7}$N nucleus.

and the more "radioactive" that isotope is said to be. The number of decays that occur in the short time interval Δt is designated ΔN because each decay that occurs corresponds to a decrease by one in the number N of parent nuclei present. That is, radioactive decay is a "one-shot" process, Fig. 41–10. Once a particular parent nucleus decays into its daughter, it cannot do it again.

If we take the limit $\Delta t \to 0$ in Eq. 41–4, ΔN will be small compared to N, and we can write the equation in infinitesimal form as

$$dN = -\lambda N\, dt. \qquad \textbf{(41–5)}$$

We can determine N as a function of t by rearranging this equation to

$$\frac{dN}{N} = -\lambda\, dt$$

and then integrating from $t = 0$ to $t = t$:

$$\int_{N_0}^{N} \frac{dN}{N} = -\int_0^t \lambda\, dt,$$

where N_0 is the number of parent nuclei present at $t = 0$ and N is the number remaining at time t. The integration gives

$$\ln \frac{N}{N_0} = -\lambda t$$

or

$$N = N_0 e^{-\lambda t}. \qquad \textbf{(41–6)}$$

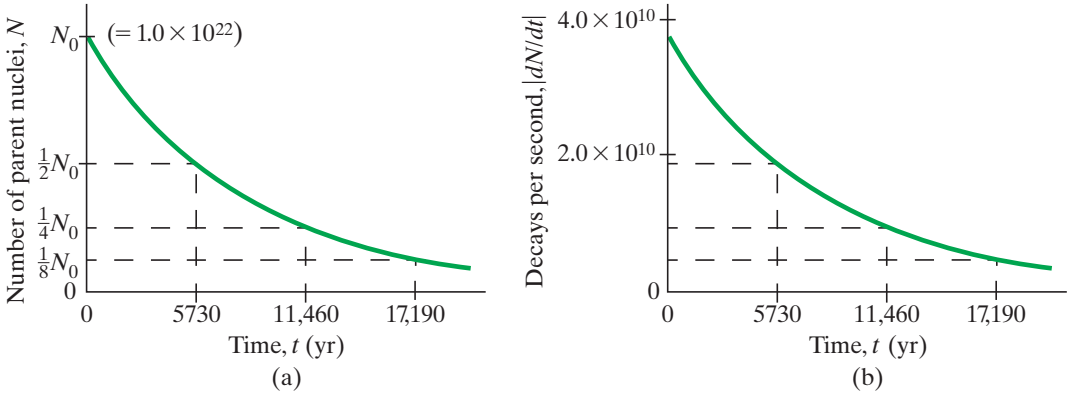
Equation 41–6 is called the **radioactive decay law**. It tells us that the number of radioactive nuclei in a given sample decreases exponentially in time. This is shown in Fig. 41–11a for the case of $^{14}_{6}$C whose decay constant is $\lambda = 3.83 \times 10^{-12}\ \text{s}^{-1}$.

The rate of decay in a pure sample, or number of decays per second, is

$$\left|\frac{dN}{dt}\right|,$$

also called the **activity** of the sample. We use absolute value signs to make activity a positive number (dN/dt is negative because the number of parent nuclei N is decreasing). The symbol R is also used for activity, $R = |dN/dt|$.

FIGURE 41–11 (a) The number N of parent nuclei in a given sample of $^{14}_{6}$C decreases exponentially. We assume a sample that has $N_0 = 1.00 \times 10^{22}$ parent nuclei. (b) The number of decays per second also decreases exponentially. The half-life (Eq. 41–8) of $^{14}_{6}$C is 5730 yr, which means that the number of parent nuclei, N, and the rate of decay, $|dN/dt|$, decrease by half every 5730 yr.



From Eqs. 41–5 and 41–6,

$$\left|\frac{dN}{dt}\right| = \lambda N = \lambda N_0 e^{-\lambda t}. \tag{41–7a}$$

At $t = 0$, the activity is

$$\left|\frac{dN}{dt}\right|_0 = \lambda N_0. \tag{41–7b}$$

Hence, at any other time t the activity is

$$\left|\frac{dN}{dt}\right| = \left|\frac{dN}{dt}\right|_0 e^{-\lambda t}, \tag{41–7c}$$

so the activity decreases exponentially in time at the same rate as for N (Fig. 41–11b). Equation 41–7c is sometimes referred to as the **radioactive decay law** (so is Eq. 41–6), and can be written using R to represent activity, $R = |dN/dt|$. The activity R of a pure sample after a time t is

$$R = \left|\frac{dN}{dt}\right| = R_0 e^{-\lambda t}. \tag{41–7d}$$

Half-life

The rate of decay of any isotope is often specified by giving its half-life rather than the decay constant λ. The **half-life** of an isotope is defined as the time it takes for half the original amount of the parent isotope in a given sample to decay. For example, the half-life of $^{14}_{6}\text{C}$ is about 5730 years. If at some time a piece of petrified wood contains, say, 1.00×10^{22} nuclei of $^{14}_{6}\text{C}$, then 5730 years later it will contain half as many, 0.50×10^{22} parent nuclei. After another 5730 years it will contain 0.25×10^{22} nuclei, and so on. This is characteristic of the exponential function, and is shown in Fig. 41–11a. Since the rate of decay $|dN/dt|$ is proportional to N, it too decreases by a factor of 2 every half-life, Fig. 41–11b.

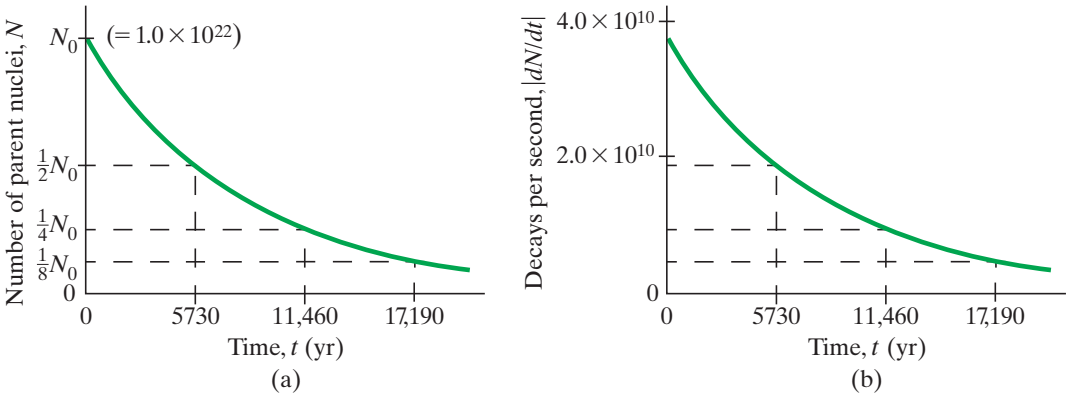


FIGURE 41–11 (Repeated.) (a) The number N of parent nuclei in a given sample of $^{14}_{6}\text{C}$ decreases exponentially. We assume a sample that has $N_0 = 1.00 \times 10^{22}$ parent nuclei. (b) The number of decays per second also decreases exponentially. The half-life of (Eq. 41–8) of $^{14}_{6}\text{C}$ is 5730 yr, which means that the number of parent nuclei, N, and the rate of decay, $|dN/dt|$, decrease by half every 5730 yr.

EXERCISE D The half-life of $^{22}_{11}\text{Na}$ is 2.6 years. How much $^{22}_{11}\text{Na}$ will be left of a pure 1.0-μg sample of $^{22}_{11}\text{Na}$ after 7.8 yr? (*a*) None. (*b*) $\frac{1}{8}\,\mu$g. (*c*) $\frac{1}{4}\,\mu$g. (*d*) $\frac{1}{2}\,\mu$g. (*e*) 0.693 μg.

EXERCISE E Return to the Chapter-Opening Question, page 1226, and answer it again now. Try to explain why you may have answered differently the first time.

The half-lives of known radioactive isotopes vary from very short $(\approx 10^{-22}\,\text{s})$ to more than 10^{23} yr $(> 10^{30}\,\text{s})$. The half-lives of many isotopes are given in Appendix G. The half-life (which we designate $T_{\frac{1}{2}}$) bears an inverse relationship

PSE5_CH41_v1.8_23Jun-20

to the decay constant. The longer the half-life of an isotope, the more slowly it decays, and hence λ is smaller. Conversely, very active isotopes (large λ) have very short half-lives. The precise relation is obtained from Eq. 41–6 by setting $N = N_0/2$ at $t = T_{\frac{1}{2}}$:

$$\frac{N_0}{2} = N_0 e^{-\lambda T_{\frac{1}{2}}} \quad \text{or} \quad e^{\lambda T_{\frac{1}{2}}} = 2.$$

We take natural logs of both sides ("ln" and "e" are inverse operations, meaning $\ln(e^x) = x$) and find

$$\ln\left(e^{\lambda T_{\frac{1}{2}}}\right) = \ln 2, \quad \text{so} \quad \lambda T_{\frac{1}{2}} = \ln 2$$

and

$$T_{\frac{1}{2}} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda}. \tag{41–8}$$

We can then write Eq. 41–6 as

$$N = N_0 e^{-0.693 t/T_{\frac{1}{2}}}.$$

Sometimes the **mean life** τ of an isotope is quoted, which is defined as $\tau = 1/\lambda$ (see also Problem 86), so then Eq. 41–6 can be written

$$N = N_0 e^{-t/\tau}$$

just as for RC and LR circuits (Chapters 26 and 30, where τ is called the time constant). Since

$$\tau = \frac{1}{\lambda} = \frac{T_{\frac{1}{2}}}{0.693} \tag{41–9a}$$

the mean life and half-life differ by a factor 0.693, so confusing them can cause serious error. The radioactive decay law, Eq. 41–7d, can be written as

$$R = R_0 e^{-t/\tau}. \tag{41–9b}$$

CAUTION
Do not confuse half-life and mean life

EXAMPLE 41–9 **Sample activity.** The isotope $^{14}_{6}\text{C}$ has a half-life of 5730 yr. If a sample contains 1.00×10^{22} carbon-14 nuclei, what is the activity of the sample?

APPROACH We first use the half-life to find the decay constant (Eq. 41–8), and use that to find the activity, Eq. 41–7b or 41–5. The number of seconds in a year is $(60\,\text{s/min})(60\,\text{min/h})(24\,\text{h/d})\left(365\tfrac{1}{4}\,\text{d/yr}\right) = 3.156 \times 10^7\,\text{s}$.

SOLUTION The decay constant λ from Eq. 41–8 is

$$\lambda = \frac{0.693}{T_{\frac{1}{2}}} = \frac{0.693}{(5730\,\text{yr})(3.156 \times 10^7\,\text{s/yr})} = 3.83 \times 10^{-12}\,\text{s}^{-1}.$$

From Eq. 41–7b, the activity or rate of decay is

$$\left|\frac{dN}{dt}\right|_0 = \lambda N_0 = (3.83 \times 10^{-12}\,\text{s}^{-1})(1.00 \times 10^{22}) = 3.83 \times 10^{10}\,\text{decays/s}.$$

Notice that the graph of Fig. 41–11b starts at this value, corresponding to the original value of $N_0 = 1.0 \times 10^{22}$ nuclei in Fig. 41–11a.

NOTE The unit "decays/s" is often written simply as s^{-1} since "decays" is not a unit but refers only to the number. This simple unit of activity is called the becquerel: 1 Bq = 1 decay/s, as discussed in Chapter 42.

EXERCISE F Determine the decay constant for radium $(T_{\frac{1}{2}} = 1600\,\text{yr})$.

CONCEPTUAL EXAMPLE 41–10 **Safety: Activity versus half-life.** One might think that a short half-life material is safer than a long half-life material because it will not last as long. Is that true?

RESPONSE No. A shorter half-life means the activity is higher and thus more "radioactive" and can cause more biological damage. In contrast, a longer half-life for the same sample size N means a lower activity but we have to worry about it for longer and find safe storage until it reaches a safe (low) level of activity.

EXAMPLE 41–11 **A sample of radioactive $^{13}_{7}$N.** A laboratory has 1.49 μg of pure $^{13}_{7}$N, which has a half-life of 10.0 min (600 s). (*a*) How many nuclei are present initially? (*b*) What is the activity (rate of decay) initially? (*c*) What is the activity after 1.00 h? (*d*) After approximately how long will the activity drop to less than one per second $(= 1\,\text{s}^{-1})$?

APPROACH We use the definition of the mole and Avogadro's number (Sections 17–7 and 17–9) to find the number of nuclei. For (*b*) we get λ from the given half-life and use Eq. 41–7b for the activity. For (*c*) and (*d*) we use Eq. 41–7c.

SOLUTION (*a*) The atomic mass is 13.0, so 13.0 g will contain 6.02×10^{23} nuclei (Avogadro's number). We have only 1.49×10^{-6} g, so the number of nuclei N_0 that we have initially is given by the ratio

$$\frac{N_0}{6.02 \times 10^{23}} = \frac{1.49 \times 10^{-6}\,\text{g}}{13.0\,\text{g}}.$$

Solving, we find $N_0 = 6.90 \times 10^{16}$ nuclei.

(*b*) From Eq. 41–8,

$$\lambda = 0.693/T_{\frac{1}{2}} = (0.693)/(600\,\text{s}) = 1.155 \times 10^{-3}\,\text{s}^{-1}.$$

Then, at $t = 0$ (Eq. 41–7b),

$$\left|\frac{dN}{dt}\right|_0 = \lambda N_0 = (1.155 \times 10^{-3}\,\text{s}^{-1})(6.90 \times 10^{16}) = 7.97 \times 10^{13}\,\text{decays/s}.$$

(*c*) After $1.00\,\text{h} = 3600\,\text{s}$, the magnitude of the activity (the rate of decay) will be (Eq. 41–7c)

$$\left|\frac{dN}{dt}\right| = \left|\frac{dN}{dt}\right|_0 e^{-\lambda t}$$
$$= (7.97 \times 10^{13}\,\text{s}^{-1})\, e^{-(1.155\times10^{-3}\,\text{s}^{-1})(3600\,\text{s})} = 1.25 \times 10^{12}\,\text{s}^{-1}.$$

(*d*) We want to determine the time t when $|dN/dt| = 1.00\,\text{s}^{-1}$. From Eq. 41–7c, we have

$$e^{-\lambda t} = \frac{|dN/dt|}{|dN/dt|_0} = \frac{1.00\,\text{s}^{-1}}{7.97 \times 10^{13}\,\text{s}^{-1}} = 1.25 \times 10^{-14}.$$

We take the natural log (ln) of both sides $(\ln e^{-\lambda t} = -\lambda t)$ and divide by λ to find

$$t = -\frac{\ln(1.25 \times 10^{-14})}{\lambda} = 2.77 \times 10^{4}\,\text{s} = 7.70\,\text{h}.$$

Easy Alternate Solution to (c) 1.00 h = 60.0 minutes is 6 half-lives, so the activity will decrease to $(\frac{1}{2})(\frac{1}{2})(\frac{1}{2})(\frac{1}{2})(\frac{1}{2})(\frac{1}{2}) = (\frac{1}{2})^6 = \frac{1}{64}$ of its original value, or $(7.97 \times 10^{13})/(64) = 1.25 \times 10^{12}$ per second.

41–9 Decay Series

It is often the case that one radioactive isotope decays to another isotope that is also radioactive. Sometimes this daughter decays to yet a third isotope which also is radioactive. Such successive decays are said to form a **decay series**. An important example is illustrated in Fig. 41–12. As can be seen, $^{238}_{92}\text{U}$ decays by α emission to $^{234}_{90}\text{Th}$, which in turn decays by β decay to $^{234}_{91}\text{Pa}$. The series continues as shown, with several possible branches near the bottom, ending at the stable lead isotope, $^{206}_{82}\text{Pb}$. The two last decays can be

$$^{206}_{81}\text{Tl} \rightarrow {}^{206}_{82}\text{Pb} + e^- + \bar{\nu}, \qquad (T_{\frac{1}{2}} = 4.2\,\text{min})$$

or

$$^{210}_{84}\text{Po} \rightarrow {}^{206}_{82}\text{Pb} + \alpha. \qquad (T_{\frac{1}{2}} = 138\,\text{days})$$

Other radioactive series also exist.

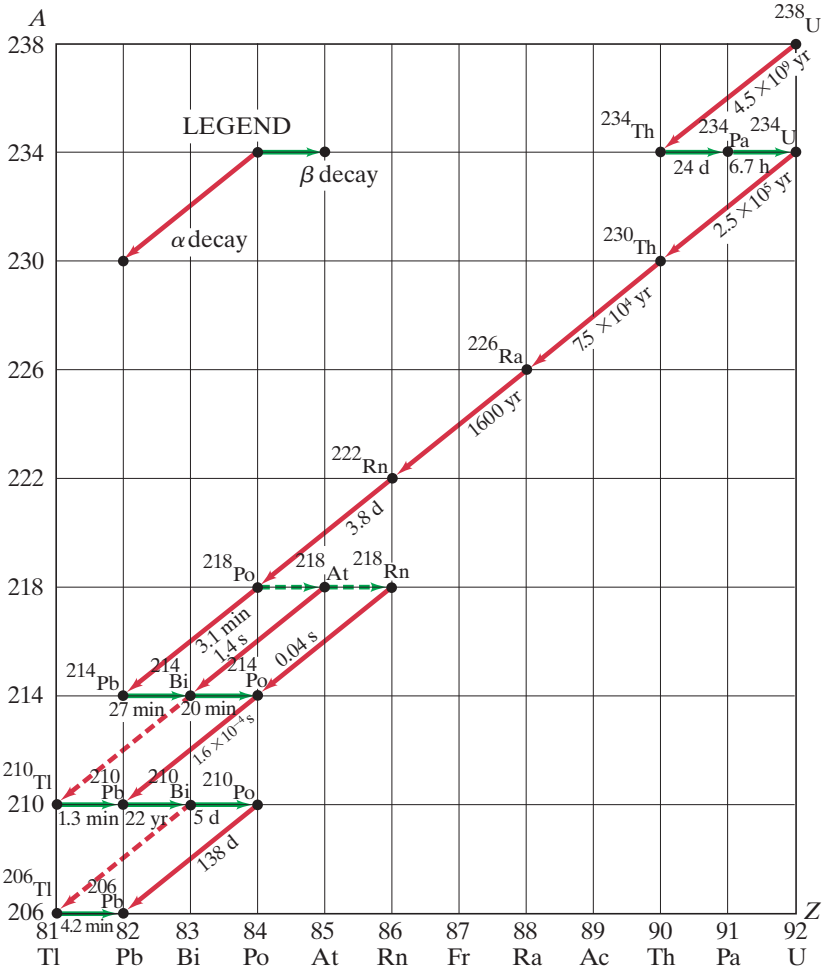


FIGURE 41–12 Decay series beginning with $^{238}_{92}\text{U}$. Nuclei in the series are specified by a dot representing A and Z values. Half-lives are given in seconds (s), minutes (min), hours (h), days (d), or years (yr). Note that a horizontal arrow represents β decay (A does not change), whereas a diagonal line represents α decay (A changes by 4, Z changes by 2). For the four nuclides shown that can decay by both α and β decay, the more prominent decay (in these four cases, $> 99.9\%$) is shown as a solid arrow and the less common decay ($< 0.1\%$) as a dashed arrow.

Because of such decay series, certain radioactive elements are found in nature that otherwise would not be. When the solar system (including Earth) was formed about 5 billion years ago, it is believed that nearly all nuclides were present, having been formed (by fusion and neutron capture, Sections 42–4 and 44–2) in a nearby supernova explosion (Section 44–2). Many isotopes with short half-lives decayed quickly and no longer are detected in nature today. But long-lived isotopes, such as $^{238}_{92}\text{U}$ with a half-life of 4.5×10^9 yr, still do exist in nature today. Indeed, about half of the original $^{238}_{92}\text{U}$ still remains. We might expect, however, that radium $\left(^{226}_{88}\text{Ra}\right)$, with a half-life of 1600 yr, would have disappeared from the Earth long ago. Indeed, the original $^{226}_{88}\text{Ra}$ nuclei must by now have all decayed.

However, because $^{238}_{92}\text{U}$ decays (in several steps) to $^{226}_{88}\text{Ra}$, the supply of $^{226}_{88}\text{Ra}$ is continually replenished, which is why it is still found on Earth today. The same can be said for many other radioactive nuclides.

CONCEPTUAL EXAMPLE 41–12 **Decay chain.** In the decay chain of Fig. 41–12, if we look at the decay of $^{234}_{92}\text{U}$, we see four successive nuclides with half-lives of 250,000 yr, 75,000 yr, 1600 yr, and a little under 4 days. Each decay in the chain has an alpha particle of a characteristic energy, and so we can monitor the radioactive decay rate of each nuclide. Given a sample that was pure $^{234}_{92}\text{U}$ a million years ago, which alpha decay would you expect to have the highest activity rate in the sample?

RESPONSE The first instinct is to say that the process with the shortest half-life would show the highest activity. Surprisingly, perhaps, the activities of the four nuclides in this sample are all the same. The reason is that in each case the decay of the parent acts as a bottleneck to the decay of the daughter. Compared to the 1600-yr half-life of $^{226}_{88}\text{Ra}$, for example, its daughter $^{222}_{86}\text{Rn}$ decays almost immediately, but it cannot decay until it is made. (This is like an automobile assembly line: if worker A takes 20 minutes to do a task and then worker B takes only 1 minute to do the next task, worker B still does only one car every 20 minutes.)

41–10 Radioactive Dating

Radioactive decay has many interesting applications. One is the technique of *radioactive dating* by which the age of ancient materials can be determined.

PHYSICS APPLIED
Carbon-14 dating

The age of any object made from once-living matter, such as wood, can be determined using the natural radioactivity of $^{14}_{6}\text{C}$. All living plants absorb carbon dioxide (CO_2) from the air and use it to synthesize organic molecules. The vast majority of these carbon atoms are $^{12}_{6}\text{C}$, but a small fraction, about 1.3×10^{-12}, is the radioactive isotope $^{14}_{6}\text{C}$. The ratio of $^{14}_{6}\text{C}$ to $^{12}_{6}\text{C}$ in the atmosphere has remained roughly constant over many thousands of years, in spite of the fact that $^{14}_{6}\text{C}$ decays with a half-life of about 5730 yr. This is because energetic nuclei in the cosmic radiation, which impinges on the Earth from outer space, strike nuclei of atoms in the atmosphere and break those nuclei into pieces, releasing free neutrons. Those neutrons can collide with nitrogen nuclei in the atmosphere to produce the nuclear transformation $\text{n} + {}^{14}_{7}\text{N} \rightarrow {}^{14}_{6}\text{C} + \text{p}$. That is, a neutron strikes and is absorbed by a $^{14}_{7}\text{N}$ nucleus, and a proton is knocked out in the process. The remaining nucleus is $^{14}_{6}\text{C}$. This continual production of $^{14}_{6}\text{C}$ in the atmosphere roughly balances the loss of $^{14}_{6}\text{C}$ by radioactive decay.

As long as a plant or tree is alive, it continually uses the carbon from carbon dioxide in the air to build new tissue and to replace old. Animals eat plants, so they too are continually receiving a fresh supply of carbon for their tissues. Organisms cannot distinguish[†] $^{14}_{6}\text{C}$ from $^{12}_{6}\text{C}$, and because the ratio of $^{14}_{6}\text{C}$ to $^{12}_{6}\text{C}$ in the atmosphere remains nearly constant, the ratio of the two isotopes within the living organism remains nearly constant as well. When an organism dies, carbon dioxide is no longer taken in and utilized. Because the $^{14}_{6}\text{C}$ decays radioactively, the ratio of $^{14}_{6}\text{C}$ to $^{12}_{6}\text{C}$ in a dead organism decreases over time. The half-life of $^{14}_{6}\text{C}$ is about 5730 yr, so the $^{14}_{6}\text{C}/^{12}_{6}\text{C}$ ratio decreases by half every 5730 yr. If, for example, the $^{14}_{6}\text{C}/^{12}_{6}\text{C}$ ratio of an ancient wooden tool is half of what it is in living trees, then the object must have been made from a tree that was felled about 5730 years ago.

Actually, corrections must be made for the fact that the $^{14}_{6}\text{C}/^{12}_{6}\text{C}$ ratio in the atmosphere has not remained precisely constant over time. The determination of what this ratio has been over the centuries has required techniques such as comparing the expected ratio to the actual ratio for objects whose age is known, such as very old trees whose annual rings can be counted reasonably accurately.

[†]Organisms operate almost exclusively via chemical reactions—which involve only the outer orbital electrons of the atom; extra neutrons in the nucleus have essentially no effect.

PSE5_CH41_v1.8_23Jun-20

EXAMPLE 41–13 **An ancient animal.** The mass of carbon in an animal bone fragment found in an archeological site is 200 g. If the bone registers an activity of 16 decays/s, what is its age?

PHYSICS APPLIED
Archeological dating

APPROACH First we determine how many $^{14}_{6}\text{C}$ atoms there were in our 200-g sample when the animal was alive, given the known fraction of $^{14}_{6}\text{C}$ to $^{12}_{6}\text{C}$, 1.3×10^{-12}. Then we use Eq. 41–7b to find the activity back then, and Eq. 41–7c to find out how long ago that was by solving for the time t.

SOLUTION The 200 g of carbon is nearly all $^{12}_{6}\text{C}$; 12.0 g of $^{12}_{6}\text{C}$ contains 6.02×10^{23} atoms, so 200 g contains

$$\left(\frac{6.02 \times 10^{23}\ \text{atoms/mol}}{12\ \text{g/mol}}\right)(200\ \text{g}) \;=\; 1.00 \times 10^{25}\ \text{atoms}.$$

When the animal was alive, the ratio of $^{14}_{6}\text{C}$ to $^{12}_{6}\text{C}$ in the bone was 1.3×10^{-12}. The number of $^{14}_{6}\text{C}$ nuclei at that time was

$$N_0 \;=\; (1.00 \times 10^{25}\ \text{atoms})(1.3 \times 10^{-12}) \;=\; 1.3 \times 10^{13}\ \text{atoms}.$$

From Eq. 41–7b, with $\lambda = 3.83 \times 10^{-12}\ \text{s}^{-1}$ for $^{14}_{6}\text{C}$ (Example 41–9), the magnitude of the activity when the animal was still alive $(t = 0)$ was

$$\left|\frac{dN}{dt}\right|_0 \;=\; \lambda N_0 \;=\; (3.83 \times 10^{-12}\ \text{s}^{-1})(1.3 \times 10^{13}) \;=\; 50\ \text{s}^{-1}.$$

From Eq. 41–7c,

$$\left|\frac{dN}{dt}\right| \;=\; \left|\frac{dN}{dt}\right|_0 e^{-\lambda t},$$

where $|dN/dt|$, its activity now, is given as $16\ \text{s}^{-1}$. Then

$$16\ \text{s}^{-1} \;=\; (50\ \text{s}^{-1})e^{\lambda t}$$

or

$$e^{-\lambda t} \;=\; \frac{50}{16}.$$

We take the natural log (ln) of both sides (and divide by λ) to get

$$t \;=\; \frac{1}{\lambda}\ln\left(\frac{50}{16}\right) \;=\; \frac{1}{3.83 \times 10^{-12}\ \text{s}^{-1}}\ln\left(\frac{50}{16}\right)$$

$$\;=\; 2.98 \times 10^{11}\ \text{s} \;=\; 9400\ \text{yr},$$

which is the time elapsed since the death of the animal.

Geological Time Scale Dating

Carbon dating is useful only for determining the age of objects less than about 60,000 years old. The amount of $^{14}_{6}\text{C}$ remaining in objects older than that is usually too small to measure accurately, although new techniques are allowing detection of even smaller amounts of $^{14}_{6}\text{C}$, pushing the time frame further back. On the other hand, radioactive isotopes with longer half-lives can be used in certain circumstances to obtain the age of older objects. For example, the decay of $^{238}_{92}\text{U}$, because of its long half-life of 4.5×10^{9} years, is useful in determining the ages of rocks on a geologic time scale. When molten material on Earth long ago solidified into rock as the temperature dropped, different compounds solidified according to the melting points, and thus different compounds separated to some extent. Uranium present in a material became fixed in position and the daughter nuclei that result from the decay of uranium were also fixed in that position. Thus, by measuring the amount of $^{238}_{92}\text{U}$ remaining in the material relative to the amount of daughter nuclei, the time when the rock solidified can be determined.

PHYSICS APPLIED
Geological dating

PHYSICS APPLIED
Oldest Earth rocks and earliest life

Radioactive dating methods using $^{238}_{92}U$ and other isotopes have shown the age of the oldest Earth rocks to be about 4×10^9 yr. The age of rocks in which the oldest fossilized organisms are embedded indicates that life appeared more than $3\frac{1}{2}$ billion years ago. The earliest fossilized remains of mammals are found in rocks 200 million years old, and humanlike creatures seem to have appeared more than 2 million years ago. Radioactive dating has been indispensable for the reconstruction of Earth's history.

41–11 Detection of Particles

Individual particles such as electrons, protons, α particles, neutrons, and γ rays are not detected directly by our senses. Consequently, a variety of instruments have been developed to detect them.

Counters

One of the most common detectors is the **Geiger counter**. As shown in Fig. 41–13, it consists of a cylindrical metal tube filled with a certain type of gas. A long wire runs down the center and is kept at a high positive voltage $(\approx 10^3\,\text{V})$ with respect to the outer cylinder. The voltage is just slightly less than that required to ionize the gas atoms. When a charged particle enters through the thin "window" at one end of the tube, it ionizes a few atoms of the gas. The freed electrons are attracted toward the positive wire, and as they are accelerated they strike and ionize additional atoms. An "avalanche" of electrons is quickly produced, and when it reaches the wire anode, it produces a voltage pulse. The pulse, after being amplified, can be sent to an electronic counter, which counts how many particles have been detected. Or the pulses can be sent to a loudspeaker and each detection of a particle is heard as a "click." Only a fraction of the radiation emitted by a sample is detected by any detector.

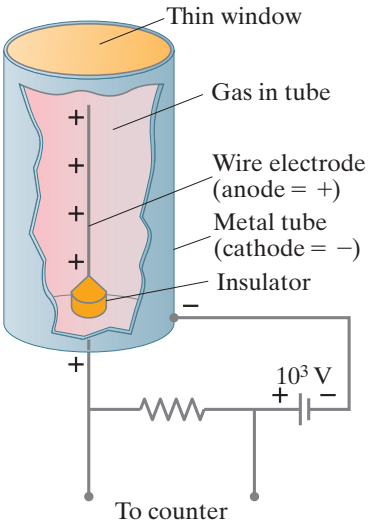
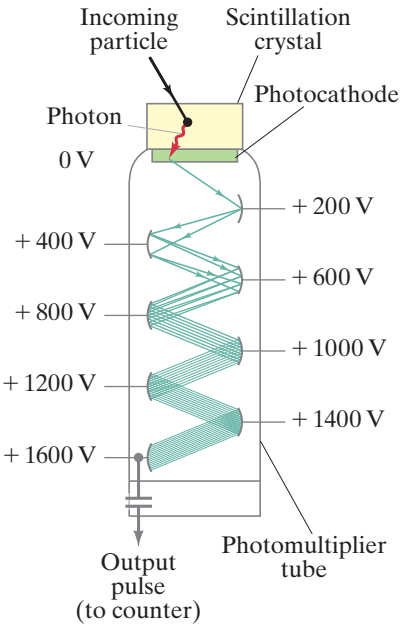


FIGURE 41–13 Diagram of a Geiger counter.

A **scintillation counter** makes use of a solid, liquid, or gas known as a **scintillator** or **phosphor**. The atoms of a scintillator are easily excited when struck by an incoming particle and emit visible light when they return to their ground states. Typical scintillators are crystals of NaI and certain plastics. One face of a solid scintillator is cemented to a photomultiplier tube, and the whole is wrapped with opaque material to keep it light-tight (in the dark) or is placed within a light-tight container. The **photomultiplier** (PM) **tube** converts the energy of the scintillator-emitted photon(s) into an electric signal. A PM tube is a vacuum tube containing several electrodes (typically 8 to 14), called *dynodes*, which are maintained at successively higher voltages as shown in Fig. 41–14. At its top surface is a photoelectric surface, called the *photocathode*, whose work function (Section 37–2) is low enough that an electron is easily released when struck by a photon from the scintillator. Such an electron is accelerated toward the positive voltage of the first dynode. When it strikes the first dynode, the electron has acquired sufficient kinetic energy so that it can eject two to five more electrons. These, in turn, are accelerated toward the higher voltage second dynode, and a multiplication process begins. The number of electrons striking the last dynode may be 10^6 or more. Thus the passage of a particle through the scintillator results in an electric signal at the output of the PM tube that can be sent to an electronic counter just as for a Geiger tube. Solid scintillators are much more dense than the gas of a Geiger counter, and so are much more efficient detectors—especially for γ rays, which interact less with matter than do α or β particles. Scintillators that can measure the total energy deposited are much used today and are called **calorimeters**.

FIGURE 41–14 Scintillation counter with a photomultiplier tube.



In tracer work (Section 42–8), **liquid scintillators** are often used. Radioactive samples taken at different times or from different parts of an organism are placed directly in small bottles containing the liquid scintillator. This is particularly convenient for detection of β rays from ^3_1H and $^{14}_6C$, which have very low energies

and have difficulty passing through the outer covering of a crystal scintillator or Geiger tube. A PM tube is still used to produce the electric signal from the liquid scintillator.

A **semiconductor detector** consists of a reverse-biased *pn* junction diode (Sections 40–8 and 40–9). A charged particle passing through the junction can excite electrons into the conduction band, leaving holes in the valence band. The freed charges produce a short electric current pulse that can be counted as for Geiger and scintillation counters.

Hospital workers and others who work around radiation may carry *film badges* which detect the accumulation of radiation exposure. The film inside is periodically replaced and developed, the darkness of the developed film being related to total exposure (see Section 42–6).

Visualization

The devices discussed so far are used for counting the number of particles (or decays of a radioactive isotope). There are also devices that allow the track of charged particles to be *seen*. Very important are semiconductor detectors. **Silicon wafer semiconductors** have their surface etched into separate tiny pixels, each providing particle position information. They are much used in elementary particle physics (Chapter 43) to track the positions of particles produced and to determine their point of origin and/or their momentum (with the help of a magnetic field). The pixel arrangement can be CCD or CMOS (Section 33–5), the latter able to incorporate electronics inside, allowing fast readout.

One of the oldest tracking devices is the **photographic emulsion**, which can be small and portable, used now particularly for cosmic-ray studies from balloons. A charged particle passing through an emulsion ionizes the atoms along its path. These points undergo a chemical change, and when the emulsion is developed (like film) the particle's path is revealed.

In a **cloud chamber**, used in the early days of nuclear physics, a gas is cooled to a temperature slightly below its usual condensation point ("supercooled"). Tiny droplets form around ions produced when a charged particle passes through (Fig. 41–15). Light scattering from these droplets reveals the track of the particle.

The **bubble chamber**, invented in 1952 by D. A. Glaser (1926–2013), makes use of a superheated liquid kept close to its normal boiling point. Bubbles characteristic of boiling form around ions produced by the passage of a charged particle, revealing paths of particles that recently passed through. Because a bubble chamber uses a liquid, often liquid hydrogen, many more interactions can occur than in a cloud chamber. A magnetic field applied across the chamber makes charged particle paths curve (Chapter 27) and allows the momentum of charged particles to be determined from the radius of curvature of their paths.

A **multiwire**[†] **chamber** consists of a set of closely spaced fine wires immersed in a gas (Fig. 41–16). Many wires are grounded, and the others between are kept at very high voltage. A charged particle passing through produces ions in the gas. Freed electrons drift toward the nearest high voltage wire, creating an "avalanche" of many more ions, and producing an electric pulse or signal at that wire. The positions of the particles are determined electronically by the position of the wire and by the time it takes the pulses to reach "readout" electronics at the ends of the wires. The paths of the particles are reconstructed electronically by computers which can "draw" a picture of the tracks, as shown in Fig. 43–15, Chapter 43. An external magnetic field curves the paths, allowing the momentum of the particles to be measured.

Many detectors are also **calorimeters** which measure the energy of the particles.

[†]Also called *wire drift chamber* or *wire proportional chamber*.

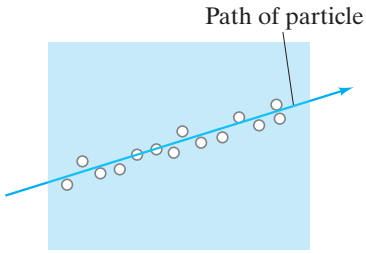
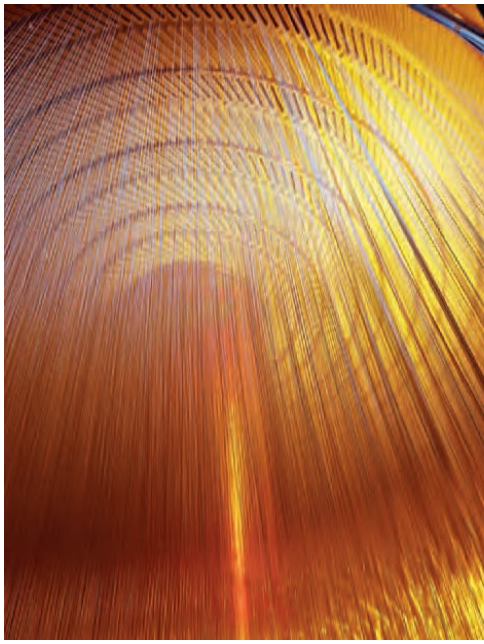


FIGURE 41–15 In a cloud chamber or bubble chamber, droplets or bubbles are formed around ions produced by the passage of a charged particle.

FIGURE 41–16 Multiwire chamber inside the Collider Detector at Fermilab (CDF). Figure 43–15 in Chapter 43 was done with this detector.



Summary

Nuclear physics is the study of atomic nuclei. Nuclei contain **protons** and **neutrons**, which are collectively known as **nucleons**. The total number of nucleons, A, is the nucleus's **atomic mass number**. The number of protons, Z, is the **atomic number**. The number of neutrons equals $A - Z$. **Isotopes** are nuclei with the same Z, but with different numbers of neutrons. For an element X, an isotope of given Z and A is represented by

$$^{A}_{Z}\mathrm{X}.$$

The nuclear radius is approximately proportional to $A^{\frac{1}{3}}$, indicating that all nuclei have about the same density. Nuclear masses are specified in **unified atomic mass units** (u), where the mass of $^{12}_{6}\mathrm{C}$ (including its 6 electrons) is defined as exactly 12.000000 u. In terms of the energy equivalent (because $E = mc^2$),

$$1\,\mathrm{u} = 931.5\,\mathrm{MeV}/c^2 = 1.6605 \times 10^{-27}\,\mathrm{kg}.$$

The mass of a stable nucleus is less than the sum of the masses of its constituent nucleons. The difference in mass (times c^2) is the **total binding energy**. It represents the energy needed to break the nucleus into its constituent nucleons. The **binding energy per nucleon** averages about 8 MeV per nucleon, and is lowest for low mass and high mass nuclei.

Unstable nuclei undergo **radioactive decay**; they change into other nuclei with the emission of an α, β, or γ particle. An α particle is a $^{4}_{2}\mathrm{He}$ nucleus; a β particle is an electron or positron; and a γ ray is a high-energy photon. In β decay, a **neutrino** is also emitted. The transformation of **parent** nuclei into **daughter** nuclei is called **transmutation** of the elements. Radioactive decay occurs spontaneously only when the mass of the products is less than the mass of the parent nucleus. The loss in mass appears as kinetic energy of the products.

Alpha decay occurs via the purely quantum mechanical process of **tunneling** through a barrier.

Nuclei are held together by the **strong nuclear force**. The **weak nuclear force** makes itself apparent in β decay. These two forces, plus the gravitational and electromagnetic forces, are the four known types of force.

Electric charge, linear and angular momentum, mass–energy, and **nucleon number** are **conserved** in all decays.

Radioactive decay is a statistical process. For a given type of radioactive nucleus, the number of nuclei that decay (ΔN) in a time Δt is proportional to the number N of parent nuclei present:

$$\Delta N = -\lambda N\,\Delta t; \qquad \textbf{(41–4a)}$$

the minus sign means N *decreases* in time.

The proportionality constant λ is called the **decay constant** and is characteristic of the given nucleus. The number N of nuclei remaining after a time t decreases exponentially

$$N = N_0\,e^{-\lambda t}, \qquad \textbf{(41–6)}$$

as does the **activity**, $R = |dN/dt|$:

$$\left|\frac{dN}{dt}\right| = \left|\frac{dN}{dt}\right|_0 e^{-\lambda t}. \qquad \textbf{(41–7c)}$$

The **half-life,** $T_{\frac{1}{2}}$, is the time required for half the nuclei of a radioactive sample to decay. It is related to the decay constant by

$$T_{\frac{1}{2}} = \frac{0.693}{\lambda}. \qquad \textbf{(41–8)}$$

Radioactive dating is the use of radioactive decay to determine the age of old objects, such as once-living biological material ($^{14}_{6}\mathrm{C}$) or geological formations ($^{238}_{92}\mathrm{U}$).

Particle **detectors** include **Geiger counters**, **scintillators** with attached **photomultiplier tubes**, and **semiconductor detectors**. Detectors that can image particle tracks include **semiconductors**, photographic **emulsions**, **bubble chambers**, and **wire drift chambers**.

Questions

1. What do different isotopes of a given element have in common? How are they different?
2. Identify the element that has 88 nucleons and 50 neutrons.
3. What are the elements represented by the X in the following: (*a*) $^{232}_{92}\mathrm{X}$; (*b*) $^{18}_{7}\mathrm{X}$; (*c*) $^{1}_{1}\mathrm{X}$; (*d*) $^{86}_{38}\mathrm{X}$; (*e*) $^{252}_{97}\mathrm{X}$?
4. How many protons and how many neutrons do each of the isotopes in Question 3 have?
5. Why are the atomic masses of many elements (see the Periodic Table) not close to whole numbers?
6. How do we know there is such a thing as the strong nuclear force?
7. What are the similarities and the differences between the strong nuclear force and the electric force?
8. What is the experimental evidence in favor of radioactivity being a nuclear process?
9. The isotope $^{64}_{29}\mathrm{Cu}$ is unusual in that it can decay by γ, β^-, and β^+ emission. What is the resulting nuclide for each case?
10. A $^{238}_{92}\mathrm{U}$ nucleus decays via α decay to a nucleus containing how many neutrons?
11. Why are nuclei much more likely to emit an α particle than separate protons and neutrons?
12. Describe, in as many ways as you can, the difference between α, β, and γ rays.
13. What element is formed by the radioactive decay of (*a*) $^{24}_{11}\mathrm{Na}$ (β^-); (*b*) $^{22}_{11}\mathrm{Na}$ (β^+); (*c*) $^{210}_{84}\mathrm{Po}$ (α)?
14. What element is formed by the decay of (*a*) $^{32}_{15}\mathrm{P}$ (β^-); (*b*) $^{35}_{16}\mathrm{S}$ (β^-); (*c*) $^{211}_{83}\mathrm{Bi}$ (α)?
15. Fill in the missing particle or nucleus:
    (*a*) $^{45}_{20}\mathrm{Ca} \rightarrow\ ?\ +\ \mathrm{e}^-\ +\ \bar{\nu}$
    (*b*) $^{58}_{29}\mathrm{Cu}^* \rightarrow\ ?\ +\ \gamma$
    (*c*) $^{46}_{24}\mathrm{Cr} \rightarrow\ ^{46}_{23}\mathrm{V}\ +\ ?$
    (*d*) $^{234}_{94}\mathrm{Pu} \rightarrow\ ?\ +\ \alpha$
    (*e*) $^{239}_{93}\mathrm{Np} \rightarrow\ ^{239}_{94}\mathrm{Pu}\ +\ ?$
16. Immediately after a $^{238}_{92}\mathrm{U}$ nucleus decays to $^{234}_{90}\mathrm{Th} + ^{4}_{2}\mathrm{He}$, the daughter thorium nucleus may still have 92 electrons circling it. Since thorium normally holds only 90 electrons, what do you suppose happens to the two extra ones?

17. When a nucleus undergoes either β^- or β^+ decay, what happens to the energy levels of the atomic electrons? What is likely to happen to these electrons following the decay?
18. The alpha particles from a given alpha-emitting nuclide are generally monoenergetic; that is, they all have the same kinetic energy. But the beta particles from a beta-emitting nuclide have a spectrum of energies. Explain the difference between these two cases.
19. Do isotopes that undergo electron capture generally lie above or below the stable nuclides in Fig. 41–2?
20. Can hydrogen or deuterium emit an α particle? Explain.
21. Why are many artificially produced radioactive isotopes rare in nature?
22. An isotope has a half-life of one month. After two months, will a given sample of this isotope have completely decayed? If not, how much remains?
23. Why are none of the elements with $Z > 92$ stable?
24. A proton strikes a ${}^{6}_{3}\text{Li}$ nucleus. As a result, an α particle and another particle are released. What is the other particle?
25. Can ${}^{14}_{6}\text{C}$ dating be used to measure the age of stone walls and tablets of ancient civilizations? Explain.
26. Explain the absence of β^+ emitters in the radioactive decay series of Fig. 41–12.
27. As ${}^{222}_{86}\text{Rn}$ decays into ${}^{206}_{82}\text{Pb}$, how many alpha and beta particles are emitted? Does it matter which path in the decay series is chosen? Why or why not?
28. A ${}^{238}\text{U}$ nucleus (initially at rest) decays into a ${}^{234}\text{Th}$ nucleus and an alpha particle. Which has the greater (i) momentum, (ii) velocity, (iii) kinetic energy? Explain.
    (*a*) The ${}^{234}\text{Th}$ nucleus.
    (*b*) The alpha particle.
    (*c*) Both the same.
29. Describe how the potential energy curve for an α particle in an α-emitting nucleus differs from that for a stable nucleus.
30. Suppose we discovered that several thousand years ago cosmic rays had bombarded the Earth's atmosphere a lot more than we had thought. Compared to previous calculations of the carbon-dated age of organic matter, we would now calculate it to be older, younger, or the same age as previously calculated? Explain.
31. (*a*) Under what circumstances could a fermium nucleus decay into an einsteinium nucleus? (*b*) What about the reverse, an Es nucleus decaying into Fm?
32. In both internal conversion and β decay, an electron is emitted. How could you determine which decay process occurred?
33. Carbon dating is useful only for determining the age of objects less than about _____ years old. (*a*) 4.5 million, or (*b*) 1.2 million, or (*c*) 600,000, or (*d*) 60,000, or (*e*) 6000.

MisConceptual Questions

1. Elements of the Periodic Table are distinguished by
    (*a*) the number of protons in the nucleus.
    (*b*) the number of neutrons in the nucleus.
    (*c*) the number of electrons in the atom.
    (*d*) Both (*a*) and (*b*).
    (*e*) (*a*), (*b*), and (*c*).
2. A nucleus has
    (*a*) more energy than its component neutrons and protons have.
    (*b*) less energy than its component neutrons and protons have.
    (*c*) the same energy as its component neutrons and protons have.
    (*d*) more energy than its component neutrons and protons have when the nucleus is at rest but less energy than when it is moving.
3. Which of the following will generally create a more stable nucleus?
    (*a*) Having more nucleons.
    (*b*) Having more protons than neutrons.
    (*c*) Having a larger binding energy per nucleon.
    (*d*) Having the same number of electrons as protons.
    (*e*) Having a larger total binding energy.
4. There are 82 protons in a lead nucleus. Why doesn't the lead nucleus burst apart?
    (*a*) Coulomb repulsive force doesn't act inside the nucleus.
    (*b*) Gravity overpowers the Coulomb repulsive force inside the nucleus.
    (*c*) The negatively charged neutrons balance the positively charged protons.
    (*d*) Protons lose their positive charge inside the nucleus.
    (*e*) The strong nuclear force holds the nucleus together.
5. The half-life of a radioactive nucleus is
    (*a*) half the time it takes for the entire substance to fully decay.
    (*b*) the time it takes for half of the substance to decay.
    (*c*) the same as the decay constant.
    (*d*) Both (*a*) and (*b*) (they are the same).
    (*e*) All of the above.
6. As a radioactive sample decays,
    (*a*) the half-life increases.
    (*b*) the half-life decreases.
    (*c*) the activity remains the same.
    (*d*) the number of radioactive nuclei increases.
    (*e*) None of the above.
7. If the half-life of a radioactive sample is 10 years, then it should take _____ years for the sample to decay completely.
    (*a*) 10
    (*b*) 20
    (*c*) 40
    (*d*) Cannot be determined.
8. After three half-lives, what fraction of the original radioactive material is left?
    (*a*) None.
    (*b*) $\frac{1}{16}$.
    (*c*) $\frac{1}{8}$.
    (*d*) $\frac{1}{4}$.
    (*e*) $\frac{3}{4}$.
    (*f*) $\frac{7}{8}$.
    (*g*) $\frac{1}{3}$.

9. Technicium $^{98}_{43}$Tc has a half-life of 4.2×10^6 yr. Strontium $^{90}_{38}$Sr has a half-life of 28.79 yr. Which statements are true?
(*a*) The decay constant of Sr is greater than the decay constant of Tc.
(*b*) The activity of 100 g of Sr is less than the activity of 100 g of Tc.
(*c*) The long half-life of Tc means that it decays by alpha decay.
(*d*) A Tc atom has a higher probability of decaying in 1 yr than a Sr atom.

10. A material having which decay constant would have the shortest half-life?
(*a*) 100/second.
(*b*) 5/year.
(*c*) 8/century.
(*d*) 10^9/day.

11. Radon has a half-life of about 1600 years. The Earth is several billion years old, so why do we still find radon on this planet?
(*a*) Ice-age temperatures preserved some of it.
(*b*) Heavier unstable isotopes decay into it.
(*c*) It is created in lightning strikes.
(*d*) It is replenished by cosmic rays.
(*e*) Its half-life has increased over time.
(*f*) Its half-life has decreased over time.

12. What has greater mass?
(*a*) A neutron and a proton that are far from each other (unbound).
(*b*) A neutron and a proton that are bound together in a hydrogen (deuterium) nucleus.
(*c*) Both the same.

Problems

[See Appendix G for masses]

41–1 Nuclear Properties

1. (I) A rho meson has a mass of $775\ \text{MeV}/c^2$. What is this in atomic mass units?

2. (I) What is the approximate radius of an α particle $\left(^4_2\text{He}\right)$?

3. (I) By what % is the radius of $^{214}_{82}$Pb greater than the radius of $^{206}_{82}$Pb?

4. (II) (*a*) What is the approximate radius of a $^{120}_{50}$Sn nucleus? (*b*) Approximately what is the value of A for a nucleus whose radius is 3.7×10^{-15} m?

5. (II) What is the mass of a bare α particle (without electrons) in MeV/c^2?

6. (II) Suppose two alpha particles were held together so they were just "touching" (use Eq. 41–1). Estimate the electrostatic repulsive force each would exert on the other. What would be the acceleration of an alpha particle subjected to this force?

7. (II) (*a*) What would be the radius of the Earth if it had its actual mass but had the density of nuclei? (*b*) By what factor would the radius of a $^{238}_{92}$U nucleus increase if it had the Earth's density?

8. (II) What stable nucleus has approximately half the radius of a uranium nucleus? [*Hint*: Find A and use Appendix G to get Z.]

9. (II) If an alpha particle were released from rest near the surface of a $^{263}_{104}$Rf nucleus, what would its kinetic energy be when far away?

10. (II) (*a*) What is the fraction of the hydrogen atom's mass (^1_1H) that is in the nucleus? (*b*) What is the fraction of the hydrogen atom's volume that is occupied by the nucleus?

11. (II) Approximately how many nucleons are there in a 1.0-g object? Does it matter what the object is made of? Why or why not?

12. (II) How much kinetic energy, in MeV, must an α particle have to just "touch" the surface of a $^{232}_{90}$Th nucleus?

41–2 Binding Energy

13. (I) Estimate the total binding energy for $^{63}_{29}$Cu, using Fig. 41–1.

14. (I) Use Fig. 41–1 to estimate the total binding energy of (*a*) $^{238}_{92}$U, and (*b*) $^{84}_{36}$Kr.

15. (II) Calculate the binding energy per nucleon for a $^{18}_{8}$O nucleus, using Appendix G.

16. (II) Use Appendix G to calculate the binding energy of $^2_1$H (deuterium).

17. (II) Determine the binding energy of the last neutron in a $^{31}_{15}$P nucleus.

18. (II) Calculate the total binding energy, and the binding energy per nucleon, for (*a*) $^7_4$Be, (*b*) $^{197}_{79}$Au. Use Appendix G.

19. (II) Compare the average binding energy of a nucleon in $^{31}_{15}$P to that in $^{32}_{15}$P, using Appendix G.

20. (III) How much energy is required to remove (*a*) a proton, (*b*) a neutron, from $^{16}_{8}$O? Explain the difference in your answers.

21. (III) (*a*) Show that the nucleus $^8_4$Be (mass = 8.005305 u) is unstable and will decay into two α particles. (*b*) Is $^{12}_6$C stable against decay into three α particles? Show why or why not.

41–3 to 41–7 Radioactive Decay

22. (I) The $^7_3$Li nucleus has an excited state 0.48 MeV above the ground state. What wavelength gamma photon is emitted when the nucleus decays from the excited state to the ground state?

23. (I) Calculate the energy released when tritium, $^3_1$H, decays by β^- emission.

24. (II) Show that the decay $^{11}_6\text{C} \rightarrow {}^{10}_5\text{B} + \text{p}$ is not possible because energy would not be conserved.

25. (II) What is the maximum kinetic energy of an electron emitted in the β decay of a free neutron?

26. (II) Give the result of a calculation that shows whether or not the following decays are possible:
(*a*) $^{233}_{92}\text{U} \rightarrow {}^{232}_{92}\text{U} + \text{n}$;
(*b*) $^{14}_{7}\text{N} \rightarrow {}^{13}_{7}\text{N} + \text{n}$;
(*c*) $^{40}_{19}\text{K} \rightarrow {}^{39}_{19}\text{K} + \text{n}$.

27. (II) $^{24}_{11}$Na is radioactive. (*a*) Is it a β^- or β^+ emitter? (*b*) Write down the decay reaction, and estimate the maximum kinetic energy of the emitted β.

27'. (II) When $^{23}_{10}\text{Ne}$ (mass = 22.9944 u) decays to $^{23}_{11}\text{Na}$ (mass = 22.9898 u), what is the maximum kinetic energy of the emitted electron? What is its minimum energy? What is the energy of the neutrino in each case? Ignore recoil of the daughter nucleus.

28. (II) A $^{238}_{92}\text{U}$ nucleus emits an α particle with kinetic energy = 4.20 MeV. (*a*) What is the daughter nucleus, and (*b*) what is the approximate atomic mass (in u) of the daughter atom? Ignore recoil of the daughter nucleus.

29. (II) Calculate the maximum kinetic energy of the β particle emitted during the decay of $^{60}_{27}\text{Co}$.

30. (II) How much energy is released in electron capture by beryllium: $^{7}_{4}\text{Be} + e^- \rightarrow {}^{7}_{3}\text{Li} + \nu$?

31. (II) The isotope $^{218}_{84}\text{Po}$ can decay by either α or β^- emission. What is the energy release in each case? The mass of $^{218}_{84}\text{Po}$ is 218.008973 u.

32. (II) The nuclide $^{32}_{15}\text{P}$ decays by emitting an electron whose maximum kinetic energy can be 1.71 MeV. (*a*) What is the daughter nucleus? (*b*) Calculate the daughter's atomic mass (in u).

33. (II) A photon with a wavelength of 1.05×10^{-13} m is emitted from an atom. Calculate its energy and explain whether it is a γ ray from the nucleus or a photon from the atom.

34. (II) How much recoil energy does a $^{40}_{19}\text{K}$ nucleus get when it emits a 1.46-MeV gamma ray?

35. (II) Determine the maximum kinetic energy of β^+ particles released when $^{13}_{7}\text{N}$ decays to $^{13}_{6}\text{C}$. What is the maximum energy the neutrino can have? What is the minimum energy of each?

36. (III) Show that when a nucleus decays by β^+ decay, the total energy released is equal to

$$(M_\text{P} - M_\text{D} - 2m_\text{e})c^2,$$

where M_P and M_D are the masses of the parent and daughter atoms (neutral), and m_e is the mass of an electron or positron.

37. (III) When $^{238}_{92}\text{U}$ decays, the α particle emitted has 4.20 MeV of kinetic energy. Calculate the recoil kinetic energy of the daughter nucleus and the Q-value of the decay.

38. (III) What is the energy of the α particle emitted in the decay $^{226}_{88}\text{Ra} \rightarrow {}^{222}_{86}\text{Rn} + \alpha$? Take into account the recoil of the daughter nucleus.

41–8 to 41–11 Half-Life, Decay Rates, Decay Series, Dating

39. (I) (*a*) What is the decay constant of $^{238}_{92}\text{U}$ whose half-life is 4.5×10^9 yr? (*b*) The decay constant of a given nucleus is $3.8 \times 10^{-5}\ \text{s}^{-1}$. What is its half-life?

40. (I) A radioactive material produces 1120 decays per minute at one time, and 4.2 h later produces 140 decays per minute. What is its half-life?

41. (I) What fraction of a sample of $^{68}_{32}\text{Ge}$, whose half-life is about 9 months, will remain after 2.2 yr?

42. (I) What is the activity of a sample of $^{14}_{6}\text{C}$ that contains 5.6×10^{20} nuclei?

43. (I) What fraction of a radioactive sample is left after exactly 5 half-lives?

44. (II) The iodine isotope $^{131}_{53}\text{I}$ is used in hospitals for diagnosis of thyroid function. If 846 μg are ingested by a patient, determine the activity (*a*) immediately, (*b*) 1.50 h later when the thyroid is being tested, and (*c*) 3.0 months later. Use Appendix G.

45. (II) A sample of $^{60}_{27}\text{Co}$ and a sample of $^{131}_{53}\text{I}$ both have N_0 atoms at $t = 0$. How long will it take until both have the same activity? (Use Appendix G for half-life data.)

46. (II) How many nuclei of $^{238}_{92}\text{U}$ remain in a rock if the activity registers 370 decays per second?

47. (II) In a series of decays, the nuclide $^{235}_{92}\text{U}$ becomes $^{207}_{82}\text{Pb}$. How many α and β^- particles are emitted in this series?

48. (II) $^{124}_{55}\text{Cs}$ has a half-life of 30.8 s. (*a*) If we have 9.6 μg initially, how many Cs nuclei are present? (*b*) How many are present 2.6 min later? (*c*) What is the activity at this time? (*d*) After how much time will the activity drop to less than about 1 per second?

49. (II) Calculate the mass of a sample of pure $^{40}_{19}\text{K}$ with an initial decay rate of $1.6 \times 10^5\ \text{s}^{-1}$. The half-life of $^{40}_{19}\text{K}$ is 1.248×10^9 yr.

50. (II) Calculate the activity of a pure 7.8-μg sample of $^{32}_{15}\text{P}$ $\left(T_{\frac{1}{2}} = 1.23 \times 10^6\ \text{s}\right)$.

51. (II) A sample of $^{233}_{92}\text{U}$ $\left(T_{\frac{1}{2}} = 1.59 \times 10^5\ \text{yr}\right)$ contains 3.50×10^{18} nuclei. (*a*) What is the decay constant? (*b*) Approximately how many disintegrations will occur per minute?

52. (II) The activity of a sample drops by a factor of 6.0 in 9.8 minutes. What is its half-life?

53. (II) A 315-g sample of pure carbon contains 1.3 parts in 10^{12} (atoms) of $^{14}_{6}\text{C}$. How many disintegrations occur per second?

54. (II) A sample of $^{238}_{92}\text{U}$ is decaying at a rate of 4.75×10^2 decays/s. What is the mass of the sample?

55. (II) **Rubidium–strontium dating.** The rubidium isotope $^{87}_{37}\text{Rb}$, a β emitter with a half-life of 4.75×10^{10} yr, is used to determine the age of rocks and fossils. Rocks containing fossils of ancient animals contain a ratio of $^{87}_{38}\text{Sr}$ to $^{87}_{37}\text{Rb}$ of 0.0260. Assuming that there was no $^{87}_{38}\text{Sr}$ present when the rocks were formed, estimate the age of these fossils.

56. (II) Two of the naturally occurring radioactive decay sequences start with $^{232}_{90}\text{Th}$ and with $^{235}_{92}\text{U}$. The first five decays of these two sequences are:

$$\alpha, \beta, \beta, \alpha, \alpha$$

and

$$\alpha, \beta, \alpha, \beta, \alpha.$$

Determine the resulting intermediate daughter nuclei in each case.

57. (II) An ancient wooden club is found that contains 67 g of carbon and has an activity of 7.0 decays per second. Determine its age assuming that in living trees the ratio of $^{14}\text{C}/^{12}\text{C}$ atoms is about 1.3×10^{-12}.

58. (II) Use Fig. 41–12 and calculate the relative decay rates for α decay of $^{218}_{84}\text{Po}$ and $^{214}_{84}\text{Po}$.

59. (II) The activity of a radioactive source decreases by 5.5% in 35.0 hours. What is the half-life of this source?

60. (III) $^{7}_{4}\text{Be}$ decays with a half-life of about 53 d. It is produced in the upper atmosphere, and filters down onto the Earth's surface. If a plant leaf is detected to have 320 decays/s of $^{7}_{4}\text{Be}$, (*a*) how long do we have to wait for the decay rate to drop to 25 per second? (*b*) Estimate the initial mass of $^{7}_{4}\text{Be}$ on the leaf.

61. (III) At $t = 0$, a pure sample of radioactive nuclei contains N_0 nuclei whose decay constant is λ. Determine a formula for the number of daughter nuclei, N_D, as a function of time; assume the daughter is stable and that $N_\text{D} = 0$ at $t = 0$.

General Problems

62. Which radioactive isotope of lead is being produced if the measured activity of a sample drops to 1.050% of its original activity in 4.00 h?

63. An old wooden tool is found to contain only 6.2% of the $^{14}_{6}C$ that an equal mass of fresh wood would. How old is the tool?

64. A neutron star consists of neutrons at approximately nuclear density. Estimate, for a 15-km-diameter neutron star, (*a*) its mass number, (*b*) its mass (kg), and (*c*) the acceleration of gravity at its surface.

65. **Tritium dating**. The $^{3}_{1}H$ isotope of hydrogen, which is called *tritium* (because it contains three nucleons), has a half-life of 12.3 yr. It can be used to measure the age of objects up to about 100 yr. It is produced in the upper atmosphere by cosmic rays and brought to Earth by rain. As an application, determine approximately the age of a bottle of wine whose $^{3}_{1}H$ radiation is about $\frac{1}{10}$ that present in new wine.

66. Some elementary particle theories (Section 43–11) suggest that the proton may be unstable, with a half-life $\geq 10^{33}$ yr. (*a*) How long would you expect to wait for one proton in your body to decay (approximate your body as all water)? (*b*) Of the roughly 7 billion people on Earth, about how many would have a proton in their body decay in a 70 yr lifetime?

67. The original experiments which established that an atom has a heavy, positive nucleus were done by shooting alpha particles through gold foil. The alpha particles had a kinetic energy of 7.7 MeV. What is the closest they could get to the center of a gold nucleus? How does this compare with the size of the nucleus?

68. How long must you wait (in half-lives) for a radioactive sample to drop to 1.50% of its original activity?

69. If the potassium isotope $^{40}_{19}K$ gives 48 decays/s in a liter of milk, estimate how much $^{40}_{19}K$ and regular $^{39}_{19}K$ are in a liter of milk. Use Appendix G.

70. Strontium-90 is produced as a nuclear fission product of uranium in both reactors and atomic bombs. Look at its location in the Periodic Table to see what other elements it might be similar to chemically, and tell why you think it might be dangerous to ingest. It has too many neutrons to be stable, and it decays with a half-life of about 29 yr. How long will we have to wait for the amount of $^{90}_{38}Sr$ on the Earth's surface to reach 1% of its current level, assuming no new material is scattered about? Write down the decay reaction, including the daughter nucleus. The daughter is radioactive: write down its decay.

71. The activity of a sample of $^{35}_{16}S$ $\left(T_{\frac{1}{2}} = 87.37 \text{ days}\right)$ is 3.94×10^4 decays per second. What is the mass of the sample?

72. The nuclide $^{191}_{76}Os$ decays with β^- energy of 0.14 MeV accompanied by γ rays of energy 0.042 MeV and 0.129 MeV. (*a*) What is the daughter nucleus? (*b*) Draw an energy-level diagram showing the ground states of the parent and daughter and excited states of the daughter. (*c*) To which of the daughter states does β^- decay of $^{191}_{76}Os$ occur?

73. Determine the activities of (*a*) 1.0 g of $^{32}_{15}P$ $\left(T_{\frac{1}{2}} = 14.3 \text{ days}\right)$ and (*b*) 1.0 g of $^{232}_{90}Th$ $\left(T_{\frac{1}{2}} = 1.40 \times 10^{10} \text{ yr}\right)$.

74. Use Fig. 41–1 to estimate the total binding energy for copper and then estimate the energy, in joules, needed to break a 3.0-g copper penny into its constituent nucleons.

75. Instead of giving atomic masses for nuclides as in Appendix G, some Tables give the **mass excess**, Δ, defined as $\Delta = M - A$, where A is the atomic mass number and M is the mass in u. Determine the mass excess, in u and in MeV/c^2, for: (*a*) $^{4}_{2}He$; (*b*) $^{12}_{6}C$; (*c*) $^{86}_{38}Sr$; (*d*) $^{235}_{92}U$. (*e*) From a glance at Appendix G, can you make a generalization about the sign of Δ as a function of Z or A? [Note: The quantity $A - M$ is called the **mass defect**.]

76. When water is placed near an intense neutron source, the neutrons can be slowed down to almost zero speed by collisions with the water molecules, and are eventually captured by a hydrogen nucleus to form the stable isotope called **deuterium**, $^{2}_{1}H$, giving off a gamma ray. What is the energy of the gamma ray?

77. The practical limit for carbon-14 dating is about 60,000 years. If a bone contains 1.0 kg of carbon, and the animal died 60,000 years ago, what is the activity today?

78. Using Section 41–2 and Appendix G, determine the energy required to remove one neutron from $^{4}_{2}He$. How many times greater is this energy than the binding energy of the last neutron in $^{13}_{6}C$?

79. (*a*) If all of the atoms of the Earth were to collapse and simply become nuclei, what would be the Earth's new radius? (*b*) If all of the atoms of the Sun were to collapse and simply become nuclei, what would be the Sun's new radius?

80. (*a*) An 84-gram sample of natural carbon contains the usual fraction of $^{14}_{6}C$. Estimate roughly how long it will take before there is only one $^{14}_{6}C$ nucleus left. (*b*) How does the answer in (*a*) change if the sample is 340 grams? What does this tell you about the limits of carbon dating?

81. If the mass of the proton were just a little closer to the mass of the neutron, the following reaction would be possible even at low collision energies:

$$e^- + p \rightarrow n + \nu.$$

(*a*) Why would this situation be catastrophic? (See last paragraph of Chapter 44.) (*b*) By what percentage would the proton's mass have to be increased to make this reaction possible?

82. What is the ratio of the kinetic energies for an alpha particle and a beta particle if both make tracks with the same radius of curvature in a magnetic field, oriented perpendicular to the paths of the particles?

83. A 1.00-g sample of natural samarium emits α particles at a rate of 120 s^{-1} due to the presence of $^{147}_{62}Sm$. The natural abundance of $^{147}_{62}Sm$ is 15%. Calculate the half-life for this decay process.

84. Almost all of naturally occurring uranium is $^{238}_{92}U$ with a half-life of 4.468×10^9 yr. Most of the rest of natural uranium is $^{235}_{92}U$ with a half-life of 7.04×10^8 yr. Today a sample contains 0.720% $^{235}_{92}U$. (*a*) What was this percentage 1.0 billion years ago? (*b*) What percentage of uranium will be $^{235}_{92}U$ 100 million years from now?

85. In 1991, the frozen remains of a Neolithic-age man, nicknamed Otzi, were found in the Italian Alps by hikers. The body was well preserved, as were his bow, arrows, knife, axe, other tools, and clothing. The date of his death can be determined using carbon-14 dating. (*a*) What is the decay constant for $^{14}_{6}\text{C}$? (*b*) How many $^{14}_{6}\text{C}$ atoms per gram of $^{12}_{6}\text{C}$ are there in a living organism? (*c*) What is the activity per gram in naturally occurring carbon for a living organism? (*d*) For Otzi, the activity per gram of carbon was measured to be 0.121. How long ago did he live?

86. (*a*) Show that the **mean life** of a radioactive nuclide, defined as

$$\tau = \frac{\int_0^\infty t\,N(t)\,dt}{\int_0^\infty N(t)\,dt},$$

is $\tau = 1/\lambda$. (*b*) What fraction of the original number of nuclei remains after one mean life?

87. A banana contains about 400 mg of potassium, of which a small fraction is the radioactive isotope $^{40}_{19}\text{K}$ (see Appendix G). Estimate the activity of an average banana due to $^{40}_{19}\text{K}$.

88. (*a*) In α decay, of, say, a $^{226}_{88}\text{Ra}$ nucleus, show that the nucleus carries away a fraction $1/\left(1+\frac{1}{4}A_\text{D}\right)$ of the total energy available, where A_D is the mass number of the daughter nucleus. [*Hint*: Use conservation of momentum as well as conservation of energy.] (*b*) Approximately what percentage of the energy available is thus carried off by the α particle when $^{226}_{88}\text{Ra}$ decays?

89. Decay series, such as that shown in Fig. 41–12, can be classified into four families, depending on whether the mass numbers have the form $4n$, $4n+1$, $4n+2$, or $4n+3$, where n is an integer. Justify this statement and show that for a nuclide in any family, all its daughters will be in the same family.

90. Show, using the decays given in Section 41–5, that the neutrino has either spin $\frac{1}{2}$ or $\frac{3}{2}$. (*a*) Calculate the kinetic energy of the α particle emitted when $^{236}_{92}\text{U}$ decays. (*b*) Use Eq. 41–1 to estimate the radius of an α particle and a $^{232}_{90}\text{Th}$ nucleus. Use this to estimate (*c*) the maximum height of the Coulomb barrier, and (*d*) its width AB in Fig. 41–7.

91. Using the uncertainty principle and the radius of a nucleus, estimate the minimum possible kinetic energy of a nucleon in, say, iron. Ignore relativistic corrections. [*Hint*: A particle can have a momentum at least as large as its momentum uncertainty.]

92. In Fig. 41–17, a nucleus decays and emits a particle that enters a region with a uniform magnetic field of 0.012 T directed into the page. The path of the detected particle is shown. (*a*) What type of radioactive decay is this? (*b*) If the radius of the circular arc is 4.7 mm, what is the velocity of the particle?

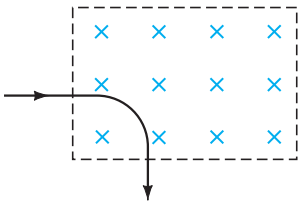


FIGURE 41–17 Problem 92.

93. A nucleus of mass 256 u, initially at rest, emits an α particle with a kinetic energy of 4.5 MeV. What is the kinetic energy of the recoiling daughter nucleus?

94. Some radioactive isotopes have half-lives that are greater than the age of the universe (like gadolinium or samarium). The only way to determine these half-lives is to monitor the decay rate of a sample that contains these isotopes. For example, suppose we find an asteroid that currently contains about 15,000 kg of $^{152}_{64}\text{Gd}$ (gadolinium) and we detect an activity of 1 decay/s. What is the half-life of gadolinium (in years)?

95. A laboratory has a 2.20-μg sample of radioactive $^{13}_{7}\text{N}$ whose decay constant $\lambda = 1.16\times10^{-3}\,\text{s}^{-1}$. Calculate the initial number of nuclei, N_0, present in the sample. Use the radioactive decay law, $N = N_0 e^{-\lambda t}$, to determine the number of nuclei N present at time t for $t = 0$ to 30 minutes (1800 s) in steps of 0.5 min (30 s). Make a graph of N versus t and from the graph determine the half-life of the sample.

ANSWERS TO EXERCISES

A: 0.042130 u.
B: 7.98 MeV/nucleon.
C: (*b*).
D: (*b*) $\frac{1}{8}\,\mu$g.
E: (*a*).
F: $1.37\times10^{-11}\,\text{s}^{-1}$.

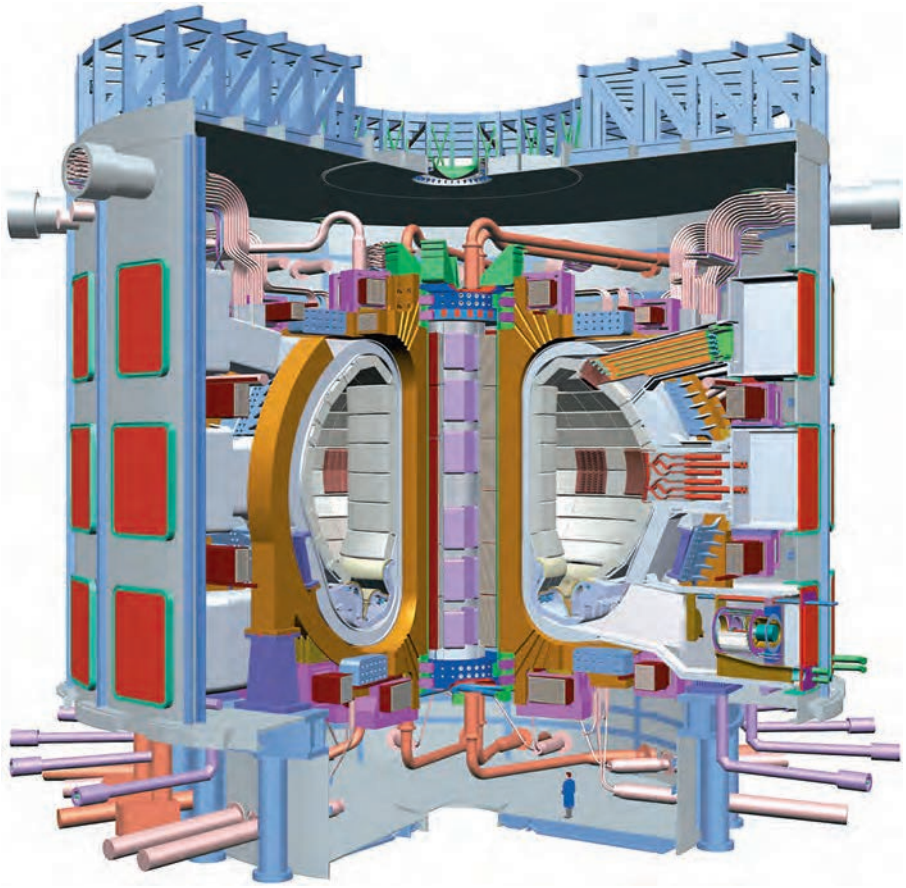


Diagram of ITER (International Thermonuclear Experimental Reactor), which will hopefully begin operation in the mid-2020s. Inside its cavity, over 12 m in diameter, a plasma of electrons and light nuclei will be heated to high temperatures that rival the Sun. Confining a plasma by magnetic fields has proved difficult, and intense research is needed if the fusion of smaller nuclei is to fulfill its promise as a source of abundant and relatively clean power.

This Chapter covers the basic physics topics of nuclear reactions, nuclear fission, nuclear fusion, and how we obtain nuclear energy. We also examine the health aspects of radiation dosimetry, therapy, and imaging by CAT, PET, SPET, and MRI.

CHAPTER 42

Nuclear Energy; Effects and Uses of Radiation

CONTENTS

42 – 1 Nuclear Reactions and the Transmutation of Elements
42 – 2 Cross Section
42 – 3 Nuclear Fission; Nuclear Reactors
42 – 4 Nuclear Fusion
42 – 5 Passage of Radiation Through Matter; Biological Damage
42 – 6 Measurement of Radiation—Dosimetry
*42 – 7 Radiation Therapy
*42 – 8 Tracers in Research and Medicine
*42 – 9 Emission Tomography: PET and SPECT
*42–10 Nuclear Magnetic Resonance (NMR) and Magnetic Resonance Imaging (MRI)

CHAPTER-OPENING QUESTIONS—Guess now!

1. The Sun is powered by
- **(a)** nuclear alpha decay.
- **(b)** nuclear beta decay.
- **(c)** nuclear gamma decay.
- **(d)** nuclear fission.
- **(e)** nuclear fusion.

2. Which radiation induces the most biological damage for a given amount of energy deposited in tissue?
- **(a)** Alpha particles.
- **(b)** Gamma radiation.
- **(c)** Beta radiation.
- **(d)** They all do the same damage for the same deposited energy.
- **(e)** It depends on the type of tissue.

We continue our study of nuclear physics in this Chapter. We begin with a discussion of nuclear reactions, and then we examine the important huge energy-releasing processes of fission and fusion. We also deal with the effects of nuclear radiation passing through matter, particularly biological matter, and how radiation is used medically for therapy, diagnosis, and imaging techniques.

42–1 Nuclear Reactions and the Transmutation of Elements

When a nucleus undergoes α or β decay, the daughter nucleus is a different element from the parent. The transformation of one element into another, called **transmutation**, also occurs via nuclear reactions. A **nuclear reaction** is said to occur when a nucleus is struck by another nucleus, or by a simpler particle such as a γ ray, neutron, or proton, and an interaction takes place. Ernest Rutherford was the first to report seeing a nuclear reaction. In 1919 he observed that some of the α particles passing through nitrogen gas were absorbed and protons emitted. He concluded that nitrogen nuclei had been transformed into oxygen nuclei via the reaction

$$^4_2\text{He} + {}^{14}_7\text{N} \rightarrow {}^{17}_8\text{O} + {}^1_1\text{H},$$

where $^4_2$He is an α particle, and $^1_1$H is a proton.

Since then, a great many nuclear reactions have been observed. Indeed, many of the radioactive isotopes used in the laboratory are made by means of nuclear reactions. Nuclear reactions can be made to occur in the laboratory, but they also occur regularly in nature. In Chapter 41 we saw an example: $^{14}_6$C is continually being made in the atmosphere via the reaction $\text{n} + {}^{14}_7\text{N} \rightarrow {}^{14}_6\text{C} + \text{p}$.

Nuclear reactions are sometimes written in a shortened form: for example,

$$\text{n} + {}^{14}_7\text{N} \rightarrow {}^{14}_6\text{C} + \text{p}$$

can be written

$$^{14}_7\text{N}\ (\text{n}, \text{p})\ {}^{14}_6\text{C}.$$

The symbols outside the parentheses on the left and right represent the initial and final nuclei, respectively. The symbols inside the parentheses represent the bombarding particle (first) and the emitted small particle (second).

In any nuclear reaction, both electric charge and nucleon number are conserved. These conservation laws are often useful, as the following Example shows.

CONCEPTUAL EXAMPLE 42–1 **Deuterium reaction.** A neutron is observed to strike an $^{16}_8$O nucleus, and a deuteron is given off. (A **deuteron**, or **deuterium**, is the isotope of hydrogen containing one proton and one neutron, $^2_1$H; it is sometimes given the symbol d or D.) What is the nucleus that results?

RESPONSE We have the reaction $\text{n} + {}^{16}_8\text{O} \rightarrow ? + {}^2_1\text{H}$. The total number of nucleons initially is $1 + 16 = 17$, and the total charge is $0 + 8 = 8$. The same totals apply after the reaction. The $^2_1$H carries off 2 nucleons and 1 charge, so the unknown product nucleus must have $Z = 7$ and $A = 15$. From the Periodic Table, we find that it is nitrogen that has $Z = 7$, so the nucleus produced is $^{15}_7$N.

EXERCISE A Determine the resulting nucleus in the reaction $\text{n} + {}^{137}_{56}\text{Ba} \rightarrow ? + \gamma$.

Energy and momentum are also conserved in nuclear reactions, and can be used to determine whether or not a given reaction can occur. For example, if the total mass of the final products is less than the total mass of the initial particles, this decrease in mass (recall $\Delta E = \Delta m\, c^2$) is converted to kinetic energy (K) of the outgoing particles. But if the total mass of the products is greater than the total mass of the initial reactants, the reaction requires energy. The reaction will then not occur unless the bombarding particle has sufficient kinetic energy. Consider a nuclear reaction of the general form

$$\text{a} + \text{X} \rightarrow \text{Y} + \text{b}, \qquad \textbf{(42–1)}$$

where particle a is a moving projectile particle (or small nucleus) that strikes nucleus X, producing nucleus Y and particle b (typically, p, n, α, γ).

We define the **reaction energy**, or ***Q*-value**, in terms of the masses involved, as

$$Q = (M_a + M_X - M_b - M_Y)c^2. \quad \textbf{(42–2a)}$$

For a γ ray, $M = 0$. If energy is released by the reaction, $Q > 0$. If energy is required, $Q < 0$.

Because energy is conserved, Q has to be equal to the change in kinetic energy (final minus initial):

$$Q = K_b + K_Y - K_a - K_X. \quad \textbf{(42–2b)}$$

If X is a target nucleus at rest (or nearly so) struck by incoming particle a, then $K_X = 0$. For $Q > 0$, the reaction is said to be *exothermic* or *exoergic*; energy is released in the reaction, so the total kinetic energy is greater after the reaction than before. If Q is negative $(Q < 0)$, the reaction is said to be *endothermic* or *endoergic*: the final total kinetic energy is less than the initial kinetic energy, and an energy input is required to make the reaction happen. The energy input comes from the kinetic energy of the initial colliding particles (a and X).

EXAMPLE 42–2 **A slow-neutron reaction.** The nuclear reaction

$$n + {}^{10}_{5}B \rightarrow {}^{7}_{3}Li + {}^{4}_{2}He$$

is observed to occur even when very slow-moving neutrons (mass $M_n = 1.0087$ u) strike boron atoms at rest. For a particular reaction in which $K_n \approx 0$, the outgoing helium $(M_{He} = 4.0026\ \text{u})$ is observed to have a speed of 9.30×10^6 m/s. Determine (*a*) the kinetic energy of the lithium $(M_{Li} = 7.0160\ \text{u})$, and (*b*) the *Q*-value of the reaction.

APPROACH Since the neutron and boron are both essentially at rest, the total momentum before the reaction is zero; momentum is conserved and so must be zero afterward as well. Thus,

$$M_{Li} v_{Li} = M_{He} v_{He}.$$

We solve this for v_{Li} and substitute it into the equation for kinetic energy. In (*b*) we use Eq. 42–2b.

SOLUTION (*a*) We can use classical kinetic energy with little error, rather than relativistic formulas, because $v_{He} = 9.30 \times 10^6$ m/s is not close to the speed of light *c*, and v_{Li} will be even less because $M_{Li} > M_{He}$. Thus we can write the kinetic energy of the lithium atom, using the momentum equation just above, as

$$K_{Li} = \frac{1}{2} M_{Li} v_{Li}^2 = \frac{1}{2} M_{Li} \left(\frac{M_{He} v_{He}}{M_{Li}} \right)^2 = \frac{M_{He}^2 v_{He}^2}{2 M_{Li}}.$$

We put in numbers, changing the mass in u to kg and recall that 1.60×10^{-13} J = 1 MeV:

$$K_{Li} = \frac{(4.0026\ \text{u})^2 (1.66 \times 10^{-27}\ \text{kg/u})^2 (9.30 \times 10^6\ \text{m/s})^2}{2 (7.0160\ \text{u}) (1.66 \times 10^{-27}\ \text{kg/u})}$$

$$= 1.64 \times 10^{-13}\ \text{J} = 1.02\ \text{MeV}.$$

(*b*) We are given the data $K_a = K_X = 0$ in Eq. 42–2b, so $Q = K_{Li} + K_{He}$, where

$$K_{He} = \tfrac{1}{2} M_{He} v_{He}^2$$

$$= \tfrac{1}{2} (4.0026\ \text{u}) (1.66 \times 10^{-27}\ \text{kg/u}) (9.30 \times 10^6\ \text{m/s})^2$$

$$= 2.87 \times 10^{-13}\ \text{J} = 1.80\ \text{MeV}.$$

Hence, $Q = 1.02\ \text{MeV} + 1.80\ \text{MeV} = 2.82\ \text{MeV}$.

PSE5_CH42_v1.6_23Jun-20

EXAMPLE 42–3 **Will the reaction "go"?** Can the reaction

$$p + {}^{13}_{6}C \rightarrow {}^{13}_{7}N + n$$

occur when ${}^{13}_{6}C$ is bombarded by 2.0-MeV protons?

APPROACH The reaction will "go" if the reaction is exothermic $(Q > 0)$ and even if $Q < 0$ if the input momentum and kinetic energy are sufficient. First we calculate Q from the difference between final and initial masses using Eq. 42–2a, and look up the masses in Appendix G.

SOLUTION The total masses before and after the reaction are:

Before		After	
$M({}^{13}_{6}C)$	13.003355	$M({}^{13}_{7}N)$	13.005739
$M({}^{1}_{1}H)$	1.007825	$M(n)$	1.008665
	14.011180		14.014404

(We must use the mass of the ${}^{1}_{1}H$ atom rather than that of the bare proton because the masses of ${}^{13}_{6}C$ and ${}^{13}_{7}N$ include the electrons, and we must include an equal number of electron masses on each side of the equation since none are created or destroyed.) The products have an excess mass of

$$(14.014404 - 14.011180)\,u = 0.003224\,u \times 931.5\,MeV/u = 3.00\,MeV.$$

Thus $Q = -3.00\,MeV$, and the reaction is endothermic. This reaction requires energy, and the 2.0-MeV protons do not have enough to make it go.

NOTE The incoming proton in this Example would need more than 3.00 MeV of kinetic energy to make this reaction go; 3.00 MeV would be enough to conserve energy, but a proton of this energy would produce the ${}^{13}_{7}N$ and n with no kinetic energy and hence no momentum. Since an incident 3.0-MeV proton has momentum, conservation of momentum would be violated. A calculation using conservation of energy *and* of momentum, as we did in Examples 41–7 and 42–2, shows that the minimum proton energy, called the **threshold energy**, is 3.23 MeV in this case (= Problem 15).

Neutron Physics

The artificial transmutation of elements took a great leap forward in the 1930s when Enrico Fermi realized that neutrons would be the most effective projectiles for causing nuclear reactions and in particular for producing new elements. Because neutrons have no net electric charge, they are not repelled by positively charged nuclei as are protons or alpha particles. Hence the probability of a neutron reaching the nucleus and causing a reaction is much greater than for charged projectiles,[†] particularly at low energies. Between 1934 and 1936, Fermi and his co-workers in Rome produced many previously unknown isotopes by bombarding different elements with neutrons. Fermi realized that if the heaviest known element, uranium, is bombarded with neutrons, it might be possible to produce new elements with atomic numbers greater than that of uranium. After several years of hard work, it was suspected that two new elements had been produced, neptunium $(Z = 93)$ and plutonium $(Z = 94)$. The full confirmation that such "transuranic" elements could be produced came several years later at the University of California, Berkeley. The reactions are shown in Fig. 42–1.

It was soon shown that what Fermi had actually observed when he bombarded uranium was an even stranger process—one that was destined to play an extraordinary role in the world at large. We discuss it in Section 42–3.

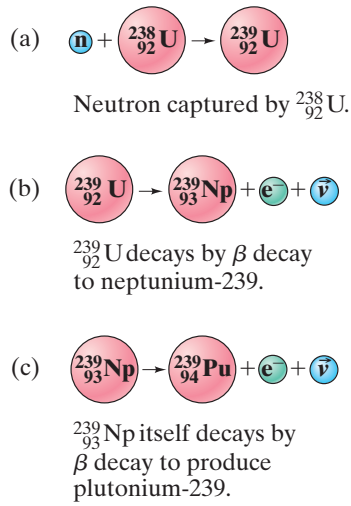


FIGURE 42–1 Neptunium and plutonium are produced in this series of reactions, after bombardment of ${}^{238}_{92}U$ by neutrons.

[†]That is, positively charged particles. Electrons rarely cause nuclear reactions because they do not interact via the strong nuclear force.

42–2 Cross Section

Some reactions have a higher probability of occurring than others. The reaction probability is specified by a quantity called the collision **cross section**. Although the size of a nucleus, like that of an atom, is not a clearly defined quantity since the edges are not distinct like those of a tennis ball or baseball, we can nonetheless define a *cross section* for nuclei undergoing collisions by using an analogy. Suppose that projectile particles strike a stationary target of total area A and thickness ℓ, as shown in Fig. 42–2. Assume also that the target is made up of identical objects (such as marbles or nuclei), each of which has a cross-sectional area σ, and we assume the incoming projectiles are small by comparison. We assume that the target objects are fairly far apart and the thickness ℓ is so small that we don't have to worry about overlapping. This is often a reasonable assumption because nuclei have diameters on the order of 10^{-14} m but are at least 10^{-10} m (atomic size) apart even in solids. If there are n nuclei per unit volume, the total cross-sectional area of all these tiny targets is

$$A' = nA\ell\sigma$$

since $nA\ell = (n)(\text{volume})$ is the total number of targets and σ is the cross-sectional area of each. If $A' \ll A$, most of the incident projectile particles will pass through the target without colliding. If R_0 is the rate at which the projectile particles strike the target (number/second), the rate at which collisions occur, R, is

$$R = R_0\frac{A'}{A} = R_0\frac{nA\ell\sigma}{A}$$

so

$$R = R_0 n\ell\sigma.$$

Thus, by measuring the collision rate, R, we can determine σ:

$$\sigma = \frac{R}{R_0 n\ell}. \qquad \textbf{(42–3)}$$

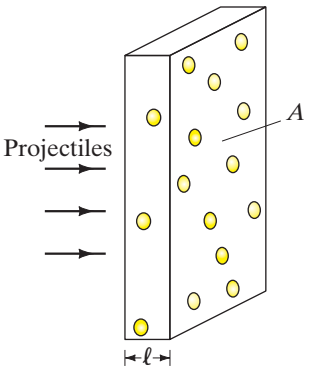


FIGURE 42–2 Projectile particles strike a target of area A and thickness ℓ made up of n nuclei per unit volume.

If nuclei were simple billiard balls, and R the number of particles that are deflected per second, σ would represent the real cross-sectional area of each ball. But nuclei are complicated objects that cannot be considered to have distinct boundaries. Furthermore, collisions can be either elastic or inelastic, and reactions can occur in which the nature of the particles can change. By measuring R for each possible process, we can determine an **effective cross section**, σ, for each process. None of these cross sections is necessarily related to a geometric cross-sectional area. Rather, σ is an "effective" target area. It is a *measure of the probability of a collision or of a particular reaction occurring* per target nucleus, independent of the dimensions of the entire target. The concept of cross section is useful because σ depends only on the properties of the interacting particles, whereas R depends on the thickness and area of the physical (macroscopic) target, on the number of particles in the incident beam, and so on.

When a given pair of particles interact, we define their **elastic cross section** σ_{el} using Eq. 42–3, where R for a given experimental setup is the rate of elastic collisions (or **elastic scattering**), by which we mean collisions for which the final particles are the same as the initial particles (a = b, X = Y in Eq. 42–1) and $Q = 0$. Similarly, the inelastic cross section, σ_{inel}, is related to the rate of inelastic collisions, or **inelastic scattering**, which involves the same final and initial particles but $Q \neq 0$, usually because excited states are involved. For each reaction in which the final particles are different than the initial particles, there is a particular cross section. For protons (p) incident on ${}^{13}_{6}\text{C}$, for example, we could have various reactions, such as $\text{p} + {}^{13}_{6}\text{C} \rightarrow {}^{13}_{7}\text{N} + \text{n}$ or $\text{p} + {}^{13}_{6}\text{C} \rightarrow {}^{10}_{5}\text{B} + {}^{4}_{2}\text{He}$, and so on. The sum of all the separate reaction cross sections (for a given pair of initial particles) is called the **total reaction cross section**, σ_{R}. The **total cross section**, σ_{T}, is

$$\sigma_{\text{T}} = \sigma_{\text{el}} + \sigma_{\text{inel}} + \sigma_{\text{R}}$$

and is a measure of all possible interactions or collisions starting with given initial particles. Said another way, σ_{T} is a measure of how many of the incident particles interact in some way and hence are eliminated from the incident beam.

PSE5_CH42_v1.6_23Jun-20

We can also define **differential cross sections**, which represent the probability of the deflected (or emitted) particles leaving at particular angles.

It is said that when one of the first nuclear cross sections was measured, a physicist, surprised that it was as large as it was $(\approx 10^{-28}\,\text{m}^2)$, remarked, "it's as big as a barn." Ever since then nuclear cross sections have been measured in "barns," where 1 barn $(\text{bn}) = 10^{-28}\,\text{m}^2$.

The value of σ for a given reaction depends on, among other things, the incident kinetic energy. Typical nuclear cross sections are on the order of barns, but they can vary from millibarns to kilobarns or more. Figure 42–3 shows the cross section for neutron capture in cadmium $\left(\text{n} + {}^{114}_{48}\text{Cd} \rightarrow {}^{115}_{48}\text{Cd} + \gamma\right)$ as a function of neutron kinetic energy. Neutron cross sections for most materials are greater at low energies, as in Fig. 42–3. To produce nuclear reactions at a high rate it is therefore desirable that the bombarding neutrons have low energy. Neutrons that have been slowed down and have reached equilibrium with matter at room temperature $\left(\frac{3}{2}kT \approx 0.04\,\text{eV} \text{ at } T = 300\,\text{K}\right)$ are called **thermal neutrons**.

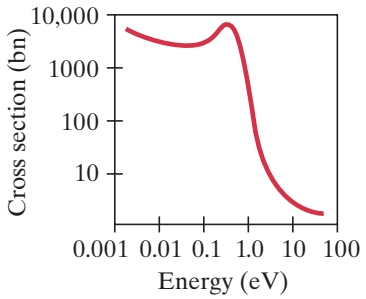


FIGURE 42–3 The neutron cross section for cadmium-114 as a function of incoming neutron kinetic energy. It is extraordinarily large for $K \lesssim 1\,\text{eV}$. Note that both scales are logarithmic.

EXAMPLE 42–4 **Using cross section.** The reaction

$$\text{p} + {}^{56}_{26}\text{Fe} \rightarrow {}^{56}_{27}\text{Co} + \text{n}$$

has a cross section of 0.65 bn for a particular incident proton energy. Suppose the iron target has an area of $1.5\,\text{cm}^2$, and is $2.0\,\mu\text{m}$ thick. The density of iron is $7.8 \times 10^3\,\text{kg/m}^3$. If the protons are incident at a rate of 2.0×10^{13} particles/s, calculate the rate at which neutrons are produced.

APPROACH We use Eq. 42–3 in the form: $R = R_0 n\ell\sigma$.

SOLUTION We are given $R_0 = 2.0 \times 10^{13}$ particles/s, $\ell = 2.0 \times 10^{-6}\,\text{m}$, and $\sigma = 0.65$ bn. Recalling from Chapter 17 that one mole (mass = 56 g for iron) contains 6.02×10^{23} atoms, then the number of iron atoms per unit volume is

$$n = \left(6.02 \times 10^{23}\,\text{atoms/mole}\right)\frac{\left(7.8 \times 10^3\,\text{kg/m}^3\right)}{\left(56 \times 10^{-3}\,\text{kg/mole}\right)} = 8.4 \times 10^{28}\,\text{atoms/m}^3.$$

Then we find that the rate at which neutrons are produced is

$$\begin{aligned} R &= R_0 n\ell\sigma \\ &= \left(2.0 \times 10^{13}\text{particles/s}\right)\left(8.4 \times 10^{28}\text{atoms/m}^3\right)\left(2.0 \times 10^{-6}\text{m}\right)\left(0.65 \times 10^{-28}\text{m}^2\right) \\ &= 2.2 \times 10^{8}\text{particles/s}. \end{aligned}$$

42–3 Nuclear Fission; Nuclear Reactors

In 1938, the German scientists Otto Hahn and Fritz Strassmann made an amazing discovery. Following up on Fermi's work, they found that uranium bombarded by neutrons sometimes produced smaller nuclei that were roughly half the size of the original uranium nucleus. Lise Meitner and Otto Frisch quickly realized what had happened: the uranium nucleus, after absorbing a neutron, actually had split into two roughly equal pieces. This was startling, for until then the known nuclear reactions involved knocking out only a tiny fragment (for example, n, p, or α) from a nucleus.

Nuclear Fission and Chain Reactions

This new phenomenon was named **nuclear fission** because of its resemblance to biological fission (cell division). It occurs much more readily for ${}^{235}_{92}\text{U}$ than for the more common ${}^{238}_{92}\text{U}$. The process can be visualized by imagining the uranium nucleus to be like a liquid drop. According to this **liquid-drop model**, the neutron absorbed by the ${}^{235}_{92}\text{U}$ nucleus gives the nucleus extra internal energy (like heating a drop of water).

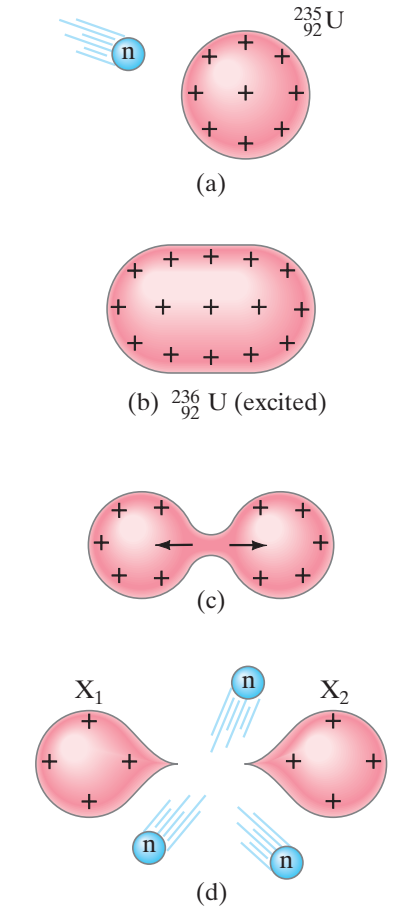
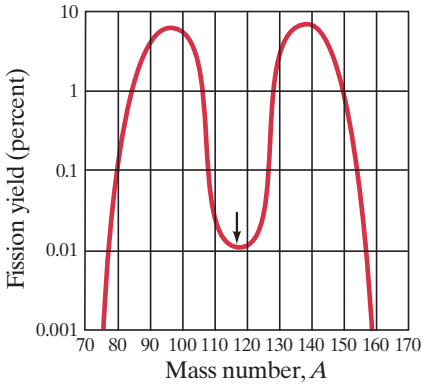


FIGURE 42–4 Fission of a $^{235}_{92}$U nucleus after capture of a neutron, according to the liquid-drop model.

FIGURE 42–5 Mass distribution of fission fragments from $^{235}_{92}$U + n. The small arrow indicates equal mass fragments $\left(\frac{1}{2} \times (236 - 2) = 117\right.$, assuming 2 neutrons are liberated$\left.\right)$. Note that the vertical scale is logarithmic.



This intermediate state, or **compound nucleus**, is $^{236}_{92}$U (because of the absorbed neutron). The extra energy of this nucleus—it is in an excited state—appears as increased motion of the individual nucleons, which causes the nucleus to take on abnormal elongated shapes, Fig. 42–4. When the nucleus elongates (in this model) into the shape shown in Fig. 42–4c, the attraction of the two ends via the short-range nuclear force is greatly weakened by the increased separation distance. Then the electric repulsive force becomes dominant, and the nucleus splits in two (Fig.42–4d). The two resulting nuclei, X_1 and X_2, are called **fission fragments**, and in the process a number of neutrons (typically two or three) are also given off. The reaction can be written

$$\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{236}_{92}\text{U} \rightarrow \text{X}_1 + \text{X}_2 + \text{neutrons}. \qquad \textbf{(42–4)}$$

The compound nucleus, $^{236}_{92}$U, exists for less than 10^{-12} s, so the process occurs very quickly. The two fission fragments, X_1 and X_2, rarely split the original uranium precisely half and half, but more often as about 40%–60%. A typical fission reaction is

$$\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{141}_{56}\text{Ba} + {}^{92}_{36}\text{Kr} + 3\text{n}, \qquad \textbf{(42–5)}$$

although many others also occur. Indeed Figure 42–5 shows the measured distribution of $^{235}_{92}$U fission fragments according to mass. Only rarely (about 1 in 10^4) does a fission result in equal mass fragments (arrow in Fig. 42–5).

CONCEPTUAL EXAMPLE 42–5 **Counting nucleons.** Identify the element X in the fission reaction $\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{A}_{Z}\text{X} + {}^{93}_{38}\text{Sr} + 2\text{n}$.

RESPONSE The number of nucleons is conserved (Section 41–7). The uranium nucleus with 235 nucleons plus the incoming neutron make $235 + 1 = 236$ nucleons. So there must be 236 nucleons after the reaction. The Sr has 93 nucleons, and the two neutrons make 95 nucleons, so X has $A = 236 - 95 = 141$. Electric charge is also conserved: before the reaction, the total charge is $92e$. After the reaction the total charge is $(Z + 38)e$ and must equal $92e$. Thus $Z = 92 - 38 = 54$. The element with $Z = 54$ is xenon (see Appendix G or the Periodic Table inside the back cover), so the isotope is $^{141}_{54}$Xe.

EXERCISE B In the fission reaction $\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{137}_{53}\text{I} + {}^{96}_{39}\text{Y} + \text{neutrons}$, how many neutrons are released?

A tremendous amount of energy is released in a fission reaction because the mass of $^{235}_{92}$U is considerably greater than the total mass of the fission fragments plus released neutrons. This can be seen from the binding-energy-per-nucleon curve of Fig. 41–1; the binding energy per nucleon for uranium is about 7.6 MeV/nucleon, but for fission fragments that have intermediate mass (in the center portion of the graph, $A \approx 100$), the average binding energy per nucleon is about 8.5 MeV/nucleon. Since the fission fragments are more tightly bound, the sum of their masses is less than the mass of the uranium. The difference in mass, or energy, between the original uranium nucleus and the fission fragments is about $8.5 - 7.6 = 0.9$ MeV per nucleon. Because there are 236 nucleons involved in each fission, the total energy released per fission is

$$(0.9\,\text{MeV/nucleon})\,(236\,\text{nucleons}) \approx 200\,\text{MeV}. \qquad \textbf{(42–6)}$$

This is an enormous amount of energy for one single nuclear fission. At a practical level, the energy from one fission is tiny. But if many such fissions could occur in a short time, an enormous amount of energy at the macroscopic level would be available. A number of physicists, including Fermi, recognized that the neutrons released in each fission (Eqs. 42–4 and 42–5) could be used to create a **chain reaction**. That is, one neutron initially causes one fission of a uranium

PSE5_CH42_v1.6_23Jun-20

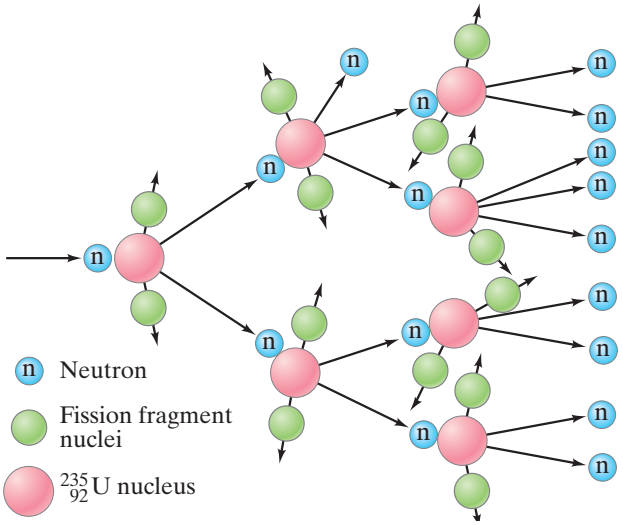


FIGURE 42–6 Chain reaction.

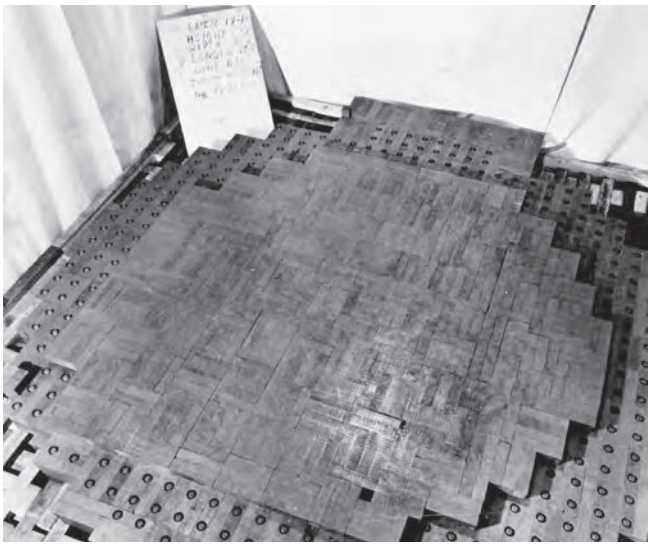


FIGURE 42–7 This is the only photograph of the first nuclear reactor, built by Fermi under the grandstand of Stagg Field at the University of Chicago. It is shown here under construction as a layer of graphite (used as moderator) was being placed over a layer of natural uranium. On December 2, 1942, Fermi slowly withdrew the cadmium control rods and the reactor went critical. This first self-sustaining chain reaction was announced to Washington, via telephone, by Arthur Compton who witnessed the event and reported: "The Italian navigator has just landed in the new world."

nucleus; the two or three neutrons released can go on to cause additional fissions, so the process multiplies as shown schematically in Fig. 42–6.

If a **self-sustaining chain reaction** was actually possible in practice, the enormous energy available in fission could be released on a larger scale. Fermi and his co-workers (at the University of Chicago) showed it was possible by constructing the first **nuclear reactor** in 1942 (Fig. 42–7).

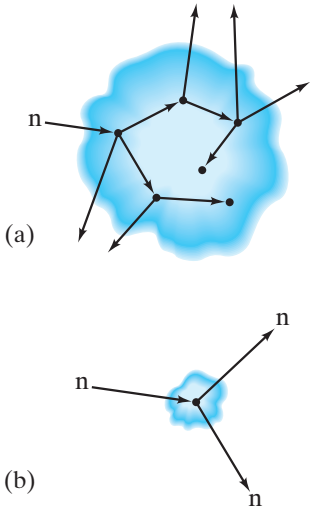
Nuclear Reactors

Several difficulties have to be overcome to make any nuclear reactor function. First, the probability that a $^{235}_{92}$U nucleus will absorb a neutron is large only for slow neutrons, but the neutrons emitted during a fission (which are needed to sustain a chain reaction) are moving very fast. A substance known as a **moderator** must be used to slow down the neutrons. The most effective moderator will consist of atoms whose mass is as close as possible to that of the neutrons. (To see why this is true, recall from Chapter 9 that a billiard ball striking an equal mass ball at rest can itself be stopped in one collision; but a billiard ball striking a heavy object bounces off with nearly unchanged speed.) The best moderator would thus contain $^{1}_{1}$H atoms. Unfortunately, $^{1}_{1}$H tends to absorb neutrons. But the isotope of hydrogen called *deuterium*, $^{2}_{1}$H, does not absorb many neutrons and is thus an almost ideal moderator. Either $^{1}_{1}$H or $^{2}_{1}$H can be used in the form of water. In the latter case, it is **heavy water**, in which the hydrogen atoms have been replaced by deuterium. Another common moderator is *graphite*, which consists of $^{12}_{6}$C atoms.

A second problem is that the neutrons produced in one fission may be absorbed and produce other nuclear reactions with other nuclei in the reactor, rather than produce further fissions. In a "light-water" reactor, the $^{1}_{1}$H nuclei absorb neutrons, as does $^{238}_{92}$U to form $^{239}_{92}$U in the reaction $n + {}^{238}_{92}U \rightarrow {}^{239}_{92}U + \gamma$. Naturally occurring uranium[†] contains 99.3% $^{238}_{92}$U and only 0.7% fissionable $^{235}_{92}$U. To increase the probability of fission of $^{235}_{92}$U nuclei, natural uranium can be **enriched** to increase the percentage of $^{235}_{92}$U by using processes such as diffusion or centrifugation. Enrichment is not usually necessary for reactors using heavy water as the moderator, because heavy water doesn't absorb neutrons.[‡]

The third problem is that some neutrons will escape through the surface of the reactor core before they can cause further fissions (Fig. 42–8). Thus the mass of fuel must be sufficiently large for a self-sustaining chain reaction to take place. The minimum mass of uranium needed is called the **critical mass**.

FIGURE 42–8 If the amount of uranium exceeds the critical mass, as in (a), a sustained chain reaction is possible. If the mass is less than critical, as in (b), too many neutrons escape before additional fissions occur, and the chain reaction is not sustained.



[†]$^{238}_{92}$U will fission, but only with fast neutrons ($^{238}_{92}$U is more stable than $^{235}_{92}$U). The probability of absorbing a fast neutron and producing a fission is too low to produce a self-sustaining chain reaction.

[‡]Other types of power or research reactors need 3% to 20% of $^{235}_{92}$U. Atom bombs need greater enrichment: $\geq 80\%$.

The value of the critical mass depends on the moderator, the fuel ($^{239}_{94}$Pu may be used instead of $^{235}_{92}$U), and how much the fuel is enriched, if at all. Typical values are on the order of a few kilograms (that is, neither grams nor thousands of kilograms). Critical mass depends also on the average number of neutrons released per fission: 2.5 for $^{235}_{92}$U, 2.9 for $^{239}_{94}$Pu and so the critical mass for $^{239}_{94}$Pu is smaller.

To have a self-sustaining chain reaction, on average at least one neutron produced in each fission must go on to produce another fission. The average number of neutrons per fission that do go on to produce further fissions is called the **neutron multiplication factor**, f. For a self-sustaining chain reaction, we must have $f \geq 1$. If $f < 1$, the reactor is "subcritical." If $f > 1$, it is "supercritical" (and could become dangerously explosive). Reactors are equipped with movable **control rods** (good neutron absorbers like cadmium or boron), whose function is to absorb neutrons and maintain the reactor at just barely "critical," $f = 1$.

The release of neutrons and subsequent fissions occur so quickly that manipulation of the control rods to maintain $f = 1$ would not be possible if it weren't for a small percentage ($\approx 1\%$) of so-called **delayed neutrons**. They come from the decay of neutron-rich fission fragments (or their daughters) having lifetimes on the order of seconds—sufficient to allow enough reaction time to operate the control rods and maintain $f = 1$.

Nuclear reactors have been built for use in research and to produce electric power. Fission produces many neutrons and a "research reactor" is basically an intense source of neutrons. These neutrons can be used as projectiles in nuclear reactions to produce nuclides not found in nature, including isotopes used as tracers and for therapy. A "power reactor" is used to produce electric power. The energy released in the fission process appears as heat, which is used to boil water and produce steam to drive a turbine connected to an electric generator (Fig. 42–9). The **core** of a nuclear reactor consists of the fuel and a moderator (water in most U.S. commercial reactors). The fuel is usually uranium enriched so that it contains 2 to 4 percent $^{235}_{92}$U. Water at high pressure or other liquid (such as liquid sodium) is allowed to flow through the core. The thermal energy it absorbs is used to produce steam in the heat exchanger, so the fissionable fuel acts as the heat input for a heat engine (Chapter 20).

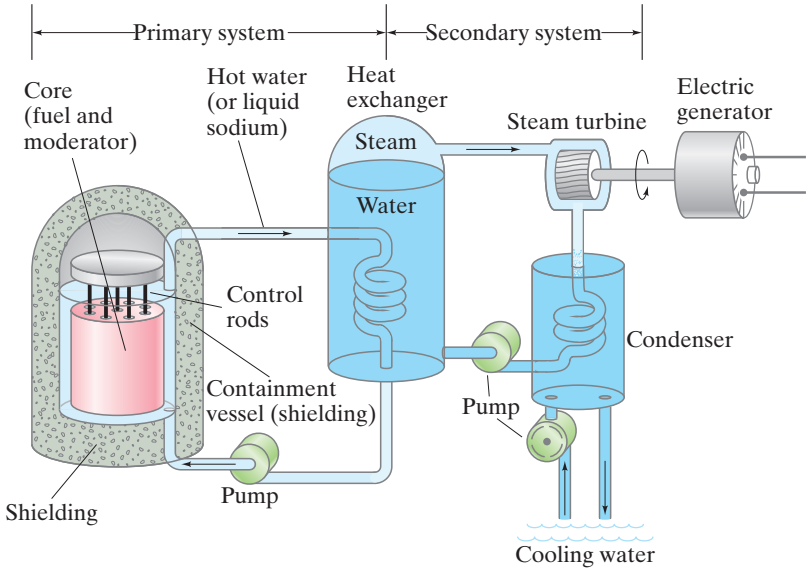


FIGURE 42–9 A nuclear reactor. The heat generated by the fission process in the fuel rods is carried off by hot water or liquid sodium and is used to boil water to steam in the heat exchanger. The steam drives a turbine to generate electricity and is then cooled in the condenser (to reduce pressure on the back side of the turbine blades).

There are problems associated with nuclear power plants. Besides the usual thermal pollution associated with any heat engine (Section 20–11), there is the serious problem of disposal of the radioactive fission fragments produced in the reactor, plus radioactive nuclides produced by neutrons interacting with the structural parts of the reactor. Fission fragments, like their uranium or plutonium parents, have about 50% more neutrons than protons. Nuclei with atomic number in the typical range for fission fragments ($Z \approx 30$ to 60) are stable only if they have more nearly equal numbers of protons and neutrons (see Fig. 41–2). Hence the highly neutron-rich fission fragments are very unstable and decay radioactively. The accidental release

PSE5_CH42_v1.6_23Jun-20

of highly radioactive fission fragments into the atmosphere poses a serious threat to human health (Section 42–5), as does possible leakage of the radioactive wastes when they are disposed of. The accidents at Three Mile Island, Pennsylvania (1979), at Chernobyl, Russia (1986), and at Fukushima, Japan (2011), have illustrated some of the dangers and have shown that nuclear plants must be located, constructed, maintained, and operated with great care and precision (Fig. 42–10).

Finally, the lifetime of nuclear power plants is limited to 30-some years, due to buildup of radioactivity and the fact that the structural materials themselves are weakened by the intense conditions inside. The cost of "decommissioning" a power plant is very great.

So-called **breeder reactors** were proposed as a solution to the problem of limited supplies of fissionable uranium, $^{235}_{92}\text{U}$. A breeder reactor is one in which some of the neutrons produced in the fission of $^{235}_{92}\text{U}$ are absorbed by $^{238}_{92}\text{U}$, and $^{239}_{94}\text{Pu}$ is produced via the set of reactions shown in Fig. 42–1. $^{239}_{94}\text{Pu}$ is fissionable with slow neutrons, so after separation it can be used as a fuel in a nuclear reactor. Thus a breeder reactor "breeds" new fuel† $\left(^{239}_{94}\text{Pu}\right)$ from otherwise useless $^{238}_{92}\text{U}$. Natural uranium is 99.3 percent $^{238}_{92}\text{U}$, which in a breeder becomes useful fissionable $^{239}_{94}\text{Pu}$, thus increasing the supply of fissionable fuel by more than a factor of 100. But breeder reactors have the same problems as other reactors, plus other serious problems. Not only is plutonium a serious health hazard in itself (radioactive with a half-life of 24,000 years), but plutonium produced in a reactor can readily be used in a bomb, increasing the danger of nuclear proliferation and theft of fuel to produce a bomb.

Nuclear power presents risks. Other large-scale energy-conversion methods, such as conventional oil and coal-burning steam plants, also present health and environmental hazards; some of them were discussed in Section 20–11, and include air pollution, oil spills, and the release of CO_2 gas which can trap heat as in a greenhouse to raise the Earth's temperature. The solution to the world's needs for energy is not only technological, but also economic and political. A major factor surely is to "conserve"—to minimize our energy use. "Reduce, reuse, recycle."



FIGURE 42–10 Smoke rising from Fukushina, Japan, after the terrible nuclear power plant meltdown in 2011.

EXAMPLE 42–6 **Uranium fuel amount.** Estimate the minimum amount of $^{235}_{92}\text{U}$ that needs to undergo fission in order to run a 1000-MW power reactor per year of continuous operation. Assume an efficiency (Chapter 20) of about 33%.

APPROACH At 33% efficiency, we need $3 \times 1000\,\text{MW} = 3000 \times 10^6\,\text{J/s}$ input. Each fission releases about 200 MeV (Eq. 42–6), so we divide the energy for a year by 200 MeV to get the number of fissions needed per year. Then we multiply by the mass of one uranium atom.

SOLUTION For 1000-MW output, the total power generation needs to be 3000 MW, of which 2000 MW is dumped as "waste" heat. Thus the total energy release in 1 yr $\left(3 \times 10^7\,\text{s}\right)$ from fission needs to be about

$$\left(3 \times 10^9\,\text{J/s}\right)\left(3 \times 10^7\,\text{s}\right) \approx 10^{17}\,\text{J}.$$

If each fission releases 200 MeV of energy, the number of fissions required for a year is

$$\frac{\left(10^{17}\,\text{J}\right)}{\left(2 \times 10^8\,\text{eV/fission}\right)\left(1.6 \times 10^{-19}\,\text{J/eV}\right)} \approx 3 \times 10^{27}\,\text{fissions}.$$

The mass of a single uranium atom is about $(235\,\text{u})\left(1.66 \times 10^{-27}\,\text{kg/u}\right) \approx 4 \times 10^{-25}\,\text{kg}$, so the total uranium mass needed is

$$\left(4 \times 10^{-25}\,\text{kg/fission}\right)\left(3 \times 10^{27}\,\text{fissions}\right) \approx 1000\,\text{kg},$$

or about a ton of $^{235}_{92}\text{U}$.

NOTE Because $^{235}_{92}\text{U}$ makes up only 0.7% of natural uranium, the yearly requirement for uranium is on the order of a hundred tons. This is orders of magnitude less than coal, both in mass and volume. Coal releases $2.8 \times 10^7\,\text{J/kg}$, whereas $^{235}_{92}\text{U}$ can release $10^{17}\,\text{J}$ per ton, as we just calculated, or $10^{17}\,\text{J}/10^3\,\text{kg} = 10^{14}\,\text{J/kg}$. For natural uranium, the figure is 100 times less, $10^{12}\,\text{J/kg}$.

PHYSICS APPLIED
Energy in coal vs. uranium

†A breeder reactor does *not* produce more fuel than it uses.



FIGURE 42–11 J. Robert Oppenheimer, on the left, with General Leslie Groves, who was the administrative head of Los Alamos during World War II. The photograph was taken at the Trinity site in the New Mexico desert, where the first atomic bomb was exploded.

FIGURE 42–12 Photo taken a month after the bomb was dropped on Nagasaki. The shacks were constructed afterwards from debris in the ruins. The bombs dropped on Hiroshima and Nagasaki were each equivalent to about 20,000 tons of the common explosive TNT $(\sim 10^{14}\,\text{J})$.



EXERCISE C A nuclear-powered submarine needs 6000-kW input power. How many $^{235}_{92}\text{U}$ fissions is this per second?

Atom Bomb

The first use of fission, however, was not to produce electric power. Instead, it was first used as a fission bomb (called the "atomic bomb"). In early 1940, with Europe already at war, Germany's leader, Adolf Hitler, banned the sale of uranium from the Czech mines he had recently taken over. Research into the fission process suddenly was enshrouded in secrecy. Physicists in the United States were alarmed. A group of them approached Einstein—a man whose name was a household word—to send a letter to President Franklin Roosevelt about the possibilities of using nuclear fission for a bomb far more powerful than any previously known, and inform him that Germany might already have begun development of such a bomb. Roosevelt responded by authorizing the program known as the Manhattan Project, to see if a bomb could be built. Work began in earnest after Fermi's demonstration in 1942 that a sustained chain reaction was possible. A new secret laboratory was developed on an isolated mesa in New Mexico known as Los Alamos. Under the direction of J. Robert Oppenheimer (1904–1967; Fig. 42–11), it became the home of famous scientists from all over Europe and the United States.

To build a bomb that was subcritical during transport but that could be made supercritical (to produce a chain reaction) at just the right moment, two pieces of uranium were used, each less than the critical mass but together greater than the critical mass. The two masses, kept separate until the moment of detonation, were then forced together quickly by a kind of gun, and a chain reaction of explosive proportions occurred. A second bomb design detonated conventional explosives (TNT) surrounding a plutonium sphere to compress it by implosion to double its density, making it more than critical and causing a nuclear explosion. The first fission bomb was tested in the New Mexico desert in July 1945. It was successful. In early August, a fission bomb using uranium was dropped on Hiroshima and a second, using plutonium, was dropped on Nagasaki (Fig. 42–12), both in Japan. World War II ended shortly thereafter.

Besides its destructive power, a fission bomb produces many highly radioactive fission fragments, as does a nuclear reactor. When a fission bomb explodes, these radioactive isotopes are released into the atmosphere as **radioactive fallout**.

Testing of nuclear bombs in the atmosphere after World War II was a cause of concern, because the movement of air masses spread the fallout all over the globe. Radioactive fallout eventually settles to the Earth, particularly in rainfall, and is absorbed by plants and grasses and enters the food chain. This is a far more serious problem than the same radioactivity on the exterior of our bodies, because α and β particles are largely absorbed by clothing and the outer (dead) layer of skin. But inside our bodies as food, the isotopes are in contact with living cells. One particularly dangerous radioactive isotope is $^{90}_{38}\text{Sr}$, which is chemically much like calcium and becomes concentrated in bone, where it causes bone cancer and destroys bone marrow. The 1963 treaty signed by over 100 nations that bans nuclear weapons testing in the atmosphere was motivated because of the hazards of fallout.

42–4 Nuclear Fusion

The mass of every stable nucleus is less than the sum of the masses of its constituent protons and neutrons. For example, the mass of the helium isotope $^{4}_{2}\text{He}$ is less than the mass of two protons plus two neutrons, Example 41–3. If two protons and two neutrons were to come together to form a helium nucleus, there would be a loss of mass. This mass loss is manifested in the release of energy.

PSE5_CH42_v1.6_23Jun-20

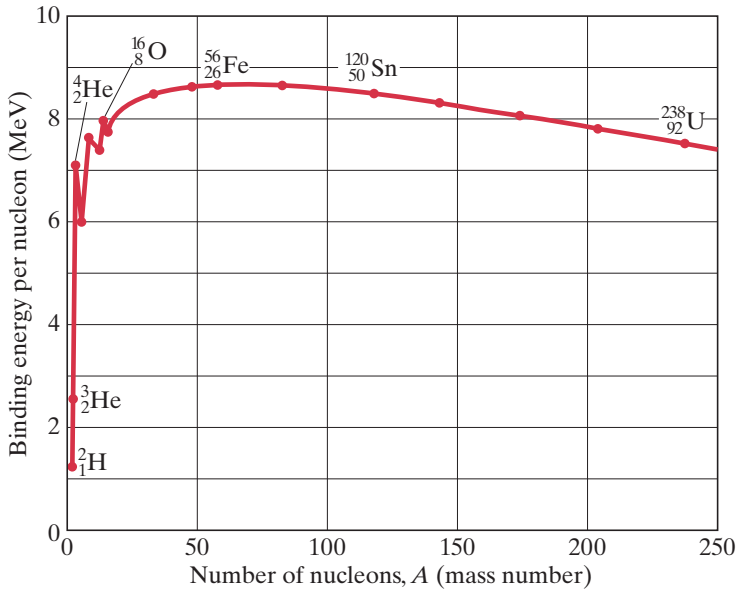


FIGURE 42–13 Average binding energy per nucleon as a function of mass number A for stable nuclei. Same as Fig. 41–1.

Nuclear Fusion; Stars

The process of building up nuclei by bringing together individual protons and neutrons, or building larger nuclei by combining small nuclei, is called **nuclear fusion**. In Fig. 42–13 (same as Fig. 41–1), we can see why small nuclei can combine to form larger ones with the release of energy: it is because the binding energy per nucleon is less for light nuclei than it is for heavier nuclei (up to about $A \approx 60$). For two positively charged nuclei to get close enough to fuse, they must have very high kinetic energy to overcome the electric repulsion. Many of the elements in the universe can be explained as being formed through the process of nuclear fusion in stars (see Chapter 44) where the temperature is extremely high, corresponding to high kinetic energy (Eq. 18–4). Today fusion is still producing the prodigious amounts of light energy (EM waves) stars emit, including our Sun.

EXAMPLE 42–7 **Fusion energy release.** One of the simplest fusion reactions involves the production of deuterium, $^{2}_{1}\text{H}$, from a neutron and a proton: $^{1}_{1}\text{H} + \text{n} \rightarrow {}^{2}_{1}\text{H} + \gamma$. How much energy is released in this reaction?

APPROACH The energy released equals the difference in mass (times c^2) between the initial and final masses.

SOLUTION From Appendix G, the initial mass is

$$1.007825\,\text{u} + 1.008665\,\text{u} = 2.016490\,\text{u},$$

and after the reaction the mass is that of the $^{2}_{1}\text{H}$, namely 2.014102 u (the γ is massless). The mass difference is

$$2.016490\,\text{u} - 2.014102\,\text{u} = 0.002388\,\text{u},$$

so the energy released is

$$(\Delta m)c^2 = (0.002388\,\text{u})\,(931.5\,\text{MeV/u}) = 2.22\,\text{MeV},$$

and it is carried off as kinetic energy of the $^{2}_{1}\text{H}$ nucleus and the γ ray.

The energy output of our Sun is believed to be due principally to the following sequence of fusion reactions:

$$^{1}_{1}\text{H} + {}^{1}_{1}\text{H} \rightarrow {}^{2}_{1}\text{H} + \text{e}^{+} + \nu \qquad (0.42\,\text{MeV}) \qquad \textbf{(42–7a)}$$

$$^{1}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{2}\text{He} + \gamma \qquad (5.49\,\text{MeV}) \qquad \textbf{(42–7b)}$$

$$^{3}_{2}\text{He} + {}^{3}_{2}\text{He} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{1}\text{H} + {}^{1}_{1}\text{H} \qquad (12.86\,\text{MeV}) \qquad \textbf{(42–7c)}$$

Proton–proton chain

where the energy released (Q-value) for each reaction is given in parentheses.

(They include keeping track of atomic electrons as for β-decay, Section 41–5). These reactions are between nuclei, without electrons at these very high temperatures; the first reaction can be written as

$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{d} + \mathrm{e}^+ + \nu$$

where p = proton and d = deuteron. The net effect of this sequence, which is called the **proton–proton chain**, is for four protons to combine to form one $^4_2\mathrm{He}$ nucleus plus two positrons, two neutrinos, and two gamma rays:

$$4\,{}^1_1\mathrm{H} \rightarrow {}^4_2\mathrm{He} + 2\mathrm{e}^+ + 2\nu + 2\gamma. \qquad \textbf{(42–8)}$$

Note that it takes two of each of the first two reactions (Eqs. 42–7a and b) to produce the two $^3_2\mathrm{He}$ for the third reaction. Also, each of the two e^+ formed (Eq. 42–7a) quickly annihilates with an electron to produce 2 γ rays (Section 37–5) with total energy $2m_\mathrm{e}c^2 = 1.02\,\mathrm{MeV}$. So the total energy release for the net reaction, Eq. 42–8, is

$$2(0.42\,\mathrm{MeV}) + 2(1.02\,\mathrm{MeV}) + 2(5.49\,\mathrm{MeV}) + 12.86\,\mathrm{MeV} = 26.7\,\mathrm{MeV}.$$

The first reaction, the formation of deuterium from two protons (Eq. 42–7a), has a very low probability and so limits the rate at which the Sun produces energy. (Thank goodness; this is why the Sun and other stars have a long lifetime and are still shining brightly.)

EXERCISE D Return to the first Chapter-Opening Question, page 1256, and answer it again now. Try to explain why you may have answered it differently the first time.

EXAMPLE 42–8 | **ESTIMATE** **Estimating fusion energy.** Estimate the energy released if the following reaction occurred:

$${}^2_1\mathrm{H} + {}^2_1\mathrm{H} \rightarrow {}^4_2\mathrm{He}.$$

APPROACH We use Fig. 42–13 for a quick estimate.

SOLUTION We see in Fig. 42–13 that each $^2_1\mathrm{H}$ has a binding energy of about $1\frac{1}{4}\,\mathrm{MeV/nucleon}$, which for 2 nuclei of mass 2 is $4\times(1\frac{1}{4}) \approx 5\,\mathrm{MeV}$. The $^4_2\mathrm{He}$ has a binding energy per nucleon (Fig. 42–13) of about 7 MeV for a total of $4\times 7\,\mathrm{MeV} \approx 28\,\mathrm{MeV}$. Hence the energy release is about $28\,\mathrm{MeV} - 5\,\mathrm{MeV} \approx 23\,\mathrm{MeV}$.

In stars hotter than the Sun, it is more likely that the energy output comes principally from the **carbon** (or **CNO**) **cycle**, which comprises the following sequence of reactions:

Carbon cycle

$$\begin{aligned}
{}^{12}_6\mathrm{C} + {}^1_1\mathrm{H} &\rightarrow {}^{13}_7\mathrm{N} + \gamma \\
{}^{13}_7\mathrm{N} &\rightarrow {}^{13}_6\mathrm{C} + \mathrm{e}^+ + \nu \\
{}^{13}_6\mathrm{C} + {}^1_1\mathrm{H} &\rightarrow {}^{14}_7\mathrm{N} + \gamma \\
{}^{14}_7\mathrm{N} + {}^1_1\mathrm{H} &\rightarrow {}^{15}_8\mathrm{O} + \gamma \\
{}^{15}_8\mathrm{O} &\rightarrow {}^{15}_7\mathrm{N} + \mathrm{e}^+ + \nu \\
{}^{15}_7\mathrm{N} + {}^1_1\mathrm{H} &\rightarrow {}^{12}_6\mathrm{C} + {}^4_2\mathrm{He}.
\end{aligned}$$

No net carbon is consumed in this cycle and the net effect is the same as the proton–proton chain, Eq. 42–8 (plus one extra γ). The theory of the proton–proton chain and of the carbon cycle as the source of energy for the Sun and stars was first worked out by Hans Bethe (1906–2005) in 1939.

PSE5_CH42_v1.6_23Jun-20

CONCEPTUAL EXAMPLE 42–9 **Stellar fusion.** What is the heaviest element likely to be produced in fusion processes in stars?

RESPONSE Fusion is possible if the final products have more binding energy (less mass) than the reactants, because then there is a net release of energy. Since the binding energy curve in Fig. 42–13 (or Fig. 41–1) peaks near $A \approx 56$ to 58 which corresponds to iron or nickel, it would not be energetically favorable to produce elements heavier than that. Nevertheless, in the center of massive stars or in supernova explosions, there is enough initial kinetic energy available to drive endothermic reactions that produce heavier elements, as well.

EXERCISE E If the Sun is generating a constant amount of energy via fusion, the mass of the Sun must be (*a*) increasing, (*b*) decreasing, (*c*) constant, (*d*) irregular.

Possible Fusion Reactors

PHYSICS APPLIED
Fusion energy reactors

The possibility of utilizing the energy released in fusion to make a power reactor is very attractive. The fusion reactions most likely to succeed in a reactor involve the isotopes of hydrogen, ${}^2_1\text{H}$ (deuterium) and ${}^3_1\text{H}$ (tritium), and are as follows, with the energy released given in parentheses:

$$ {}^2_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_1\text{H} + {}^1_1\text{H} \qquad (4.03\ \text{MeV}) \qquad \textbf{(42–9a)} $$

$$ {}^2_1\text{H} + {}^2_1\text{H} \rightarrow {}^3_2\text{He} + \text{n} \qquad (3.27\ \text{MeV}) \qquad \textbf{(42–9b)} $$

$$ {}^2_1\text{H} + {}^3_1\text{H} \rightarrow {}^4_2\text{He} + \text{n}. \qquad (17.59\ \text{MeV}) \qquad \textbf{(42–9c)} $$

Comparing these energy yields with that for the fission of ${}^{235}_{92}\text{U}$, we can see that the energy released in fusion reactions can be greater for a given mass of fuel than in fission. Furthermore, as fuel, a fusion reactor could use deuterium, which is very plentiful in the water of the oceans (the natural abundance of ${}^2_1\text{H}$ is 0.0115% on average, or about 1 g of deuterium per 80 L of water). The simple proton–proton reaction of Eq. 42–7a, which could use a much more plentiful source of fuel, ${}^1_1\text{H}$, has such a small probability of occurring that it cannot be considered a possibility on Earth.

Although a useful fusion reactor has not yet been achieved, considerable progress has been made in overcoming the inherent difficulties. The problems are associated with the fact that all nuclei have a positive charge and repel each other. However, if they can be brought close enough together so that the short-range attractive strong nuclear force can come into play, it can pull the nuclei together and fusion can occur. For the nuclei to get close enough together, they must have large kinetic energy to overcome the electric repulsion. High kinetic energies are readily attainable with particle accelerators (Chapter 43), but the number of particles involved is too small. To produce realistic amounts of energy, we must deal with matter in bulk, for which high kinetic energy means higher temperatures. Indeed, very high temperatures are required for sustained fusion to occur, and fusion devices are often referred to as **thermonuclear devices**. The interiors of the Sun and other stars are very hot, many millions of degrees, so the nuclei are moving fast enough for fusion to take place, and the energy released keeps the temperature high so that further fusion reactions can occur. The Sun and the stars represent huge self-sustaining thermonuclear reactors that stay together because of their great gravitational mass. But on Earth, containment of the fast-moving nuclei at the high temperatures and densities required has proven difficult.

It was realized after World War II that the temperature produced within a fission (or "atomic") bomb was close to 10^8 K. This suggested that a fission bomb could be used to ignite a fusion bomb (popularly known as a thermonuclear or hydrogen bomb) to release the vast energy of fusion. The uncontrollable release of fusion energy in an H-bomb (in 1952) was relatively less difficult to obtain. But to realize usable energy from fusion at a slow and controlled rate has turned out to be a serious challenge.

EXAMPLE 42–10 **ESTIMATE** **Temperature needed for d–t fusion.** Estimate the temperature required for deuterium–tritium fusion (d–t) to occur.

APPROACH We assume the nuclei approach head-on, each with kinetic energy K, and that the nuclear force comes into play when the distance between their centers equals the sum of their nuclear radii. The electrostatic potential energy (Chapter 23) of the two particles at this distance equals the minimum total kinetic energy of the two particles when far apart. The average kinetic energy is related to Kelvin temperature by Eq. 18–4.

SOLUTION The radii of the two nuclei ($A_d = 2$ and $A_t = 3$) are given by Eq. 41–1: $r_d \approx 1.5\,\text{fm}$, $r_t \approx 1.7\,\text{fm}$, so $r_d + r_t = 3.2 \times 10^{-15}\,\text{m}$. We equate the kinetic energy of the two initial particles to the potential energy when at this distance:

$$2K \approx \frac{1}{4\pi\epsilon_0}\frac{e^2}{(r_d + r_t)}$$

$$\approx \left(9.0 \times 10^9 \frac{\text{N}\cdot\text{m}^2}{\text{C}^2}\right)\frac{(1.6 \times 10^{-19}\,\text{C})^2}{(3.2 \times 10^{-15}\,\text{m})(1.6 \times 10^{-19}\,\text{J/eV})} \approx 0.45\,\text{MeV}.$$

Thus, $K \approx 0.22\,\text{MeV}$, and if we ask that the average kinetic energy be this high, then from Eq. 18–4, $\frac{3}{2}kT = \overline{K}$, we have a temperature of

$$T = \frac{2\overline{K}}{3k} = \frac{2(0.22\,\text{MeV})(1.6 \times 10^{-13}\,\text{J/MeV})}{3(1.38 \times 10^{-23}\,\text{J/K})} \approx 2 \times 10^9\,\text{K}.$$

NOTE More careful calculations show that the temperature required for fusion is actually about an order of magnitude less than this rough estimate, partly because it is not necessary that the *average* kinetic energy be 0.22 MeV—a small percentage of nuclei with this much energy (in the high-energy tail of the Maxwell distribution, Fig. 18–3) would be sufficient. Reasonable estimates for a usable fusion reactor are in the range $T \gtrsim 1$ to $4 \times 10^8\,\text{K}$.

A high temperature is required for a fusion reactor. But there must also be a high density of nuclei to ensure a sufficiently high collision rate. A real difficulty with controlled fusion is to contain nuclei long enough and at a high enough density for sufficient reactions to occur so that a usable amount of energy is obtained. At the temperatures needed for fusion, the atoms are ionized, and the resulting collection of nuclei and electrons is referred to as a **plasma**. Ordinary materials vaporize at a few thousand degrees at most, and hence cannot be used to contain a high-temperature plasma. Two major containment techniques are *magnetic confinement* and *inertial confinement*.

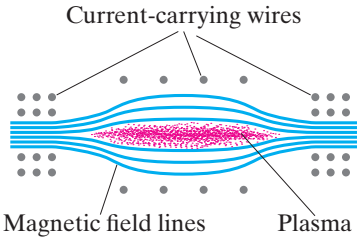


FIGURE 42–14 "Magnetic bottle" used to confine a plasma.

In **magnetic confinement**, magnetic fields are used to try to contain the hot plasma. A simple approach is the "magnetic bottle" shown in Fig. 42–14. The paths of the charged particles in the plasma are bent by the magnetic field; where magnetic field lines are close together, the force on the particles reflects them back toward the center. Unfortunately, magnetic bottles develop "leaks" and the charged particles slip out before sufficient fusion takes place. The most promising design today is the **tokamak**, first developed in Russia. A tokamak (Fig. 42–15) is

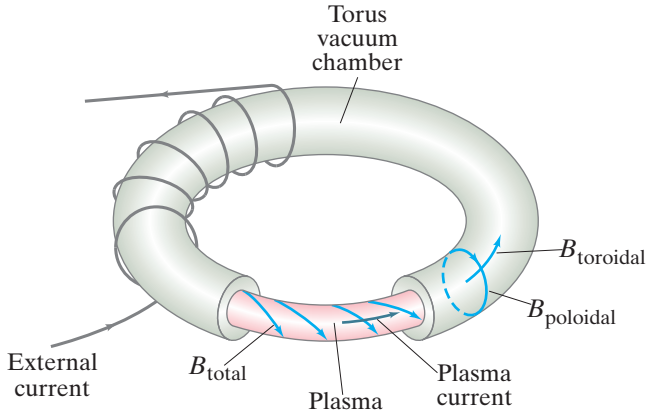


FIGURE 42–15 Tokamak configuration, showing the total $\vec{\mathbf{B}}$ field due to external current plus current in the plasma itself.

PSE5_CH42_v1.6_23Jun-20

toroid-shaped (a torus, which is like a donut) and involves complicated magnetic fields: current-carrying conductors produce a magnetic field directed along the axis of the toroid ("toroidal" field); an additional field is produced by currents within the plasma itself ("poloidal" field). The combination produces a helical field as shown in Fig. 42–15, confining the plasma, at least briefly, so it doesn't touch the vacuum chamber's metal walls (Fig. 42–16a).

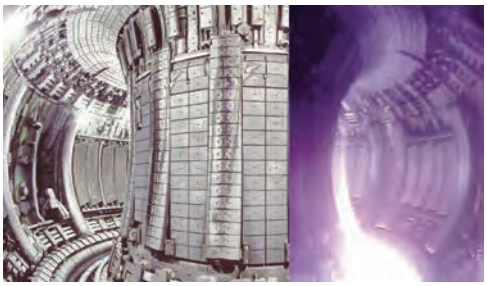
In 1957, J. D. Lawson showed that the product of ion density n $(= \text{ions/m}^3)$ and confinement time τ must exceed a minimum value of approximately

$$n\tau \gtrsim 3 \times 10^{20}\,\text{s/m}^3.$$

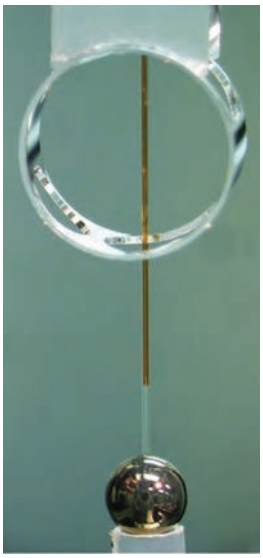
This **Lawson criterion** must be reached to produce **ignition**, meaning fusion that continues after all external heating is turned off. Practically, it is expected to be achieved with $n \approx 1$ to $3 \times 10^{20}\,\text{m}^{-3}$ and $\tau \approx 1$ to 3 s. To reach **break-even**, the point at which the energy output due to fusion is equal to the energy input to heat the plasma, requires an $n\tau$ about an order of magnitude less. The break-even point was very closely approached in the 1990s at the Tokamak Fusion Test Reactor (TFTR) at Princeton, and the very high temperature needed for ignition $(4 \times 10^8\,\text{K})$ was exceeded—although not both of these at the same time.

Magnetic confinement fusion research continues throughout the world. This research will help us in developing the huge multinational test device (European Union, India, Japan, South Korea, Russia, China, and the U.S.), called ITER (International Thermonuclear Experimental Reactor) situated in France. It is expected that ITER will produce temperatures above 10^8 K for extended periods (minutes or hours) and to begin running in the 2020s, with an expected power output of about 500 MW, 10 times the input energy. ITER is planned to be the final research step before building a working reactor.

The second method for containing the fuel for fusion is **inertial confinement fusion** (ICF): a small pellet or capsule of deuterium and tritium (Fig. 42–16b) is struck simultaneously from hundreds of directions by very intense laser beams. The intense influx of energy heats and ionizes the pellet into a plasma, compressing it and heating it to temperatures at which fusion can occur $(> 10^8\,\text{K})$. The confinement time is on the order of 10^{-11} to 10^{-9} s, during which time the ions do not move appreciably because of their own inertia, and fusion can take place.



(a)



(b)

FIGURE 42–16 (a) Tokamak: split image view of the Joint European Torus (JET) located near Oxford, England. Interior, on the left, and an actual plasma in there $(\text{T} \approx 1 \times 10^8\,\text{K})$ on the right. (b) A 2-mm-diameter round d–t (deuterium–tritium) inertial target, being filled through a thin glass tube from above, at the National Ignition Facility (NIF), Lawrence Livermore National Laboratory, California.

42–5 Passage of Radiation Through Matter; Biological Damage

When we speak of *radiation*, we include α, β, γ, and X-rays, as well as protons, neutrons, and other particles such as pions (see Chapter 43). Because charged particles can ionize the atoms or molecules of any material they pass through, they are referred to as **ionizing radiation**. And because radiation produces ionization, it can cause considerable damage to materials, particularly to biological tissue.

Charged particles, such as α and β rays and protons, cause ionization because of electric forces. That is, when they pass through a material, they can attract or repel electrons strongly enough to remove them from the atoms of the material. Since the α and β rays emitted by radioactive substances have energies on the order of 1 MeV $(10^4$ to 10^7 eV), whereas ionization of atoms and molecules requires on the order of 10 eV (Chapter 37), we see that a single α or β particle can cause thousands of ionizations.

Neutral particles also give rise to ionization when they pass through materials. For example, X-ray and γ-ray photons can ionize atoms by knocking out electrons by means of the photoelectric and Compton effects (Chapter 37). Furthermore, if a γ ray has sufficient energy (greater than 1.02 MeV), it can undergo pair production: an electron and a positron are produced (Section 37–5). The charged particles produced in all of these processes can themselves go on to produce further ionization. Neutrons, on the other hand, interact with matter mainly by collisions with nuclei, with which they interact strongly.

Often the nucleus is broken apart by such a collision, altering the molecule of which it was a part. The fragments produced can in turn cause ionization.

Radiation passing through matter can do considerable damage. Metals and other structural materials become brittle and their strength can be weakened if the radiation is very intense, as in nuclear reactor power plants and for space vehicles that must pass through areas of intense cosmic radiation.

*Biological Damage

PHYSICS APPLIED
Biological radiation damage

The radiation damage produced in biological organisms is due primarily to ionization produced in cells. Several related processes can occur. Ions or radicals are produced that are highly reactive and take part in chemical reactions that interfere with the normal operation of the cell. All forms of radiation can ionize atoms by knocking out electrons. If these are bonding electrons, the molecule may break apart, or its structure may be altered so it does not perform its normal function or may perform a harmful function. In the case of proteins, the loss of one molecule is not serious if there are other copies of the protein in the cell and additional copies can be made from the gene that codes for it. However, large doses of radiation may damage so many molecules that new copies cannot be made quickly enough, and the cell dies.

Damage to the DNA is more serious, since a cell may have only one copy. Each alteration in the DNA can affect a gene and alter the molecule that gene codes for (Section 40–3), so that needed proteins or other molecules may not be made at all. Again the cell may die. The death of a single cell is not normally a problem, since the body can replace it with a new one. (There are exceptions, such as neurons, which are mostly not replaceable, so their loss is serious.) But if many cells die, the organism may not be able to recover. On the other hand, a cell may survive but be defective. It may go on dividing and produce many more defective cells, to the detriment of the whole organism. Thus radiation can cause cancer—the rapid uncontrolled production of cells.

The possible damage done by the medical use of X-rays and other radiation must be balanced against the medical benefits and prolongation of life as a result of their diagnostic use.

42–6 Measurement of Radiation—Dosimetry

Although the passage of ionizing radiation through the human body can cause considerable damage, radiation can also be used to treat certain diseases, particularly cancer, often by using very narrow beams directed at a cancerous tumor in order to destroy it (Section 42–7). It is therefore important to be able to quantify the amount, or **dose**, of radiation. This is the subject of **dosimetry**.

PHYSICS APPLIED
Dosimetry

The strength of a source can be specified at a given time by stating the **source activity**: how many nuclear decays (or disintegrations) occur per second. The traditional unit is the **curie** (Ci), defined as

$$1\ \text{Ci} = 3.70 \times 10^{10}\ \text{decays per second.}$$

(This number comes from the original definition as the activity of exactly one gram of radium.) Although the curie is still in common use, the SI unit for source activity is the **becquerel** (Bq), defined as

$$1\ \text{Bq} = 1\ \text{decay/s.}$$

CAUTION
In the lab, activity will be less than written on the bottle—note the date

Commercial suppliers of **radionuclides** (radioactive nuclides) used as tracers specify the activity at a given time. Because the activity decreases over time, more so for short-lived isotopes, it is important to take this decrease into account.

The magnitude of the source activity, $|dN/dt|$, is related to the number of radioactive nuclei present, N, and to the half-life, $T_{\frac{1}{2}}$, by (see Section 41–8):

$$\left|\frac{dN}{dt}\right| = \lambda N = \frac{0.693}{T_{\frac{1}{2}}} N.$$

PSE5_CH42_v1.6_23Jun-20

EXAMPLE 42–11 **Radioactivity taken up by cells.** In a certain experiment, 0.016 μCi of $^{32}_{15}$P is injected into a medium containing a culture of bacteria. After 1.0 h the cells are washed and a 70% efficient detector (counts 70% of emitted β rays) records 720 counts per minute from the cells. What percentage of the original $^{32}_{15}$P was taken up by the cells?

APPROACH The half-life of $^{32}_{15}$P is about 14 days (Appendix F), so we can ignore any loss of activity over 1 hour. From the given activity, we find how many β rays are emitted. We can compare 70% of this to the $(720/\text{min})/(60\,\text{s/min}) = 12$ per second detected.

SOLUTION The total number of decays per second originally was $(0.016 \times 10^{-6})(3.7 \times 10^{10}) = 590$. The counter could be expected to count 70% of this, or 410 per second. Since it counted $720/60 = 12$ per second, then $12/410 = 0.029$ or 2.9% was incorporated into the cells.

Another type of measurement is the exposure or **absorbed dose**—that is, the *effect* the radiation has on the absorbing material. The earliest unit of dosage was the **roentgen** (R), defined in terms of the amount of ionization produced by the radiation $(1\,\text{R} = 1.6 \times 10^{12}$ ion pairs per gram of dry air at standard conditions$)$. Today, 1 R is defined as the amount of X-ray or γ radiation that deposits 0.878×10^{-2} J of energy per kilogram of air. The roentgen was largely superseded by another unit of absorbed dose applicable to any type of radiation, the **rad**: *1 rad is that amount of radiation which deposits energy per unit mass of* 1.00×10^{-2} *J/kg in any absorbing material.* (This is quite close to the roentgen for X- and γ rays.) The proper SI unit for absorbed dose is the **gray** (Gy):

$$1\,\text{Gy} = 1\,\text{J/kg} = 100\,\text{rad}. \qquad \textbf{(42–10)}$$

The absorbed dose depends not only on the energy per particle and on the strength of a given source or of a radiation beam (number of particles per second), but also on the type of material that is absorbing the radiation. Bone, for example, absorbs more of X-ray or γ radiation normally used than does flesh, so the same beam passing through a human body deposits a greater dose (in rads or grays) in bone than in flesh.

The gray and the rad are physical units of dose—the energy deposited per unit mass of material. They are, however, not the most meaningful units for measuring the biological damage produced by radiation because equal doses of different types of radiation cause differing amounts of damage. For example, 1 rad of α radiation does 10 to 20 times the amount of damage as 1 rad of β or γ rays. This difference arises largely because α rays (and other heavy particles such as protons and neutrons) move much more slowly than β and γ rays of equal energy due to their greater mass. Hence, ionizing collisions occur closer together, so more irreparable damage can be done. The **relative biological effectiveness** (RBE) of a given type of radiation is defined as the number of rads of X-ray or γ radiation that produces the same biological damage as 1 rad of the given radiation. For example, 1 rad of slow neutrons does the same damage as 5 rads of X-rays. Table 42–1 gives the RBE for several types of radiation. The numbers are approximate because they depend somewhat on the energy of the particles and on the type of damage that is used as the criterion.

The **effective dose** can be given as the product of the dose in rads and the RBE, and this unit is known as the **rem** (which stands for *rad equivalent man*):

$$\text{effective dose (in rem)} = \text{dose (in rad)} \times \text{RBE}. \qquad \textbf{(42–11a)}$$

This unit is being replaced by the SI unit for "effective dose," the **sievert** (Sv):

$$\text{effective dose (Sv)} = \text{dose (Gy)} \times \text{RBE} \qquad \textbf{(42–11b)}$$

so

$$1\,\text{Sv} = 100\,\text{rem} \quad \text{or} \quad 1\,\text{rem} = 10\,\text{mSv}.$$

TABLE 42–1 Relative Biological Effectiveness (RBE)

Type	RBE
X- and γ rays	1
β (electrons)	1
Protons	2
Slow neutrons	5
Fast neutrons	≈10
α particles and heavy ions	≈20

By these definitions, 1 rem (or 1 Sv) of any type of radiation does approximately the same amount of biological damage. For example, 50 rem of fast neutrons does the same damage as 50 rem of γ rays. But note that 50 rem of fast neutrons is only 5 rads, whereas 50 rem of γ rays is 50 rads.

EXERCISE F Return to the second Chapter-Opening Question, page 1256, and answer it again now. Try to explain why you may have answered it differently the first time.

Human Exposure to Radiation

PHYSICS APPLIED
Radon

We are constantly exposed to low-level radiation from natural sources: cosmic rays, natural radioactivity in rocks and soil, and naturally occurring radioactive isotopes in our food, such as $^{40}_{19}$K. **Radon**, $^{222}_{86}$Rn, is of considerable concern today. It is the product of radium decay and is an intermediate in the decay series from uranium (see Fig. 41–12). Most intermediates remain in the rocks where formed, but radon is a gas that can escape from rock (and from building material like concrete) to enter the air we breathe, and damage the interior of the lung.

PHYSICS APPLIED
Human radiation exposure

The **natural radioactive background** averages about 0.30 rem (300 mrem) per year per person in the U.S., although there are large variations. From medical X-rays and scans, the average person receives about 50 to 60 mrem per year, giving an average total dose of about 360 mrem (3.6 mSv) per person. U.S. government regulators suggest an upper limit of allowed radiation for an individual in the general populace at about 100 mrem (1 mSv) per year in addition to natural background. It is believed that even low doses of radiation increase the chances of cancer or genetic defects; there is *no safe level* or threshold of radiation exposure.

PHYSICS APPLIED
Radiation worker exposure
Film badge

The upper limit for people who work around radiation—in hospitals, in power plants, in research—has been set higher, a maximum of 20 mSv (2 rem) whole-body dose, averaged over some years (a maximum of 50 mSv (5 rem/yr) in any one year). To monitor exposure, those people who work around radiation generally carry some type of dosimeter, one common type being a **radiation film badge** which is a piece of film wrapped in light-tight material. The passage of ionizing radiation through the film changes it so that the film is darkened upon development, and thus indicates the received dose. Newer types include the *thermoluminescent dosimeter* (TLD). Dosimeters and badges do not protect the worker, but high levels detected suggest reassignment or modified work practices to reduce radiation exposure to acceptable levels.

PHYSICS APPLIED
Radiation sickness

Large doses of radiation can cause unpleasant symptoms such as nausea, fatigue, and loss of body hair, because of cellular damage. Such effects are sometimes referred to as **radiation sickness**. Very large doses can be fatal, although the time span of the dose is important. A brief dose of 10 Sv (1000 rem) is nearly always fatal. A 3-Sv (300-rem) dose in a short period of time is fatal in about 50% of patients within a month. However, the body possesses remarkable repair processes, so that a 3-Sv dose spread over several weeks is usually not fatal. It will, nonetheless, cause considerable damage to the body.

The effects of low doses over a long time are difficult to determine and are not well known as yet.

CONCEPTUAL EXAMPLE 42–12 **Limiting the dose.** A worker in an environment with a radioactive source is warned that she is accumulating a dose too quickly and will have to lower her exposure by a factor of ten to continue working for the rest of the year. If the worker is able to work farther away from the source, how much farther away is necessary?

RESPONSE If the energy is radiated uniformly in all directions, then the intensity (dose/area) should decrease as the distance squared, just as it does for sound and light. If she can work four times farther away, the exposure lowers by a factor of sixteen, enough to make her safe.

PSE5_CH42_v1.6_23Jun-20

EXAMPLE 42–13 **Whole-body dose.** What whole-body dose is received by a 70-kg laboratory worker exposed to a 40-mCi $^{60}_{27}$Co source, assuming the person's body has cross-sectional area 1.5 m^2 and is normally about 4.0 m from the source for 4.0 h per day? $^{60}_{27}$Co emits γ rays of energy 1.33 MeV and 1.17 MeV in quick succession. Approximately 50% of the γ rays interact in the body and deposit all their energy. (The rest pass through.)

APPROACH Of the given energy emitted, only a fraction passes through the worker, equal to *her* area divided by the total area (or $4\pi r^2$) over a full sphere of radius 4.0 m (Fig. 42–17).

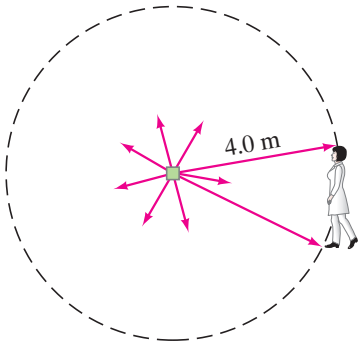


FIGURE 42–17 Radiation spreads out in all directions. A person 4.0 m away intercepts only a fraction: her cross-sectional area divided by the area of a sphere of radius 4.0 m. Example 42–13.

SOLUTION The total γ-ray energy per decay is $(1.33 + 1.17)$ MeV $= 2.50$ MeV, so the total energy emitted by the source per second is

$$(0.040\ \text{Ci})\left(3.7 \times 10^{10}\ \text{decays/Ci} \cdot \text{s}\right)(2.50\ \text{MeV}) \;=\; 3.7 \times 10^{9}\ \text{MeV/s}.$$

The proportion of this energy intercepted by the body is its 1.5-m^2 area divided by the area of a sphere of radius 4.0 m (Fig. 42–17):

$$\frac{1.5\ \text{m}^2}{4\pi r^2} \;=\; \frac{1.5\ \text{m}^2}{4\pi\,(4.0\ \text{m})^2} \;=\; 7.5 \times 10^{-3}.$$

So the rate energy is deposited in the body (remembering that only 50% of the γ rays interact in the body) is

$$E = \left(\tfrac{1}{2}\right)\left(7.5 \times 10^{-3}\right)\left(3.7 \times 10^{9}\ \text{MeV/s}\right)\left(1.6 \times 10^{-13}\ \text{J/MeV}\right) \;=\; 2.2 \times 10^{-6}\ \text{J/s}.$$

Since $1\ \text{Gy} = 1\ \text{J/kg}$, the whole-body dose rate for this 70-kg person is $\left(2.2 \times 10^{-6}\ \text{J/s}\right)/(70\ \text{kg}) = 3.1 \times 10^{-8}\ \text{Gy/s}$. In 4.0 h, this amounts to a dose of

$$(4.0\ \text{h})\,(3600\ \text{s/h})\left(3.1 \times 10^{-8}\ \text{Gy/s}\right) \;=\; 4.5 \times 10^{-4}\ \text{Gy}.$$

RBE ≈ 1 for gammas, so the effective dose is 450 μSv (Eqs. 42–11b and 42–10) or:

$$(100\ \text{rad/Gy})\left(4.5 \times 10^{-4}\ \text{Gy}\right)(1\ \text{rem/rad}) \;=\; 45\ \text{mrem} \;=\; 0.45\ \text{mSv}.$$

NOTE This 45-mrem effective dose is almost 50% of the normal allowed dose for a whole year $(100\ \text{mrem/yr})$, or 1% of the maximum one-year allowance for radiation workers. This worker should not receive such a large dose every day and should seek ways to reduce it (shield the source, vary the work, work farther from the source, work less time this close to source, etc.).

We have assumed that the intensity of radiation decreases as the square of the distance. It actually falls off faster than $1/r^2$ because of absorption in the air, so our answers are a slight overestimate of dose received.

PHYSICS APPLIED
Radon exposure

EXAMPLE 42–14 **Radon exposure.** In the U.S., yearly deaths from radon exposure (the second leading cause of lung cancer) are estimated to be on the order of 20,000 and maybe much more. The Environmental Protection Agency recommends taking action to reduce the radon concentration in living areas if it exceeds 4 pCi/L of air. In some areas 50% of houses exceed this level from naturally occurring radon in the soil. Estimate (*a*) the number of decays/s in 1.0 m^3 of air and (*b*) the mass of radon that emits 4.0 pCi of $^{222}_{86}Rn$ radiation.

APPROACH We can use the definition of the curie to determine how many decays per second correspond to 4 pCi, then Eq. 41–7b to determine how many nuclei of radon it takes to have this activity $\left|dN/dt\right|$.

SOLUTION (*a*) We saw at the start of this Section that $1\text{ Ci} = 3.70 \times 10^{10}$ decays/s. Thus

$$\left|\frac{dN}{dt}\right| = 4.0\text{ pCi} = (4.0 \times 10^{-12}\text{ Ci})(3.70 \times 10^{10}\text{ decays/s/Ci})$$
$$= 0.148\text{ s}^{-1}$$

per liter of air. In 1 m^3 of air $(1\text{ m}^3 = 10^6\text{ cm}^3 = 10^3\text{ L})$ there would be $(0.148\text{ s}^{-1})(1000) \approx 150$ decays/s.

(*b*) From Eqs. 41–7a and 41–8

$$\frac{dN}{dt} = \lambda N = \frac{0.693}{T_{\frac{1}{2}}} N.$$

Appendix G tells us $T_{\frac{1}{2}} = 3.8235$ days for radon, so

$$N = \left|\frac{dN}{dt}\right| \frac{T_{\frac{1}{2}}}{0.693}$$
$$= (0.148\text{ s}^{-1}) \frac{(3.8235\text{ days})(8.64 \times 10^4\text{ s/day})}{0.693}$$
$$= 7.06 \times 10^4\text{ atoms of radon-222.}$$

The molar mass (222 u) and Avogadro's number are used to find the mass:

$$m = \frac{(7.06 \times 10^4\text{ atoms})(222\text{ g/mol})}{6.02 \times 10^{23}\text{ atoms/mol}} = 2.6 \times 10^{-17}\text{ g}$$

or 26 attograms in 1 L of air at the limit of 4 pCi/L. This 2.6×10^{-17} g/L is 2.6×10^{-14} grams of radon per m^3 of air.

NOTE Each radon atom emits 4 α particles and 4 β particles, each one capable of causing many harmful ionizations, before the sequence of decays reaches a stable element.

PHYSICS APPLIED
Radiation therapy

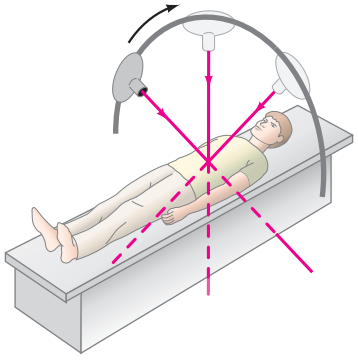


FIGURE 42–18 Radiation source rotates so that the beam always passes through the diseased tissue, but minimizes the dose in the rest of the body.

*42–7 Radiation Therapy

The medical application of radioactivity and radiation to human beings involves two basic aspects: (1) **radiation therapy**—the treatment of disease (mainly cancer)—which we discuss in this Section; and (2) the *diagnosis* of disease, which we discuss in the following Sections of this Chapter.

Radiation can cause cancer. It can also be used to treat it. Rapidly growing cancer cells are especially susceptible to destruction by radiation. Nonetheless, large doses are needed to kill the cancer cells, and some of the surrounding normal cells are inevitably killed as well. It is for this reason that cancer patients receiving radiation therapy often suffer side effects characteristic of radiation sickness. To minimize the destruction of normal cells, a narrow beam of γ or X-rays is often used when a cancerous tumor is well localized. The beam is directed at the tumor, and the source (or body) is rotated so that the beam passes through various parts of the body to keep the dose at any one place as low as possible—except at the tumor and its immediate surroundings, where the beam passes at all times (Fig. 42–18).

The radiation may be from a radioactive source such as $^{60}_{27}$Co, or it may be from an X-ray machine that produces photons in the range 200 keV to 5 MeV. Protons, neutrons, electrons, and pions, which are produced in particle accelerators (Section 43–1), are also being used in cancer therapy.

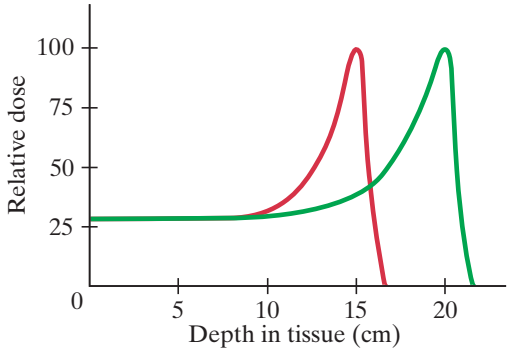


FIGURE 42–19 Energy deposited in tissue as a function of depth for 170-MeV protons (red curve) and 190-MeV protons (green). The peak of each curve is often called the Bragg peak.

Protons used to kill tumors have a special property that makes them particularly useful. As shown in Fig. 42–19, when protons enter tissue, most of their energy is deposited at the end of their path. The protons' initial kinetic energy can be chosen so that most of the energy is deposited at the depth of the tumor itself, to destroy it. The incoming protons deposit only a small amount of energy in the tissue in front of the tumor, and none at all behind the tumor, thus having less negative effect on healthy tissue than X- or γ rays. Because tumors have physical size, even several centimeters in diameter, a range of proton energies is often used. Heavier ions, such as α particles or carbon ions, are similarly useful. This **proton therapy** technique is more than a half century old, but the necessity of having a large accelerator has meant that few hospitals have used the technique until now.

PHYSICS APPLIED
Proton therapy

Another form of treatment is to insert a tiny radioactive source directly inside a tumor, which will eventually kill the majority of the cells. A similar technique is used to treat cancer of the thyroid with the radioactive isotope $^{131}_{53}$I. The thyroid gland concentrates iodine present in the bloodstream, particularly in any area where abnormal growth is taking place. Its intense radioactivity can destroy the defective cells.

Another application of radiation is for sterilizing bandages, surgical equipment, and even packaged foods such as ground beef, chicken, and produce, because bacteria and viruses can be killed or deactivated by large doses of radiation.

*42–8 Tracers in Research and Medicine

Radioactive isotopes are used in biological and medical research as **tracers**. A given compound is artificially synthesized incorporating a radioactive isotope such as $^{14}_{6}$C or $^{3}_{1}$H. Such "tagged" molecules can then be traced as they move through an organism or as they undergo chemical reactions. The presence of these tagged molecules (or parts of them, if they undergo chemical change) can be detected by a Geiger or scintillation counter, which detects emitted radiation (see Section 41–11). How food molecules are digested, and to what parts of the body they are diverted, can be traced in this way.

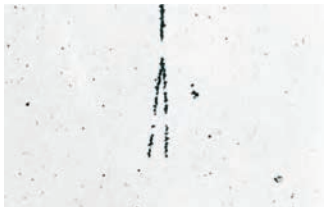
Radioactive tracers have been used to determine how amino acids and other essential compounds are synthesized by organisms. The permeability of cell walls to various molecules and ions can be determined using radioactive tracers: the tagged molecule or ion is injected into the extracellular fluid, and the radioactivity present inside and outside the cells is measured as a function of time.

In a technique known as **autoradiography**, the position of the radioactive isotopes is detected on film. For example, the distribution of carbohydrates produced in the leaves of plants from absorbed CO_2 can be observed by keeping the plant in an atmosphere where the carbon atom in the CO_2 is $^{14}_{6}$C. After a time, a leaf is placed firmly on a photographic plate and the emitted radiation darkens the film most strongly where the isotope is most strongly concentrated (Fig. 42–20a).

FIGURE 42–20 (a) Autoradiograph of a leaf exposed for 30 s to $^{14}CO_2$. Only the tissue where the CO_2 has been taken up, to be used in photosynthesis (Example 37–7), has become radioactive. The non-metabolizing tissue of the veins is free of $^{14}_{6}$C and does not blacken the X-ray sheet. (b) Autoradiograph of chromosomal DNA. The dashed arrays of film grains show the Y-shaped growing point of replicating DNA.



(a)



(b)

Autoradiography using labeled nucleotides (components of DNA) has revealed much about the details of DNA replication (Fig. 42–20b). Today gamma cameras are used in a similar way.

PHYSICS APPLIED
Medical diagnosis

For medical diagnosis, the radionuclide commonly used today is $^{99m}_{43}$Tc, a long-lived excited state of technetium-99 (the "m" in the symbol stands for "metastable" state). It is formed when $^{99}_{42}$Mo decays. The great usefulness of $^{99m}_{43}$Tc derives from its convenient half-life of 6 h (short, but not too short) and the fact that it can combine with a large variety of compounds. The compound to be labeled with the radionuclide is so chosen because it concentrates in the organ or region of the anatomy to be studied. Detectors outside the body then record, or image, the distribution of the radioactively labeled compound. The detection could be done by a single detector (Fig. 42–21a) which is moved across the body, measuring the intensity of radioactivity at a large number of points. The image represents the relative intensity of radioactivity at each point. The relative radioactivity is a diagnostic tool. For example, high or low radioactivity may represent overactivity or underactivity of an organ or part of an organ, or in another case may represent a lesion or tumor. More complex **gamma cameras** make use of many detectors which simultaneously record the radioactivity at many points. The measured intensities can be displayed on a TV or computer monitor. The image is sometimes called a scintigram (after scintillator), Fig. 42–21b. Gamma cameras are relatively inexpensive, but their resolution is limited—by non-perfect collimation.[†] Yet they allow "dynamic" studies: images that change in time, like a movie.

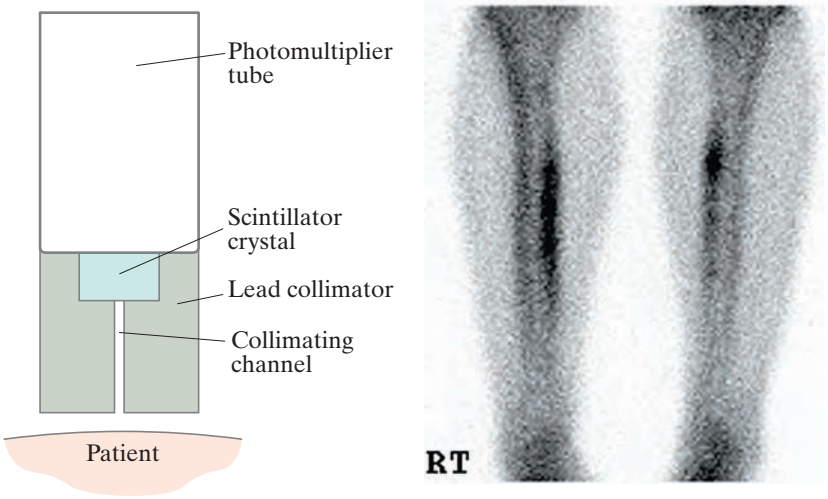


FIGURE 42–21 (a) Collimated gamma-ray detector for scanning (moving) over a patient. The collimator selects γ rays that come in a (nearly) straight line from the patient. Without the collimator, γ rays from all parts of the body could strike the scintillator, producing a poor image. Detectors today usually have many collimator tubes and are called *gamma cameras*. (b) Gamma camera image (scintigram), of both legs of a patient with shin splints, detecting γs from $^{99m}_{43}$Tc.

*42–9 Emission Tomography: PET and SPECT

PHYSICS APPLIED
Medical imaging

The images formed using the standard techniques of nuclear medicine, as briefly discussed in the previous Section, are produced from radioactive tracer sources within the *volume* of the body. It is also possible to image the radioactive emissions from a single plane or slice through the body using the computed tomography techniques discussed in Section 35–11. A gamma camera measures the radioactive intensity from the tracer at many points and angles around the patient. The data are processed in much the same way as for X-ray CT scans (Section 35–11).

PHYSICS APPLIED
Emission tomography (SPECT, PET)

This technique is referred to as **single photon emission computed tomography** (SPECT), or simply SPET (single photon emission tomography).

[†]To "collimate" means to "make parallel," usually by blocking non-parallel rays with a narrow tube inside lead, as in Fig. 42–21a.

Another important technique is **positron emission tomography** (PET), which makes use of positron emitters such as $^{11}_{6}\text{C}$, $^{13}_{7}\text{N}$, $^{15}_{8}\text{O}$, and $^{18}_{9}\text{F}$ whose half-lives are short. These isotopes are incorporated into molecules that, when inhaled or injected, accumulate in the organ or region of the body to be studied. When such a nuclide undergoes β^+ decays, the emitted positron travels at most a few millimeters before it collides with a normal electron. In this collision, the positron and electron are annihilated, producing two γ rays $(e^+ + e^- \rightarrow 2\gamma)$, each having an energy of 511 keV $(= m_e c^2)$. The two γ rays fly off in opposite directions $(180° \pm 0.25°)$ since they must have almost exactly equal and opposite momenta to conserve momentum (the momenta of the initial e^+ and e^- are essentially zero compared to the momenta afterward of the γ rays). Because the photons travel along the same line in opposite directions, their detection in coincidence by rings of detectors surrounding the patient (Fig. 42–22) readily establishes the line along which the emission took place. If the difference in time of arrival of the two photons could be determined accurately, the actual position of the emitting nuclide along that line could be calculated. Present-day electronics can measure times to at best ± 300 ps, so at the γ ray's speed $(c = 3 \times 10^8 \text{ m/s})$, the actual position could be determined to an accuracy on the order of about $d = vt \approx (3 \times 10^8 \text{ m/s})(300 \times 10^{-12} \text{ s}) \approx 10 \text{ cm}$, which is not very useful. Although there may be future potential for *time-of-flight* measurements to determine position, today computed tomography techniques are used instead, similar to those for X-ray CT, which can reconstruct PET images with a resolution on the order of 2–5 mm. One big advantage of PET is that no collimators are needed (as for detection of a single photon—see Fig. 42–21a). Thus, fewer photons are "wasted" and lower doses can be administered to the patient with PET.

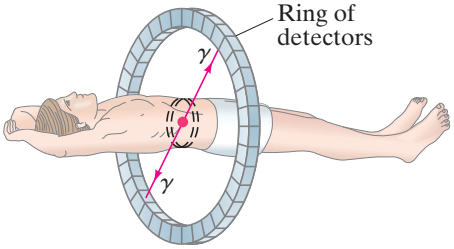
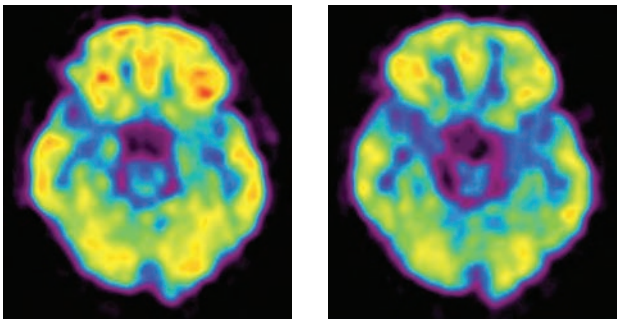


FIGURE 42–22 Positron emission tomography (PET) system showing a ring of detectors to detect the two annihilation γ rays $(e^+ + e^- \rightarrow 2\gamma)$ emitted at 180° to each other.

Both PET and SPECT systems can give images related to biochemistry, metabolism, and function. This is to be compared to X-ray CT scans, whose images reflect shape and structure—that is, the anatomy of the imaged region.

Figure 42–23 shows PET scans of the same person's brain (a) when using a cell phone near the ear and (b) with the cell phone off. The bright red spots in (a) indicate a higher rate of glucose metabolism, suggesting excitability of brain tissue (the glucose was tagged with a radioactive tracer). Emfs from the cell phone antenna thus seem to affect metabolism and may be harmful to us!

The colors shown here are faked (only visible light has colors). The original images are various shades of gray, representing intensity (or counts).



(a) (b)

FIGURE 42–23 False-color PET scans of a horizontal section through a brain showing glucose metabolism rates (red is high) by a person (a) using a cell phone near the ear, and (b) with the cell phone off.

*42–10 Nuclear Magnetic Resonance (NMR); Magnetic Resonance Imaging (MRI)

Nuclear magnetic resonance (NMR) is a phenomenon which soon after its discovery in 1946 became a powerful research tool in a variety of fields from physics to chemistry and biochemistry. It is also an important medical imaging technique. We first briefly discuss the phenomenon, and then look at its applications.

*Nuclear Magnetic Resonance (NMR)

We saw in Chapter 39 (Section 39–2) that when atoms are placed in a magnetic field B, atomic energy levels are split into several closely spaced levels (the Zeeman effect) according to the angular momentum or spin of the state. The splitting is proportional to B and to the magnetic moment, μ. Nuclei too have magnetic moments (Section 41–1), and we examine mainly the simplest nucleus, hydrogen $\left({}^{1}_{1}\mathrm{H}\right)$ because it is the one most used, even for medical imaging. The ${}^{1}_{1}\mathrm{H}$ nucleus consists of a single proton. Its spin angular momentum (and its magnetic moment), like that of the electron, can take on only two values when placed in a magnetic field: we call these "spin up" (parallel to the field) and "spin down" (antiparallel to the field) as suggested in Fig. 42–24. When a magnetic field is present, an energy state of a nucleus splits into two levels as shown in Fig. 42–25, with the spin-up state (parallel to field) having the lower energy. The spin-down state acquires an additional energy $\mu_p B_T$ and the spin-up state has its energy changed by $-\mu_p B_T$ (see Eq. 27–12, Example 27–12, and Section 39–7), where B_T is the total magnetic field *at the nucleus*. The difference in energy between the two states (Fig. 42–25) is thus

$$\Delta E = 2\mu_p B_T,$$

where μ_p is the magnetic moment of the proton. For materials other than hydrogen, the splitting is still proportional to B_T:

$$\Delta E = kB_T$$

where k is a constant that is different for different nuclides.

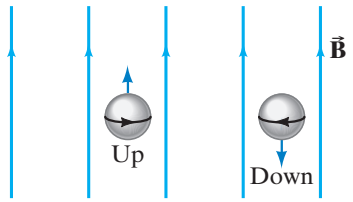
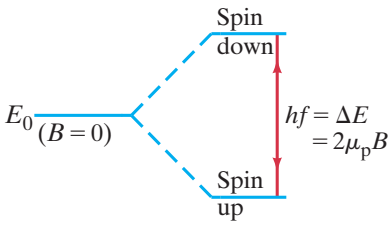


FIGURE 42–24 Schematic picture of a proton in a magnetic field $\vec{\mathbf{B}}$ (pointing upward) with the two possible states of proton spin, up and down.

FIGURE 42–25 Energy E_0 in the absence of a magnetic field splits into two levels in the presence of a magnetic field. The vertical red line with arrowheads represents transitions between the two states.



In a standard **nuclear magnetic resonance** (NMR) setup, the sample to be examined is placed in a static magnetic field. A radiofrequency (RF) pulse of electromagnetic radiation (that is, photons) is applied to the sample. If the frequency, f, of this pulse corresponds precisely to the energy difference between the two energy levels (Fig. 42–25), so that

$$hf = \Delta E = 2\mu_p B_T, \qquad \textbf{(42–12)}$$

then the photons of the RF beam will be absorbed, exciting many of the nuclei from the lower state to the upper state. This is a resonance phenomenon because there is significant absorption only if f is very near $f = 2\mu_p B_T/h$. Hence the name "nuclear magnetic resonance." For free ${}^{1}_{1}\mathrm{H}$ nuclei, the frequency is 42.58 MHz for a field $B_T = 1.0\,\mathrm{T}$ (Example 42–15). If the H atoms are bound in a molecule, the total magnetic field B_T at the H nuclei will be the sum of the external applied field (B_{ext}) plus the local magnetic field (B_{local}) due to electrons and nuclei of neighboring atoms. Since f is proportional to B_T, the value of f for a given external field will be slightly different for bound H atoms than for free atoms:

$$hf = 2\mu_p(B_{ext} + B_{local}).$$

This small change in frequency can be measured, and is called the "chemical shift." A great deal has been learned about the structure of molecules and bonds using this NMR technique.

EXAMPLE 42–15 **NMR for free protons.** Calculate the resonant frequency for free protons in a 1.000-T magnetic field.

APPROACH We use Eq. 42–12, where the magnetic moment of the proton (Section 41–1) is

$$\mu_p = 2.7928\mu_N = 2.7928\left(\frac{e\hbar}{2m_p}\right) = 2.7928\left(\frac{eh}{4\pi m_p}\right).$$

SOLUTION We solve for f in Eq. 42–12 and find

$$f = \frac{\Delta E}{h} = \frac{2\mu_p B}{h}$$

$$= (2.7928)\left(\frac{eB}{2\pi m_p}\right) = 2.7928\left[\frac{(1.6022\times10^{-19}\,\mathrm{C})(1.000\,\mathrm{T})}{2\pi(1.6726\times10^{-27}\,\mathrm{kg})}\right]$$

$$= 42.58\,\mathrm{MHz}.$$

*Magnetic Resonance Imaging (MRI)

PHYSICS APPLIED
NMR imaging (MRI)

For producing medically useful NMR images—now commonly called MRI, or **magnetic resonance imaging**—the element most used is hydrogen since it is the commonest element in the human body and gives the strongest NMR signals. The experimental apparatus is shown in Fig. 42–26. The large coils set up the static magnetic field, and the RF coils produce the RF pulse of electromagnetic waves (photons) that cause the nuclei to jump from the lower state to the upper one (Fig. 42–24). These same coils (or another coil) can detect the absorption of energy or the emitted radiation (also of frequency $f = \Delta E/h$) when the nuclei jump back down to the lower state.



FIGURE 42–26 Typical MRI imaging setup: (a) diagram; (b) photograph; (c) xxxx.

The formation of a two-dimensional or three-dimensional image can be done using techniques similar to those for computed tomography (Section 35–11). The simplest thing to measure for creating an image is the intensity of absorbed and/or reemitted radiation from many different points of the body, and this would be a measure of the density of H atoms at each point. But how do we determine from what part of the body a given photon comes? One technique is to give the static magnetic field a gradient; that is, instead of applying a uniform magnetic field, B_T, the field is made to vary with position across the width of the sample (or patient), Fig. 42–27. Because the frequency absorbed by the H nuclei is proportional to B_T (Eq. 42–12), only one plane within the body will have the proper value of B_T to absorb photons of a particular frequency f. By varying f, absorption by different planes can be measured. Alternately, if the field gradient is applied *after* the RF pulse, the frequency of the emitted photons will be a measure of where they were emitted. If a magnetic field gradient in one direction is applied during excitation (absorption of photons) and photons of a single frequency are transmitted, only H nuclei in one thin slice will be excited. By applying a gradient during reemission in a direction perpendicular to the first, the frequency f of the reemitted radiation will represent depth in that slice. Other ways of varying the magnetic field throughout the volume of the body can be used in order to correlate NMR frequency with position.

A reconstructed image based on the density of H atoms (that is, the intensity of absorbed or emitted radiation) is not very interesting. More useful are images based on the rate at which the nuclei decay back to the ground state, and such images can produce resolution of 1 mm or better. This NMR technique (sometimes called **spin-echo**) produces images of great diagnostic value, both in the delineation of structure (anatomy) and in the study of metabolic processes. An NMR image is shown in Fig. 42–28, color enhanced—no medical imaging uses visible light, so the colors shown are added. The original images, those looked at by your doctor, are various shades of gray, representing intensity (or counts).

FIGURE 42–27 A static field that is stronger at the bottom than at the top. The frequency of absorbed or emitted radiation is proportional to B in NMR.

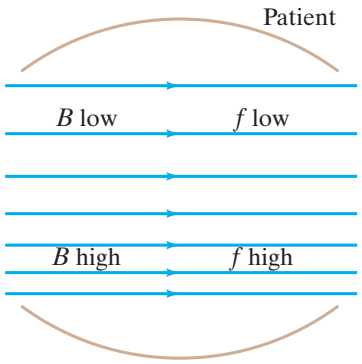
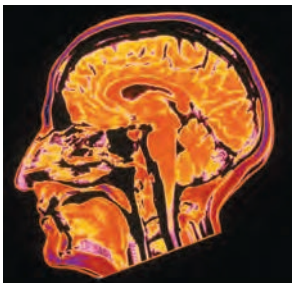


FIGURE 42–28 False-color NMR image (MRI) through the head showing structures in the brain.



NMR imaging is considered to be noninvasive. We can calculate the energy of the photons involved: as shown in Example 42–15, in a 1.0-T magnetic field, $f = 42.58\,\text{MHz}$ for ^1_1H. This corresponds to an energy of $hf = (6.6 \times 10^{-34}\,\text{J}\cdot\text{s})(43 \times 10^6\,\text{Hz}) \approx 3 \times 10^{-26}\,\text{J}$ or about $10^{-7}\,\text{eV}$. Since molecular bonds are on the order of 1 eV, the RF photons can cause little cellular disruption. This should be compared to X- or γ rays, whose energies are 10^4 to $10^6\,\text{eV}$ and thus can cause significant damage. The static magnetic fields, though often large (typically about 2.0 T), are believed to be harmless (except for people who wear heart pacemakers).

Table 42–2 lists the major techniques we have discussed for imaging the interior of the human body, along with the optimum resolution attainable today. Resolution is only one factor that must be considered, because the different imaging techniques provide different types of information that are useful for different types of diagnosis.

TABLE 42–2 Medical Imaging Techniques

Technique	Where Discussed in This Book	Optimal Resolution
Conventional X-ray	Section 35–11	$\frac{1}{2}$ mm
CT scan, X-ray	Section 35–11	0.2 mm
Nuclear medicine (tracers)	Section 42–8	1 cm
SPECT (single photon emission)	Section 42–9	1 cm
PET (positron emission)	Section 42–9	2–5 mm
MRI (NMR)	Section 42–10	$\frac{1}{2}$–1 mm
Ultrasound	Section 16–9	0.3–2 mm

Summary

A **nuclear reaction** occurs when two nuclei collide and two or more other nuclei (or particles) are produced. In this process, as in radioactivity, **transmutation** (change) of elements occurs.

The **reaction energy** or ***Q*-value** of a reaction $a + X \rightarrow Y + b$ is

$$Q = (M_a + M_X - M_b - M_Y)c^2 \qquad \textbf{(42–2a)}$$

$$= K_b + K_Y - K_a - K_X. \qquad \textbf{(42–2b)}$$

The effective **cross section** σ for a reaction is a measure of the reaction probability per target nucleus.

In **fission**, a heavy nucleus such as uranium splits into two intermediate-sized nuclei after being struck by a neutron. $^{235}_{92}\text{U}$ is fissionable by slow neutrons, whereas some fissionable nuclei require fast neutrons. Much energy is released in fission ($\approx 200\,\text{MeV}$ per fission) because the binding energy per nucleon is lower for heavy nuclei than it is for intermediate-sized nuclei, so the mass of a heavy nucleus is greater than the total mass of its fission products. The fission process releases neutrons, so that a **chain reaction** is possible. The **critical mass** is the minimum mass of fuel needed so that enough emitted neutrons go on to produce more fissions and sustain a chain reaction. In a **nuclear reactor** or nuclear weapon, a **moderator** is used to slow down the released neutrons.

The **fusion** process, in which small nuclei combine to form larger ones, also releases energy. The energy from our Sun originates in the fusion reactions known as the **proton–proton chain** in which four protons fuse to form a ^4_2He nucleus producing 25 MeV of energy. A useful fusion reactor for power generation has not yet proved possible because of the difficulty in containing the fuel (e.g., deuterium) long enough at the extremely high temperature required ($\approx 10^8\,\text{K}$). Nonetheless, progress has been made in confining the collection of charged ions known as a **plasma**. The two main methods are **magnetic confinement**, using a magnetic field in a device such as the donut-shaped **tokamak**, and **inertial confinement** in which intense laser beams compress a fuel pellet of deuterium and tritium.

Radiation can cause damage to materials, including biological tissue. Quantifying amounts of radiation is the subject of **dosimetry**. The **curie** (Ci) and the **becquerel** (Bq) are units that measure the **source activity** or rate of decay of a sample: $1\,\text{Ci} = 3.70 \times 10^{10}$ decays per second, whereas 1 Bq = 1 decay/s. The **absorbed dose**, often specified in **rads**, measures the amount of energy deposited per unit mass of absorbing material: 1 rad is the amount of radiation that deposits energy at the rate of $10^{-2}\,\text{J/kg}$ of material. The SI unit of absorbed dose is the **gray**: 1 Gy = 1 J/kg = 100 rad. The **effective dose** is often specified by the **rem** = rad × RBE, where RBE is the "relative biological effectiveness" of a given type of radiation; 1 rem of any type of radiation does approximately the same amount of biological damage. The average dose received per person per year in the United States is about 360 mrem. The SI unit for effective dose is the **sievert**: 1 Sv = 100 rem.

[*Nuclear radiation is used in medicine for cancer therapy, and for imaging of biological structure and processes. Tomographic imaging of the human body, which can provide 3-dimensional detail, includes several types: PET, SPET (= SPECT), MRI, and CT scans (discussed in Chapter 35). MRI makes use of **nuclear magnetic resonance** (NMR).]

PSE5_CH42_v1.6_23Jun-20

Questions

1. Fill in the missing particles or nuclei:
(a) $n + {}^{232}_{90}Th \rightarrow ? + \gamma$;
(b) $n + {}^{137}_{56}Ba \rightarrow {}^{137}_{55}Cs + ?$;
(c) $d + {}^{2}_{1}H \rightarrow {}^{4}_{2}He + ?$;
(d) $d + {}^{2}_{1}H \rightarrow {}^{4}_{2}He + ?$;
where d stands for deuterium.

2. When ${}^{22}_{11}Na$ is bombarded by deuterons $\left({}^{2}_{1}H\right)$, an α particle is emitted. What is the resulting nuclide? Write down the reaction equation.

3. The isotope ${}^{32}_{15}P$ is produced by the reaction: $n + ? \rightarrow {}^{32}_{15}P + p$. What must be the target nucleus?

4. What is the Q-value for radioactive decay reactions?
(a) $Q < 0$. (b) $Q > 0$. (c) $Q = 0$.
(d) The sign of Q depends on the nucleus.

5. Why are neutrons such good projectiles for producing nuclear reactions?

6. The energy from nuclear fission appears in the form of thermal energy—but the thermal energy of what?

7. (a) If ${}^{235}_{92}U$ released only 1.5 neutrons per fission on average (instead of 2.5), would a chain reaction be possible? (b) If so, how would the chain reaction be different than if 3 neutrons were released per fission?

8. Why can't uranium be enriched by chemical means?

9. How can a neutron, with practically no kinetic energy, excite a nucleus to the extent shown in Fig. 42–4?

10. A reactor that uses highly enriched uranium can use ordinary water (instead of heavy water) as a moderator and still have a self-sustaining chain reaction. Explain.

11. Why would a porous block of uranium be more likely to explode if kept under water rather than in air?

12. Why are neutrons released in a fission reaction?

13. What is the reason for the "secondary system" in a nuclear reactor, Fig. 42–9? That is, why is the water heated by the fuel in a nuclear reactor not used directly to drive the turbines?

14. Why must the fission process release neutrons if it is to be useful?

15. ${}^{238}_{92}U$ releases an average of 2.5 neutrons per fission compared to 2.9 for ${}^{239}_{94}Pu$. Which of these two nuclei do you think would have the smaller critical mass? Explain.

16. What is the basic difference between fission and fusion?

17. Why do gamma particles penetrate matter more easily than beta particles do?

18. Light energy emitted by the Sun and stars comes from the fusion process. What conditions in the interior of stars make this possible?

19. A higher temperature is required for deuterium–deuterium ignition than for deuterium–tritium. Explain.

20. How do stars, and our Sun, maintain confinement of the plasma for fusion?

21. People who work around metals that emit alpha particles are trained that there is little danger from proximity or touching the material, but they must take extreme precautions against ingesting it. Why? (Eating and drinking while working are forbidden.)

***22.** How might radioactive tracers be used to find a leak in a pipe?

23. Radiation is sometimes used to sterilize medical supplies and even food. Explain how it works.

24. What is the difference between absorbed dose and effective dose? What are the SI units for each?

25. What keeps the Sun and stars together, allowing them to sustain fusion?

26. State three problems that must be overcome to make a functioning fission nuclear reactor.

27. Why can small nuclei combine to form larger ones, releasing energy in the process?

28. (a) State three environmental problems or dangers that do or could result from the operation of a nuclear fission reactor; (b) describe an additional problem or danger associated with a breeder reactor.

29. Why does the first reaction in the proton–proton chain limit the rate at which the Sun produces energy?

30. What are the heaviest elements for which energy is released if the elements are created by fusion of lighter elements?

***31.** Explain how each of the following can cause damage to materials: beta particles, alpha particles, energetic neutrons, and gamma rays.

MisConceptual Questions

1. In a nuclear reaction, which of the following is *not* conserved?
(a) Energy.
(b) Momentum.
(c) Electric charge.
(d) Nucleon number.
(e) None of the above.

2. Fission fragments are typically
(a) β^+ emitters.
(b) β^- emitters.
(c) Both.
(d) Neither.

3. Which of the following properties would decrease the critical mass needed to sustain a nuclear chain reaction?
(a) Low boiling point.
(b) High melting point.
(c) More neutrons released per fission.
(d) Low nuclear density.
(e) Filled valence shell.
(f) All of the above.

4. Rather than having a maximum at about $A \approx 60$, as shown in Fig. 42–13, suppose the average binding energy per nucleon continually increased with increasing mass number. Then,
(a) fission would still be possible, but not fusion.
(b) fusion would still be possible, but not fission.
(c) both fission and fusion would still be possible.
(d) neither fission nor fusion would be possible.

5. Why is a moderator needed in a normal uranium fission reactor?
(a) To increase the rate of neutron capture by uranium-235.
(b) To increase the rate of neutron capture by uranium-238.
(c) To increase the rate of production of plutonium-239.
(d) To increase the critical mass of the fission fuel.
(e) To provide more neutrons for the reaction.
(f) All of the above.

6. For neutrons to induce nuclear fusion
(*a*) neutrons must be moving fast to provide energy to the reaction.
(*b*) neutrons must be moving slow to be captured by the nucleus.
(*c*) the speed of neutrons is irrelevant.

7. A primary difficulty in energy production by fusion is
(*a*) the scarcity of necessary fuel.
(*b*) the disposal of radioactive by-products produced.
(*c*) the high temperatures necessary to overcome the electrical repulsion of protons.
(*d*) the fact that it is possible in volcanic regions only.

8. If two hydrogen nuclei, ${}^{2}_{1}\text{H}$, each of mass m_{H}, fuse together and form a helium nucleus of mass m_{He},
(*a*) $m_{\text{He}} < 2m_{\text{H}}$.
(*b*) $m_{\text{He}} = 2m_{\text{H}}$.
(*c*) $m_{\text{He}} > 2m_{\text{H}}$.
(*d*) All of the above are possible.

9. Which radiation induces the most biological damage for a given amount of energy deposited in tissue?
(*a*) Alpha particles.
(*b*) Gamma radiation.
(*c*) Beta radiation.
(*d*) All do the same damage for the same deposited energy.

10. Which would produce the most energy in a single reaction?
(*a*) The fission reaction associated with uranium-235.
(*b*) The fusion reaction of the Sun (two hydrogen nuclei fused to one helium nucleus).
(*c*) Both (*a*) and (*b*) are about the same.
(*d*) Need more information.

11. The fuel necessary for fusion-produced energy could be derived from
(*a*) water.
(*b*) superconductors.
(*c*) uranium.
(*d*) helium.
(*e*) sunlight.

12. Which of the following is true?
(*a*) Any amount of radiation is harmful to living tissue.
(*b*) Radiation is a natural part of the environment.
(*c*) Radiation can be used for imaging.
(*d*) All of the above are true.

13. Which of the following would reduce the cell damage due to radiation for a lab technician who works with radioactive isotopes in a hospital or lab?
(*a*) Increase the worker's distance from the radiation source.
(*b*) Decrease the time the worker is exposed to the radiation.
(*c*) Use shielding to reduce the amount of radiation that strikes the worker.
(*d*) Have the worker wear a radiation badge when working with the radioactive isotopes.
(*e*) All of the above except (*d*).
(*f*) All of the above.

14. If the same dose in rads of each type of radiation was provided over the same amount of time, which type would be most harmful?
(*a*) X-rays.
(*b*) γ rays.
(*c*) β rays.
(*d*) α particles.

15. ${}^{235}_{92}\text{U}$ releases an average of 2.5 neutrons per fission compared to 2.9 for ${}^{239}_{94}\text{Pu}$. Which has the smaller critical mass?
(*a*) ${}^{235}_{92}\text{U}$.
(*b*) ${}^{239}_{94}\text{Pu}$.
(*c*) Both the same.

Problems

(NOTE: Masses are found in Appendix G.)

42–1 Nuclear Reactions, Transmutation

1. (I) Natural aluminum is all ${}^{27}_{13}\text{Al}$. If it absorbs a neutron, what does it become? Does it decay by β^+ or β^-? What will be the product nucleus?

2. (I) Determine whether the reaction ${}^{2}_{1}\text{H} + {}^{2}_{1}\text{H} \rightarrow {}^{3}_{2}\text{He} + \text{n}$ requires a threshold energy, and why.

3. (I) Is the reaction $\text{n} + {}^{238}_{92}\text{U} \rightarrow {}^{239}_{92}\text{U} + \gamma$ possible with slow neutrons? Explain.

4. (II) (*a*) Complete the following nuclear reaction, $\text{p} + ? \rightarrow {}^{32}_{16}\text{S} + \gamma$. (*b*) What is the *Q*-value?

5. (II) The reaction $\text{p} + {}^{18}_{8}\text{O} \rightarrow {}^{18}_{9}\text{F} + \text{n}$ requires an input of energy equal to 2.438 MeV. What is the mass of ${}^{18}_{9}\text{F}$?

6. (II) (*a*) Can the reaction $\text{n} + {}^{24}_{12}\text{Mg} \rightarrow {}^{23}_{11}\text{Na} + \text{d}$ occur if the bombarding particles have 19.00 MeV of kinetic energy? (d stands for deuterium, ${}^{2}_{1}\text{H}$.) (*b*) If so, how much energy is released? If not, what kinetic energy is needed?

7. (II) (*a*) Can the reaction $\text{p} + {}^{7}_{3}\text{Li} \rightarrow {}^{4}_{2}\text{He} + \alpha$ occur if the incident proton has kinetic energy = 2800 keV? (*b*) If so, what is the total kinetic energy of the products? If not, what kinetic energy is needed?

8. (II) In the reaction $\alpha + {}^{14}_{7}\text{N} \rightarrow {}^{17}_{8}\text{O} + \text{p}$, the incident α particles have 8.95 MeV of kinetic energy. The mass of ${}^{17}_{8}\text{O}$ is 16.999132 u. (*a*) Can this reaction occur? (*b*) If so, what is the total kinetic energy of the products? If not, what kinetic energy is needed?

9. (II) Calculate the *Q*-value for the "capture" reaction $\alpha + {}^{16}_{8}\text{O} \rightarrow {}^{20}_{10}\text{Ne} + \gamma$.

10. (II) Calculate the total kinetic energy of the products of the reaction $\text{d} + {}^{13}_{6}\text{C} \rightarrow {}^{14}_{7}\text{N} + \text{n}$ if the incoming deuteron has kinetic energy $K = 45.3$ MeV.

11. (II) Radioactive ${}^{14}_{6}\text{C}$ is produced in the atmosphere when a neutron is absorbed by ${}^{14}_{7}\text{N}$. Write the reaction and find its *Q*-value.

12. (II) An example of a **stripping** nuclear reaction is $\text{d} + {}^{6}_{3}\text{Li} \rightarrow \text{X} + \text{p}$. (*a*) What is X, the resulting nucleus? (*b*) Why is it called a "stripping" reaction? (*c*) What is the *Q*-value of this reaction? Is the reaction endothermic or exothermic?

13. (II) An example of a **pick-up** nuclear reaction is ${}^{3}_{2}\text{He} + {}^{12}_{6}\text{C} \rightarrow \text{X} + \alpha$. (*a*) Why is it called a "pick-up" reaction? (*b*) What is the resulting nucleus? (*c*) What is the *Q*-value of this reaction? Is the reaction endothermic or exothermic?

14. (II) Calculate the energy released (or energy input required) for the reaction $\alpha + {}^{9}_{4}\text{Be} \rightarrow {}^{12}_{6}\text{C} + \text{n}$.

15. (III) Use conservation of energy and momentum to show that a bombarding proton must have an energy of 3.23 MeV in order to make the reaction ${}^{13}_{6}\text{C}(\text{p}, \text{n}){}^{13}_{7}\text{N}$ occur. (See Example 42–3.)

16. (III) How much kinetic energy (if any) would the proton require for the reaction ${}^{14}_{6}\text{C}(\text{p}, \text{n}){}^{14}_{7}\text{N}$ to proceed?

42–2 Cross Section

17. (I) The cross section for the reaction $\text{n} + {}^{10}_{5}\text{B} \rightarrow {}^{7}_{3}\text{Li} + {}^{4}_{2}\text{He}$ is about 40 bn for an incident neutron of low energy (kinetic energy ≈ 0). The boron is contained in a gas with $n = 1.3 \times 10^{21}$ nuclei/m^3 and the target has thickness $\ell = 12.0$ cm. What fraction of incident neutrons will be scattered?

18. (I) What is the effective cross section for the collision of two hard spheres of radius R_1 and R_2?

19. (II) When the target is thick, the rate at which projectile particles collide with nuclei in the rear of the target is less than in the front of the target, since some scattering (i.e., collisions) takes place in the front layers. Let R_0 be the rate at which incident particles strike the front of the target, and R_x be the rate at a distance x into the target ($R_x = R_0$ at $x = 0$). Then show that the rate at which particles are scattered (and therefore lost from the incident beam) in a thickness dx is $-dR_x = R_x n\sigma\, dx$, where the minus sign means that R_x is decreasing and n is the number of nuclei per unit volume. Then show that $R_x = R_0 e^{-n\sigma x}$, where σ is the total cross section. If the thickness of the target is ℓ, what does $R_x = R_0 e^{-n\sigma\ell}$ represent?

20. (II) A 1.3-cm-thick lead target reduces a beam of gamma rays to 25% of its original intensity. What thickness of lead will allow only one γ in 10^6 to penetrate (see Problem 19)?

21. (II) Use Fig. 42–3 to estimate what thickness of ${}^{114}_{48}\text{Cd}$ ($\rho = 8650$ kg/m^3) will cause a 3.0% reaction rate ($R/R_0 = 0.030$) for (*a*) 0.10-eV neutrons (*b*) 5.0-eV neutrons.

42–3 Nuclear Fission

22. (I) What is the energy released in the fission reaction of Eq. 42–5? (The masses of ${}^{141}_{56}\text{Ba}$ and ${}^{92}_{36}\text{Kr}$ are 140.914411 u and 91.926156 u, respectively.)

23. (I) Calculate the energy released in the fission reaction $\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{88}_{38}\text{Sr} + {}^{136}_{54}\text{Xe} + 12\text{n}$. Use Appendix G, and assume the initial kinetic energy of the neutron is very small.

24. (I) How many fissions take place per second in a 1000-MW reactor? Assume 200 MeV is released per fission.

25. (I) The energy produced by a fission reactor is about 200 MeV per fission. What fraction of the mass of a ${}^{235}_{92}\text{U}$ nucleus is this?

26. (II) Suppose that the average electric power consumption, day and night, in a typical house is 1200 W. What initial mass of ${}^{235}_{92}\text{U}$ would have to undergo fission to supply the electrical needs of such a house for a year? (Assume 200 MeV is released per fission, as well as 100% efficiency.)

27. (II) Consider the fission reaction

$$ {}^{235}_{92}\text{U} + \text{n} \rightarrow {}^{133}_{51}\text{Sb} + {}^{98}_{41}\text{Nb} + ?\text{n}. $$

(*a*) How many neutrons are produced in this reaction? (*b*) Calculate the energy release. The atomic masses for Sb and Nb isotopes are 132.915250 u and 97.910328 u, respectively.

28. (II) How much mass of ${}^{235}_{92}\text{U}$ is required to produce the same amount of energy as burning 1.0 kg of coal (about 3×10^7 J)?

29. (II) What initial mass of ${}^{235}_{92}\text{U}$ is required to operate an 880-MW reactor for 1 yr? Assume 32% efficiency.

30. (II) If a 1.0-MeV neutron emitted in a fission reaction loses one-half of its kinetic energy in each collision with moderator nuclei, how many collisions must it make to reach thermal energy ($\frac{3}{2}kT = 0.040$ eV)?

31. (II) Assuming a fission of ${}^{236}_{92}\text{U}$ into two roughly equal fragments, estimate the electric potential energy just as the fragments separate from each other. Assume that the fragments are spherical (see Eq. 41–1) and compare your calculation to the nuclear fission energy released, about 200 MeV.

32. (II) Estimate the ratio of the height of the Coulomb barrier for α decay to that for fission of ${}^{236}_{92}\text{U}$. (Both are described by a potential energy diagram of the shape shown in Fig. 41–7.)

33. (III) Suppose that the neutron multiplication factor is 1.0004. If the average time between successive fissions in a chain of reactions is 1.0 ms, by what factor will the reaction rate increase in 1.0 s?

42–4 Nuclear Fusion

34. (I) What is the average kinetic energy of protons at the center of a star where the temperature is 2×10^7 K? [*Hint*: See Eq. 18–4.]

35. (II) Show that the energy released in the fusion reaction ${}^{2}_{1}\text{H} + {}^{3}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + \text{n}$ is 17.57 MeV.

36. (II) Show that the energy released when two deuterium nuclei fuse to form ${}^{3}_{2}\text{He}$ with the release of a neutron is 3.23 MeV.

37. (II) Verify the Q-value stated for each of the reactions of Eqs. 42–7. [*Hint*: Be careful with electrons (included in mass values for p, d, t in Appendix G).]

38. (II) (*a*) Calculate the energy release per gram of fuel for the reactions of Eqs. 42–9a, b, and c. (*b*) Calculate the energy release per gram of uranium ${}^{235}_{92}\text{U}$ in fission, and give its ratio to each reaction in (*a*).

39. (II) How much energy is released when ${}^{238}_{92}\text{U}$ absorbs a slow neutron (kinetic energy ≈ 0) and becomes ${}^{239}_{92}\text{U}$?

40. (II) If a typical house requires 1200 W of electric power on average, what minimum amount of deuterium fuel would have to be used in a year to supply these electrical needs? Assume the reaction of Eq. 42–9b.

41. (II) If ${}^{6}_{3}\text{Li}$ is struck by a slow neutron, it can form ${}^{4}_{2}\text{He}$ and another isotope. (*a*) What is the second isotope? (This is a method of generating this isotope.) (*b*) How much energy is released in the process?

42. (II) Suppose a fusion reactor ran on "d–d" reactions, Eqs. 42–9a and b in equal amounts. Estimate how much natural water, for fuel, would be needed per hour to run a 1150-MW reactor, assuming 33% efficiency.

43. (II) How much energy (J) is contained in 1.00 kg of water if its natural deuterium is used in the fusion reaction of Eq. 42–9a? Compare to the energy obtained from the burning of 1.0 kg of gasoline, about 5×10^7 J.

44. (II) Determine the energy output for each of the 6 reactions and decays of the carbon cycle.

45. (III) Show that the energies carried off by the $^4_2$He nucleus and the neutron for the reaction of Eq. 42–9c are about 3.5 MeV and 14 MeV, respectively. Are these fixed values, independent of the plasma temperature?

46. (III) (*a*) Give the ratio of the energy needed for the first reaction of the *carbon cycle* to the energy needed for a deuterium–tritium reaction (Example 42–10). (*b*) If a deuterium–tritium reaction actually requires a temperature $T \approx 3 \times 10^8$ K, estimate the temperature needed for the first carbon-cycle reaction.

47. (III) The energy output of massive stars is believed to be due to the *carbon cycle* (see text). (*a*) Show that no carbon is consumed in this cycle and that the net effect is the same as for the proton–proton cycle. (*b*) What is the total energy release? (*c*) Why might the carbon cycle require a higher temperature $(\approx 2 \times 10^7\,\text{K})$ than the proton–proton chain $(\approx 1.5 \times 10^7\,\text{K})$?

42–6 Dosimetry

48. (I) 210 rads of α-particle radiation is equivalent to how many rads of X-rays in terms of biological damage?

49. (I) A dose of 4.0 Sv of γ rays in a short period would be lethal to about half the people subjected to it. How many grays is this?

50. (I) How much energy is deposited in the body of a 65-kg adult exposed to a 2.5-Gy dose?

51. (I) How many rads of slow neutrons will do as much biological damage as 75 rads of fast neutrons?

52. (II) A cancer patient is undergoing radiation therapy in which protons with an energy of 1.2 MeV are incident on a 0.20-kg tumor. (*a*) If the patient receives an effective dose of 1.0 rem, what is the absorbed dose? (*b*) How many protons are absorbed by the tumor? Assume RBE ≈ 1.

53. (II) A 0.042-μCi sample of $^{32}_{15}$P is injected into an animal for tracer studies. If a Geiger counter intercepts 25% of the emitted β particles, what will be the counting rate, assumed 85% efficient?

54. (II) About 35 eV is required to produce one ion pair in air. Show that this is consistent with the two definitions of the roentgen given in the text.

55. (II) A 1.4-mCi source of $^{32}_{15}$P (in $NaHPO_4$), a β emitter, is implanted in a tumor where it is to administer 36 Gy. The half-life of $^{32}_{15}$P is 14.3 days, and 1.0 mCi delivers about 10 mGy/min. Approximately how long should the source remain implanted?

56. (II) What is the mass of a 3.00-μCi $^{14}_6$C source?

57. (II) $^{57}_{27}$Co emits 122-keV γ rays. If a 63-kg person swallowed 1.55 μCi of $^{57}_{27}$Co, what would be the dose rate (Gy/day) averaged over the whole body? Assume that 50% of the γ-ray energy is deposited in the body. [*Hint*: Determine the rate of energy deposited in the body and use the definition of the gray.]

58. (II) Ionizing radiation can be used on meat products to reduce the levels of microbial pathogens. Refrigerated meat is limited to 4.5 kGy. If 1.2-MeV electrons irradiate 4 kg of beef, how many electrons would it take to reach the allowable limit?

59. (III) Huge amounts of radioactive $^{131}_{53}$I were released in the accident at Chernobyl in 1986. Chemically, iodine goes to the human thyroid. (It can be used for diagnosis and treatment of thyroid problems.) In a normal thyroid, $^{131}_{53}$I absorption can cause damage to the thyroid. (*a*) Write down the reaction for the decay of $^{131}_{53}$I. (*b*) Its half-life is 8.0 d; how long would it take for ingested $^{131}_{53}$I to become 5.0% of the initial value? (*c*) Absorbing 1 mCi of $^{131}_{53}$I can be harmful; what mass of iodine is this?

60. (III) Assume a liter of milk typically has an activity of 2000 pCi due to $^{40}_{19}$K. If a person drinks two glasses (0.5 L) per day, estimate the total effective dose (in Sv and in rem) received in a year. As a crude model, assume the milk stays in the stomach 12 hr and is then released. Assume also that very roughly 10% of the 1.5 MeV released per decay is absorbed by the body. Compare your result to the normal allowed dose of 100 mrem per year. Make your estimate for (*a*) a 60-kg adult, and (*b*) a 6-kg baby.

61. (II) Radon gas, $^{222}_{86}$Rn, is considered a serious health hazard (see discussion in text). It decays by α-emission. (*a*) What is the daughter nucleus? (*b*) Is the daughter nucleus stable or radioactive? If the latter, how does it decay, and what is its half-life? (See Fig. 41–12.) (*c*) Is the daughter nucleus also a noble gas, or is it chemically reactive? (*d*) Suppose 1.4 ng of $^{222}_{86}$Rn seeps into a basement. What will be its activity? If the basement is then sealed, what will be the activity 1 month later?

*42–9 NMR

***62.** (II) Calculate the wavelength of photons needed to produce NMR transitions in free protons in a 1.500-T field. In what region of the spectrum is this wavelength?

***63.** (II) Carbon-13 has a magnetic moment $\mu = 0.7023\mu_N$ (see Eq. 41–2). What magnetic field would be necessary if $^{13}_6$C were to be detected in a proton NMR spectrometer operating at 42.58 MHz? (This large field necessitates that a $^{13}_6$C spectrometer operate at a lower frequency.)

General Problems

64. J. Chadwick discovered the neutron by bombarding $^{9}_{4}\text{Be}$ with the popular projectile of the day, alpha particles. (*a*) If one of the reaction products was the then unknown neutron, what was the other product? (*b*) What is the *Q*-value of this reaction?

65. Fusion temperatures are often given in keV. Determine the conversion factor from kelvins to keV using, as is common in this field, $\overline{K} = kT$ without the factor $\frac{3}{2}$.

66. One means of enriching uranium is by diffusion of the gas UF_6. Calculate the ratio of the speeds of molecules of this gas containing $^{235}_{92}\text{U}$ and $^{238}_{92}\text{U}$, on which this process depends.

67. (*a*) What mass of $^{235}_{92}\text{U}$ was actually fissioned in the first atomic bomb, whose energy was the equivalent of about 20 kilotons of TNT (1 kiloton of TNT releases 5×10^{12} J)? (*b*) What was the actual mass transformed to energy?

68. The average yearly background radiation in a certain town consists of 26 mrad of X-rays and γ rays plus 3.6 mrad of particles having a RBE of 10. How many rem will a person receive per year on the average?

69. Deuterium makes up 0.0115% of natural hydrogen on average. Make a rough estimate of the total deuterium in the Earth's oceans and estimate the total energy released if all of it were used in fusion reactors.

70. A shielded γ-ray source yields a dose rate of 0.058 rad/h at a distance of 1.0 m for an average-sized person. If workers are allowed a maximum dose of 5.0 rem in 1 year, how close to the source may they operate, assuming a 35-h work week? Assume that the intensity of radiation falls off as the square of the distance. (It actually falls off more rapidly than $1/r^2$ because of absorption in the air, so your answer will give a better-than-permissible value.)

71. Radon gas, $^{222}_{86}\text{Rn}$, is formed by α decay. (*a*) Write the decay equation. (*b*) Ignoring the kinetic energy of the daughter nucleus (it's so massive), estimate the kinetic energy of the α particle produced. (*c*) Estimate the momentum of the alpha and of the daughter nucleus. (*d*) Estimate the kinetic energy of the daughter, and show that your approximation in (*b*) was valid.

72. Consider a system of nuclear power plants that produce 2200 MW. (*a*) What total mass of $^{235}_{92}\text{U}$ fuel would be required to operate these plants for 1 yr, assuming that 200 MeV is released per fission? (*b*) Typically 6% of the $^{235}_{92}\text{U}$ nuclei that fission produce $^{90}_{38}\text{Sr}$, a β^- emitter with a half-life of 29 yr. What is the total radioactivity of the $^{90}_{38}\text{Sr}$, in curies, produced in 1 yr? (Neglect the fact that some of it decays during the 1-yr period.)

73. In the net reaction, Eq. 42–8, for the proton–proton chain in the Sun, the neutrinos escape from the Sun with energy of about 0.5 MeV. The remaining energy, 26.2 MeV, is available to heat the Sun. Use this value to calculate the "heat of combustion" per kilogram of hydrogen fuel and compare it to the heat of combustion of coal, about 3×10^7 J/kg.

74. Estimate how much total energy would be released via fission if 1.0 kg of uranium were enriched to 5% of the isotope $^{235}_{92}\text{U}$.

75. Some stars, in a later stage of evolution, may begin to fuse two $^{12}_{6}\text{C}$ nuclei into one $^{24}_{12}\text{Mg}$ nucleus. (*a*) How much energy would be released in such a reaction? (*b*) What kinetic energy must two carbon nuclei each have when far apart, if they can then approach each other to within 6.0 fm, center-to-center? (*c*) Approximately what temperature would this require?

76. Energy reaches Earth from the Sun at a rate of about 1300 W/m^2. Calculate (*a*) the total power output of the Sun, and (*b*) the number of protons consumed per second in the reaction of Eq. 42–8, assuming that this is the source of all the Sun's energy. (*c*) Assuming that the Sun's mass of 2.0×10^{30} kg was originally all protons and that all could be involved in nuclear reactions in the Sun's core, how long would you expect the Sun to "glow" at its present rate? See Problem 73. [*Hint*: Use $1/r^2$ law in (*a*).]

77. Estimate how many solar neutrinos pass through a 180-m^2 ceiling of a room, at latitude 45°, for an hour around midnight on midsummer night. [*Hint*: See Problems 73 and 76.]

78. An average adult body contains about 0.10 μCi of $^{40}_{19}\text{K}$, which comes from food. (*a*) How many decays occur per second? (*b*) The potassium decay produces beta particles with energies of around 1.4 MeV. Estimate the dose per year in sieverts for a 62-kg adult. Is this a significant fraction of the 3.6-mSv/yr background rate?

79. When the nuclear reactor accident occurred at Chernobyl in 1986, 2.0×10^7 Ci were released into the atmosphere. Assuming that this radiation was distributed uniformly over the surface of the Earth, what was the activity per square meter? (The actual activity was not uniform; even within Europe wet areas received more radioactivity from rainfall.)

80. A star with a large helium abundance can burn helium in the reaction $^{4}_{2}\text{He} + {}^{4}_{2}\text{He} + {}^{4}_{2}\text{He} \rightarrow {}^{12}_{6}\text{C}$. What is the *Q*-value for this reaction?

81. A 1.2-μCi $^{137}_{55}\text{Cs}$ source is used for 1.6 hours by a 58-kg worker. Radioactive $^{137}_{55}\text{Cs}$ decays by β^- decay with a half-life of 30 yr. The average energy of the emitted betas is about 190 keV per decay. The β decay is quickly followed by a γ with an energy of 660 keV. Assuming the person absorbs *all* emitted energy, what effective dose (in rem) is received?

82. A large amount of $^{90}_{38}\text{Sr}$ was released during the Chernobyl nuclear reactor accident in 1986. The $^{90}_{38}\text{Sr}$ enters the body through the food chain. How long will it take for 95% of the $^{90}_{38}\text{Sr}$ released during the accident to decay? See Appendix G.

83. Three radioactive sources have the same activity, 35 mCi. Source A emits 1.0-MeV γ rays, source B emits 2.0-MeV γ rays, and source C emits 2.0-MeV alphas. What is the relative danger of these sources?

84. A 57-kg patient is to be given a medical test involving the ingestion of $^{99m}_{43}\text{Tc}$ (Section 42–8) which decays by emitting a 140-keV gamma. The half-life for this decay is 6 hours. Assuming that about half the gamma photons exit the body without interacting with anything, what must be the initial activity of the Tc sample if the whole-body dose cannot exceed 50 mrem? Make the rough approximation that biological elimination of Tc can be ignored.

85. The early scattering experiments performed around 1910 in Ernest Rutherford's laboratory in England produced the first evidence that an atom consists of a heavy nucleus surrounded by electrons. In one of these experiments, α particles struck a gold-foil target 4.0×10^{-5} cm thick in which there were 5.9×10^{28} gold atoms per cubic meter. Although most α particles either passed straight through the foil or were scattered at small angles, approximately 1.6×10^{-3} percent were scattered at angles greater than 90°—that is, in the backward direction. (*a*) Calculate the cross section, in barns, for backward scattering. (*b*) Rutherford concluded that such backward scattering could occur only if an atom consisted of a very tiny, massive, and positively charged nucleus with electrons orbiting some distance away. Assuming that backward scattering occurs for nearly direct collisions (i.e., $\sigma \approx$ area of nucleus), estimate the diameter of a gold nucleus.

86. Suppose a future fusion reactor would be able to put out 1000 MW of electrical power continuously. Assume the reactor will produce energy solely through the reaction given in Eq. 42–9a and will convert this energy to electrical energy with an efficiency of 33%. Estimate the minimum amount of deuterium needed to run this facility per year.

87. Centuries ago, paint generally contained a different amount of cobalt $\left({}^{59}_{27}\text{Co}\right)$ than paint today. A certain "old" painting is suspected of being a new forgery, and an examiner has decided to use **neutron activation analysis** to test this hypothesis. After placing the painting in a neutron flux of 5.0×10^{12} neutrons/cm^2/s for 5.0 minutes, an activity of 62 decays/s of ${}^{60}_{27}\text{Co}$ $\left(T_{\frac{1}{2}} = 5.27\,\text{yr}\right)$ is observed. Assuming ${}^{59}_{27}\text{Co}$ has a cross section of 19 bn, how much cobalt (in grams) does the paint contain?

88. If a 72-kg power plant worker has been exposed to the maximum slow-neutron radiation for a given year, how much total energy (in J) has that worker absorbed? What if he were exposed to fast protons?

89. Consider the fission reaction

$$\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{92}_{38}\text{Sr} + \text{X} + 3\text{n}.$$

(*a*) What is X? (*b*) If this were part of a chain reaction in a fission power reactor running at "barely critical," what would happen on average to the three produced neutrons? (*c*) (optional) What is the Q-value of this reaction? [*Hint*: Mass values can be found at www.nist.gov/pml/data/comp.cfm.]

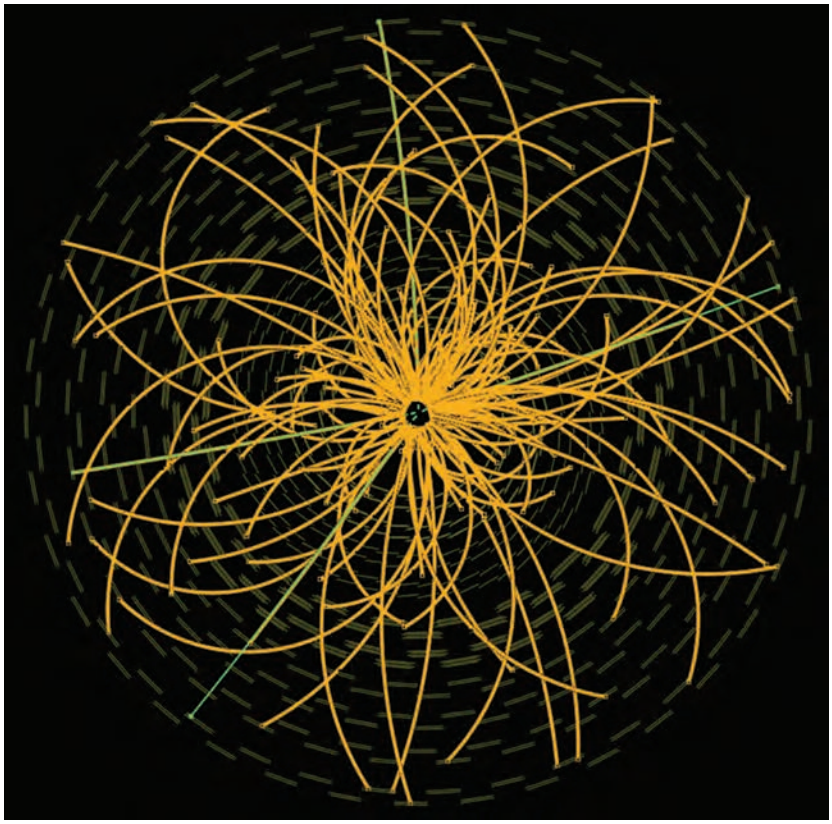
90. Consider the effort by humans to harness nuclear fusion as a viable energy source. (*a*) What are some advantages of using nuclear fusion rather than nuclear fission? (*b*) What is the major technological problem with using controlled nuclear fusion as a source of energy? (*c*) Discuss two different approaches to solving this problem. (*d*) What fuel is necessary in a nuclear fusion reaction? (*e*) Write a nuclear reaction using two nuclei of the fuel in part (*d*) to create a third nucleus. (*f*) Calculate the Q-value of the reaction in part (*e*).

91. Show, using the laws of conservation of energy and momentum, that for a nuclear reaction requiring energy, the minimum kinetic energy of the bombarding particle (the **threshold energy**) is equal to $\left[-Qm_{\text{pr}}/\left(m_{\text{pr}} - m_{\text{b}}\right)\right]$, where $-Q$ is the energy required (difference in total mass between products and reactants), m_{b} is the mass of the bombarding particle, and m_{pr} is the total mass of the products. Assume the target nucleus is at rest before an interaction takes place, and that all speeds are nonrelativistic.

ANSWERS TO EXERCISES

A: ${}^{138}_{56}\text{Ba}$.
B: 3 neutrons.
C: 2×10^{17}.
D: (*e*).
E: (*b*).
F: (*a*).

This photo is a computer reconstruction of particles produced due to a 7 TeV proton–proton collision at the Large Hadron Collider (LHC). It is a candidate for having produced the long-sought Higgs boson (plus other particles).



The Higgs in this case could have decayed (very quickly $\sim 10^{-22}$ s) into two Z bosons (which are carriers of the weak force):

$$H^0 \rightarrow Z^0 + Z^0.$$

We don't see the tracks of the Z^0 particles because (1) they are neutral and (2) they decay too quickly $(\sim 10^{-24}\,\text{s})$, in this case:

$$Z^0 \rightarrow e^- + e^+.$$

The tracks of the 2 electrons and 2 positrons are shown as green lines. The Higgs is thought to play a fundamental role in the Standard Model of particle physics, importantly providing mass to fundamental particles.

The CMS detector of this photo uses a combination of the detector types discussed in Section 41–11. A magnetic field causes particles to move in curved paths so the momentum of each can be measured (Section 27–4). Tracks of particles with very large momentum, such as our electrons here, are barely curved.

We will study elementary particle physics from its beginnings until today, including antiparticles, neutrinos, quarks, the Standard Model, and theories that go beyond. We start with the great machines that accelerate particles so they can collide at high energies.

CHAPTER 43

Elementary Particles

CHAPTER-OPENING QUESTIONS—Guess now!

1. Physicists reserve the term "fundamental particle" for particles with a special property. What do you think that special property is?

(a) Particles that are massless.
(b) Particles that possess the minimum allowable electric charge.
(c) Particles that have no known internal structure.
(d) Particles that produce no force on other objects.

2. The fundamental particles as we see them today, besides the long-sought-for Higgs boson, are

(a) atoms and electrons.
(b) protons, neutrons, and electrons.
(c) protons, neutrons, electrons, and photons.
(d) quarks, leptons, and gauge bosons (carriers of force).
(e) hadrons, leptons, and gauge bosons.

In the final two Chapters of this book we discuss two of the most exciting areas of contemporary physics: elementary particles in this Chapter, and cosmology and astrophysics in Chapter 44. These are subjects at the forefront of knowledge—elementary particles treats the smallest objects in the universe; cosmology treats the largest (and oldest) aspects of the universe. The reader who wants an understanding of the great beauties of present-day science (and its limits) will want to read these Chapters. So will those who want to be good citizens, even if there is not time to cover them in a physics course.

CONTENTS

43 – 1 High-Energy Particles and Accelerators
43 – 2 Beginnings of Elementary Particle Physics—Particle Exchange
43 – 3 Particles and Antiparticles
43 – 4 Particle Interactions and Conservation Laws
43 – 5 Neutrinos
43 – 6 Particle Classification
43 – 7 Particle Stability and Resonances
43 – 8 Strangeness? Charm? Towards a New Model
43 – 9 Quarks
43–10 The Standard Model: QCD and Electroweak Theory
43–11 Grand Unified Theories
43–12 Strings and Supersymmetry

In this penultimate Chapter we discuss *elementary particle* physics, which represents the human endeavor to understand the basic building blocks of all matter, and the fundamental forces that govern their interactions.

Almost a century ago, by the 1930s, it was accepted that all atoms can be considered to be made up of neutrons, protons, and electrons. The basic constituents of the universe were no longer considered to be atoms (as they had been for 2000 years) but rather the proton, neutron, and electron. Besides these three "elementary particles," there were several others: the positron (a positive electron), the neutrino, and the γ particle (or photon), for a total of six elementary particles.

By the 1950s and 1960s many new types of particles similar to the neutron and proton were discovered, as well as many "midsized" particles called *mesons* whose masses were mostly less than nucleon masses but more than the electron mass. (Other mesons, found later, have masses greater than nucleons.) Physicists felt that these particles could not all be fundamental, and must be made up of even smaller constituents (later confirmed by experiment), which were given the name *quarks*.

By the term **fundamental particle**, we mean a particle that is so simple, so basic, that it has no internal structure[†] (is not made up of smaller subunits)—see Chapter-Opening Question 1.

Today, the fundamental constituents of matter are considered to be **quarks** (they make up protons and neutrons as well as mesons) and **leptons** (a class that includes electrons, positrons, and neutrinos). There are also the "carriers of force" known as **gauge bosons**, including the photon, gluons, and W and Z bosons. In addition there is the **Higgs** boson, predicted in the 1960s but whose first suggestions of experimental detection were finally announced in 2012. The theory that describes our present view is called the **Standard Model**. How we came to our present understanding of elementary particles and the Standard Model is the subject of this Chapter.

One of the exciting developments of the last few years is an emerging synthesis between the study of elementary particles and astrophysics (Chapter 44). In fact, detailed observations in astrophysics have led to the conclusion that the greater part of the mass–energy content of the universe is not ordinary matter but two mysterious and invisible forms known as "dark matter" and "dark energy" which cannot be explained by the Standard Model in its present form.

Indeed, we are now aware that the Standard Model is not sufficient. There are problems and important questions still unanswered, and we will mention some of them in this Chapter and how we hope to answer them.

43–1 High-Energy Particles and Accelerators

In the late 1940s, after World War II, it was found that if the incoming particle in a nuclear reaction (Section 42–1) has sufficient energy, new types of particles can be produced. The earliest experiments used **cosmic rays**—particles that impinge on the Earth from space. In the laboratory, various types of particle accelerators have been constructed to accelerate protons or electrons to high energies so they can collide with other particles—often protons (the hydrogen nucleus). Heavy ions, up to uranium (U), have also been accelerated. These **high-energy accelerators** have been used to probe more deeply into matter, to produce and study new particles, and to give us information about the basic forces and constituents of nature. The particles produced in high-energy collisions can be detected by a variety of special detectors, discussed in Section 41–11, including scintillation counters, multiwire chambers, and semiconductors. The rate of production of any group of particles is quantified using the concept of *cross section*, Section 42–2. Because the projectile particles are at high energy, this field is sometimes called **high-energy physics**.

[†]Recall from Section 17–1 that the word "atom" comes from the Greek meaning "indivisible." Atoms have a substructure (protons, neutrons) so are not fundamental. Yet an atom is still the smallest "piece" of an element that has the characteristics of that material.

PSE5_CH43_v1.6_23Jun-20

Wavelength and Resolution

Particles accelerated to high energy can probe the interior of nuclei and nucleons or other particles they strike. An important factor is that faster-moving projectiles can reveal finer detail. The wavelength of projectile particles is given by de Broglie's wavelength formula (Eq. 37–7),

$$\lambda = \frac{h}{p}, \tag{43–1}$$

showing that the greater the momentum p of the bombarding particle, the shorter its wavelength. As discussed in Chapter 35 on diffraction, resolution of details in images is limited by the wavelength: the shorter the wavelength, the finer the detail that can be obtained. This is one reason why particle accelerators of higher and higher energy have been built in recent years: to probe ever deeper into the structure of matter, to smaller and smaller scales.

EXAMPLE 43–1 **High resolution with electrons.** What is the wavelength, and hence the expected resolution, for 1.3-GeV electrons?

APPROACH Because 1.3 GeV is much larger than the electron mass, we must be dealing with relativistic speeds. The momentum of the electrons is found from Eq. 36–14, and the wavelength is $\lambda = h/p$.

SOLUTION Each electron has kinetic energy $K = 1.3\,\text{GeV} = 1300\,\text{MeV}$, which is about 2500 times the rest energy of the electron $(mc^2 = 0.51\,\text{MeV})$. Thus we can ignore the term $(mc^2)^2$ in Eq. 36–14, $E^2 = p^2c^2 + m^2c^4$, and we solve for p:

$$p = \sqrt{\frac{E^2 - m^2c^4}{c^2}} \approx \sqrt{\frac{E^2}{c^2}} = \frac{E}{c}.$$

Therefore the de Broglie wavelength is

$$\lambda = \frac{h}{p} = \frac{hc}{E},$$

where $E = 1.3\,\text{GeV}$. Hence

$$\lambda = \frac{(6.63 \times 10^{-34}\,\text{J}\cdot\text{s})(3.0 \times 10^{8}\,\text{m/s})}{(1.3 \times 10^{9}\,\text{eV})(1.6 \times 10^{-19}\,\text{J/eV})} = 0.96 \times 10^{-15}\,\text{m},$$

or 0.96 fm. This resolution of about 1 fm is on the order of the size of nuclei (see Eq. 41–1).

NOTE The maximum possible resolution of this beam of electrons is far greater than for a light beam in a light microscope $(\lambda \approx 500\,\text{nm})$.

EXERCISE A What is the wavelength of a proton with $K = 1.00\,\text{TeV}$?

Another major reason today for building high-energy accelerators is that new particles of greater mass can be produced at higher collision energies, transforming the kinetic energy of the colliding particles into massive particles by $E = mc^2$, as we will discuss shortly. Now we look at particle accelerators.

Cyclotron

The cyclotron was developed in 1930 by E. O. Lawrence (1901–1958; Fig. 43–1) at the University of California, Berkeley. It uses a magnetic field to maintain charged ions—usually protons—in nearly circular paths. Although particle physicists no longer use simple cyclotrons, they are used in medicine for treating cancer, and their operating principles are useful for understanding modern accelerators.



FIGURE 43–1 Ernest O. Lawrence, left, with Donald Cooksey and the "dees" of an early cyclotron.

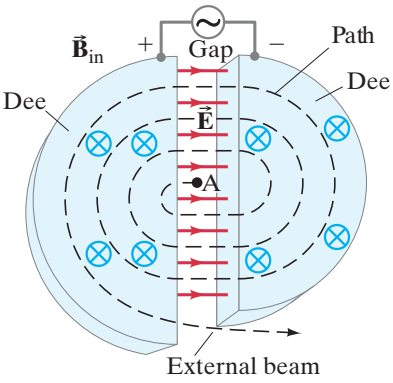


FIGURE 43–2 Diagram of a cyclotron. The magnetic field, applied by a large electromagnet, points into the page. The protons start at A, the ion source. The red electric field lines shown are for the alternating electric field in the gap at a certain moment.

The protons move in a vacuum inside two D-shaped cavities, as shown in Fig. 43–2. Each time they pass into the gap between the "dees," a voltage accelerates them (the electric force), increasing their speed and increasing the radius of curvature of their path in the magnetic field. After many revolutions, the protons acquire high kinetic energy and reach the outer edge of the cyclotron where they strike a target. The protons speed up only when they are in the gap *between* the dees, and the voltage must be alternating. When protons are moving to the right across the gap in Fig. 43–2, the right dee must be electrically negative and the left one positive. A half-cycle later, the protons are moving to the left, so the left dee must be negative in order to accelerate them.

The frequency, f, of the applied voltage must be equal to the frequency of the circulating protons. When ions of charge q are circulating *within* the hollow dees, the net force F on each is due to the magnetic field B, so $F = qvB$, where v is the speed of the ion at a given moment (Eq. 27–5). The magnetic force is perpendicular to both $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$, and does not speed up the ions but causes them to move in circles; the acceleration within the dees is centripetal and equals v^2/r, where r is the radius of the ion's path at a given moment. We use Newton's second law, $F = ma$, and find that

$$F = ma$$
$$qvB = \frac{mv^2}{r}$$

when the protons are within the dees (not the gap), so their (constant) speed at radius r is

$$v = \frac{qBr}{m}.$$

The time required for a complete revolution is the period T and is equal to

$$T = \frac{\text{distance}}{\text{speed}} = \frac{2\pi r}{qBr/m} = \frac{2\pi m}{qB}.$$

Hence the frequency of revolution f is

$$f = \frac{1}{T} = \frac{qB}{2\pi m}. \qquad \textbf{(43–2)}$$

This is known as the **cyclotron frequency**.

EXAMPLE 43–2 **Cyclotron.** A small cyclotron of maximum radius $R = 0.25$ m accelerates protons in a 1.7 T magnetic field. Calculate (*a*) the frequency needed for the applied alternating voltage, and (*b*) the kinetic energy of protons when they leave the cyclotron.

APPROACH The frequency of the protons revolving within the dees (Eq. 43–2) must equal the frequency of the voltage applied across the gap if the protons are going to increase in speed.

SOLUTION (*a*) From Eq. 43–2,

$$f = \frac{qB}{2\pi m} = \frac{(1.6\times10^{-19}\,\text{C})(1.7\,\text{T})}{(6.28)(1.67\times10^{-27}\,\text{kg})} = 2.6\times10^{7}\,\text{Hz} = 26\,\text{MHz},$$

which is in the radio-wave region of the EM spectrum (Fig. 31–12).

(*b*) The protons leave the cyclotron at $r = R = 0.25$ m. From $qvB = mv^2/r$ (see above), we have $v = qBr/m$, so their kinetic energy is

$$K = \frac{1}{2}mv^2 = \frac{1}{2}m\frac{q^2B^2R^2}{m^2} = \frac{q^2B^2R^2}{2m}$$
$$= \frac{(1.6\times10^{-19}\,\text{C})^2(1.7\,\text{T})^2(0.25\,\text{m})^2}{(2)(1.67\times10^{-27}\,\text{kg})} = 1.4\times10^{-12}\,\text{J} = 8.7\,\text{MeV}.$$

NOTE The kinetic energy is much less than the rest energy of the proton (938 MeV), so relativity is not needed.

NOTE The magnitude of the voltage applied to the dees does not appear in the formula for K, and so does not affect the final energy. But the higher this voltage, the fewer the revolutions required to bring the protons to full energy.

An important aspect of the cyclotron is that the frequency of the applied voltage, as given by Eq. 43–2, does not depend on the radius r of the particle's path. Thus the frequency does not have to be changed as the protons or ions start from the source and are accelerated to paths of larger and larger radii. But this is only true at nonrelativistic energies. At higher speeds, the momentum (Eq. 36–8) is $p = \gamma mv = mv/\sqrt{1 - v^2/c^2}$, so m in Eq. 43–2 has to be replaced by γm and the cyclotron frequency f (Eq. 43–2) depends on speed v because γ does. To keep the particles in sync, machines called **synchrocyclotrons** reduce the frequency in time to correspond to the increase of γm (in Eq. 43–2) as a packet of charged particles increases in speed more slowly at larger orbits.

Synchrotron

Another way to accelerate relativistic particles is to increase the magnetic field B in time so as to keep f (Eq. 43–2) constant as the particles speed up. Such devices are called **synchrotrons**; the particles move in a circle of fixed radius, which can be very large. The larger the radius, the greater the kinetic energy the particles can have for a given magnetic field strength (see argument on previous page). The biggest synchrotron of all is at the European Center for Nuclear Research (CERN) in Geneva, Switzerland, the Large Hadron Collider (LHC). It is 4.3 km in radius, and 27 km in circumference, and accelerates protons to 6.5 TeV.

The *Tevatron* accelerator at Fermilab (Fermi National Accelerator Laboratory, near Chicago, Illinois), has a radius of 1.0 km.[†] The Tevatron accelerated protons to about $1000\,\text{GeV} = 1\,\text{TeV}$ (hence its name, $1\,\text{TeV} = 10^{12}\,\text{eV}$). It was shut down in 2011, but Fermilab continues as a research facility, especially for studying properties of neutrinos.

These large synchrotrons use a narrow ring of magnets (see Fig. 43–3) with each magnet placed at the same radius from the center of the circle. The magnets are interrupted by gaps where high voltage accelerates the particles to higher speeds. Another way to describe the acceleration is to say the particles "surf" on a traveling electromagnetic wave within radiofrequency (RF) cavities. (The particles are first given considerable energy in smaller accelerators, "injectors," before being injected into the large ring of the large synchrotron.)

One problem of any accelerator is that accelerating electric charges radiate electromagnetic energy (see Chapter 31). Since ions or electrons are accelerated in an accelerator, we can expect considerable energy to be lost by radiation. The effect increases with energy and is especially important in circular machines where centripetal acceleration is present, such as synchrotrons, and hence is called **synchrotron radiation**. Synchrotron radiation can be useful, however. Intense beams of photons (γ rays) are sometimes needed, and they are often obtained from an electron synchrotron. Strong sources of such photons are referred to as **light sources** and are used all over the world as γ-ray sources.



FIGURE 43–3 The interior of the tunnel of the main accelerator at Fermilab, showing (red) the ring of superconducting magnets used to keep particles moving in a circular path at the 1-TeV Tevatron.

[†]Robert Wilson, who helped design the Tevatron, and founded the field of proton therapy (Section 42–7), expressed his views on accelerators and national security in this exchange with Senator John Pastore during testimony before a Congressional Committee in 1969:

> Pastore: "Is there anything connected with the hopes of this accelerator [the Tevatron] that in any way involves the security of the country?"
>
> Robert Wilson: "No sir, I don't believe so."
>
> Pastore: "Nothing at all?"
>
> Wilson: "Nothing at all. …"
>
> Pastore: "It has no value in that respect?"
>
> Wilson: "It has only to do with the respect with which we regard one another, the dignity of men, our love of culture. … It has to do with are we good painters, good sculptors, great poets? I mean all the things we really venerate in our country and are patriotic about … it has nothing to do directly with defending our country except to make it worth defending."

Linear Accelerators

In a **linear accelerator** (linac), electrons or ions are accelerated along a straight-line path, Fig. 43–4, passing through a series of tubular conductors. Voltage applied to the tubes is alternating so that when electrons (say) reach a gap, the tube in front of them is positive and the one they just left is negative. At low speeds, the particles cover less distance in the same amount of time, so the tubes are shorter at first. Electrons, with their small mass, get close to the speed of light quickly, $v \approx c$, and the tubes are nearly equal in length. Linear accelerators are particularly important for accelerating electrons to avoid loss of energy due to synchrotron radiation. The largest electron linear accelerator has been at Stanford University (Stanford Linear Accelerator Center, or SLAC), about 3 km (2 mi) long, able to accelerate electrons to 50 GeV. Linacs accelerating protons are used as injectors into circular machines to provide initial kinetic energy. Many hospitals have 10- or 20-MeV electron linacs that strike a metal foil to produce γ-ray photons to irradiate tumors.

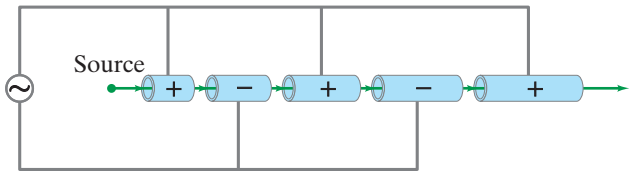
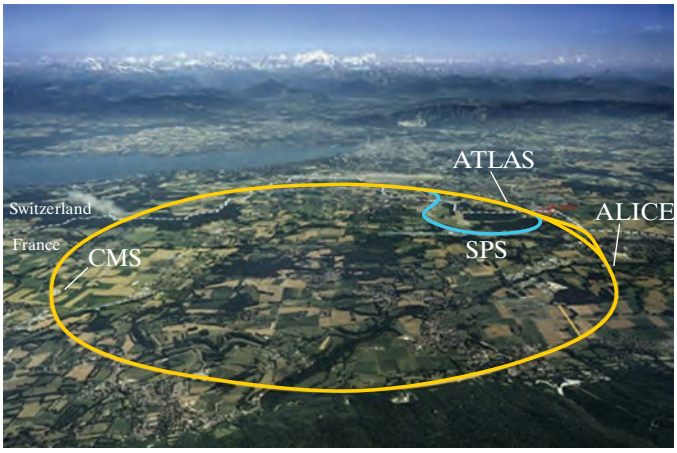


FIGURE 43–4 Diagram of a simple linear accelerator.

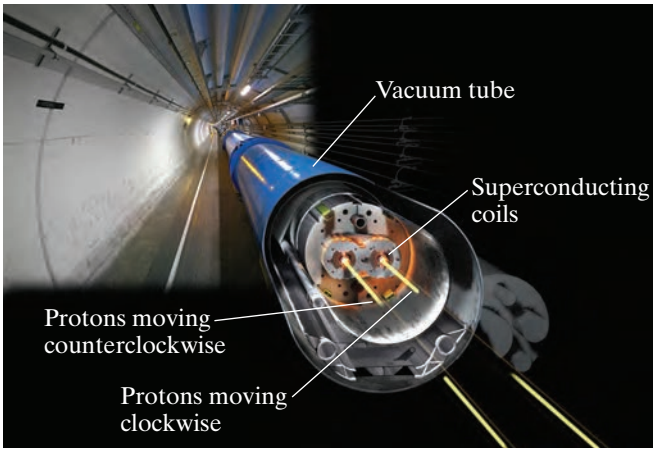
Colliding Beams

High-energy physics experiments were once done by aiming a beam of particles from an accelerator at a stationary target. But to obtain the maximum possible collision energy from a given accelerator, two beams of particles are now accelerated to very high energy and are steered so that they collide head-on. One way to accomplish such **colliding beams** with a single accelerator is through the use of **storage rings**, in which oppositely circulating beams can be repeatedly brought into collision with one another at particular points. For example, in the experiments that provided strong evidence for the top quark (Section 43–9 and Fig. 43–15), the Fermilab Tevatron accelerated protons and antiprotons each to 900 GeV, so that the combined energy of head-on collisions was 1.8 TeV.

The largest collider is the Large Hadron Collider (LHC) at CERN, with a circumference of 27 km (Fig. 43–5). The two colliding beams are designed to each carry protons at up to 6.5 or 7 TeV for a total interaction energy of 13 TeV. For the experiments in 2011 and 2012 the total interaction energy was 7 TeV and 8 TeV. The protons for each of the beams, moving in opposite directions, are accelerated in several stages. The penultimate is SPS (Super Proton Synchrotron), seen in Fig. 43–5a which accelerates protons from 28 GeV to the 450 GeV, and then they are injected into the LHC itself.



(a)



(b)

FIGURE 43–5 (a) The large circle represents the position of the tunnel, about 100 m below the ground at CERN (near Geneva) on the French-Swiss border, which houses the LHC. The smaller circle shows the position of the Super Proton Synchrotron used for accelerating protons prior to injection into the LHC. (b) Circulating proton beams, in opposite directions, inside the vacuum tube within the LHC tunnel.

Figure 43–6 shows part of one of the detectors (ATLAS) as it was being constructed at the LHC. The detectors within ATLAS include silicon semiconductor detectors with huge numbers of pixels used to track particle paths and find their point of interaction, and to measure their radius of curvature in a magnetic field and thus determine their momentum (Section 27–4). Their energy is determined in "calorimeters" utilizing plastic, liquid, or dense metal compound crystal scintillators (Section 41–11).

In the planning stage is the International Linear Collider (ILC) which would have colliding beams of electrons and positrons $(e^-$ and $e^+)$ at around 0.3 to 1 TeV. It would utilize semiconductor detectors using CMOS (Section 33–5) with embedded transistors to allow fast readout.

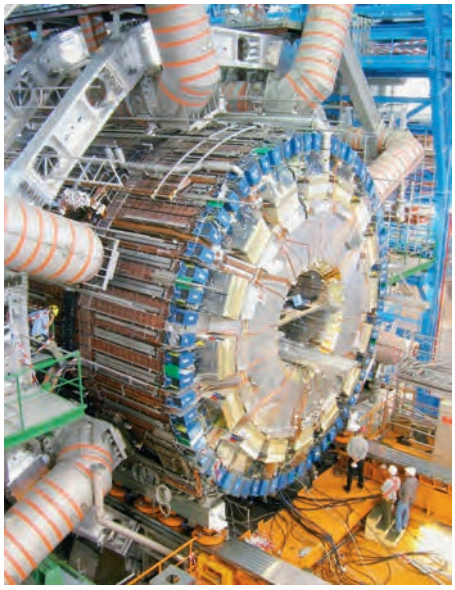


FIGURE 43–6 ATLAS, one of the large complex detectors at the LHC, is shown here as it was being built. In 2012 it was used to provide evidence for the Higgs boson. Note the people near the bottom. From the outside, the CMS detector at the LHC looks similar.

EXAMPLE 43–3 **Protons at relativistic speeds.** Determine the energy required to accelerate a proton in a high-energy accelerator (*a*) from rest to $v = 0.900c$, and (*b*) from $v = 0.900c$ to $v = 0.999c$. (*c*) What is the kinetic energy achieved by the proton in each case?

APPROACH We use the work-energy principle, which is still valid relativistically as mentioned in Section 36–11: $W = \Delta K$.

SOLUTION The kinetic energy of a proton of mass m is given by Eq. 36–10,

$$K = (\gamma - 1)mc^2,$$

where the relativistic factor γ is

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

The work-energy theorem becomes

$$W = \Delta K = (\gamma_2 - 1)mc^2 - (\gamma_1 - 1)mc^2 = (\gamma_2 - \gamma_1)mc^2$$

where γ_1 and γ_2 are for the initial and final speeds.

(*a*) For $v = v_1 = 0$, $\gamma_1 = 1$; and for $v_2 = 0.900c$

$$\gamma_2 = \frac{1}{\sqrt{1 - (0.900)^2}} = 2.29.$$

For a proton, $mc^2 = 938\,\text{MeV}$, so the work (or energy) needed to accelerate it from rest to $v_2 = 0.900c$ is

$$W = \Delta K = (\gamma_2 - \gamma_1)mc^2$$
$$= (2.29 - 1.00)(938\,\text{MeV}) = 1.21\,\text{GeV}.$$

(*b*) To go from $v_2 = 0.900c$ to $v_3 = 0.999c$, we need

$$\gamma_3 = \frac{1}{\sqrt{1 - (0.999)^2}} = 22.4.$$

So the work needed to accelerate a proton from $0.900c$ to $0.999c$ is

$$W = \Delta K = (\gamma_3 - \gamma_2)mc^2$$
$$= (22.4 - 2.29)(938\,\text{MeV}) = 18.9\,\text{GeV},$$

which is 15 times as much.

(*c*) The kinetic energy reached by the proton in (*a*) is just equal to the work done on it, $K = 1.21\,\text{GeV}$. The final kinetic energy of the proton in (*b*), moving at $v_3 = 0.999c$, is

$$K = (\gamma_3 - 1)mc^2 = (21.4)(938\,\text{MeV}) = 20.1\,\text{GeV}.$$

NOTE This result makes sense because, starting from rest, we did work

$$W = 1.21\,\text{GeV} + 18.9\,\text{GeV} = 20.1\,\text{GeV}$$

on it.

EXAMPLE 43–4 **Speed of a 6.5-TeV proton.** What is the speed of a 6.5-TeV proton produced at the LHC?

APPROACH The kinetic energy $K = 6.5\,\text{TeV} = 6.5 \times 10^{12}\,\text{eV}$ is much greater than the mass of the proton, $0.938 \times 10^{9}\,\text{eV}$, so relativistic calculations must be used. In particular, we use Eq. 36–10:

$$K = (\gamma - 1)mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2.$$

SOLUTION Compared to $K = 6.5 \times 10^{12}\,\text{eV}$, the rest energy $(\approx 10^{-3}\,\text{TeV})$ can be neglected, so we write

$$K = \frac{mc^2}{\sqrt{1 - v^2/c^2}}.$$

Then

$$1 - \frac{v^2}{c^2} = \left(\frac{mc^2}{K}\right)^2$$

or

$$\frac{v}{c} = \sqrt{1 - \left(\frac{mc^2}{K}\right)^2} = \sqrt{1 - \left(\frac{938 \times 10^6\,\text{eV}}{6.5 \times 10^{12}\,\text{eV}}\right)^2} = \sqrt{1 - 2.08 \times 10^{-8}}.$$

Using the binomial theorem

$$v = \left(1 - \tfrac{1}{2}(2.08 \times 10^{-8})\right)c = 0.999\,999\,9896c.$$

The proton is traveling at a speed extremely close to c, the speed of light.

43–2 Beginnings of Elementary Particle Physics—Particle Exchange

The accepted model for elementary particles today views *quarks* and *leptons* as the fundamental constituents of ordinary matter. To understand our present-day view of elementary particles, it is necessary to understand the ideas leading up to its formulation.

Elementary particle physics might be said to have begun in 1935 when the Japanese physicist Hideki Yukawa (1907–1981) predicted the existence of a new particle that would in some way mediate the strong nuclear force. To understand Yukawa's idea, we first consider the electromagnetic force. When we first discussed electricity, we saw that the electric force acts over a distance, without contact. To better perceive how a force can act over a distance, we used the idea of a **field**. The force that one charged particle exerts on a second can be said to be due to the electric field set up by the first. Similarly, the magnetic field can be said to carry the magnetic force. Later (Chapter 31), we saw that electromagnetic (EM) fields can travel through space as waves. Finally, in Chapter 37, we saw that electromagnetic radiation (light) can be considered as either a wave or as a collection of particles called *photons*. Because of this wave–particle duality, it is possible to imagine that the electromagnetic force between charged particles is due to

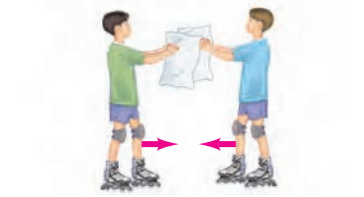
(1) the EM field set up by one charged particle and felt by the other, or

(2) an exchange of photons (γ particles) between them.

It is (2) that we want to concentrate on here, and a crude analogy for how an exchange of particles could give rise to a force is suggested in Fig. 43–7. In part (a), two children start throwing heavy pillows at each other; each throw and each catch results in the child being pushed backward by the impulse. This is the equivalent of a repulsive force. On the other hand, if the two children exchange pillows by grabbing them out of the other person's hand, they will be pulled toward each other, as when an attractive force acts.



(a) Repulsive force (children throwing pillows)



(b) Attractive force (children grabbing pillows from each other's hands)

FIGURE 43–7 Forces equivalent to particle exchange. (a) Repulsive force (children on roller skates throwing pillows at each other). (b) Attractive force (children grabbing pillows from each other's hands).

PSE5_CH43_v1.6_23Jun-20

For the electromagnetic force, it is photons that are exchanged between two charged particles that give rise to the force between them. A simple diagram describing this photon exchange is shown in Fig. 43–8. Such a diagram, called a **Feynman diagram** after its inventor, the American physicist Richard Feynman (1918–1988), is based on the theory of **quantum electrodynamics** (QED).

Figure 43–8 represents the simplest case in QED, in which a single photon is exchanged. One of the charged particles emits the photon and recoils somewhat as a result; and the second particle absorbs the photon. In any collision or *interaction*, energy and momentum are transferred from one charged particle to the other, carried by the photon. The photon is absorbed by the second particle after it is emitted by the first particle and is not observable. Hence the photon is referred to as a **virtual** photon, in contrast to one that is free and can be detected by instruments. The photon is said to **mediate**, or **carry**, the electromagnetic force.

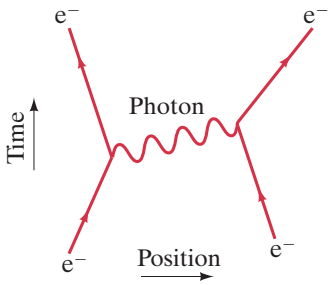
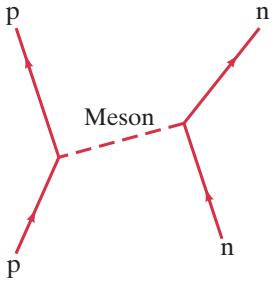


FIGURE 43–8 Feynman diagram showing a photon acting as the carrier of the electromagnetic force between two electrons. This is sort of an x vs. t graph, with t increasing upward. Starting at the bottom, two electrons approach each other (the distance between them decreases in time). As they get close, momentum and energy get transferred from one to the other, carried by a photon (or, perhaps, by more than one), and the two electrons bounce apart.

By analogy with photon exchange that mediates the electromagnetic force, Yukawa argued that there ought to be a particle that mediates the strong nuclear force—the force that holds nucleons together in the nucleus. Yukawa called this predicted particle a **meson** (meaning "medium mass"). Figure 43–9 is a Feynman diagram showing the original model of meson exchange: a meson carrying the strong force between a neutron and a proton.

FIGURE 43–9 Early model showing meson exchange when a proton and neutron interact via the strong nuclear force. (Today, as we shall see shortly, we view the strong force as carried by gluons between quarks.)



A rough estimate of the mass of the meson can be made as follows. Suppose the proton on the left in Fig. 43–9 is at rest. For it to emit a meson would require energy (to make the meson's mass) which, coming from nowhere, would violate conservation of energy. But the uncertainty principle allows nonconservation of energy by an amount ΔE if it occurs only for a time Δt given by $(\Delta E)(\Delta t) \approx h/2\pi$. We set ΔE equal to the energy needed to create the mass m of the meson: $\Delta E = mc^2$. Conservation of energy is violated only as long as the meson exists, which is the time Δt required for the meson to pass from one nucleon to the other, where it is absorbed and disappears. If we assume the meson travels at relativistic speed, close to the speed of light c, then Δt need be at most about $\Delta t = d/c$, where d is the maximum distance that can separate the interacting nucleons. Thus we can write

$$\Delta E\,\Delta t \approx \frac{h}{2\pi}$$

$$mc^2\left(\frac{d}{c}\right) \approx \frac{h}{2\pi}$$

or

$$mc^2 \approx \frac{hc}{2\pi d}. \qquad \textbf{(43–3)}$$

The range of the strong nuclear force (the maximum distance away it can be felt) is small—not much more than the size of a nucleon or small nucleus (see Eq. 41–1)—so let us take $d \approx 1.5 \times 10^{-15}\,\text{m}$. Then from Eq. 43–3,

$$mc^2 \approx \frac{hc}{2\pi d} = \frac{(6.6 \times 10^{-34}\,\text{J}\cdot\text{s})(3.0 \times 10^{8}\,\text{m/s})}{(6.28)(1.5 \times 10^{-15}\,\text{m})} \approx 2.1 \times 10^{-11}\,\text{J} = 130\,\text{MeV}.$$

The mass of the predicted meson, roughly $130\,\text{MeV}/c^2$, is about 250 times the electron mass of $0.51\,\text{MeV}/c^2$.

EXERCISE B What effect does an increase in the mass of the virtual exchange particle have on the range of the force it mediates? (*a*) Decreases it; (*b*) increases it; (*c*) has no appreciable effect; (*d*) decreases the range for charged particles and increases the range for neutral particles.

Note that since the electromagnetic force has infinite range, Eq. 43–3 with $d = \infty$ tells us that the exchanged particle for the electromagnetic force, the photon, will have zero mass, which it does.

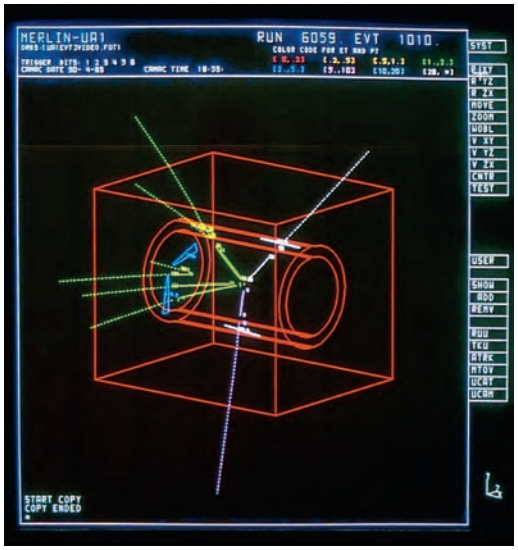
The particle predicted by Yukawa was discovered in cosmic rays by C. F. Powell and G. Occhialini in 1947, and is called the "π" or pi meson, or simply the **pion**. It comes in three charge states: $+e$, $-e$, or 0, where $e = 1.6 \times 10^{-19}$ C. The π^+ and π^- have mass of 139.6 MeV/c^2 and the π^0 a mass of 135.0 MeV/c^2, all close to Yukawa's prediction. All three interact strongly with matter. Reactions observed in the laboratory, using a particle accelerator, include

$$\begin{aligned} \mathrm{p} + \mathrm{p} &\rightarrow \mathrm{p} + \mathrm{p} + \pi^0, \\ \mathrm{p} + \mathrm{p} &\rightarrow \mathrm{p} + \mathrm{n} + \pi^+. \end{aligned} \qquad \textbf{(43–4)}$$

The incident proton from the accelerator must have sufficient energy to produce the additional mass of the free pion.

Yukawa's theory of pion exchange as carrier of the strong force has been superseded by *quantum chromodynamics* in which protons, neutrons, and other strongly interacting particles are made up of basic entities called *quarks*, and the basic carriers of the strong force are *gluons*, as we shall discuss shortly. But the basic idea of the earlier theory, that forces can be understood as the exchange of particles, remains valid.

There are four known types of force—or interactions—in nature. The electromagnetic force is carried by the photon, the strong force by gluons. What about the other two: the weak force and gravity? These too are believed to be mediated by particles. The particles that transmit the weak force are referred to as the W^+, W^-, and Z^0, and were detected in 1983 (Fig. 43–10). The quantum (or carrier) of the gravitational force has been named the **graviton**, but its existence has not been detected and it may not be detectable.

(a) 

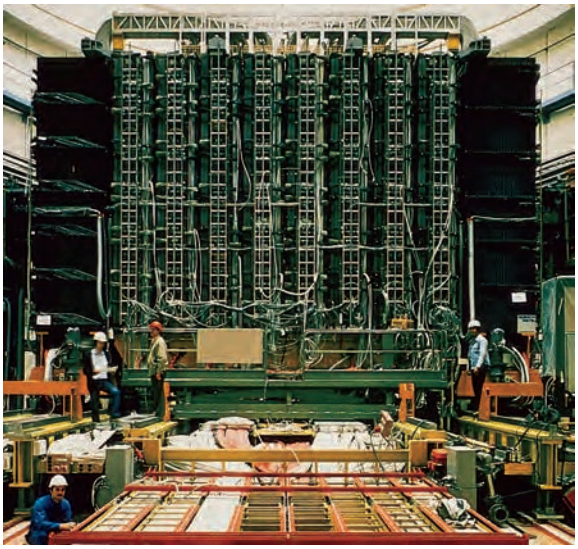
(b) 

FIGURE 43–10 (a) Computer reconstruction of a Z-particle decay into an electron and a positron $(Z^0 \rightarrow e^+ + e^-)$ whose tracks are shown in white, which took place in the UA1 detector at CERN. (b) Photo of the UA1 detector at CERN as it was being built in the 1980s.

A comparison of the four forces is given in Table 43–1, where they are listed according to their (approximate) relative strengths. Although gravity may be the most obvious force in daily life (because of the huge mass of the Earth), on a nuclear scale gravity is by far the weakest of the four forces, and its effect at the particle level can nearly always be ignored.

TABLE 43–1 The Four Forces in Nature

Type	Relative Strength (approx., for 2 protons in nucleus)	Field Particle
Strong	1	Gluons (= g)
Electromagnetic	10^{-2}	Photon ($= \gamma$)
Weak	10^{-6}	$W^{\pm}$ and Z^0
Gravitational	10^{-38}	Graviton (?)

PSE5_CH43_v1.6_23Jun-20

43–3 Particles and Antiparticles

The positron, as we discussed in Sections 37–5 (pair production) and 41–5 (β^+ decay), is basically a positive electron. That is, many of its properties are the same as for the electron, such as mass, but it has the opposite electric charge ($+e$). Other quantum numbers that we will discuss shortly are also reversed. The positron is said to be the **antiparticle** to the electron.

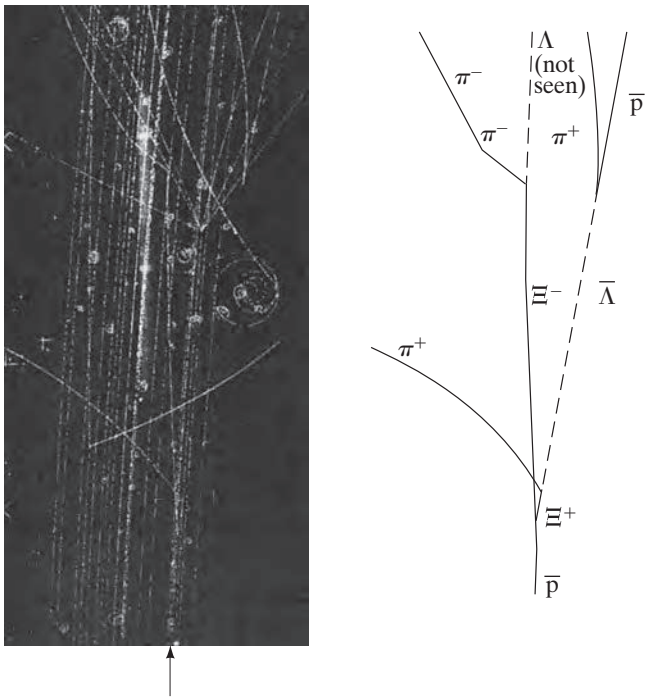


FIGURE 43–11 Liquid-hydrogen bubble-chamber photograph of an antiproton ($\bar{p}$) colliding with a proton at rest, producing a Xi—anti-Xi pair ($\bar{p} + p \rightarrow \Xi^- + \Xi^+$) that subsequently decay into other particles. The drawing indicates the assignment of particles to each track, which is based on how or if that particle decays, and on mass values estimated from measurement of momentum (curvature of track in magnetic field) and energy (thickness of track, for example). Neutral particle paths are shown by dashed lines since neutral particles rarely ionize atoms, around which bubbles form, and hence leave no tracks. 1950s.

The positron was first detected as a curved path in a cloud chamber in a magnetic field by Carl Anderson in 1932. It was predicted that other particles also would have antiparticles. It was decades before another type was found. Finally, in 1955 the antiparticle to the proton, the **antiproton** ($\bar{p}$), which carries a negative charge (Fig. 43–11), was discovered at the University of California, Berkeley, by Emilio Segrè (1905–1989, Fig. 43–12) and Owen Chamberlain (1920–2006). A bar, such as over the p, is used to indicate the antiparticle ($\bar{p}$). Soon after, the antineutron ($\bar{n}$) was found. All particles have antiparticles. But a few, like the photon, the π^0, and the Higgs, do not have distinct antiparticles—we say that they are their own antiparticles.

Antiparticles are produced in nuclear reactions when there is sufficient energy available to produce the required mass, and they do not live very long in the presence of matter, because matter and antimatter annihilate each other. For example, a positron is stable when by itself but rarely survives for long; as soon as it encounters an electron, the two annihilate each other. The energy of their vanished mass, plus any kinetic energy they possessed, is usually converted into the energy of two γ rays. Annihilation also occurs for all other particle–antiparticle pairs.

Antimatter is a term referring to material that would be made up of "antiatoms" in which antiprotons and antineutrons would form the nucleus around which positrons (antielectrons) would move. The term is also used for antiparticles in general. If there were pockets of antimatter in the universe, a huge explosion would occur when it encountered normal matter. It is believed that antimatter was prevalent in the very early universe (Section 44–7), but the vast majority annihilated with more prevalent ordinary matter.

FIGURE 43–12 Emilio Segrè: he worked with Fermi in the 1930s, later discovered the first "man-made" element, technetium, and other elements, and then the antiproton. The inscription below the photo is from a book by Segrè given to this book's author.



Negative Sea of Electrons; Vacuum State

The original idea for antiparticles came from a relativistic wave equation developed in 1928 by the Englishman P. A. M. Dirac (1902–1984). Recall that the non-relativistic Schrödinger equation took conservation of energy as a starting point. The Dirac equation too was based in part on energy conservation. We saw in Chapter 36 that the total energy E of a particle with mass m and momentum p and zero potential energy is given by Eq. 36–14, $E^2 = p^2c^2 + m^2c^4$. Thus

$$E = \pm\sqrt{p^2c^2 + m^2c^4}.$$

Dirac applied his new equation and found that it included solutions with both + and − signs. He could not ignore the solution with the negative sign, which we might have thought unphysical. If those negative energy states are real, then we would expect normal free electrons to drop down into those states, emitting photons—never experimentally seen. To deal with this difficulty, Dirac postulated that all those negative energy states are *normally occupied*. That is, what we thought was the **vacuum** is instead a vast "**sea**" **of electrons** in negative energy states (Fig. 43–13a). These electrons are not normally observable. But if a photon strikes one of these negative energy electrons, that electron can be knocked up to a normal $(E > mc^2)$ energy state as shown in Fig. 43–13b. Note in Fig. 43–13 that there are no energy states between $E = -mc^2$ and $E = +mc^2$ (because p^2 cannot be negative in the equation $E = \pm\sqrt{p^2c^2 + m^2c^4}$). The photon that knocks an e^- from the negative sea up to a normal state (Fig. 43–13b) must have an energy greater than $2mc^2$. What is left behind is a hole (as in semiconductors, Sections 40–7 and 40–8) with positive charge. We call that "hole" a **positron**, and it can move around as a free particle with positive energy. Thus Fig. 43–13b represents (Section 37–5) **pair production**: $\gamma \rightarrow e^-e^+$.

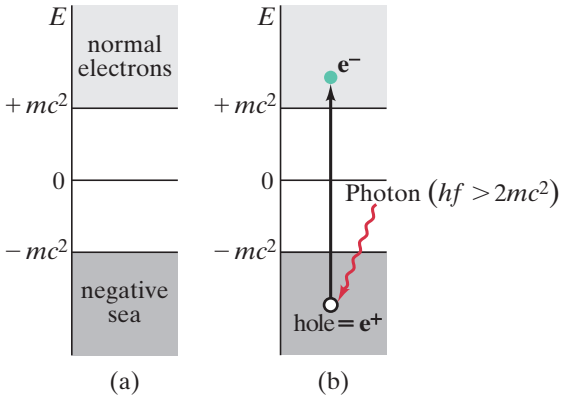


FIGURE 43–13 (a) Possible energy states for an electron. Note the vast sea of fully occupied electron states at $E < -mc^2$. (b) An electron in the negative sea is hit by a photon $(E > 2mc^2)$ and knocks it up to a normal positive energy state. The positive "hole" left behind acts like a positive electron—it is a positron.

The vast sea of electrons with negative energy in Fig. 43–13 is the vacuum (or **vacuum state**). According to quantum mechanics, the vacuum is not empty, but contains electrons and other particles as well. The uncertainty principle allows a particle to jump briefly up to a normal energy, thus creating a **particle–antiparticle** pair. It is possible that they could be the source of the recently discovered *dark energy* that fills the universe (Chapter 44). We still have a lot to learn.

43–4 Particle Interactions and Conservation Laws

One of the important uses of high-energy accelerators and colliders is to study the interactions of elementary particles with each other. As a means of ordering this subnuclear world, the conservation laws are indispensable. The laws of conservation of energy, of momentum, of angular momentum, and of electric charge are found to hold precisely in all particle interactions.

PSE5_CH43_v1.6_23Jun-20

A study of particle interactions has revealed a number of new conservation laws which (just like the old ones) are ordering principles: they help to explain why some reactions occur and others do not. For example, the following reaction has never been observed:

$$\mathrm{p} + \mathrm{n} \not\rightarrow \mathrm{p} + \mathrm{p} + \bar{\mathrm{p}}$$

even though charge, energy, and so on, are conserved ($\not\rightarrow$ means the reaction does not occur). To understand why such a reaction does not occur, physicists hypothesized a new conservation law, the conservation of **baryon number**. (Baryon number is a generalization of nucleon number, which we saw earlier is conserved in nuclear reactions and decays.) All nucleons are defined to have baryon number $B = +1$, and all antinucleons (antiprotons, antineutrons) have $B = -1$. All other types of particles, such as photons, mesons, and electrons and other leptons, have $B = 0$. The reaction shown at the start of this paragraph does not conserve baryon number since the left side has $B = (+1) + (+1) = +2$, and the right has $B = (+1) + (+1) + (-1) = +1$. On the other hand, the following reaction does conserve B and *does* occur if the incoming proton has sufficient energy:

$$\begin{aligned} \mathrm{p} + \mathrm{p} &\rightarrow \mathrm{p} + \mathrm{p} + \bar{\mathrm{p}} + \mathrm{p}, \\ B = +1 + 1 &= +1 + 1 - 1 + 1. \end{aligned}$$

As indicated, $B = +2$ on both sides of this equation. From these and other reactions, the **conservation of baryon number** has been established as a basic principle of physics.

Also useful are conservation laws for the three **lepton numbers**, associated with weak interactions including decays. In ordinary β decay, an electron or positron is emitted along with a neutrino or antineutrino. In another type of decay, a particle known as a "μ" or **muon** can be emitted instead of an electron. The muon (discovered in 1937) seems to be much like an electron, except its mass is 207 times larger $(106\,\mathrm{MeV}/c^2)$. The neutrino (ν_e) that accompanies an emitted electron is found to be different from the neutrino (ν_μ) that accompanies an emitted muon. Each of these neutrinos has an antiparticle: $\bar{\nu}_\mathrm{e}$ and $\bar{\nu}_\mu$. In ordinary β decay we have, for example,

CAUTION
The different types of neutrinos are not identical

$$\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu}_\mathrm{e}$$

but not $\mathrm{n} \not\rightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu}_\mu$. To explain why these do not occur, the concept of **electron lepton number**, L_e, was invented. If the electron (e^-) and the electron neutrino (ν_e) are assigned $L_\mathrm{e} = +1$, and e^+ and $\bar{\nu}_\mathrm{e}$ are assigned $L_\mathrm{e} = -1$, whereas all other particles have $L_\mathrm{e} = 0$, then all observed decays conserve L_e. For example, in $\mathrm{n} \rightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu}_\mathrm{e}$, initially $L_\mathrm{e} = 0$, and afterward $L_\mathrm{e} = 0 + (+1) + (-1) = 0$. Decays that do not conserve L_e, even though they would obey the other conservation laws, are not observed to occur.

In a decay involving muons, such as

$$\pi^+ \rightarrow \mu^+ + \nu_\mu,$$

a second quantum number, **muon lepton number** (L_μ), is conserved. The μ^- and ν_μ are assigned $L_\mu = +1$, and their antiparticles μ^+ and $\bar{\nu}_\mu$ have $L_\mu = -1$, whereas all other particles have $L_\mu = 0$. L_μ too is conserved in interactions and decays. Similar assignments can be made for the **tau lepton number**, L_τ, associated with the τ lepton (discovered in 1976 with mass more than 3000 times the electron mass) and its neutrino, ν_τ.

Antiparticles have not only opposite electric charge from their particles, but also opposite B, L_e, L_μ, and L_τ. For example, a neutron has $B = +1$, an antineutron has $B = -1$ (and all the L's are zero).

CONCEPTUAL EXAMPLE 43–5 **Lepton number in muon decay.** Which of the following decay schemes is possible for muon decay: (*a*) $\mu^- \rightarrow e^- + \bar{\nu}_e$; (*b*) $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$; (*c*) $\mu^- \rightarrow e^- + \nu_e$? All of these particles have $L_\tau = 0$.

RESPONSE A μ^- has $L_\mu = +1$ and $L_e = 0$. This is the initial state for all decays given, and the final state must also have $L_\mu = +1$, $L_e = 0$. In (*a*), the final state has $L_\mu = 0 + 0 = 0$, and $L_e = +1 - 1 = 0$; L_μ would not be conserved and indeed this decay is not observed to occur. The final state of (*b*) has $L_\mu = 0 + 0 + 1 = +1$ and $L_e = +1 - 1 + 0 = 0$, so both L_μ and L_e are conserved. This is in fact the most common decay mode of the μ^-. Lastly, (*c*) does not occur because L_e (= +2 in the final state) is not conserved, nor is L_μ.

EXAMPLE 43–6 **Energy and momentum are conserved.** In addition to the "number" conservation laws which help explain the decay schemes of particles, we can also apply the laws of conservation of energy and momentum. The decay of a Σ^+ particle at rest with mass 1189 MeV/c^2 (Table 43–2 in Section 43–6) can yield a proton $(m_p = 938\ \text{MeV}/c^2)$ and a neutral pion, π^0 $(m_{\pi^0} = 135\ \text{MeV}/c^2)$:

$$\Sigma^+ \rightarrow p + \pi^0.$$

Determine the kinetic energies of the proton and π^0.

APPROACH We find the energy release from the change in mass $(E = mc^2)$ as we did for nuclear processes (Eq. 41–3 or 42–2a). Then we apply conservation of energy and momentum, using relativistic formulas as the energies are large.

SOLUTION The energy released, $Q = K_p + K_{\pi^0}$ (the sum of the kinetic energies of p and π^0) equals the change in mass $\times\, c^2$:

$$Q = [m_{\Sigma^+} - (m_p + m_{\pi^0})]c^2 = [1189 - (938 + 135)]\ \text{MeV} = 116\ \text{MeV}.$$

Next we apply conservation of momentum: the initial particle Σ^+ is at rest, so the π^0 and p have opposite momentum but are equal in magnitude: $p_{\pi^0} = p_p$. We square this equation, $p_{\pi^0}^2 = p_p^2$, which becomes, using Eq. 36–14 $(p^2c^2 = E^2 - m^2c^4)$,

$$E_{\pi^0}^2 - m_{\pi^0}^2c^4 = E_p^2 - m_p^2c^4.$$

Solving for $E_{\pi^0}^2$:

$$E_{\pi^0}^2 = E_p^2 - m_p^2c^4 + m_{\pi^0}^2c^4.$$

We substitute Eq. 36–11a, $E = K + mc^2$, for both the π^0 and the p:

$$(K_{\pi^0} + m_{\pi^0}c^2)^2 = (K_p + m_pc^2)^2 - m_p^2c^4 + m_{\pi^0}^2c^4$$

$$K_{\pi^0}^2 + 2K_{\pi^0}m_{\pi^0}c^2 + \cancel{m_{\pi^0}^2c^4} = K_p^2 + 2K_pm_pc^2 + \cancel{m_p^2c^4} - \cancel{m_p^2c^4} + \cancel{m_{\pi^0}^2c^4}.$$

Next (after cancelling as shown) we substitute $K_p = Q - K_{\pi^0}$:

$$\cancel{K_{\pi^0}^2} + 2K_{\pi^0}m_{\pi^0}c^2 = Q^2 - 2QK_{\pi^0} - \cancel{K_{\pi^0}^2} + 2Qm_pc^2 - 2K_{\pi^0}m_pc^2.$$

After cancelling as shown, we solve for K_{π^0}:

$$K_{\pi^0} = \frac{Q^2 + 2Qm_pc^2}{2m_{\pi^0}c^2 + 2Q + 2m_pc^2} = \frac{(116\ \text{MeV})^2 + 2(116\ \text{MeV})(938\ \text{MeV})}{2(135\ \text{MeV}) + 2(116\ \text{MeV}) + 2(938\ \text{MeV})}$$

which gives $K_{\pi^0} = 97$ MeV. Then $K_p = 116\ \text{MeV} - 97\ \text{MeV} = 19\ \text{MeV}$.

43–5 Neutrinos

We first met neutrinos with regard to β decay in Section 41–5. The study of neutrinos is a "hot" subject today. Experiments are being carried out in deep underground laboratories, sometimes in deep mine shafts. The thick layer of earth above is meant to filter out all other "background" particles, leaving mainly the very weakly interacting neutrinos to arrive at the detectors.

Some very important results have come to the fore in recent years. First there was the **solar neutrino problem**. The energy output of the Sun is believed to be due to the nuclear fusion reactions discussed in Chapter 42, Eqs. 42–7 and 42–8.

PSE5_CH43_v1.6_23Jun-20

The neutrinos emitted in these reactions are all ν_e (accompanied by e^+). But the rate at which ν_e arrive at Earth was measured starting in the late 1960s to be much less than expected based on the power output of the Sun. It was then proposed that, on the long trip between Sun and Earth, ν_e might turn into ν_μ or ν_τ. Subsequent experiments, definitive only in 2001, confirmed this hypothesis. Thus the three neutrinos, ν_e, ν_μ, ν_τ, can change into one another in certain circumstances, a phenomenon called **neutrino flavor oscillation**.[†] (Each of the three neutrino types is called, whimsically, a different "flavor.") This result suggests that the lepton numbers L_e, L_μ, and L_τ are not perfectly conserved. But the sum, $L_e + L_\mu + L_\tau$, is believed to be always conserved.

The second exceptional result has long been speculated on: are neutrinos massless as originally thought, or do they have a nonzero mass? Rough upper limits on the masses have been made. The latest astrophysical experiments show that the sum of all three neutrino masses combined is less than about $0.23\ \text{eV}/c^2$. But can all the masses be zero? Not if there are the flavor oscillations discussed above. It seems that at most, one type could have zero mass, and it is likely that at least one neutrino type has a mass of at least $0.04\ \text{eV}/c^2$.

As a result of neutrino oscillations, the three types of neutrino may not be exactly what we thought they were (e, μ, τ). Instead, it seems the three basic neutrinos, called 1, 2, and 3, are combinations of ν_e, ν_μ, and ν_τ.

Another outstanding question is whether or not neutrinos are in the category called **Majorana particles**,[‡] meaning they would be their own antiparticles, like γ, π^0, and Higgs. If so, a lot of other questions (and answers) would appear.

*Neutrino Mass Estimate from a Supernova

The explosion of a supernova in the outer parts of our Galaxy observed in 1987 (Section 44–2) released vast numbers of neutrinos and offered an opportunity to estimate electron neutrino mass. If neutrinos do have mass, then their speed would be less than c, and neutrinos of different energy would take different times to travel the 170,000 light-years from the supernova to Earth. To get an idea of how such a measurement could be done, suppose two neutrinos from "SN1987A" were emitted at the same time and were actually detected on Earth $\left(\text{via the reaction } \bar{\nu}_e + p \rightarrow n + e^+\right)$ 10 seconds apart, with measured kinetic energies of about 20 MeV and 10 MeV. From other laboratory measurements we expect the neutrino mass to be less than 100 eV; and since our neutrinos have kinetic energy of 20 MeV and 10 MeV, we can make the approximation $m_\nu c^2 \ll E$, so that E (the total energy) is essentially equal to the kinetic energy. We use Eq. 36–11b, which tells us

$$E = \frac{m_\nu c^2}{\sqrt{1 - v^2/c^2}}.$$

We solve this for v, the velocity of a neutrino with energy E:

$$v = c\left(1 - \frac{m_\nu^2 c^4}{E^2}\right)^{\frac{1}{2}} = c\left(1 - \frac{m_\nu^2 c^4}{2E^2} + \cdots\right),$$

where we have used the binomial expansion $(1-x)^{\frac{1}{2}} = 1 - \frac{1}{2}x + \cdots$, and we ignore higher-order terms since $m_\nu^2 c^4 \ll E^2$. The time t for a neutrino to travel a distance d $(= 170{,}000\ \text{ly})$ is

$$t = \frac{d}{v} = \frac{d}{c\left(1 - \frac{m_\nu^2 c^4}{2E^2}\right)} \approx \frac{d}{c}\left(1 + \frac{m_\nu^2 c^4}{2E^2}\right),$$

where again we used the binomial expansion $\left[(1-x)^{-1} = 1 + x + \cdots\right]$.

[†]Neutrino oscillations had first been proposed in 1957 by Bruno Pontecorvo. He also proposed that the electron and muon neutrinos are different species; and he also suggested a way to confirm the existence of neutrinos by detecting $\bar{\nu}_e$ emitted in huge numbers by a nuclear reactor, an experiment carried out by Frederick Reines and Clyde Cowan in the 1950s. The experimentalists who confirmed these two predictions were awarded the Nobel Prize, but not the theorist who proposed them.

[‡]The brilliant young physicist Ettore Majorana (1906–1938) disappeared from a ship under mysterious circumstances in 1938 at the age of 31.

The difference in arrival times for our two neutrinos of energies $E_1 = 20\,\text{MeV}$ and $E_2 = 10\,\text{MeV}$ is

$$t_2 - t_1 = \frac{d}{c}\frac{m_\nu^2 c^4}{2}\left(\frac{1}{E_2^2} - \frac{1}{E_1^2}\right).$$

We solve this for $m_\nu c^2$ and set $t_2 - t_1 = 10\,\text{s}$:

$$\begin{aligned} m_\nu c^2 &= \left[\frac{2c(t_2 - t_1)}{d}\frac{E_1^2 E_2^2}{E_1^2 - E_2^2}\right]^{\frac{1}{2}} \\ &= \left[\frac{2(3.0\times10^8\,\text{m/s})(10\,\text{s})}{(1.7\times10^5\,\text{ly})(1.0\times10^{16}\,\text{m/ly})}\frac{(400\,\text{MeV}^2)(100\,\text{MeV}^2)}{(400\,\text{MeV}^2 - 100\,\text{MeV}^2)}\right]^{\frac{1}{2}} \\ &= 22\times10^{-6}\,\text{MeV} = 22\,\text{eV}. \end{aligned}$$

This calculation, with its optimistic assumptions, estimates the mass of the neutrino to be $22\,\text{eV}/c^2$. But there would be experimental uncertainties, and even worse there is the unwarranted assumption that the two neutrinos were emitted at the same time.

Theoretical models of supernova explosions suggest that the neutrinos are emitted in a burst that lasts from a second or two up to perhaps 10 s. If we assume the neutrinos are not emitted simultaneously but rather at any time over a 10-s interval, then that 10-s difference in arrival times could be due to a 10-s difference in their emission time. In this case the data would be consistent with zero neutrino mass, and it put an approximate *upper limit* of $22\,\text{eV}/c^2$.

The actual detection of these neutrinos was brilliant—it was a rare event that allowed us to detect something other than EM radiation from beyond the solar system, and was an exceptional confirmation of theory. In the experiments, the most sensitive detector consisted of several thousand tons of water in an underground chamber. It detected 11 events in 12 seconds, probably via the reaction $\bar{\nu}_e + p \rightarrow n + e^+$. There was not a clear correlation between energy and time of arrival. Nonetheless, a careful analysis of that experiment set a rough upper limit on the electron antineutrino mass of about $4\,\text{eV}/c^2$. More recent results (mentioned earlier in this Section) are much more definitive—they provide evidence that mass is much smaller, and that it is *not zero*—but precise neutrino masses still elude us.

43–6 Particle Classification

In the decades after the discovery of the π meson in the late 1940s, hundreds of other subnuclear particles were discovered. One way to categorize the particles is according to their interactions, since not all particles interact via all four of the forces known in nature (though all interact via gravity). Table 43–2 (next page) lists some of the more common particles classified in this way along with many of their properties. At the top of Table 43–2 are the so-called "fundamental" particles which we believe have no internal structure. Below them are some of the "composite" particles made up of quarks according to the Standard Model.

The **fundamental particles** include quarks and the **gauge bosons** (so-named after the theory that describes them, *gauge theory*) that include the gluons, the photon, and the W and Z particles; these are the particles that mediate (or "carry") the strong, electromagnetic, and weak interactions, respectively.

Also fundamental are the **leptons**, which are particles that do not interact via the strong force but do interact via the weak nuclear force. Leptons that carry electric charge also interact via the electromagnetic force. The leptons include the electron, the muon, and the tau, and three types of neutrino: the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ). Each lepton has an antiparticle. Finally, the recently detected Higgs boson is also considered to be fundamental, with no internal structure.

The second category of particle in Table 43–2 is the **hadrons**, which are **composite** particles (made up of quarks as we will discuss shortly). Hadrons are

PSE5_CH43_v1.6_23Jun-20

TABLE 43–2 Particles (selected)[†]

Category	Forces involved	Particle name	Symbol	Anti-particle	Spin	Mass (MeV/c^2)	B	L_e	L_μ	L_τ	S	Mean life (s)	Principal Decay Modes
							[antiparticles have opposite sign]						
Fundamental													
Gauge bosons (force carriers)	str	Gluons	g	Self	1	0	0	0	0	0	0	Stable	
	em	Photon	γ	Self	1	0	0	0	0	0	0	Stable	
	w, em	W	W^+	W^-	1	80.385×10^3	0	0	0	0	0	3×10^{-25}	$e\nu_e$, $\mu\nu_\mu$, $\tau\nu_\tau$, hadrons
	w	Z	Z^0	Self	1	91.188×10^3	0	0	0	0	0	3×10^{-25}	e^+e^-, $\mu^+\mu^-$, $\tau^+\tau^-$, hadrons
Higgs boson	w, str	Higgs	H^0	Self	0	125.7×10^3	0	0	0	0	0	1.6×10^{-22}	$b\bar{b}$, Z^0Z^0, W^+W^-, $g\bar{g}$, $\tau\bar{\tau}$, $\gamma\gamma$
Leptons	w, em[‡]	Electron	e^-	e^+	$\frac{1}{2}$	0.511	0	+1	0	0	0	Stable	
		Neutrino (e)	ν_e	$\bar{\nu}_e$	$\frac{1}{2}$	0 ($<0.23\,eV/c^2$)[‡]	0	+1	0	0	0	Stable	
		Muon	μ^-	μ^+	$\frac{1}{2}$	105.7	0	0	+1	0	0	2.20×10^{-6}	$e^-\bar{\nu}_e\nu_\mu$
		Neutrino (μ)	ν_μ	$\bar{\nu}_\mu$	$\frac{1}{2}$	0 ($<0.23\,eV/c^2$)[‡]	0	0	+1	0	0	Stable	
		Tau	τ^-	τ^+	$\frac{1}{2}$	1777	0	0	0	+1	0	2.90×10^{-13}	$\mu^-\bar{\nu}_\mu\nu_\tau$, $e^-\bar{\nu}_e\nu_\tau$, hadrons $+\nu_\tau$
		Neutrino (τ)	ν_τ	$\bar{\nu}_\tau$	$\frac{1}{2}$	0 ($<0.23\,eV/c^2$)[‡]	0	0	0	+1	0	Stable	
Quarks	w, em, str	(see Table 43–3)											
Hadrons (composite), selected													
Mesons (quark–antiquark)	str, em, w	Pion	π^+	π^-	0	139.6	0	0	0	0	0	2.60×10^{-8}	$\mu^+\nu_\mu$
			π^0	Self	0	135.0	0	0	0	0	0	0.85×10^{-16}	2γ
		Kaon	K^+	K^-	0	493.7	0	0	0	0	+1	1.24×10^{-8}	$\mu^+\nu_\mu$, $\pi^+\pi^0$
			K^0_S	$\overline{K}^0_S$	0	497.6	0	0	0	0	+1	0.895×10^{-10}	$\pi^+\pi^-$, $2\pi^0$
			K^0_L	$\overline{K}^0_L$	0	497.6	0	0	0	0	+1	5.12×10^{-8}	$\pi^\pm e^\mp\ (\bar{\nu}_e)$, $\pi^\pm\mu^\mp(\bar{\nu}_\mu)$, 3π
		Eta	η^0	Self	0	547.9	0	0	0	0	0	5.1×10^{-19}	2γ, $3\pi^0$, $\pi^+\pi^-\pi^0$
		Rho	ρ^0	Self	1	775	0	0	0	0	0	4.4×10^{-24}	$\pi^+\pi^-$, $2\pi^0$
			ρ^+	ρ^-	1	775	0	0	0	0	0	4.4×10^{-24}	$\pi^+\pi^0$
		and others											
Baryons (3 quarks)	str, em, w	Proton	p	$\bar{p}$	$\frac{1}{2}$	938.3	+1	0	0	0	0	Stable	
		Neutron	n	$\bar{n}$	$\frac{1}{2}$	939.6	+1	0	0	0	0	882	$pe^-\bar{\nu}_e$
		Lambda	Λ^0	$\overline{\Lambda}^0$	$\frac{1}{2}$	1115.7	+1	0	0	0	−1	2.63×10^{-10}	$p\pi^-$, $n\pi^0$
		Sigma	Σ^+	$\overline{\Sigma}^-$	$\frac{1}{2}$	1189.4	+1	0	0	0	−1	0.80×10^{-10}	$p\pi^0$, $n\pi^+$
			Σ^0	$\overline{\Sigma}^0$	$\frac{1}{2}$	1192.6	+1	0	0	0	−1	7.4×10^{-20}	$\Lambda^0\gamma$
			Σ^-	$\overline{\Sigma}^+$	$\frac{1}{2}$	1197.4	+1	0	0	0	−1	1.48×10^{-10}	$n\pi^-$
		Xi	Ξ^0	$\overline{\Xi}^0$	$\frac{1}{2}$	1314.9	+1	0	0	0	−2	2.90×10^{-10}	$\Lambda^0\pi^0$
			Ξ^-	Ξ^+	$\frac{1}{2}$	1321.7	+1	0	0	0	−2	1.64×10^{-10}	$\Lambda^0\pi^-$
		Omega and others	Ω^-	Ω^+	$\frac{3}{2}$	1672.5	+1	0	0	0	−3	0.82×10^{-10}	$\Xi^0\pi^-$, Λ^0K^-, $\Xi^-\pi^0$

[†]See also Table 43–4 for particles with charm and bottom. S in this Table stands for "strangeness" (see Section 43–8). More detail online at: pdglive.lbl.gov.

[‡]Neutrinos partake only in the weak interaction. Experimental upper limits on neutrino masses are given in parentheses, as obtained mainly from the Planck survey (Chapter 44). Detection of neutrino oscillations suggests that at least one type of neutrino has a nonzero mass greater than $0.04\,eV/c^2$.

particles that interact via the strong nuclear force and are said to be **strongly interacting particles**. They also interact via the other forces, but the strong force predominates at short distances. The hadrons include the proton, neutron, pion, and many other particles. They are divided into two subgroups: **baryons**, which are particles that have baryon number +1 (or −1 in the case of their antiparticles) and, as we shall see, are each made up of three quarks; and **mesons**, which have baryon number = 0, and are made up of a quark and an antiquark.

Only a few of the hundreds of hadrons (a veritable "zoo") are included in Table 43–2. Notice that the baryons Λ, Σ, Ξ, and Ω all decay to lighter-mass baryons, and eventually to a proton or neutron. All these processes conserve baryon number. Since there is no particle lighter than the proton with $B = +1$, if baryon number is strictly conserved, the proton itself does not decay and is stable. (But see Section 43–11.) Note that Table 43–2 gives the **mean life** (τ) of each particle (as is done in particle physics), not the half-life ($T_{\frac{1}{2}}$). Recall that they differ by a factor 0.693: $\tau = T_{\frac{1}{2}}/\ln 2 = T_{\frac{1}{2}}/0.693$, Eq. 41–9. The term **lifetime** in particle physics means the mean life τ (= mean lifetime).

The baryon and lepton numbers (B, L_e, L_μ, L_τ), as well as strangeness S (Section 43–8), as given in Table 43–2 are for particles; their antiparticles have opposite sign for these numbers.

PSE5_CH43_v1.6_23Jun-20

EXAMPLE 43–7 **Baryon decay.** Show that the decay modes of the Σ^+ baryon given in Table 43–2 do not violate the conservation laws we have studied up to now: energy, charge, baryon number, lepton numbers.

APPROACH Table 43–2 shows two possible decay modes, (*a*) $\Sigma^+ \rightarrow p + \pi^0$, (*b*) $\Sigma^+ \rightarrow n + \pi^+$. All the particles have lepton numbers equal to zero.

SOLUTION (*a*) Energy: for $\Sigma^+ \rightarrow p + \pi^0$ the change in mass-energy is

$$\begin{aligned}\Delta(Mc^2) &= m_\Sigma c^2 - m_p c^2 - m_{\pi^0} c^2 \\ &= 1189.4\,\text{MeV} - 938.3\,\text{MeV} - 135.0\,\text{MeV} = +116.1\,\text{MeV},\end{aligned}$$

so energy can be conserved with the resulting particles having kinetic energy.
Charge: $+e = +e + 0$, so charge is conserved.
Baryon number: $+1 = +1 + 0$, so baryon number is conserved.
(*b*) Energy: for $\Sigma^+ \rightarrow n + \pi^+$, the mass-energy change is

$$\begin{aligned}\Delta(Mc^2) &= m_\Sigma c^2 - m_n c^2 - m_{\pi^+} c^2 \\ &= 1189.4\,\text{MeV} - 939.6\,\text{MeV} - 139.6\,\text{MeV} = +110.2\,\text{MeV}.\end{aligned}$$

This reaction releases 110.2 MeV of energy as kinetic energy of the products.
Charge: $+e = 0 + e$, so charge is conserved.
Baryon number: $+1 = +1 + 0$, so baryon number is conserved.

43–7 Particle Stability and Resonances

Many particles listed in Table 43–2 are unstable. The lifetime of an unstable particle depends on which force is most active in causing the decay. When a stronger force influences a decay, that decay occurs more quickly. Decays caused by the weak force typically have lifetimes of 10^{-13} s or longer (W and Z decay more quickly). Decays via the electromagnetic force have much shorter lifetimes, typically about 10^{-16} to 10^{-19} s, and normally involve a γ (photon). Most of the unstable particles included in Table 43–2 decay either via the weak or the electromagnetic interaction.

Many particles have been found that decay via the strong interaction, with very short lifetimes, typically about 10^{-23} s. Their lifetimes are so short they do not travel far enough to be detected before decaying. The existence of such short-lived particles is inferred from their decay products. Consider the first such particle discovered (by Fermi), using a beam of π^+ particles with varying amounts of energy directed through a hydrogen target (protons). The number of interactions (π^+ scattered) plotted versus the pion's kinetic energy is shown in Fig. 43–14. The large number of interactions around 200 MeV led Fermi to

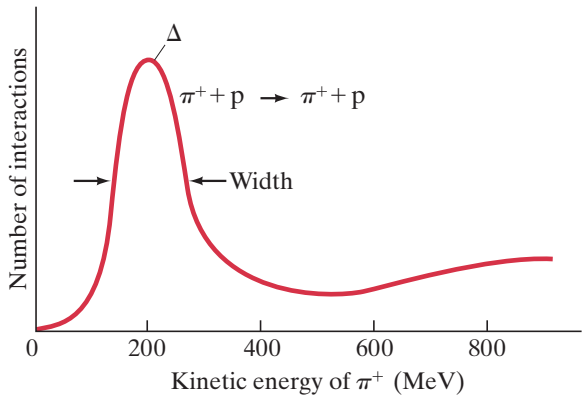


FIGURE 43–14 Number of π^+ particles scattered elastically by a proton target as a function of the incident π^+ kinetic energy. The resonance shape represents the formation of a short-lived particle, the Δ, which has a charge in this case of $+2e$ (Δ^{++}).

conclude that the π^+ and proton combined momentarily to form a short-lived particle before coming apart again, or at least that they resonated together for a short time. Indeed, the large peak in Fig. 43–14 resembles a resonance curve (see Figs. 14–24, 14–27, and 30–22), and this new "particle"—now called the Δ—is referred to as a **resonance**. Hundreds of other resonances have been found, and are regarded as excited states of lighter mass particles such as a nucleon.

The **width** of a resonance—in Fig. 43–14 the full width of the Δ peak at half the maximum is on the order of 100 MeV—is an interesting application of the uncertainty principle. If a particle lives only 10^{-23} s, then its mass (i.e., its rest energy) will be uncertain by an amount

$$\Delta E \approx \frac{h}{2\pi\,\Delta t} \approx \frac{(6.6\times10^{-34}\,\text{J}\cdot\text{s})}{(6)(10^{-23}\,\text{s})} \approx 10^{-11}\,\text{J} \approx 100\,\text{MeV},$$

which is what is observed. Actually, the lifetimes of $\approx 10^{-23}$ s for such resonances are inferred by the reverse process: from the measured width being ≈ 100 MeV.

43–8 Strangeness? Charm? Towards a New Model

In the early 1950s, the newly found particles K, Λ, and Σ were found to behave rather strangely in two ways. First, they were always produced in pairs. For example, the reaction

$$\pi^- + \text{p} \rightarrow \text{K}^0 + \Lambda^0$$

occurred with high probability, but the similar reaction $\pi^- + \text{p} \nrightarrow \text{K}^0 + \text{n}$ was never observed to occur even though it did not violate any known conservation law. The second feature of these **strange particles**, as they came to be called, was that they were produced via the strong interaction (that is, at a high interaction rate), but did not decay at a fast rate characteristic of the strong interaction (even though they decayed into strongly interacting particles).

To explain these observations, a new quantum number, **strangeness**, and a new conservation law, **conservation of strangeness**, were introduced. By assigning the strangeness numbers (S) indicated in Table 43–2, the production of strange particles in pairs was given an explanation. Antiparticles were assigned opposite strangeness from their particles. For example, in the reaction $\pi^- + \text{p} \rightarrow \text{K}^0 + \Lambda^0$, the initial state has strangeness $S = 0 + 0 = 0$, and the final state has $S = +1 - 1 = 0$, so strangeness is conserved. But for $\pi^- + \text{p} \nrightarrow \text{K}^0 + \text{n}$, the initial state has $S = 0$ and the final state has $S = +1 + 0 = +1$, so strangeness would not be conserved; and this reaction is not observed.

To explain the decay of strange particles, it is assumed that strangeness is conserved in the strong interaction but is *not conserved in the weak interaction*. Thus, strange particles were forbidden by strangeness conservation to decay to nonstrange particles of lower mass via the strong interaction, but could decay by means of the weak interaction at the observed longer lifetimes of 10^{-10} to 10^{-8} s.

CAUTION
Partially conserved quantities

The conservation of strangeness was the first example of a **partially conserved** quantity. In this case, the quantity strangeness is conserved by strong interactions but not by weak.

CONCEPTUAL EXAMPLE 43–8 **Guess the missing particle.** Using the conservation laws for particle interactions, determine the possibilities for the missing particle in the reaction

$$\pi^- + \mathrm{p} \rightarrow \mathrm{K}^0 + ?$$

in addition to $\mathrm{K}^0 + \Lambda^0$ mentioned above.

RESPONSE We write equations for the conserved numbers in this reaction, with B, L_e, S, and Q as unknowns whose determination will reveal what the possible particle might be:

Baryon number:	$0 + 1 = 0 + B$
Lepton number:	$0 + 0 = 0 + L_e$
Charge:	$-1 + 1 = 0 + Q$
Strangeness:	$0 + 0 = 1 + S.$

The unknown product particle would have to have these characteristics:

$$B = +1 \qquad L_e = 0 \qquad Q = 0 \qquad S = -1.$$

In addition to Λ^0, a neutral sigma particle, Σ^0, is also consistent with these numbers.

In the next Section we will discuss another partially conserved quantity which was given the name **charm**. The discovery in 1974 of a particle with charm helped solidify a new theory involving quarks, which we now discuss.

43–9 Quarks

One difference between leptons and hadrons is that the hadrons interact via the strong interaction, whereas the leptons do not. There is an even more fundamental difference. The six leptons $(e^-, \mu^-, \tau^-, \nu_e, \nu_\mu, \nu_\tau)$ are considered to be truly fundamental particles because they do not show any internal structure, and have no measurable size. (Attempts to determine the size of leptons have put an upper limit of about 10^{-18} m.) On the other hand, there are hundreds of hadrons, and experiments indicate they do have an internal structure. When an electron collides with another electron, it scatters off as per Coulomb's law. But electrons scattering off a proton reveal a more complex pattern, implying internal parts within the proton (= quarks).

In 1963, M. Gell-Mann and G. Zweig proposed that none of the hadrons, not even the proton and neutron, are truly fundamental, but instead are made up of combinations of three more fundamental pointlike entities called (somewhat whimsically) **quarks**.[†] Today, the quark theory is well-accepted, and quarks are considered truly fundamental particles, like leptons. The three quarks originally proposed were named **up**, **down**, and **strange**, with abbreviations u, d, s. The theory today has six quarks, just as there are six leptons—based on a presumed *symmetry* in nature. The other three quarks are called **charm**, **bottom**, and **top** (c, b, t). The names apply also to new properties of each quark (quantum numbers c, b, t) that distinguish these new quarks from the 3 original quarks (see Table 43–3).

[†]Gell-Mann chose the word from a phrase in James Joyce's *Finnegans Wake*.

TABLE 43–3 Properties of Quarks (Antiquarks have opposite sign *Q, B, S, c, b, t*)

Quarks								
Name	**Symbol**	**Mass (MeV/c^2)**	**Charge *Q***	**Baryon Number *B***	**Strangeness *S***	**Charm *c***	**Bottom *b***	**Top *t***
Up	u	2.3	$+\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	0
Down	d	4.8	$-\frac{1}{3}e$	$\frac{1}{3}$	0	0	0	0
Strange	s	95	$-\frac{1}{3}e$	$\frac{1}{3}$	−1	0	0	0
Charm	c	1275	$+\frac{2}{3}e$	$\frac{1}{3}$	0	+1	0	0
Bottom	b	4180	$-\frac{1}{3}e$	$\frac{1}{3}$	0	0	−1	0
Top†	t	173,500	$+\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	+1

†The top quark, with its extremely short lifetime of 5×10^{-25} s, does not live long enough to form hadrons.

These properties (like strangeness) are conserved in strong, but not weak, interactions. Figure 43–15 shows one of the events that provided evidence for the top quark.

All quarks have spin $\frac{1}{2}$ and an electric charge of either $+\frac{2}{3}e$ or $-\frac{1}{3}e$ (that is, a fraction of the previously thought smallest charge, e, the charge on one electron). Antiquarks have opposite sign of electric charge Q, baryon number B, strangeness S, charm c, bottom b, and top t. Other properties of quarks are shown in Table 43–3.

All hadrons are considered to be made up of combinations of quarks (plus the gluons that hold them together), and their properties are described by looking at their quark content. Mesons consist of a quark–antiquark pair. For example, a π^+ meson is a $u\bar{d}$ combination: note that for the $u\bar{d}$ pair (Table 43–3), $Q = \frac{2}{3}e + \frac{1}{3}e = +1e$, $B = \frac{1}{3} - \frac{1}{3} = 0$, $S = 0 + 0 = 0$, as they must for a π^+; and a $K^+ = u\bar{s}$, with $Q = +1$, $B = 0$, $S = +1$. A π^0 can be made of $u\bar{u}$ or $d\bar{d}$, or a combination of the two.

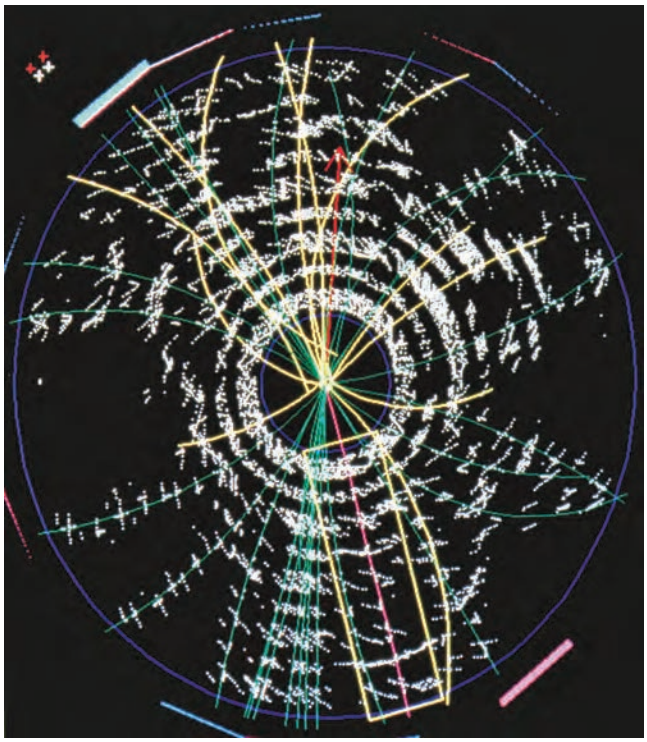
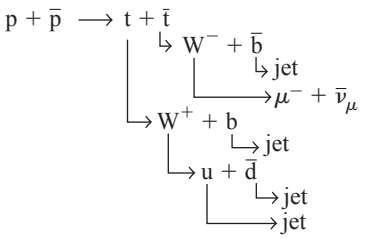


FIGURE 43–15 This computer-generated reconstruction of a proton–antiproton collision at Fermilab occurred at an energy of nearly 2 TeV. It is one of the events that provided evidence for the top quark (1995). The multiwire chamber (Section 41–11) is in a magnetic field, and the radius of curvature of the charged particle tracks is a measure of each particle's momentum (Chapter 27). The white dots represent signals seen on the electric wires of the multiwire chamber. The colored lines are particle paths. The top quark (t) has too brief a lifetime ($\approx 10^{-23}$ s) to be detected itself, so we look for its possible decay products. Analysis indicates the following interaction and subsequent decays:



The tracks in the photo include jets (groups of particles moving in roughly the same direction), and a muon (μ^-) whose track is the pink one enclosed by a yellow rectangle to make it stand out.

TABLE 43–4 Partial List of Heavy Hadrons, with Charm and Bottom ($L_e = L_\mu = L_\tau = 0$)

Category	Particle	Anti-particle	Spin	Mass (MeV/c^2)	Baryon Number B	Strangeness S	Charm c	Bottom b	Mean life (s)	Principal Decay Modes
Mesons	D^+	D^-	0	1869.6	0	0	+1	0	10.4×10^{-13}	K + others, e + others
	D^0	$\overline{D}^0$	0	1864.8	0	0	+1	0	4.1×10^{-13}	K + others, μ or e + others
	D_S^+	D_S^-	0	1968.3	0	+1	+1	0	5.0×10^{-13}	K + others
	J/ψ (3097)	Self	1	3096.9	0	0	0	0	0.71×10^{-20}	Hadrons, e^+e^-, $\mu^+\mu^-$
	Y (9460)	Self	1	9460.3	0	0	0	0	1.2×10^{-20}	Hadrons, $\mu^+\mu^-$, e^+e^-, $\tau^+\tau^-$
	B^-	B^+	0	5279.3	0	0	0	−1	1.6×10^{-12}	D^0 + others
	B^0	$\overline{B}^0$	0	5279.6	0	0	0	−1	1.5×10^{-12}	D^0 + others
Baryons	Λ_c^+	Λ_c^-	$\frac{1}{2}$	2286.5	+1	0	+1	0	2.0×10^{-13}	Hadrons (e.g., Λ + others)
	Σ_c^{++}	Σ_c^{--}	$\frac{1}{2}$	2454	+1	0	+1	0	2.9×10^{-22}	$\Lambda_c^+\pi^+$
	Σ_c^+	Σ_c^-	$\frac{1}{2}$	2453	+1	0	+1	0	$>1.4 \times 10^{-22}$	$\Lambda_c^+\pi^0$
	Σ_c^0	$\overline{\Sigma}_c^0$	$\frac{1}{2}$	2454	+1	0	+1	0	3.0×10^{-22}	$\Lambda_c^+\pi^-$
	Λ_b^0	$\overline{\Lambda}_b^0$	$\frac{1}{2}$	5619.5	+1	0	0	−1	1.5×10^{-12}	$J/\psi\Lambda^0$, $pD^0\pi^-$, $\Lambda_c^+\pi^+\pi^-\pi^-$

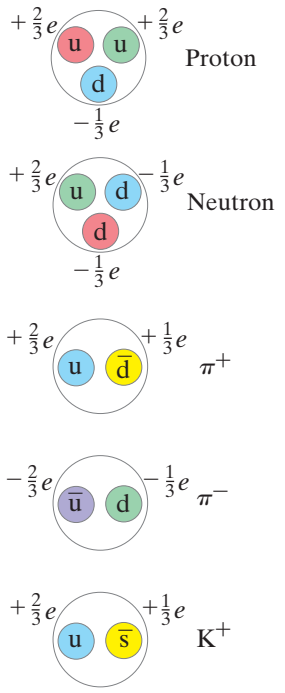


FIGURE 43–16 Quark compositions for several particles.

Baryons, on the other hand, consist of three quarks. For example, a neutron is $n = ddu$, whereas an antiproton is $\bar{p} = \bar{u}\,\bar{u}\,\bar{d}$. See Fig. 43–16. Strange particles all contain an s or $\bar{s}$ quark, whereas charm particles contain a c or $\bar{c}$ quark. A few of these hadrons are listed in Table 43–4.

Current models suggest that quarks may be so tightly bound together that they may not ever exist singly in the free state. But quarks can be detected indirectly when they turn into narrow **jets** of other particles, as in Fig. 43–15. Also, observations of very high energy electrons scattered off protons suggest that protons are indeed made up of constituents.

Today, the truly **fundamental particles** are considered to be the six quarks, the six leptons, the gauge bosons that carry the fundamental forces, and the Higgs (page 1320). See Table 43–5, where the quarks and leptons are arranged in three "families" or "generations." Ordinary matter—atoms made of protons, neutrons, and electrons—is contained in the "first generation." The others are thought to have existed in the very early universe, but are seen by us today only at powerful colliders or in cosmic rays. All of the hundreds of hadrons can be accounted for by combinations of the six quarks and six antiquarks.

EXERCISE C Return to the Chapter-Opening Questions, page 1289, and answer them again now. Try to explain why you may have answered differently the first time.

TABLE 43–5 The Fundamental Particles† of the Standard Model

Gauge bosons	γ	g	Z^0	$W^\pm$
Higgs boson	H^0			
	Quarks		and Leptons	
First generation	u	d	e	ν_e
Second generation	s	c	μ	ν_μ
Third generation	b	t	τ	ν_τ

†The graviton (G^0) has been hypothesized but not detected and may not be detectable. It is not part of the Standard Model.

CONCEPTUAL EXAMPLE 43–9 **Quark combinations.** Find the baryon number, charge, and strangeness for the following quark combinations, and identify the hadron particle that is made up of these quark combinations: (*a*) udd, (*b*) $u\bar{u}$, (*c*) uss, (*d*) sdd, and (*e*) $b\bar{u}$.

RESPONSE We use Table 43–3 to get the properties of the quarks, then Table 43–2 or 43–4 to find the particle that has these properties.

(*a*) udd has

$$Q = +\tfrac{2}{3}e - \tfrac{1}{3}e - \tfrac{1}{3}e = 0,$$

$$B = \tfrac{1}{3} + \tfrac{1}{3} + \tfrac{1}{3} = 1,$$

$$S = 0 + 0 + 0 = 0,$$

as well as $c = 0$, bottom $= 0$, top $= 0$. The only baryon $(B = +1)$ that has $Q = 0$, $S = 0$, etc., is the neutron (Table 43–2).

(*b*) $u\bar{u}$ has $Q = \frac{2}{3}e - \frac{2}{3}e = 0$, $B = \frac{1}{3} - \frac{1}{3} = 0$, and all other quantum numbers $= 0$. Sounds like a π^0 ($d\bar{d}$ also gives a π^0).

(*c*) uss has $Q = 0$, $B = +1$, $S = -2$, others $= 0$. This is a Ξ^0.

(*d*) sdd has $Q = -1$, $B = +1$, $S = -1$, so must be a Σ^-.

(*e*) $b\bar{u}$ has $Q = -1$, $B = 0$, $S = 0$, $c = 0$, bottom $= -1$, top $= 0$. This must be a B^- meson (Table 43–4).

EXERCISE D What is the quark composition of a K^- meson?

43–10 The Standard Model: QCD and Electroweak Theory

Not long after the quark theory was proposed, it was suggested that quarks have another property (or quality) called **color**, or "color charge" (analogous to electric charge). The distinction between the six types of quark (u, d, s, c, b, t) was referred to as **flavor**. According to theory, each flavor of quark can have one of three colors, usually designated red, green, and blue. (These are the three primary colors which, when added together in appropriate amounts, as on a TV screen, produce white.) Note that the names "color" and "flavor" have nothing to do with our senses, but are purely whimsical—as are other names, such as charm, in this new field. (We did, however, "color" the quarks in Fig. 43–16.) The antiquarks are colored antired, antigreen, and antiblue. Baryons are made up of three quarks, one of each color. Mesons consist of a quark–antiquark pair of a particular color and its anticolor. Both baryons and mesons are thus colorless or white.

Originally, the idea of quark color was proposed to preserve the Pauli exclusion principle (Section 39–4). Not all particles obey the exclusion principle. Those that do, such as electrons, protons, and neutrons, are called **fermions**. Those that don't are called **bosons**. These two categories are distinguished also in their spin:† bosons have integer spin (0, 1, 2, etc.) whereas fermions have half-integer spin, usually $\frac{1}{2}$ as for electrons and nucleons, but other fermions have spin $\frac{3}{2}$, $\frac{5}{2}$, etc. Matter is made up mainly of fermions, but the carriers of the forces (γ, W, Z, and gluons) are all bosons. Quarks are fermions (they have spin $\frac{1}{2}$) and therefore should obey the exclusion principle. Yet for three particular baryons (uuu, ddd, and sss), all three quarks would have the same quantum numbers, and at least two quarks have their spin in the same direction (since there are only two choices, spin up $[m_s = +\frac{1}{2}]$ or spin down $[m_s = -\frac{1}{2}]$).

†Spin is written in units of $\hbar$: $1\hbar$, $2\hbar$, etc; $\frac{1}{2}\hbar$, $\frac{3}{2}\hbar$.

This would seem to violate the exclusion principle; but if quarks have that additional quantum number *color*, which is different for each quark, it would serve to distinguish them and allow the exclusion principle to hold. Although quark color, and the resulting threefold increase in the number of quarks, was originally an *ad hoc* idea, it also served to bring the theory into better agreement with experiment, such as predicting the correct lifetime of the π^0 meson, and the measured rate of hadron production in observed e^+e^- collisions at accelerators. The idea of color soon became a central feature of the theory as determining the force binding quarks together in a hadron.

Each quark is assumed to carry a *color charge*, analogous to electric charge, and the strong force between quarks is referred to as the **color force**. This theory of the strong force is called **quantum chromodynamics** (*chroma* = color in Greek), or **QCD**, to indicate that the force acts between color charges (and not between, say, electric charges). The strong force between two hadrons is considered to be a force between the quarks that make them up, as suggested in Fig. 43–17.

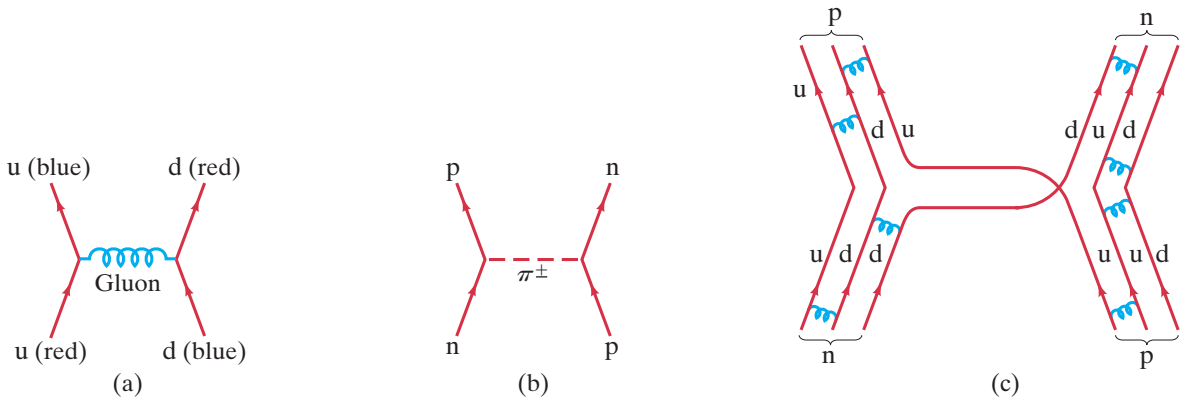


FIGURE 43–17 (a) The force between two quarks holding them together as part of a proton, for example, is carried by a gluon, which in this case involves a change in color. (b) Strong interaction $n + p \rightarrow n + p$ with the exchange of a charged π meson (+ or −, depending on whether it is considered moving to the left or to the right). (c) Quark representation of the same interaction $n + p \rightarrow n + p$. The blue coiled lines between quarks represent gluon exchanges holding the hadrons together. (The exchanged meson may be regarded as $\bar{u}d$ emitted by the n and absorbed by the p, or as $u\bar{d}$ emitted by p and absorbed by n, because a u (or d) quark going to the left in the diagram is equivalent to a $\bar{u}$ (or $\bar{d}$) going to the right.)

The particles that transmit the color force (analogous to photons for the EM force) are called **gluons** (a play on "glue"). They are included in Tables 43–2 and 43–5. There are eight gluons, according to the theory, all massless and all have color charge.[†]

You might ask what would happen if we try to see a single quark with color by reaching deep inside a hadron and extracting a single quark. Quarks are so tightly bound to other quarks that extracting one would require a tremendous amount of energy, so much that it would be sufficient to create more quarks $(E = mc^2)$. Indeed, such experiments are done at modern particle colliders and all we get is more hadrons (quark–antiquark pairs, or triplets, which we observe as mesons or baryons), never an isolated quark. This property of quarks, that they are always bound in groups that are colorless, is called **confinement**.

The color force has the interesting property that, as two quarks approach each other very closely (equivalently, have high energy), the force between them becomes small. This aspect is referred to as **asymptotic freedom**.

[†]Compare to the EM interaction, where the photon has no electric charge. Because gluons have color charge, they could attract each other and form composite particles (photons cannot). Such "glueballs" are being searched for.

PSE5_CH43_v1.6_23Jun-20

The weak force, as we have seen, is assumed to be mediated by the W^+, W^-, and Z^0 particles. It acts between the "weak charges" that each particle has. Each fundamental particle can thus have electric charge, weak charge, color charge, and gravitational mass, although one or more of these could be zero. For example, all leptons have color charge of zero, so they do not interact via the strong force.

CONCEPTUAL EXAMPLE 43–10 **Beta decay.** Draw a Feynman diagram, showing what happens in beta decay using quarks.

RESPONSE Beta decay is a result of the weak interaction, and the mediator is either a $W^\pm$ or Z^0 particle. What happens, in part, is that a neutron (udd quarks) decays into a proton (uud). Apparently a d quark (charge $-\frac{1}{3}e$) has turned into a u quark (charge $+\frac{2}{3}e$). Charge conservation means that a negatively charged particle, namely a W^-, was emitted by the d quark. Since an electron and an antineutrino appear in the final state, they must have come from the decay of the virtual W^-, as shown in Fig. 43–18.

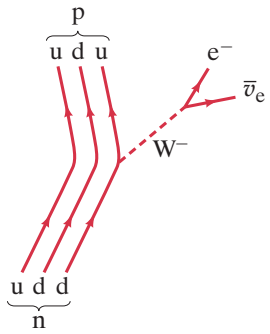


FIGURE 43–18 Quark representation of the Feynman diagram for β decay of a neutron into a proton. Example 43–10.

To summarize, the Standard Model says that the truly fundamental particles (Table 43–5) are the leptons, the quarks, the gauge bosons (photon, W and Z, and the gluons), and the Higgs boson. The photon, leptons, W^+, W^-, and Z^0 have all been observed in experiments. But only combinations of quarks (baryons and mesons) have been observed in the free state, and it seems likely that free quarks and gluons cannot be observed in isolation.

One important aspect of theoretical work is the attempt to find a **unified** basis for the different forces in nature. This was a long-held hope of Einstein, which he was never able to fulfill. A so-called **gauge theory** that unifies the weak and electromagnetic interactions was put forward in the 1960s by S. Weinberg, S. Glashow, and A. Salam. In this **electroweak theory**, the weak and electromagnetic forces are seen as two different manifestations of a single, more fundamental, *electroweak* interaction. The electroweak theory has had many successes, including the prediction of the $W^\pm$ particles as carriers of the weak force, with masses of $80.38 \pm 0.02\,\text{GeV}/c^2$ in excellent agreement with the measured values of $80.385 \pm 0.015\,\text{GeV}/c^2$ (and similar accuracy for the Z^0).

The combination of electroweak theory plus QCD for the strong interaction is referred to today as the **Standard Model** (**SM**) of fundamental particles.

EXAMPLE 43–11 **ESTIMATE** **Range of weak force.** The weak nuclear force is of very short range, meaning it acts over only a very short distance. Estimate its range using the masses (Table 43–2) of the $W^\pm$ and Z: $m \approx 80$ or $90\,\text{GeV}/c^2 \approx 10^2\,\text{GeV}/c^2$.

APPROACH We assume the $W^\pm$ or Z^0 exchange particles can exist for a time Δt given by the uncertainty principle (Section 38–3), $\Delta t \approx \hbar/\Delta E$, where $\Delta E \approx mc^2$ is the energy needed to create the virtual particle $(W^\pm, Z)$ that carries the weak force.

SOLUTION Let Δx be the distance the virtual W or Z can move before it must be reabsorbed within the time $\Delta t \approx \hbar/\Delta E$. To find an upper limit on Δx, and hence the maximum range of the weak force, we let the W or Z travel close to the speed of light, so $\Delta x \lesssim c\,\Delta t$. Recalling that $1\,\text{GeV} = 1.6 \times 10^{-10}\,\text{J}$, then

$$\Delta x \lesssim c\,\Delta t \approx \frac{c\hbar}{\Delta E} \approx \frac{(3 \times 10^8\,\text{m/s})(10^{-34}\,\text{J}\cdot\text{s})}{(10^2\,\text{GeV})(1.6 \times 10^{-10}\,\text{J/GeV})} \approx 10^{-18}\,\text{m}.$$

This is indeed a very small range.

NOTE Compare this to the range of the electromagnetic force whose range is infinite ($1/r^2$ never becomes zero for any finite r), which makes sense because the mass of its virtual exchange particle, the photon, is zero (in the denominator of the above equation).

We did a similar calculation for the strong force in Section 43–2, estimating the mass of the π meson as exchange particle. In our deeper view of the strong force, namely the color force between quarks within a nucleon, the gluons have zero mass, which implies infinite range (see formula in Example 43–11). We might have expected a range of about 10^{-15} m (nuclear size). But according to the Standard Model, the color force is weak at very close distances and increases greatly with distance (causing quark confinement). Thus its range could be very large.

Theoreticians have wondered why the W and Z have large masses rather than being massless like the photon. Peter Higgs and others in 1964 used electroweak theory to suggest an explanation by means of a **Higgs field** and its particle, the **Higgs boson**, which interact with the W and Z to "slow them down." In being forced to go slower than the speed of light, they would have to have mass ($m = 0$ only if $v = c$). Indeed, the Higgs field is thought to permeate the vacuum ("empty space") and to perhaps confer mass on particles that now have mass by slowing them down. In 2012 strong evidence was announced at CERN's Large Hadron Collider (Section 43–1) for a particle of mass 125 GeV/c^2 that is thought to be the long-sought Higgs boson of the Standard Model. Intense research continues, not only to better understand this particle, but to search for additional Higgs-like particles suggested by theories that go beyond the Standard Model such as *supersymmetry* (Section 43–12) and Majorana neutrinos.

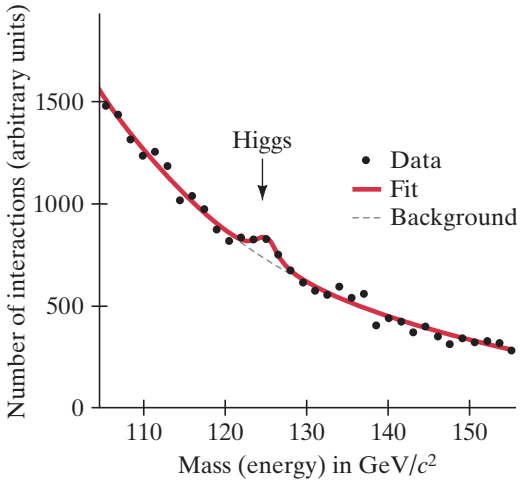


FIGURE 43–19 Evidence for the Higgs boson.

FIGURE 43–20 Fabiola Gianotti, leader of the ATLAS team (3000 physicists), at the LHC with theorist Peter Higgs, July 4, 2012, when the long hoped-for boson was announced.



Figure 43–19 shows the "resonance" bump (Section 43–7) that represents the Higgs boson as detected by the CMS team at the LHC. A second experiment, ATLAS, came up with the same mass. ATLAS is considered the largest scientific experiment ever (see Fig. 43–20).

There is no way to know if the Chapter-Opening Photo of a possible Higgs event, page 1289, is actually a Higgs or is a background event. As can be seen in Fig. 43–19, there are many more background events around 125 GeV than there are in the resonance bump representing the Higgs boson.

43–11 Grand Unified Theories

The Standard Model, for all its success, cannot explain some important issues—such as why the charge on the electron has *exactly* the same magnitude as the charge on the proton. This is crucial, because if the charge magnitudes were even a little different, atoms would not be neutral and the resulting large electric forces would surely have made life impossible. Indeed, the Standard Model is now considered to be a low-energy approximation to a more complete theory.

With the success of unified electroweak theory, theorists are trying to incorporate it and QCD for the strong (color) force into a so-called **grand unified theory** (**GUT**).

PSE5_CH43_v1.6_23Jun-20

One type of such a grand unified theory of the electromagnetic, weak, and strong forces has been proposed in which there is only one class of particle—leptons and quarks belong to the same family and are able to change freely from one type to the other—and the three forces are different aspects of a single underlying force. The unity is predicted to occur, however, only on a scale of less than about 10^{-31} m, corresponding to a typical particle energy of about 10^{16} GeV. If two elementary particles (leptons or quarks) approach each other to within this **unification scale**, the apparently fundamental distinction between them would not exist at this level, and a quark could readily change to a lepton, or vice versa. Baryon and lepton numbers would not be conserved. The weak, electromagnetic, and strong (color) force would blend to a force of a single strength.

What happens between the unification distance of 10^{-31} m and more normal (larger) distances is referred to as **symmetry breaking**. As an analogy, consider an atom in a crystal. Deep within the atom, there is much symmetry—in the innermost regions the electron cloud is spherically symmetric (Chapter 39). Farther out, this symmetry breaks down—the electron clouds are distributed preferentially along the lines (bonds) joining the atoms in the crystal. In a similar way, at 10^{-31} m the force between elementary particles is theorized to be a single force—it is symmetrical and does not single out one type of "charge" over another. But at larger distances, that symmetry is broken and we see three distinct forces. (In the "Standard Model" of electroweak interactions, Section 43–10, the symmetry breaking between the electromagnetic and the weak interactions occurs at about 10^{-18} m.)

CONCEPTUAL EXAMPLE 43–12 **Symmetry.** The table in Fig. 43–21 has four identical place settings. Four people sit down to eat. Describe the symmetry of this table and what happens to it when someone starts the meal.

RESPONSE The table has several kinds of symmetry. It is symmetric to rotations of 90°: that is, the table will look the same if everyone moved one chair to the left or to the right. It is also north–south symmetric and east–west symmetric, so that swaps across the table don't affect the way the table looks. It also doesn't matter whether any person picks up the fork to the left of the plate or the fork to the right. But once that first person picks up either fork, the choice is set for all the rest at the table as well. The symmetry has been *broken*. The underlying symmetry is still there—the blue glasses could still be chosen either way—but some choice must get made and at that moment the symmetry of the diners is broken.

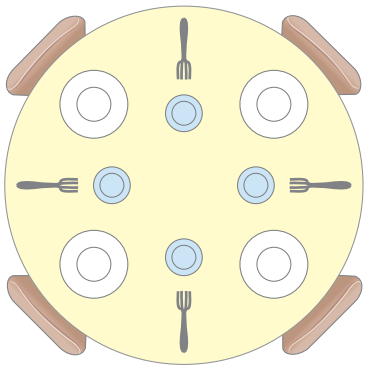


FIGURE 43–21 Symmetry around a table. Example 43–12.

Another example of symmetry breaking is a pencil standing on its point before falling. Standing, it looks the same from any horizontal direction. From above, it is a tiny circle. But when it falls to the table, it points in one particular direction—the symmetry is broken.

Proton Decay

Since unification is thought to occur at such tiny distances and huge energies, the theory is difficult to test experimentally. But it is not completely impossible. One testable prediction is the idea that the proton might decay (via, for example, $p \rightarrow \pi^0 + e^+$) and violate conservation of baryon number. This could happen if two quarks within a proton approached to within 10^{-31} m of each other. But it is very unlikely at normal temperature and energy, so the decay of a proton can only be an unlikely process. In the simplest form of GUT, the theoretical estimate of the proton mean life for the decay mode $p \rightarrow \pi^0 + e^+$ is about 10^{31} yr, and is within the realm of testability.[†] Proton decays have still not been seen, and experiments put the lower limit on the proton mean life for the above mode to be about 10^{33} yr, somewhat greater than this prediction. This may seem a disappointment, but on the other hand, it presents a challenge. Indeed more complex GUTs may resolve this conflict.

[†]This is much larger than the age of the universe ($\approx 14 \times 10^9$ yr). But we don't have to wait 10^{31} yr to see. Instead we can wait for one decay among 10^{31} protons over a year (see Eqs. 41–4 and 41–9a, $\Delta N = \lambda N\, \Delta t = N\, \Delta t/\tau$, and Example 43–13).

EXAMPLE 43–13 ESTIMATE **Proton decay.** An experiment uses 3300 tons of water waiting to see a proton decay of the type $p \rightarrow \pi^0 + e^+$. If the experiment is run for 4 years without detecting a decay, estimate the lower limit on the proton mean life.

APPROACH As with radioactive decay, the number of decays is proportional to the number of parent species (N), the time interval (Δt), and the decay constant (λ) which is related to the mean life τ by (see Eqs. 41–4 and 41–9a):

$$\Delta N = -\lambda N \, \Delta t = -\frac{N \, \Delta t}{\tau}.$$

SOLUTION Dealing only with magnitudes, we solve for τ:

$$\tau = \frac{N \, \Delta t}{\Delta N}.$$

Thus for $\Delta N < 1$ (we don't see even one decay) over the four-year trial,

$$\tau > N\,(4\,\text{yr}),$$

where N is the number of protons in 3300 tons of water. To determine N, we note that each molecule of H_2O contains $2 + 8 = 10$ protons. So one mole of water (18 g, 6×10^{23} molecules) contains $10 \times 6 \times 10^{23} = 6 \times 10^{24}$ protons in 18 g of water ($= 18\,\text{g}/1000\,\text{g} = 1/56$ of a kg), or about 3×10^{26} protons per kilogram. One ton is 10^3 kg, so the 3300 tons contains $(3.3 \times 10^6\,\text{kg})(3 \times 10^{26}\,\text{protons/kg}) \approx 1 \times 10^{33}$ protons. Then our very rough estimate for a lower limit on the proton mean life is $\tau > (10^{33})(4\,\text{yr}) \approx 4 \times 10^{33}\,\text{yr}$.

*GUT and Cosmology

An interesting prediction of unified theories relates to cosmology (Chapter 44). It was thought by many theorists that during the first 10^{-35} s after the theorized Big Bang that created the universe, the temperature was so extremely high that particles had energies corresponding to the unification scale. Baryon number would not have been conserved then, perhaps allowing an imbalance that might account for the observed predominance of matter ($B > 0$) over antimatter ($B < 0$) in the universe. The fact that we are surrounded by matter, with no significant antimatter in sight, is considered a problem in search of an explanation (not given by the Standard Model). We call this the **matter–antimatter problem**. To understand it may require still undiscovered phenomena—perhaps related to quarks or neutrinos, or the Higgs boson or supersymmetry (next Section).

Many theorists no longer think the Big Bang was sufficiently hot to create unification. Nonetheless we see that there is a deep connection between investigations at either end of the size scale: theories about the tiniest objects (elementary particles) have a strong bearing on the understanding of the universe on a large scale. We look at this more in the next Chapter.

Figure 43–22 is a rough diagram indicating how the four fundamental forces in nature might have "condensed out" (a symmetry was broken) as time went on after the Big Bang (Chapter 44), and as the mean temperature of the universe and the typical particle energy decreased.

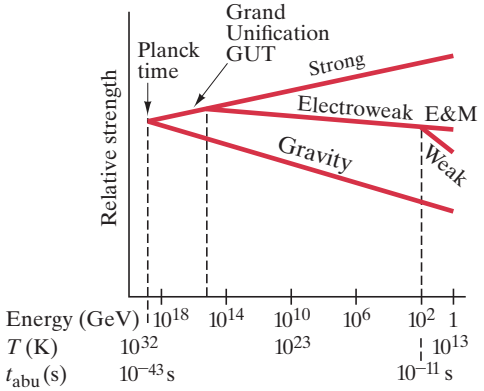


FIGURE 43–22 Time and energy plot of the four fundamental forces, perhaps unified at the "Planck time" (10^{-43} s after the birth of the universe), and how each condensed out, assuming a very hot Big Bang. The symbol t_{abu} means time after the birth of the universe. Note that the typical particle energy (and average temperature of the universe) decreases to the right, as time increases.

43–12 Strings and Supersymmetry

We have seen that the Standard Model is unable to address important experimental issues, and that theoreticians are attacking the problem as experimenters search for new data, new particles, new concepts.

Even more ambitious than grand unified theories are attempts to also incorporate gravity, and thus unify all four forces in nature into a single theory. (Such theories are sometimes referred to misleadingly as **theories of everything**.) A major attempt to unify all four forces is called **string theory**, introduced by Gabriele Veneziano in 1968: Each fundamental particle (Table 43–5) is imagined not as a point but as a one-dimensional **string**, perhaps 10^{-35} m long, which vibrates in a particular standing wave pattern. (You might say each particle is a different note on a tiny stretched string.) More sophisticated theories propose the fundamental entities as being multidimensional **branes** (after 2-D membranes).

A related idea that also goes way beyond the Standard Model is **supersymmetry**, which applied to strings is known as **superstring theory**. Supersymmetry, developed by Bruno Zumino (1923–2014) and Julius Wess (1934–2007), predicts that interactions exist that would change fermions into bosons and vice versa, and that each known fermion would have a supersymmetric boson partner. Thus, for each quark (a fermion), there would be a **squark** (a boson) or "supersymmetric" quark. For every lepton there would be a **slepton**. Likewise, for every known boson (photons and gluons, for example), there would be a supersymmetric fermion (**photinos** and **gluinos**). Supersymmetry predicts also that a *graviton*, which transmits the gravity force, has a partner, the **gravitino**. Supersymmetry (often abbreviated SUSY) offers solutions to a number of important theoretical problems. Supersymmetric particles are a candidate for the "dark matter" of the universe (discussed in Chapter 44). But why hasn't this "missing part" of the universe ever been detected? The best guess is that supersymmetric particles might be heavier than their conventional counterparts, perhaps too heavy to have been produced in today's accelerators. A search for supersymmetric particles is already being done at CERN's new Large Hadron Collider.

Versions of these theories predict other interesting properties, such as that space has 11 dimensions, but 7 of them are "coiled up" so we normally only notice the 4-D of space–time. We would like to know if and how many extra dimensions there are, and how and why they are hidden. We hope to have some answers from the LHC (Section 43–1).

Some theorists think SUSY and other theories are approximations to a more fundamental, still undiscovered, **M-theory**. Edward Witten coined the term when proposing an 11 dimensional approximation, but never said what "M" stands for (could it be W upside-down?).

The world of elementary particles is opening new vistas. What happens in the future is bound to be exciting.

Summary

Particle accelerators are used to accelerate charged particles, such as electrons and protons, so they can have very high energy collisions with other particles. High-energy particles have short wavelength and so can be used to probe the structure of matter in great detail (very small distances). High kinetic energy also allows the creation of new particles through collisions (via $E = mc^2$).

Cyclotrons and **synchrotrons** use a magnetic field to keep the particles in a circular path and accelerate them at intervals by high voltage. **Linear accelerators** accelerate particles along a line. **Colliding beams** allow higher interaction energy.

An **antiparticle** has the same mass as a particle but opposite charge. Certain other properties are also opposite: for example, the antiproton has **baryon number** (nucleon number) opposite ($B = -1$) to that for the proton ($B = +1$).

In all nuclear and particle reactions, the following conservation laws hold: momentum, angular momentum, mass–energy, electric charge, baryon number, and **lepton numbers**.

Certain particles have a property called **strangeness**, which is conserved by the strong force but not by the weak force. The properties **charm**, **bottom**, and **top** also are conserved by the strong force but not by the weak force.

Just as the electromagnetic force can be said to be due to an exchange of photons, the strong nuclear force is carried by massless **gluons**. The W and Z particles carry the weak force. These fundamental force carriers (photon, W and Z, gluons) are called **gauge bosons**.

Other particles can be classified as either *leptons* or *hadrons*. **Leptons** participate only in gravity, the weak, and the electromagnetic interactions. **Hadrons**, which today are considered **composite** particles, are made up of **quarks**, and participate in all four interactions, including the strong interaction. The hadrons can be classified as: **mesons**, each made up of 2 quarks and having baryon number zero; and **baryons**, each made up of 3 quarks and having nonzero baryon number.

Most particles, except for the photon, electron, neutrinos, and proton, decay with measurable mean lives varying from 10^{-25} s to 10^{3} s. The mean life depends on which force is predominant. Weak decays usually have mean lives greater than about 10^{-13} s. Electromagnetic decays typically have mean lives on the order of 10^{-16} to 10^{-19} s. The shortest lived particles, called **resonances**, decay via the strong interaction and live typically for only about 10^{-23} s.

Today's **Standard Model** of elementary particles considers **quarks** as the basic building blocks of the hadrons. The six quark "flavors" are called **up**, **down**, **strange**, **charm**, **bottom**, and **top**. It is expected that there are the same number of quarks as leptons (six of each), and that quarks and leptons are the truly fundamental particles along with the gauge bosons (γ, W, Z, gluons) and the Higgs boson.

Quarks are said to have **color**, and, according to **quantum chromodynamics** (QCD), the strong color force acts between their color charges and is transmitted by **gluons**. **Electroweak theory** views the weak and electromagnetic forces as two aspects of a single underlying interaction. QCD plus the electroweak theory are referred to as the **Standard Model** of the fundamental particles.

Grand unified theories of forces suggest that at very short distance $(10^{-31}\,\text{m})$ and very high energy, the weak, electromagnetic, and strong forces would appear as a single force, and the fundamental difference between quarks and leptons would disappear.

According to **string theory**, the fundamental particles may be tiny strings, 10^{-35} m long, distinguished by their standing wave pattern. **Supersymmetry** predicts that each fermion (or boson) has a corresponding boson (or fermion) partner.

Questions

1. Give a reaction between two nucleons, similar to Eq. 43–4, that could produce a π^-.
2. If a proton is moving at very high speed, so that its kinetic energy is much greater than its rest energy (mc^2), can it then decay via $\text{p} \rightarrow \text{n} + \pi^+$? Why or why not?
3. What would an "antiatom," made up of the antiparticles to the constituents of normal atoms, consist of? What might happen if *antimatter*, made of such antiatoms, came in contact with our normal world of matter?
4. What particle in a decay signals the electromagnetic interaction?
5. (*a*) Does the presence of a neutrino among the decay products of a particle necessarily mean that the decay occurs via the weak interaction? (*b*) Do all decays via the weak interaction produce a neutrino? Explain.
6. Why is it that a neutron decays via the weak interaction even though the neutron and one of its decay products (proton) are strongly interacting?
7. Which of the four interactions (strong, electromagnetic, weak, gravitational) does an electron take part in? A neutrino? A proton?
8. Verify that charge and baryon number are conserved in each of the decays shown in Table 43–2.
9. Which of the particle decays listed in Table 43–2 occur via the electromagnetic interaction?
10. Which of the particle decays listed in Table 43–2 occur by the weak interaction?
11. The Δ baryon has spin $\frac{3}{2}$, baryon number 1, and charge $Q = +2,\ +1,\ 0,$ or -1. Why is there no charge state $Q = -2$?
12. Which of the particle decays in Table 43–4 occur via the electromagnetic interaction?
13. Which of the particle decays in Table 43–4 occur by the weak interaction?
14. Quarks have spin $\frac{1}{2}$. How do you account for the fact that baryons have spin $\frac{1}{2}$ or $\frac{3}{2}$, and mesons have spin 0 or 1?
15. Suppose there were a kind of "neutrinolet" that was massless, had no color charge or electrical charge, and did not feel the weak force. Could you say that this particle even exists?
16. Is it possible for a particle to be both (*a*) a lepton and a baryon? (*b*) a baryon and a hadron? (*c*) a meson and a quark? (*d*) a hadron and a lepton? Explain.
17. Using the ideas of quantum chromodynamics, would it be possible to find particles made up of two quarks and no antiquarks? What about two quarks and two antiquarks?
18. Why can neutrons decay when they are free, but not when they are inside a stable nucleus?
19. Is the reaction $\text{e}^- + \text{p} \rightarrow \text{n} + \bar{\nu}_\text{e}$ possible? Explain.
20. Occasionally, the Λ will decay by the following reaction: $\Lambda^0 \rightarrow \text{p}^+ + \text{e}^- + \bar{\nu}_\text{e}$. Which of the four forces in nature is responsible for this decay? How do you know?
21. (*a*) What are the two major classes of particles that make up the matter of the universe? (*b*) Name six types, or flavors, of each class of particles. (*c*) What are the four known fundamental forces in the universe? (*d*) Name the particles that carry the forces in part *c*. Which force is much weaker than the other three?
22. (*a*) What property characterizes all hadrons? (*b*) What property characterizes all baryons? (*c*) What property characterizes all mesons?

PSE5_CH43_v1.6_23Jun-20

MisConceptual Questions

1. Quarks are considered to be fundamental particles because (choose all that apply)
   (*a*) they are components of many leptons.
   (*b*) they are components of all ordinary matter.
   (*c*) they have no known internal structure.
   (*d*) None of the above.
2. Which of the following forces is most important for understanding how atoms form molecules?
   (*a*) Electromagnetic. (*c*) Strong nuclear.
   (*b*) Gravitational. (*d*) Weak nuclear.
3. Which of the following are considered to be truly fundamental particles? Choose all that apply.
   (*a*) Electrons. (*c*) Neutrons. (*e*) Pions.
   (*b*) Protons. (*d*) Quarks. (*f*) Mesons.
4. There are six kinds (= flavors) of quarks: up, down, strange, charm, bottom, and top. Which flavors make up most of the known matter in the universe?
   (*a*) Up and down quarks.
   (*b*) Strange and charm quarks.
   (*c*) Bottom and top quarks.
   (*d*) All of the above.
5. Which of the following particles can not be composed of quarks?
   (*a*) Proton. (*d*) Neutron.
   (*d*) Electron. (*e*) Higgs boson.
   (*c*) π meson.
6. If gravity is the weakest force, why is it the one we notice most?
   (*a*) Our bodies are not sensitive to the other forces.
   (*b*) The other forces act only within atoms and therefore have no effect on us.
   (*c*) Gravity may be "very weak" but always attractive, and the Earth has enormous mass. The strong and weak nuclear forces have very short range. The electromagnetic force has a long range, but most matter is electrically neutral.
   (*d*) At long distances, the gravitational force is actually stronger than the other forces.
   (*e*) The other forces act only on elementary particles, not on objects our size.
7. Is it possible for a tau lepton (whose mass is almost twice that of a proton) to decay into only hadrons?
   (*a*) Yes, because it is so massive it could decay into a proton and pions.
   (*b*) Yes, it could decay into pions and nothing else.
   (*c*) No, such a decay would violate lepton number; all of its decay products must be leptons.
   (*d*) No, its decay products must include a tau neutrino but could include hadrons such as pions.
   (*e*) No, the tau lepton is too massive to decay.
8. Which of the following are today considered fundamental particles (that is, not composed of smaller components)? Choose as many as apply.
   (*a*) Atoms. (*e*) Quarks.
   (*b*) Electrons. (*f*) Photon.
   (*c*) Protons. (*g*) Higgs boson.
   (*d*) Neutrons.
9. The electron's antiparticle is called the positron. Which of the following properties, if any, are the same for electrons and positrons?
   (*a*) Mass.
   (*b*) Charge.
   (*c*) Lepton number.
   (*d*) None of the above.
10. The strong nuclear force between a neutron and a proton is due to
    (*a*) the exchange of π mesons between the neutron and the proton.
    (*b*) the conservation of baryon number.
    (*c*) the beta decay of the neutron into the proton.
    (*d*) the exchange of gluons between the quarks within the neutron and the proton.
    (*e*) Both (*a*) and (*d*) at different scales.
11. Electrons are still considered fundamental particles (in the group called leptons). But protons and neutrons are no longer considered fundamental; they have substructure and are made up of
    (*a*) pions. (*c*) quarks. (*e*) photons.
    (*b*) leptons. (*d*) bosons.
12. Which of the following will interact via the weak nuclear force *only*?
    (*a*) Quarks. (*e*) Electrons.
    (*b*) Gluons. (*f*) Muons.
    (*c*) Neutrons. (*g*) Higgs boson.
    (*d*) Neutrinos.
13. Which of the following conservation laws are not obeyed for some types of interaction?
    (*a*) Conservation of strangeness.
    (*b*) Conservation of energy.
    (*c*) Conservation of charge.
    (*d*) Conservation of momentum.
    (*e*) None of the above.
14. Which of the following are not leptons?
    (*a*) Neutrinos.
    (*b*) Electrons.
    (*c*) Mesons.
    (*d*) Positrons.

Problems

43–1 Particles and Accelerators

1. (I) What is the total energy of a proton whose kinetic energy is 5.25 GeV?
2. (I) Calculate the wavelength of 24-GeV electrons.
3. (I) If α particles are accelerated by the cyclotron of Example 43–2, what must be the frequency of the voltage applied to the dees?
4. (I) What is the time for one complete revolution for a very high-energy proton in the 27-km LHC circumference?
5. (II) What strength of magnetic field is used in a cyclotron in which protons make 3.5×10^7 revolutions per second?

6. (II) (*a*) If the cyclotron of Example 43–2 accelerated α particles, what maximum energy could they attain? What would their speed be? (*b*) Repeat for deuterons $\left({}^{2}_{1}\mathrm{H}\right)$. (*c*) In each case, what frequency of voltage is required?

7. (II) Which is better for resolving details of the nucleus: 25-MeV alpha particles or 25-MeV protons? Compare each of their wavelengths with the size of a nucleon in a nucleus.

8. (II) What is the wavelength (= minimum resolvable size) of 6.5-TeV protons at the LHC?

9. (II) The 1.0-km radius Fermilab Tevatron took about 20 seconds to bring the energies of the stored protons from 150 GeV to 1.0 TeV. The acceleration was done once per turn. Estimate the energy given to the protons on each turn. (You can assume that the speed of the protons is essentially c the whole time.)

10. (II) If 2×10^{14} protons moving at $v \approx c$, with $K = 6.5\text{TeV}$, are stored in the 4.3-km-radius ring of the LHC, (*a*) how much current (amperes) is carried by this beam? (*b*) How fast would a 1600-kg car have to move to carry the same kinetic energy as this beam?

11. (II) Protons are injected into the 4.3-km-radius Large Hadron Collider with an energy of 450 GeV. If they are accelerated by 8.0 MV each revolution, how far do they travel and approximately how much time does it take for them to reach 6.5 TeV?

12. (II) What magnetic field is required for 6.5-TeV protons in the 4.25-km-radius Large Hadron Collider (LHC)? [Use relativity.]

13. (II) A cyclotron with a radius of 1.0 m is to accelerate deuterons $\left({}^{2}_{1}\mathrm{H}\right)$ to an energy of 14 MeV. (*a*) What is the required magnetic field? (*b*) What frequency is needed for the voltage between the dees? (*c*) If the potential difference between the dees averages 22 kV, how many revolutions will the particles make before exiting? (*d*) How much time does it take for one deuteron to go from start to exit? (*e*) Estimate how far it travels during this time.

14. (II) Show that the energy of a particle (charge e) in a synchrotron, in the relativistic limit $(v \approx c)$, is given by $E\ (\text{in eV}) = Brc$, where B is the magnetic field and r is the radius of the orbit (SI units).

43–2 to 43–6 Particle Interactions, Particle Exchange

15. (I) About how much energy is released when a Λ^0 decays to $\mathrm{n} + \pi^0$? (See Table 43–2.)

16. (I) How much energy is released in the decay

$$\pi^+ \rightarrow \mu^+ + \nu_\mu ?$$

See Table 43–2.

17. (I) Estimate the range of the strong force if the mediating particle were the kaon instead of a pion.

18. (I) How much energy is required to produce a neutron–antineutron pair?

19. (II) Determine the total energy released when Σ^0 decays to Λ^0 and then to a proton.

20. (II) Two protons collide while heading toward each other with equal speeds. What minimum kinetic energy must each have if a π^0 meson is to be created in the process? (See Table 43–2.)

21. (II) What minimum kinetic energy must a proton and an antiproton each have if they are traveling at the same speed toward each other, collide, and produce a $\mathrm{K^+K^-}$ pair in addition to themselves? (See Table 43–2.)

22. (II) What are the wavelengths of the two photons produced when a proton and an antiproton at rest annihilate?

23. (II) (*a*) How much energy is released when an electron and a positron annihilate each other? (*b*) How much energy is released when a proton and an antiproton annihilate each other? (All particles have $K \approx 0$.)

24. (II) What is the total energy of a proton whose kinetic energy is 12 GeV? What is its wavelength?

25. (II) The Λ^0 cannot decay by the following reactions. What conservation laws are violated in each of the reactions?
(*a*) $\Lambda^0 \not\rightarrow \mathrm{n} + \pi^-$ (*c*) $\Lambda^0 \not\rightarrow \pi^+ + \pi^-$
(*b*) $\Lambda^0 \not\rightarrow \mathrm{p} + \mathrm{K^-}$

26. (II) What would be the wavelengths of the two photons produced when an electron and a positron, each with 390 keV of kinetic energy, annihilate in a head-on collision?

27. (II) Which of the following reactions and decays are possible? For those forbidden, explain what laws are violated.
(a) $\pi^- + \mathrm{p} \rightarrow \mathrm{n} + \eta^0$ (d) $\mathrm{p} \rightarrow \mathrm{e^+} + \nu_\mathrm{e}$
(b) $\pi^+ + \mathrm{p} \rightarrow \mathrm{n} + \pi^0$ (e) $\mu^+ \rightarrow \mathrm{e^+} + \bar{\nu}_\mu$
(c) $\pi^+ + \mathrm{p} \rightarrow \mathrm{p} + \mathrm{e^+}$ (*f*) $\mathrm{p} \rightarrow \mathrm{n} + \mathrm{e^+} + \nu_\mathrm{e}$

28. (II) Antiprotons can be produced when a proton with sufficient energy hits a stationary proton. Even if there is enough energy, which of the following reactions will not happen?

$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{p} + \bar{\mathrm{p}}$$
$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{p} + \mathrm{p} + \bar{\mathrm{p}}$$
$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{p} + \mathrm{p} + \mathrm{p} + \bar{\mathrm{p}}$$
$$\mathrm{p} + \mathrm{p} \rightarrow \mathrm{p} + \mathrm{e^+} + \mathrm{e^+} + \bar{\mathrm{p}}$$

29. (II) Calculate the kinetic energy of each of the two products in the decay $\Xi^- \rightarrow \Lambda^0 + \pi^-$. Assume the Ξ^- decays from rest.

30. (II) In the rare decay $\pi^+ \rightarrow \mathrm{e^+} + \nu_\mathrm{e}$, what is the kinetic energy of the positron? Assume the π^+ decays from rest and $m_\nu = 0$.

31. (III) For the decay $\Lambda^0 \rightarrow \mathrm{p} + \pi^-$, calculate (*a*) the Q-value (energy released), and (*b*) the kinetic energy of the p and π^-, assuming the Λ^0 decays from rest. (Use relativistic formulas.)

32. (III) Calculate the maximum kinetic energy of the electron when a muon decays from rest via $\mu^- \rightarrow \mathrm{e^-} + \bar{\nu}_\mathrm{e} + \nu_\mu$. [*Hint*: In what direction do the two neutrinos move relative to the electron in order to give the electron the maximum kinetic energy? Both energy and momentum are conserved; use relativistic formulas.]

33. (III) Could a π^+ meson be produced if a 130-MeV proton struck a proton at rest? What minimum kinetic energy must the incoming proton have?

43–7 to 43–11 Resonances, Standard Model, Quarks, QCD, GUT

34. (I) The mean life of the Σ^0 particle is 7×10^{-20} s. What is the uncertainty in its rest energy? Express your answer in MeV.

35. (I) The measured width of the ψ (3686) meson is about 300 keV. Estimate its mean life.

36. (I) The measured width of the J/ψ meson is 88 keV. Estimate its mean life.

37. (I) The $\mathrm{B^-}$ meson is a $\mathrm{b\bar{u}}$ quark combination. (*a*) Show that this is consistent for all quantum numbers. (*b*) What are the quark combinations for $\mathrm{B^+}$, $\mathrm{B^0}$, $\mathrm{\bar{B}^0}$?

38. (I) What is the energy width (or uncertainty) of (*a*) η^0, and (*b*) ρ^+? See Table 43–2.

39. (II) Which of the following decays are possible? For those that are forbidden, explain which laws are violated.
(*a*) $\Xi^0 \rightarrow \Sigma^+ + \pi^-$ (*c*) $\Sigma^0 \rightarrow \Lambda^0 + \gamma + \gamma$
(*b*) $\Omega^- \rightarrow \Sigma^0 + \pi^- + \nu$

40. (II) Calculate the Q-value for each of the reactions, Eq. 43–4, for producing a pion.

41. (II) In ordinary radioactive decay, a W particle may be created even though the decaying particle has less mass than the W particle. If you assume $\Delta E \approx$ mass of the virtual W, what is the expected lifetime of the W?

42. (II) What quark combinations produce (*a*) a Ξ^0 baryon and (*b*) a Ξ^- baryon?

43. (II) What are the quark combinations that can form (*a*) a neutron, (*b*) an antineutron, (*c*) a Λ^0, (*d*) a $\overline{\Sigma}^0$?

44. (II) What particles do the following quark combinations produce: (*a*) uud, (*b*) $\bar{u}\,\bar{u}\,\bar{s}$, (*c*) $\bar{u}s$, (*d*) $d\bar{u}$, (*e*) $\bar{c}s$?

45. (II) What is the quark combination needed to produce a D^0 meson ($Q = B = S = 0, \quad c = +1$)?

46. (II) The D_S^+ meson has $S = c = +1, \quad B = 0$. What quark combination would produce it?

47. (II) Draw a possible Feynman diagram using quarks (as in Fig. 43–17c) for the reaction $\pi^- + p \rightarrow \pi^0 + n$.

48. (II) Draw a Feynman diagram for the reaction $n + \nu_\mu \rightarrow p + \mu^-$.

General Problems

49. The mean lifetimes listed in Table 43–2 are in terms of *proper time*, measured in a reference frame where the particle is at rest. If a tau lepton is created with a kinetic energy of 920 MeV, how long would its track be as measured in the lab, on average, ignoring any collisions?

50. Which of the following reactions are possible, and by what interaction could they occur? For those forbidden, explain why.
(a) $\pi^- + p \rightarrow K^0 + p + \pi^0$
(b) $K^- + p \rightarrow \Lambda^0 + \pi^0$
(c) $K^+ + n \rightarrow \Sigma^+ + \pi^0 + \gamma$
(d) $K^+ \rightarrow \pi^0 + \pi^0 + \pi^+$
(e) $\pi^+ \rightarrow e^+ + \nu_e$

51. Which of the following reactions are possible, and by what interaction could they occur? For those forbidden, explain why.
(*a*) $\pi^- + p \rightarrow K^+ + \Sigma^-$
(*b*) $\pi^+ + p \rightarrow K^+ + \Sigma^+$
(*c*) $\pi^- + p \rightarrow \Lambda^0 + K^0 + \pi^0$
(*d*) $\pi^+ + p \rightarrow \Sigma^0 + \pi^0$
(*e*) $\pi^- + p \rightarrow p + e^- + \bar{\nu}_e$

52. One decay mode for a π^+ is $\pi^+ \rightarrow \mu^+ + \nu_\mu$. What would be the equivalent decay for a π^-? Check conservation laws.

53. Symmetry breaking occurs in the electroweak theory at about 10^{-18} m. Show that this corresponds to an energy that is on the order of the mass of the $W^\pm$.

54. How many fundamental fermions are there in a water molecule?

55. Use the quark model to describe the reaction

$$\bar{p} + n \rightarrow \pi^- + \pi^0.$$

56. Identify the missing particle in the following reactions.
(*a*) $p + p \rightarrow p + n + \pi^+ + ?$ (*b*) $p + ? \rightarrow n + \mu^+$

57. The mass of a π^0 can be measured by observing the reaction $\pi^- + p \rightarrow \pi^0 + n$ with initial kinetic energy near zero. The neutron is observed to be emitted with a kinetic energy of 0.60 MeV. Use conservation of energy and momentum to determine the π^0 mass.

58. (*a*) Show that the so-called unification distance of 10^{-31} m in grand unified theory is equivalent to an energy of about 10^{16} GeV. Use the uncertainty principle, and also de Broglie's wavelength formula, and explain how they apply. (*b*) Calculate the temperature corresponding to 10^{16} GeV.

59. Calculate the Q-value for the reaction $\pi^- + p \rightarrow \Lambda^0 + K^0$, when negative pions strike stationary protons. Estimate the minimum pion kinetic energy needed to produce this reaction. [*Hint*: Assume Λ^0 and K^0 move off with the same velocity.]

60. A proton and an antiproton annihilate each other at rest and produce two pions, π^- and π^+. What is the kinetic energy of each pion?

61. For the reaction $p + p \rightarrow 3p + \bar{p}$, where one of the initial protons is at rest, use relativistic formulas to show that the threshold energy is $6m_p c^2$, equal to three times the magnitude of the Q-value of the reaction, where m_p is the proton mass. [*Hint*: Assume all final particles have the same velocity.]

62. At about what kinetic energy (in eV) can the rest energy of a proton be ignored when calculating its wavelength, if the wavelength is to be within 2.0% of its true value? What are the corresponding wavelength and speed of the proton?

63. What fraction of the speed of light c is the speed of a 6.0-TeV proton?

64. Using the information in Section 43–1, show that the Large Hadron Collider's two colliding proton beams can resolve details that are less than 1/10,000 the size of a nucleus.

65. Searches are underway for a process called **neutrinoless double beta decay**, in which a nucleus decays by emitting two electrons. (*a*) If the parent nucleus is $^{96}_{40}$Zr, what would the daughter nucleus be? (*b*) What conservation laws would be violated during this decay? (*c*) How could $^{96}_{40}$Zr decay to the same daughter nucleus without violating any conservation laws?

66. Estimate the lifetime of the Higgs boson from the width of the "bump" in Fig. 43–19, using the uncertainty principle. [Note: This is not a realistic estimate because the underlying processes are very complicated.]

67. Show that all conservation laws hold for all the decays described in Fig. 43–15 for the decays of the top quark.

68. The Higgs boson, Section 43–10, has very probably been detected at the CERN LHC. (*a*) If a Higgs boson at rest decays into two tau leptons, what is the kinetic energy of each tau? Follow the analysis of Example 43–6. See Table 43–2. (*b*) What are the signs of the electric charges of the two tau leptons? (*c*) Could a Higgs boson decay into two Z bosons (Table 43–2)?

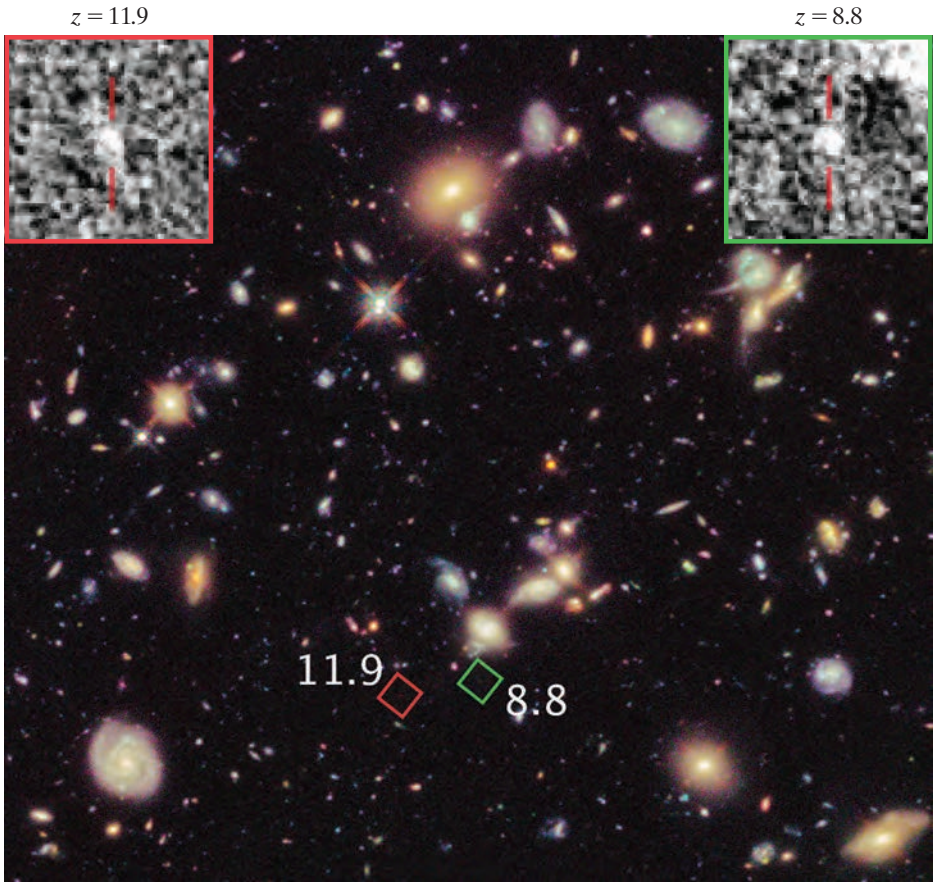
69. (*a*) Show, by conserving momentum and energy, that it is impossible for an isolated electron to radiate only a single photon. (*b*) With this result in mind, how can you defend the photon exchange diagram in Fig. 43–8?

70. A particle at rest, with a rest energy of mc^2, decays into two fragments with rest energies of $m_1 c^2$ and $m_2 c^2$. Show that the kinetic energy of fragment 1 is

$$K_1 = \frac{1}{2mc^2}\left[(mc^2 - m_1 c^2)^2 - (m_2 c^2)^2\right].$$

ANSWERS TO EXERCISES

A: 1.24×10^{-18} m = 1.24 am.
B: (*a*).
C: (*c*); (*d*).
D: $s\bar{u}$.



This Hubble eXtreme Deep Field (XDF) photograph is of a very small part of the sky. It includes what may be the most distant galaxies observable by us (small red and green squares, and shown enlarged in the corners), with $z \approx 8.8$ and 11.9, that already existed when the universe was about 0.4 billion years old. We see these galaxies as they appeared then, 13.4 billion years ago, which is when they emitted this light. The most distant galaxies were young and small and grew to become large galaxies by colliding and merging with other small galaxies.

We examine the latest theories on how stars and galaxies form and evolve, including the role of nucleosynthesis, as well as Einstein's general theory of relativity which deals with gravity and curvature of space. We take a thorough look at the evidence for the expansion of the universe, and the Standard Model of the universe evolving from an initial Big Bang. We point out some unsolved problems, including the nature of dark matter and dark energy that make up most of our universe.

CHAPTER 44

Astrophysics and Cosmology

CONTENTS

44 – 1 Stars and Galaxies
44 – 2 Stellar Evolution: Birth and Death of Stars, Nucleosynthesis
44 – 3 Distance Measurements
44 – 4 General Relativity: Gravity and the Curvature of Space
44 – 5 The Expanding Universe: Redshift and Hubble's Law
44 – 6 The Big Bang and the Cosmic Microwave Background
44 – 7 The Standard Cosmological Model: Early History of the Universe
44 – 8 Inflation: Explaining Flatness, Uniformity, and Structure
44 – 9 Dark Matter and Dark Energy
44–10 Large-Scale Structure of the Universe
44–11 Gravitational Waves—LIGO
44–12 Finally …

CHAPTER-OPENING QUESTIONS—Guess now!

1. Until recently, astronomers expected the expansion rate of the universe would be decreasing. Why?

(a) Friction.
(b) The second law of thermodynamics.
(c) Gravity.
(d) The electromagnetic force.

2. The universe began expanding right at the beginning. How long will it continue to expand?

(a) Until it runs out of room.
(b) Until friction slows it down and brings it to a stop.
(c) Until all galaxies are moving at the speed of light relative to the center.
(d) Possibly forever.

In the previous Chapter, we studied the tiniest objects in the universe—the elementary particles. Now we leap to the grandest objects in the universe—stars, galaxies, and clusters of galaxies—plus the history and structure of the universe itself. These two extreme realms, elementary particles and the cosmos, are among the most intriguing and exciting subjects in science. And, surprisingly, these two extreme realms are related in a fundamental way, as was already hinted in Chapter 43.

Use of the techniques and ideas of physics to study the night sky is often referred to as **astrophysics**. Central to our present theoretical understanding of the universe (or cosmos) is Einstein's *general theory of relativity* which represents our most complete understanding of gravitation. Many other aspects of physics are involved, from electromagnetism and thermodynamics to atomic and nuclear physics as well as elementary particles. General Relativity serves also as the foundation for modern **cosmology**, which is the study of the universe as a whole. Cosmology deals especially with the search for a theoretical framework to understand the observed universe, its origin, and its future. The questions posed by cosmology are profound and difficult; the possible answers stretch the imagination. They are questions like "Has the universe always existed, or did it have a beginning in time?" Either alternative is difficult to imagine: time going back indefinitely into the past, or an actual moment when the universe began (but, then, what was there before?). And what about the size of the universe? Is it infinite in size? It is hard to imagine infinity. Or is it finite in size? This is also hard to imagine, for if the universe is finite, it does not make sense to ask what is beyond it, because the universe is all there is.

During the last 20 years, so much progress has occurred in astrophysics and cosmology that many scientists are calling recent work a "Golden Age" for cosmology. Our survey will be qualitative, but we will nonetheless touch on the major ideas. We begin with a look at what can be seen beyond the Earth.

44–1 Stars and Galaxies

According to the ancients, the stars, except for the few that seemed to move relative to the others (the planets), were fixed on a sphere beyond the last planet. The universe was neatly self-contained, and we on Earth were at or near its center. But in the centuries following Galileo's first telescopic observations of the night sky in 1609, our view of the universe has changed dramatically. We no longer place ourselves at the center, and we view the universe as vastly larger. The distances involved are so great that we specify them in terms of the time it takes light to travel the given distance: for example,

$$1 \text{ light-second} = (3.0 \times 10^8 \text{ m/s})(1.0 \text{ s}) = 3.0 \times 10^8 \text{ m} = 300{,}000 \text{ km};$$

$$1 \text{ light-minute} = (3.0 \times 10^8 \text{ m/s})(60 \text{ s}) = 18 \times 10^6 \text{ km}.$$

The most common unit is the **light-year** (ly):

$$\begin{aligned} 1 \text{ ly} &= (2.998 \times 10^8 \text{ m/s})(3.156 \times 10^7 \text{ s/yr}) \\ &= 9.46 \times 10^{15} \text{ m} \approx 10^{16} \text{ m} \\ &\approx 10^{13} \text{ km}. \end{aligned}$$

For specifying distances to the Sun and Moon, we usually use meters or kilometers, but we could specify them in terms of light-seconds or -minutes. The Earth–Moon distance is 384,000 km, which is 1.28 light-seconds. The Earth–Sun distance is 1.50×10^{11} m, or 150,000,000 km; this is equal to 8.3 light-minutes (it takes 8.3 min for light emitted by the Sun to reach us). Far out in our solar system, Pluto is about 6×10^9 km from the Sun, or 6×10^{-4} ly.† The nearest star to us, other than the Sun, is Proxima Centauri, about 4.2 ly away.

On a clear moonless night, thousands of stars of varying degrees of brightness can be seen, as well as the long cloudy stripe known as the Milky Way (Fig. 44–1). Galileo first observed, with his telescope, that the Milky Way is comprised of countless individual stars. A century and a half later (about 1750), Thomas Wright suggested that the Milky Way was a flat disk of stars extending to great distances in a plane, which we call the **Galaxy** (Greek for "milky way").

†We can also say this is about 5 light-hours.



(a)



(b)

FIGURE 44–1 Sections of the Milky Way. In (a), the thin line is the trail of an artificial Earth satellite in this long time exposure. The dark diagonal area is due to dust absorption of visible light, blocking the view. In (b) the view is toward the center of the Galaxy (taken in summer from Arizona).

Our Galaxy has a diameter of about 100,000 light-years and a thickness of roughly 2000 ly. It has a central bulge ($\approx$ 10,000 ly in diameter) very tightly packed with stars, and spiral arms (Fig. 44–2). Our Sun, which is a star like many others, is located about halfway from the galactic center to the edge, some 28,000 ly from the center. Our Galaxy contains roughly 200 to 400 billion (4×10^{11}) stars. The Sun orbits the galactic center approximately once every 250 million years, so its speed is roughly 200 km/s relative to the center of the Galaxy. The total mass of all the stars in our Galaxy is estimated to be about 4×10^{42} kg of ordinary matter. There is also strong evidence that our Galaxy is permeated and surrounded by a massive invisible "halo" of "dark matter" (Section 44–9) that extends almost 10^6 ly from the center. At the center there seems to be a supermassive black hole whose mass is 4×10^6 times the mass of our Sun.

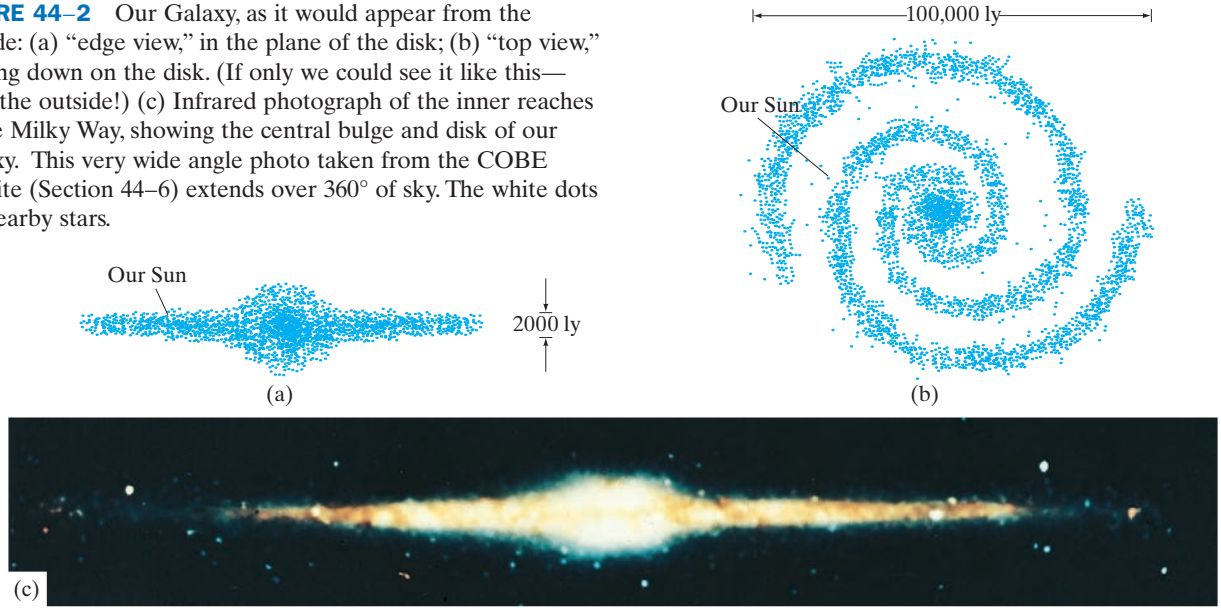


FIGURE 44–2 Our Galaxy, as it would appear from the outside: (a) "edge view," in the plane of the disk; (b) "top view," looking down on the disk. (If only we could see it like this—from the outside!) (c) Infrared photograph of the inner reaches of the Milky Way, showing the central bulge and disk of our Galaxy. This very wide angle photo taken from the COBE satellite (Section 44–6) extends over 360° of sky. The white dots are nearby stars.

EXAMPLE 44–1 **ESTIMATE** **Our Galaxy's mass.** Estimate the total mass of our Galaxy using the orbital data above for the Sun about the center of the Galaxy. Assume the mass of the Galaxy is concentrated in the central bulge.

APPROACH We assume that the Sun (including our solar system) has total mass m and moves in a circular orbit about the center of the Galaxy (total mass M), and that the mass M can be considered as being located at the center of the Galaxy. We then apply Newton's second law, $F = ma$, with a being the centripetal acceleration, $a = v^2/r$, and for F we use the universal law of gravitation (Chapter 6).

SOLUTION Our Sun and solar system orbit the center of the Galaxy, according to the best measurements as mentioned above, with a speed of about $v = 200$ km/s at a distance from the Galaxy center of about $r = 26{,}000$ ly. We use Newton's second law:

$$F = ma$$

$$G\frac{Mm}{r^2} = m\frac{v^2}{r}$$

where M is the mass of the Galaxy and m is the mass of our Sun and solar system. Solving this, we find

$$M = \frac{rv^2}{G} \approx \frac{(26{,}000\text{ ly})(10^{16}\text{ m/ly})(2 \times 10^5\text{ m/s})^2}{6.67 \times 10^{-11}\text{ N}\cdot\text{m}^2/\text{kg}^2} \approx 2 \times 10^{41}\text{ kg}.$$

NOTE In terms of *numbers* of stars, if they are like our Sun $(m = 2.0 \times 10^{30}\text{ kg})$, there would be about $(2 \times 10^{41}\text{ kg})/(2 \times 10^{30}\text{ kg}) \approx 10^{11}$ or very roughly on the order of 100 billion stars.

PSE5_CH44_v1.6_23Jun-20

In addition to stars both within and outside the Milky Way, we can see by telescope many faint cloudy patches in the sky which were all referred to once as "nebulae" (Latin for "clouds"). A few of these, such as those in the constellations Andromeda and Orion, can actually be discerned with the naked eye on a clear night. Some are **star clusters** (Fig. 44–3), groups of stars that are so numerous they appear to be a cloud. Others are glowing clouds of gas or dust (Fig. 44–4), and it is for these that we now mainly reserve the word **nebula**.

Most fascinating are those that belong to a third category: they often have fairly regular elliptical shapes. Immanuel Kant (about 1755) guessed they are faint because they are a great distance beyond our Galaxy. At first it was not universally accepted that these objects were **extragalactic**—that is, outside our Galaxy. But the very large telescopes constructed in the twentieth century revealed that individual stars could be resolved within these extragalactic objects and that many contain spiral arms. Edwin Hubble (1889–1953) did much of this observational work in the 1920s using the 2.5-m (100-inch) telescope[†] on Mt. Wilson near Los Angeles, California, then the world's largest. Hubble demonstrated that these objects were indeed extragalactic because of their great distances. The distance to our nearest large galaxy,[‡] Andromeda, is over 2 million light-years, a distance 20 times greater than the diameter of our Galaxy. It seemed logical that these nebulae must be **galaxies** similar to ours. (Note that it is usual to capitalize the word "galaxy" only when it refers to our own.) Today it is thought there are roughly 10^{11} galaxies in the observable universe—that is, roughly as many galaxies as there are stars in a galaxy. See Fig. 44–5.

Many galaxies tend to be grouped in **galaxy clusters** held together by their mutual gravitational attraction. There may be anywhere from a few dozen to many thousands of galaxies in each cluster. Furthermore, clusters themselves seem to be organized into even larger aggregates: clusters of clusters of galaxies, or **superclusters**.



FIGURE 44–3 This globular star cluster is located in the constellation Hercules.

FIGURE 44–4 This gaseous nebula, found in the constellation Carina, is about 9000 light-years from us.





(a) (b) (c)

FIGURE 44–5 Photographs of galaxies. (a) Spiral galaxy in the constellation Hydra. (b) Two galaxies: the larger and more dramatic one is known as the Whirlpool galaxy. (c) An infrared image (given "false" colors) of the same galaxies as in (b), here showing the arms of the spiral as having more substance than in the visible light photo (b); the different colors correspond to different light intensities. Visible light is scattered and absorbed by interstellar dust much more than infrared is, so infrared gives us a clearer image.

[†]2.5 m (= 100 inches) refers to the diameter of the curved objective mirror. The bigger the mirror, the more light it collects (greater brightness) and the less diffraction there is (better resolution), so more and fainter stars can be seen. See Chapters 33 and 35. Until recently, photographic films or plates were used to take long time exposures. Now large solid-state CCD or CMOS sensors (Section 33–5) are available containing hundreds of millions of pixels (compared to 10 million pixels in a good-quality digital camera).

[‡]The *Magellanic clouds* are much closer than Andromeda, but are small and are usually considered small satellite galaxies of our own Galaxy.

TABLE 44–1 Astronomical Distances

Object	Approx. Distance from Earth (ly)
Moon	4×10^{-8}
Sun	1.6×10^{-5}
Size of solar system (distance to Pluto)	6×10^{-4}
Nearest star (Proxima Centauri)	4.2
Center of our Galaxy	2.6×10^{4}
Nearest large galaxy	2.4×10^{6}
Farthest galaxies	13.4×10^{9}

The farthest detectable galaxies are more than 10^{10} ly distant. See Table 44–1.

CONCEPTUAL EXAMPLE 44–2 **Looking back in time.** Astronomers often think of their telescopes as time machines, looking back toward the origin of the universe. How far back do they look?

RESPONSE The distance in light-years measures how long in years the light has been traveling to reach us, so Table 44–1 tells us also how far back in time we are looking. For example, if we saw Proxima Centauri explode into a supernova today, then the event would have really occurred about 4.2 years ago. The most distant galaxies emitted the light we see now roughly 13×10^9 years ago. What we see was how they were then, 13×10^9 yr ago.

EXERCISE A Suppose we could place a huge mirror 1 light-year away from us. What would we see in this mirror if it is facing us on Earth? When did what we see in the mirror take place? (This might be called a "time machine.")

Besides the usual stars, clusters of stars, galaxies, and clusters and super-clusters of galaxies, the universe contains many other interesting objects. Among these are stars known as *red giants, white dwarfs, neutron stars*, exploding stars called *novae* and *supernovae*, and *black holes* whose gravity is so strong that even light cannot escape them. In addition, there is electromagnetic radiation that reaches the Earth but does not come from the bright pointlike objects we call stars: particularly important is the microwave background radiation that arrives nearly uniformly from all directions in the universe.

Finally, there are **active galactic nuclei** (**AGN**), which are very luminous point-like sources of light in the centers of distant galaxies. The most dramatic examples of AGN are **quasars** ("quasistellar objects" or QSOs), which are so luminous that the surrounding starlight of the galaxy is drowned out. Their luminosity is thought to come from matter falling into a giant black hole at a galaxy's center.

44–2 Stellar Evolution: Birth and Death of Stars, Nucleosynthesis

The stars appear unchanging. Night after night the night sky reveals no significant variations. Indeed, on a human time scale, the vast majority of stars change very little (except for novae, supernovae, and certain variable stars). Although stars *seem* fixed in relation to each other, many move sufficiently for the motion to be detected. Speeds of stars relative to neighboring stars can be hundreds of km/s, but at their great distance from us, this motion is detectable only by careful measurement. There is also a great range of brightness among stars, due to differences in the rate stars emit energy and to their different distances from us.

Luminosity and Brightness of Stars

Any star or galaxy has an **intrinsic luminosity**, L (or simply **luminosity**), which is its total power radiated in watts. Also important is the **apparent brightness**, b, defined as the power crossing unit area at the Earth perpendicular to the path of the light. Given that energy is conserved, and ignoring any absorption in space, the total emitted power L when it reaches a distance d from the star will be spread over a sphere of surface area $4\pi d^2$. If d is the distance from the star to the Earth, then L must be equal to $4\pi d^2$ times b (power per unit area at Earth). That is,

$$b = \frac{L}{4\pi d^2}. \qquad \textbf{(44–1)}$$

PSE5_CH44_v1.6_23Jun-20

EXAMPLE 44–3 **Apparent brightness.** Suppose a star has luminosity equal to that of our Sun. If it is 10 ly away from Earth, how much dimmer will it appear?

APPROACH We use the inverse square law in Eq. 44–1 to determine the relative brightness $(b \propto 1/d^2)$ since the luminosity L is the same for both stars.

SOLUTION Using the inverse square law, the star appears dimmer by a factor

$$\frac{b_{\text{star}}}{b_{\text{Sun}}} = \frac{d_{\text{Sun}}^2}{d_{\text{star}}^2} = \frac{(1.5 \times 10^8\,\text{km})^2}{(10\,\text{ly})^2(10^{13}\,\text{km/ly})^2} \approx 2 \times 10^{-12}.$$

Careful study of nearby stars has shown that the luminosity for most stars depends on the mass: *the more massive the star, the greater its luminosity*.† Another important parameter of a star is its surface temperature, which can be determined from the spectrum of electromagnetic frequencies it emits. As we saw in Chapter 37, as the temperature of a body increases, the spectrum shifts from predominantly lower frequencies (and longer wavelengths, such as red) to higher frequencies (and shorter wavelengths such as blue). Quantitatively, the relation is given by Wien's law (Eq. 37–1): the wavelength λ_P at the peak of the spectrum of light emitted by a blackbody (we often approximate stars as blackbodies) is inversely proportional to its Kelvin temperature T; that is, $\lambda_P T = 2.90 \times 10^{-3}\,\text{m} \cdot \text{K}$. The surface temperatures of stars typically range from about 3000 K (reddish) to about 50,000 K (UV).

EXAMPLE 44–4 **Determining star temperature and star size.** Suppose that the distances from Earth to two nearby stars can be reasonably estimated, and that their measured apparent brightnesses suggest the two stars have about the same luminosity, L. The spectrum of one of the stars peaks at about 700 nm (so it is reddish). The spectrum of the other peaks at about 350 nm (bluish). Use Wien's law (Eq. 37–1) and the Stefan-Boltzmann equation (Section 19–10) to determine (*a*) the surface temperature of each star, and (*b*) how much larger one star is than the other.

APPROACH We determine the surface temperature T for each star using Wien's law and each star's peak wavelength. Then, using the Stefan-Boltzmann equation (power output or luminosity $\propto AT^4$ where A = surface area of emitter), we can find the surface area ratio and relative sizes of the two stars.

SOLUTION (*a*) Wien's law (Eq. 37–1) states that $\lambda_P T = 2.90 \times 10^{-3}\,\text{m} \cdot \text{K}$. So the temperature of the reddish star is

$$T_r = \frac{2.90 \times 10^{-3}\,\text{m} \cdot \text{K}}{\lambda_P} = \frac{2.90 \times 10^{-3}\,\text{m} \cdot \text{K}}{700 \times 10^{-9}\,\text{m}} = 4140\,\text{K}.$$

The temperature of the bluish star will be double this because its peak wavelength is half (350 nm vs. 700 nm):

$$T_b = 8280\,\text{K}.$$

(*b*) The Stefan-Boltzmann equation, Eq. 19–18, states that the power radiated *per unit area* of surface from a blackbody is proportional to the fourth power of the Kelvin temperature, T^4. The temperature of the bluish star is double that of the reddish star, so the bluish one must radiate $(T_b/T_r)^4 = 2^4 = 16$ times as much energy per unit area. But we are given that they have the same luminosity (the same total power output); so the surface area of the blue star must be $\frac{1}{16}$ that of the red one. The surface area of a sphere is $4\pi r^2$, so the radius of the reddish star is $\sqrt{16} = 4$ times larger than the radius of the bluish star (or $4^3 = 64$ times the volume).

†Applies to "main-sequence" stars (see next page). The mass of a star can be determined by observing its gravitational effects on other visible objects. Many stars are part of a cluster, the simplest being a binary star in which two stars orbit around each other, allowing their masses to be determined using rotational mechanics.

H–R Diagram

An important astronomical discovery, made around 1900, was that for most stars, the color is related to the intrinsic luminosity and therefore to the mass. A useful way to present this relationship is by the so-called Hertzsprung–Russell (H–R) diagram. On the H–R diagram, the horizontal axis shows the surface temperature T and the vertical axis is the luminosity L; each star is represented by a point on the diagram, Fig. 44–6. Most stars fall along the diagonal band termed the **main sequence**. Starting at the lower right we find the coolest stars: by Wien's law, $\lambda_P T = \text{constant}$, their light output peaks at long wavelengths, so they are reddish in color. They are also the least luminous and therefore of low mass. Farther up toward the left we find hotter and more luminous stars that are whitish, like our Sun. Still farther up we find even more luminous and more massive stars, bluish in color. Stars that fall on this diagonal band are called *main-sequence stars*. There are also stars that fall outside the main sequence. Above and to the right we find extremely large stars, with high luminosities but with low (reddish) color temperature: these are called **red giants**. At the lower left, there are a few stars of low luminosity but with high temperature: these are the **white dwarfs**.

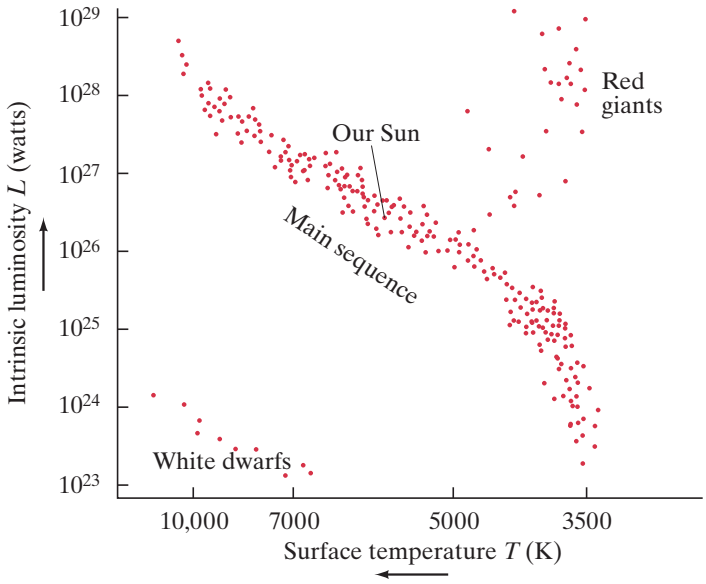


FIGURE 44–6 Hertzsprung–Russell (H–R) diagram is a logarithmic graph of luminosity vs. surface temperature T of stars (note that T increases to the left).

EXAMPLE 44–5 **ESTIMATE** **Distance to a star using the H–R diagram and color.** Suppose that detailed study of a certain star suggests that it most likely fits on the main sequence of an H–R diagram. Its measured apparent brightness is $b = 1.0 \times 10^{-12}\,\text{W/m}^2$, and the peak wavelength of its spectrum is $\lambda_P \approx 600\,\text{nm}$. Estimate its distance from us.

APPROACH We find the temperature using Wien's law, Eq. 37–1. The luminosity is estimated for a main-sequence star on the H–R diagram of Fig. 44–6, and then the distance is found using the relation between brightness and luminosity, Eq. 44–1.

SOLUTION The star's temperature, from Wien's law (Eq. 37–1), is

$$T \approx \frac{2.90 \times 10^{-3}\,\text{m}\cdot\text{K}}{600 \times 10^{-9}\,\text{m}} \approx 4800\,\text{K}.$$

A star on the main sequence of an H–R diagram at this temperature has luminosity of about $L \approx 1 \times 10^{26}\,\text{W}$, read off of Fig. 44–6. Then, from Eq. 44–1,

$$d = \sqrt{\frac{L}{4\pi b}} \approx \sqrt{\frac{1 \times 10^{26}\,\text{W}}{4(3.14)\left(1.0 \times 10^{-12}\,\text{W/m}^2\right)}} \approx 3 \times 10^{18}\,\text{m}.$$

Its distance from us in light-years is

$$d = \frac{3 \times 10^{18}\,\text{m}}{10^{16}\,\text{m/ly}} \approx 300\,\text{ly}.$$

EXERCISE B Estimate the distance to a 6000-K main-sequence star with an apparent brightness of $2.0 \times 10^{-12}\,\text{W/m}^2$.

Stellar Evolution; Nucleosynthesis

Why are there different types of stars, such as red giants and white dwarfs, as well as main-sequence stars? Were they all born this way, in the beginning? Or might each different type represent a different age in the life cycle of a star? Astronomers and astrophysicists today believe the latter is the case. Note, however, that we cannot actually follow any but the tiniest part of the life cycle of any given star because they live for ages vastly greater than ours, on the order of millions or billions of years. Nonetheless, let us follow the process of **stellar evolution** from the birth to the death of a star, as astrophysicists have theoretically reconstructed it today.

Stars are born, it is believed, when gaseous clouds (mostly hydrogen) contract due to the pull of gravity. A huge gas cloud might fragment into numerous contracting masses, each mass centered in an area where the density is only slightly greater than that at nearby points. Once such "globules" form, gravity causes each to contract in toward its center of mass. As the particles of such a *protostar* accelerate inward, their kinetic energy increases. Eventually, when the kinetic energy is sufficiently high, the Coulomb repulsion between the positive charges is not strong enough to keep all the hydrogen nuclei apart, and nuclear fusion can take place.

In a star like our Sun, the fusion of hydrogen (sometimes referred to as "burning")[†] occurs mainly via the *proton–proton chain* (Section 42–4, Eqs. 42–7), in which four protons fuse to form a ${}^4_2\text{He}$ nucleus with the release of γ rays, positrons, and neutrinos: $4\,{}^1_1\text{H} \rightarrow {}^4_2\text{He} + 2\text{e}^+ + 2\nu_\text{e} + 2\gamma$. These reactions require a temperature of about 10^7 K, corresponding to an average kinetic energy ($\approx kT$) of about 1 keV (Eq. 18–4). In more massive stars, the carbon cycle produces the same net effect: four ${}^1_1\text{H}$ produce a ${}^4_2\text{He}$—see Section 42–4. The fusion reactions take place primarily in the core of a star, where T may be on the order of 10^7 to 10^8 K. (The surface temperature is much lower—on the order of a few thousand kelvins.) The tremendous release of energy in these fusion reactions produces an outward pressure sufficient to halt the inward gravitational contraction. Our protostar, now really a young *star*, stabilizes on the *main sequence*. Exactly where the star falls along the main sequence depends on its mass. The more massive the star, the farther up (and to the left) it falls on the H–R diagram of Fig. 44–6. Our Sun required perhaps 30 million years to reach the main sequence, and is expected to remain there about 10 billion years (10^{10} yr). Although most stars are billions of years old, evidence is strong that stars are actually being born at this moment. More massive stars have shorter lives, because they are hotter and the Coulomb repulsion is more easily overcome, so they use up their fuel faster. Our Sun may remain on the main sequence for 10^{10} years, but a star ten times more massive may reside there for only 10^7 years.

As hydrogen fuses to form helium, the helium that is formed is denser and tends to accumulate in the central core where it was formed. As the core of helium grows, hydrogen continues to fuse in a shell around it: see Fig. 44–7. When much of the hydrogen within the core has been consumed, the production of energy decreases at the center and is no longer sufficient to prevent the huge gravitational forces from once again causing the core to contract and heat up. The hydrogen in the shell around the core then fuses even more fiercely because of this rise in temperature, allowing the outer envelope of the star to expand and to cool. The surface temperature, thus reduced, produces a spectrum of light that peaks at longer wavelength (reddish).

[†]The word "burn," meaning fusion, is put in quotation marks because these high-temperature fusion reactions occur via a *nuclear* process, and must not be confused with ordinary burning (of, say, paper, wood, or coal) in air, which is a *chemical* reaction, occurring at the *atomic* level (and at a much lower temperature).

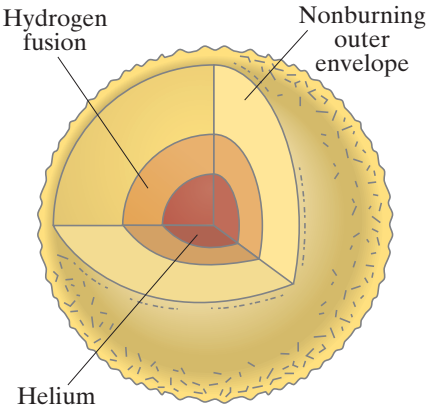


FIGURE 44–7 A shell of "burning" hydrogen (fusing to become helium) surrounds the core where the newly formed helium gravitates.

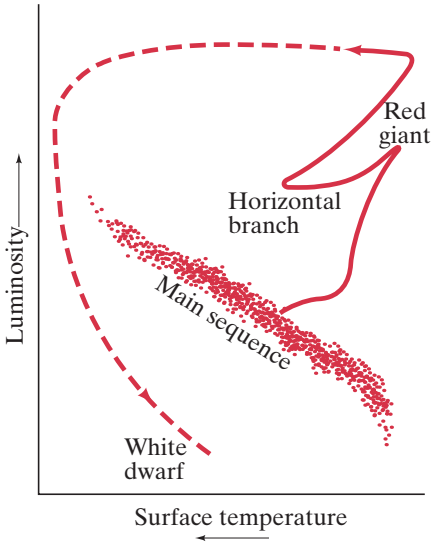


FIGURE 44–8 Evolutionary "track" of a star like our Sun represented on an H–R diagram.

This process marks a new step in the evolution of a star. The star has become redder, it has grown in size, and it has become more luminous, which means it has left the main sequence. It will have moved to the right and upward on the H–R diagram, as shown in Fig. 44–8. As it moves upward, it enters the **red giant** stage. Thus, theory explains the origin of red giants as a natural step in a star's evolution. Our Sun, for example, has been on the main sequence for about $4\frac{1}{2}$ billion years. It will probably remain there another 5 or 6 billion years. When our Sun leaves the main sequence, it is expected to grow in diameter (as it becomes a red giant) by a factor of 100 or more, perhaps swallowing up inner planets such as Mercury and possibly Venus and even Earth.

If the star is like our Sun, or larger, further fusion can occur. As the star's outer envelope expands, its core continues to shrink and heat up. When the temperature reaches about 10^8 K, even helium nuclei, in spite of their greater charge and hence greater electrical repulsion, can come close enough to each other to undergo fusion. The reactions are

$$\begin{aligned} {}^4_2\text{He} + {}^4_2\text{He} &\rightarrow {}^8_4\text{Be} \\ {}^4_2\text{He} + {}^8_4\text{Be} &\rightarrow {}^{12}_6\text{C} \end{aligned} \qquad \textbf{(44–2)}$$

with the emission of two γ rays. These two reactions must occur in quick succession (because ${}^8_4$Be is very unstable), and the net effect is

$$3\,{}^4_2\text{He} \rightarrow {}^{12}_6\text{C} + 2\gamma. \qquad (Q = 7.3\,\text{MeV})$$

This fusion of helium causes a change in the star which moves rapidly to the "horizontal branch" on the H–R diagram (Fig. 44–8). Further fusion reactions are possible, with ${}^4_2$He fusing with ${}^{12}_6$C to form ${}^{16}_8$O. In more massive stars, higher Z elements like ${}^{20}_{10}$Ne or ${}^{24}_{12}$Mg can be made. This process of creating heavier nuclei from lighter ones (or by absorption of neutrons which tends to occur at higher Z) is called **nucleosynthesis**.

Low Mass Stars—White Dwarfs

The final fate of a star depends on its mass. Stars can lose mass as parts of their outer envelope move off into space. Stars born with a mass less than about 8 solar masses (8× the mass of our Sun) eventually end up with a residual mass less than about 1.4 solar masses. A residual mass of 1.4 solar masses is known as the **Chandrasekhar limit**. For stars smaller than this, no further fusion energy can be obtained because of the large Coulomb repulsion between nuclei. The core of such a "low mass" star (original mass $\lesssim$ 8 solar masses) contracts under gravity. The outer envelope expands again and the star becomes an even brighter and larger red giant, Fig. 44–8. Eventually the outer layers escape into space, and the newly revealed surface is hotter than before. So the star moves to the left in the H–R diagram (horizontal dashed line in Fig. 44–8). Then, as the core shrinks the star cools, and typically follows the downward dashed route shown on the left in Fig. 44–8, becoming a **white dwarf**. A white dwarf with a residual mass equal to that of the Sun would be about the size of the Earth. A white dwarf contracts to the point at which the electrons start to overlap, but no further because, by the Pauli exclusion principle, no two electrons can be in the same quantum state. At this point the star is supported against further collapse by this **electron degeneracy** pressure. A white dwarf continues to lose internal energy by radiation, decreasing in temperature and becoming dimmer until it glows no more. It has then become a cold dark chunk of extremely dense material.

High Mass Stars—Supernovae, Neutron Stars, Black Holes

Stars whose original mass is greater than about 8 solar masses are thought to follow a very different scenario. A star with this great a mass can contract under gravity and heat up even further. At temperatures $T \approx 3$ or 4×10^9 K, nuclei as heavy as ${}^{56}_{26}$Fe and ${}^{56}_{28}$Ni can be made. But here the formation of heavy nuclei from lighter ones, by fusion, ends. As we saw in Fig. 41–1, the average binding energy per nucleon begins to decrease for A greater than about 60. Further fusions would *require* energy, rather than release it.

At these extremely high temperatures, well above 10^9 K, high-energy collisions can cause the breaking apart of iron and nickel nuclei into He nuclei, and eventually into protons and neutrons:

$$^{56}_{26}\text{Fe} \rightarrow 13\,^{4}_{2}\text{He} + 4\text{n}$$
$$^{4}_{2}\text{He} \rightarrow 2\text{p} + 2\text{n}.$$

These are energy-requiring (endothermic) reactions, which rob energy from the core, allowing gravitational contraction to begin. This then can force electrons and protons together to form neutrons in **inverse β decay**:

$$\text{e}^- + \text{p} \rightarrow \text{n} + \nu.$$

As a result of these reactions, the pressure in the core drops precipitously. As the core collapses under the huge gravitational forces, the tremendous mass becomes essentially an enormous nucleus made up almost exclusively of neutrons. The size of the star is no longer limited by the exclusion principle applied to electrons, but rather by **neutron degeneracy** pressure, and the star contracts rapidly to form an enormously dense **neutron star**. The core of a neutron star contracts to the point at which all neutrons are as close together as they are in an atomic nucleus. That is, the density of a neutron star is on the order of 10^{14} times greater than normal solids and liquids on Earth. A cupful of such dense matter would weigh billions of tons. A neutron star that has a mass 1.5 times that of our Sun would have a diameter of only about 20 km. (Compare this to a white dwarf with 1 solar mass whose diameter would be $\approx 10^4$ km, as mentioned on the previous page.)

The contraction of the core of a massive star would mean a great reduction in gravitational potential energy. Somehow this energy would have to be released. Indeed, it was suggested in the 1930s that the final core collapse to a neutron star could be accompanied by a catastrophic explosion known as a **supernova** (plural = supernovae). The tremendous energy release (Fig. 44–9) could form virtually all elements of the Periodic Table (see below) and blow away the entire outer envelope of the star, spreading its contents into interstellar space. The presence of heavy elements on Earth and in our solar system suggests that our solar system formed from the debris of many such supernova explosions.

The elements heavier than Ni are thought to form mainly by **neutron capture** in these exploding supernovae (rather than by fusion, as for elements up to Ni). Large numbers of free neutrons, resulting from nuclear reactions, are present inside those highly evolved stars and they can readily combine with, say, a $^{56}_{26}$Fe nucleus to form (if three are captured) $^{59}_{26}$Fe, which decays to $^{59}_{27}$Co. The $^{59}_{27}$Co can capture neutrons, also becoming neutron rich and decaying by β^- to the next higher Z element, and so on to the highest Z elements.

The final state of a neutron star depends on its mass. If the final mass is less than about three solar masses, the subsequent evolution of the neutron star is thought to resemble that of a white dwarf. If the mass is greater than this (original mass $\gtrsim 40$ solar masses), the neutron star collapses under gravity, overcoming even neutron degeneracy. Gravity would then be so strong that emitted light could not escape—it would be pulled back in by the force of gravity. Since no radiation could escape from such a "star," we could not see it—it would be black. An object may pass by it and be deflected by its gravitational field, but if the object came too close it would be swallowed up, never to escape. This is a **black hole**.

Novae and Supernovae

Novae (singular is *nova*, meaning "new" in Latin) are faint stars that have suddenly increased in brightness by as much as a factor of 10^6 and last for a month or two before fading. Novae are thought to be faint white dwarfs that have pulled mass from a nearby companion (they make up a *binary* system), as illustrated in Fig. 44–10. The captured mass of hydrogen suddenly fuses into helium at a high rate for a few weeks. Many novae (maybe all) are *recurrent*—they repeat their bright glow years later.

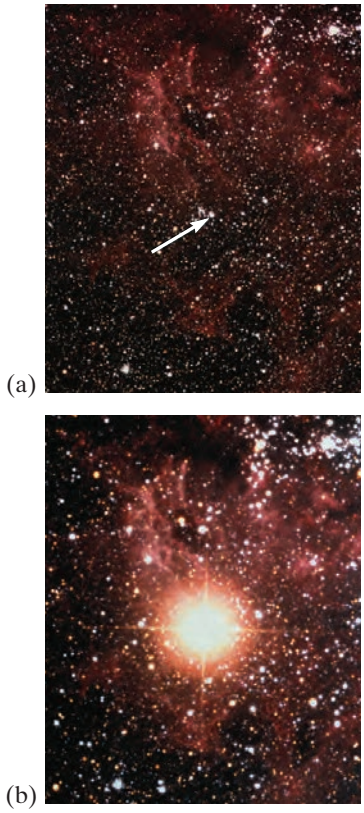
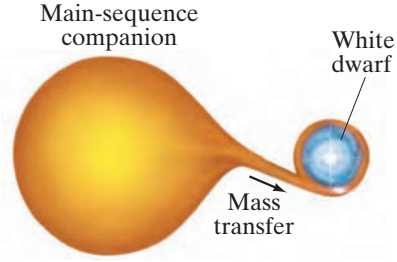


FIGURE 44–9 The star indicated by the arrow in (a) exploded as a supernova, as shown in (b). It was detected in 1987, and named SN1987A. The large bright spot in (b) indicates a huge release of energy but does not represent the physical size.

FIGURE 44–10 Hypothetical model for novae and Type Ia supernovae, showing how a white dwarf could pull mass from its normal companion.



Supernovae are also brief explosive events, but release millions of times more energy than novae, up to 10^{10} times more luminous than our Sun. The peak of brightness may exceed that of the entire galaxy in which they are located, but lasts only a few days or weeks. They slowly fade over a few months. Many supernovae form by core collapse to a neutron star as described above. See Fig. 44–9.

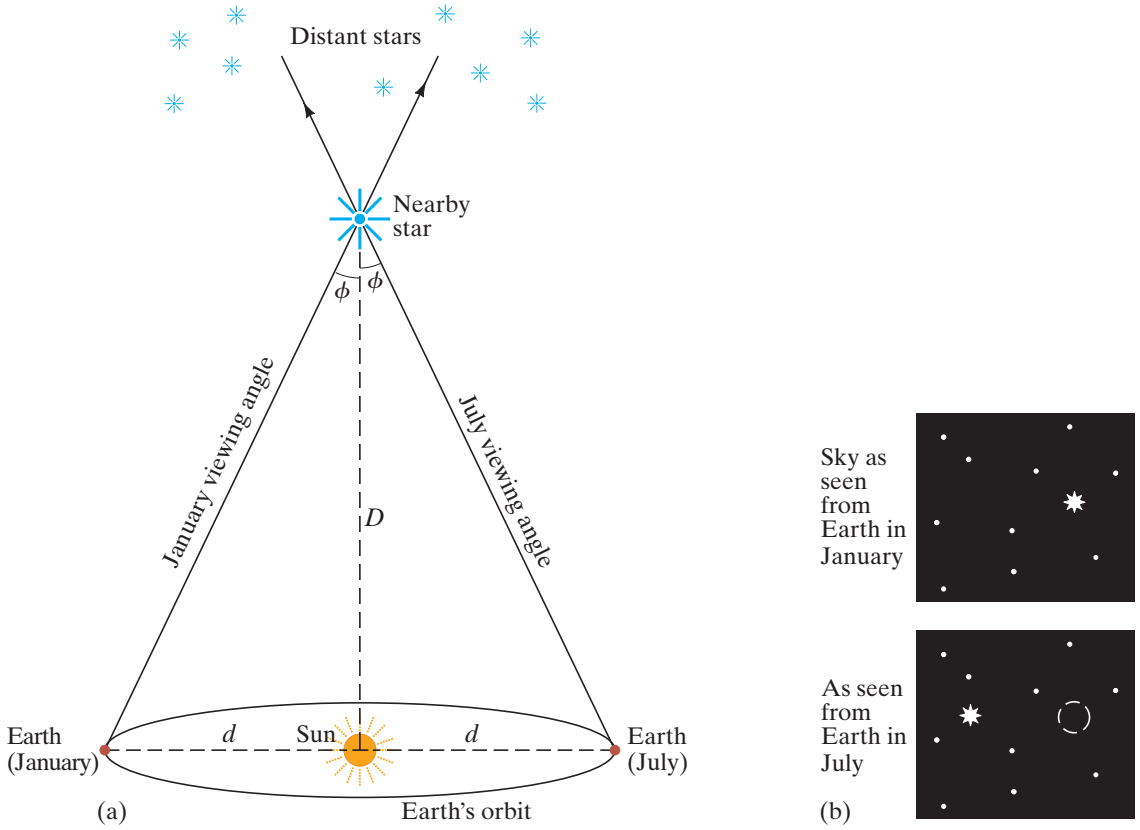
Type Ia supernovae are different. They all seem to have very nearly the same luminosity. They are believed to be binary stars, one of which is a white dwarf that pulls mass from its companion, much like for a nova, Fig. 44–10. The mass is higher, and as mass is captured and the total mass approaches the Chandrasekhar limit of 1.4 solar masses, it explodes as a "white-dwarf" supernova by undergoing a "thermonuclear runaway"—an uncontrolled chain of nuclear reactions that entirely destroys the white dwarf. Type Ia supernovae are useful to us as "standard candles" in the night sky to help us determine distance—see next Section.

44–3 Distance Measurements

Parallax

We have talked about the vast distances of objects in the universe. But how do we measure these distances? One basic technique employs simple geometry to measure the **parallax** of a star. By parallax we mean the apparent motion of a star, against the background of much more distant stars, due to the Earth's motion around the Sun. As shown in Fig. 44–11, we can measure the angle 2ϕ that the star appears to shift, relative to very distant stars, when viewed 6 months apart. If we know the distance d from Earth to Sun, we can reconstruct the right triangles shown in Fig. 44–11 and can then determine the distance D to the star. This is essentially the way the heights of mountains are determined, by "triangulation": see Example 1–7.

FIGURE 44–11 (a) Determining the distance D to a relatively nearby star using parallax. Horizontal distances are greatly exaggerated: in reality ϕ is a very small angle (less than $\frac{1}{3600}° = 1'' = 1$ second of arc). (b) Diagram of the sky showing the apparent position of the "nearby" star relative to more distant stars, at two different times (January and July). The viewing angle in January puts the star more to the right relative to distant stars, whereas in July it is more to the left (dashed circle shows January location).



PSE5_CH44_v1.6_23Jun-20

EXAMPLE 44–6 **ESTIMATE** **Distance to a star using parallax.** Estimate the distance D to a star if the angle 2ϕ in Fig. 44–11a is measured to be $2\phi = 0.00012°$.

APPROACH From trigonometry, $\tan\phi = d/D$ in Fig. 44–11a. The Sun–Earth distance is $d = 1.5 \times 10^8$ km (inside front cover).

SOLUTION The angle $\phi = 0.00006°$, or about $(0.00006°)(2\pi \text{ rad}/360°) = 1.0 \times 10^{-6}$ radians. We can use $\tan\phi \approx \phi$ because ϕ is very small. We solve for D in $\tan\phi = d/D$. The distance D to the star is

$$D = \frac{d}{\tan\phi} \approx \frac{d}{\phi} = \frac{1.5 \times 10^8 \text{ km}}{1.0 \times 10^{-6} \text{ rad}} = 1.5 \times 10^{14} \text{ km},$$

or about 15 ly.

*Parsec

Distances to stars are often specified in terms of parallax angle (ϕ in Fig. 44–11a) given in seconds of arc: 1 second $(1'')$ is $\frac{1}{60}$ of one minute $(1')$ of arc, which is $\frac{1}{60}$ of a degree, so $1'' = \frac{1}{3600}$ of a degree. The distance is then specified in **parsecs** (pc) (meaning *par*allax angle in *sec*onds of arc): $D = 1/\phi$ with ϕ in seconds of arc. In Example 44–6, $\phi = (6 \times 10^{-5})°(3600) = 0.22''$ of arc, so we would say the star is at a distance of $1/0.22'' = 4.5$ pc. One parsec is given by (recall $D = d/\phi$, and we set the Sun–Earth distance (Fig. 44–11a) as $d = 1.496 \times 10^{11}$ m):

$$1 \text{ pc} = \frac{d}{1''} = \frac{1.496 \times 10^{11} \text{ m}}{(1'')\left(\frac{1'}{60''}\right)\left(\frac{1°}{60'}\right)\left(\frac{2\pi \text{ rad}}{360°}\right)} = 3.086 \times 10^{16} \text{ m}$$

$$1 \text{ pc} = (3.086 \times 10^{16} \text{ m})\left(\frac{1 \text{ ly}}{9.46 \times 10^{15} \text{ m}}\right) = 3.26 \text{ ly}.$$

Distant Stars and Galaxies

Parallax can be used to determine the distance to stars as far away as about 100 light-years from Earth, and from an orbiting spacecraft perhaps 5 to 10 times farther. Beyond that distance, parallax angles are too small to measure. For greater distances, more subtle techniques must be employed. We might compare the apparent brightnesses of two stars, or two galaxies, and use the *inverse square law* (apparent brightness drops off as the square of the distance) to roughly estimate their relative distances. We can't expect this technique to be very precise because we don't expect any two stars, or two galaxies, to have the same intrinsic luminosity.

When comparing galaxies, a better estimate assumes that the *brightest* stars in all galaxies (or the brightest galaxies in galaxy clusters) are similar and have about the same intrinsic luminosity. Consequently, their *apparent brightness* would be a measure of how far away they were.

Another technique makes use of the H–R diagram. Measurement of a star's surface temperature (from its spectrum) places it at a certain point (within 20%) on the H–R diagram, assuming it is a main-sequence star, and then its luminosity can be estimated from the vertical axis (Fig. 44–6). Its apparent brightness and Eq. 44–1 give its approximate distance; see Example 44–5.

A better estimate comes from comparing *variable stars*, especially *Cepheid variables* whose luminosity varies over time with a period that is found to be related to their average luminosity. Thus, from their period and apparent brightness we get their distance.

Distance via SNIa, Redshift

The largest distances are estimated by comparing the apparent brightnesses of Type Ia supernovae ("SNIa"). Type Ia supernovae all have a similar origin (as described on the previous page and Fig. 44–10), and their brief explosive burst of light when the Chandrasekhar limit (1.4 solar masses) is reached leads us to believe they have nearly the same luminosity. They are thus sometimes referred to as "standard candles."

Another important technique for estimating the distance of very distant galaxies is from the "redshift" in the line spectra of elements and compounds. The amount of redshift is related to the expansion of the universe, as we shall discuss in Section 44–5. It is useful for objects farther than 10^7 to 10^8 ly away.

As we look farther and farther away, measurement techniques are less and less reliable, so there is more uncertainty in the measurements of large distances.

44–4 General Relativity: Gravity and the Curvature of Space

We have seen that the force of gravity plays an important role in the processes that occur in stars. Gravity too is important for the evolution of the universe as a whole. The reasons gravity plays a dominant role in the universe, and not one of the other of the four forces in nature, are (1) it is long-range and (2) it is always attractive. The strong and weak nuclear forces act over very short distances only, on the order of the size of a nucleus; hence they do not act over astronomical distances (they do act between nuclei and nucleons in stars to produce nuclear reactions). The electromagnetic force, like gravity, acts over great distances. But it can be either attractive or repulsive. And since the universe does not seem to contain large areas of net electric charge, a large net force does not occur. But gravity acts only as an *attractive* force between *all* masses, and there are large accumulations of mass in the universe. The force of gravity as Newton described it in his law of universal gravitation was modified by Einstein. In his general theory of relativity, Einstein developed a theory of gravity that now forms the basis of cosmological dynamics.

In the *special theory of relativity* (Chapter 36), Einstein concluded that there is no way for an observer to determine whether a given frame of reference is at rest or is moving at constant velocity in a straight line. Thus the laws of physics must be the same in different inertial reference frames. But what about the more general case of motion where reference frames can be *accelerating*?

Einstein tackled the problem of accelerating reference frames in his **general theory of relativity** and in it also developed a theory of gravity. The mathematics of General Relativity is complex, so our discussion will be mainly qualitative.

We begin with Einstein's **principle of equivalence**, which states that

> **no experiment can be performed that could distinguish between a uniform gravitational field and an equivalent uniform acceleration.**

If observers sensed that they were accelerating (as in a vehicle speeding around a sharp curve), they could not prove by any experiment that in fact they weren't simply experiencing the pull of a gravitational field. Conversely, we might think we are being pulled by gravity when in fact we are undergoing an acceleration having nothing to do with gravity.

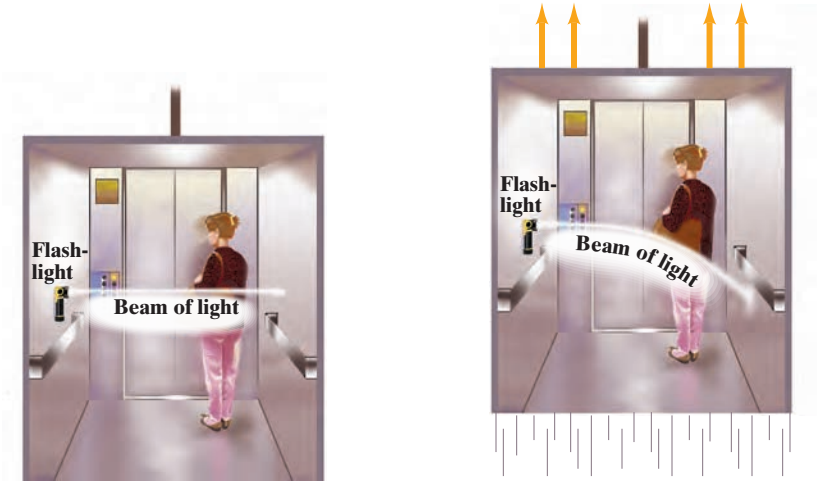
As a thought experiment, consider a person in a freely falling elevator near the Earth's surface. If our observer held out a book and let go of it, what would happen? Gravity would pull it downward toward the Earth, but at the same rate $(g = 9.8\,\text{m/s}^2)$ at which the person and elevator were falling. So the book would hover right next to the person's hand (Fig. 44–12). The effect is exactly the same as if this reference frame was at rest and *no* forces were acting. On the other hand, if the elevator was out in space where the gravitational field is essentially zero, the released book would float, just as it does in Fig. 44–12. Next, if the elevator (out in space) is accelerated upward (using rockets) at an acceleration of $9.8\,\text{m/s}^2$, the book as seen by our observer would fall to the floor with an acceleration of $9.8\,\text{m/s}^2$, just as if it were falling due to gravity at the surface of the Earth. According to the principle of equivalence, the observer could not determine whether the book fell because the elevator was accelerating upward, or because a gravitational field was acting downward and the elevator was at rest. The two descriptions are equivalent.



FIGURE 44–12 In an elevator falling freely under gravity, (a) a person releases a book; (b) the released book hovers next to the owner's hand; (b) is a few moments after (a).

PSE5_CH44_v1.6_23Jun-20

The principle of equivalence is related to the concept that there are two types of mass. Newton's second law, $F = ma$, uses **inertial mass**. We might say that inertial mass represents "resistance" to being accelerated by any type of force. The second type of mass is **gravitational mass**. When one object attracts another by the gravitational force (Newton's law of universal gravitation, $F = Gm_1 m_2/r^2$, Chapter 6), the strength of the force is proportional to the product of the *gravitational masses* of the two objects. This is much like Coulomb's law for the electric force between two objects which is proportional to the product of their electric charges. The electric charge on an object is not related to its inertial mass; so why should we expect that an object's gravitational mass (call it gravitational charge if you like) be related to its inertial mass? All along we have assumed they were the same. Why? Because no experiment—not even of high precision—has been able to discern any measurable difference between inertial mass and gravitational mass. (For example, in the absence of air resistance, all objects fall at the same acceleration, g, on Earth.) This is another way to state the equivalence principle: *gravitational mass is equivalent to inertial mass*.



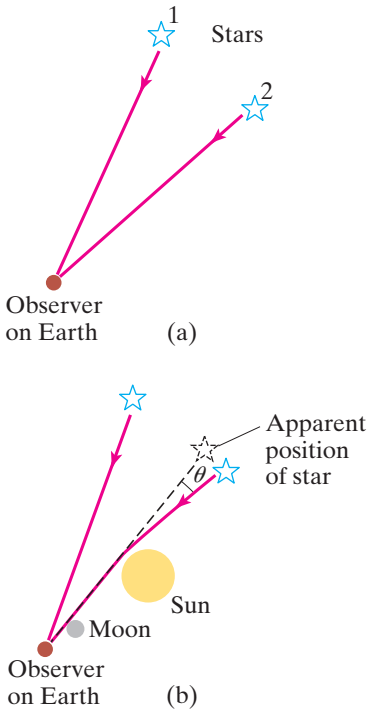
(a) (b)

FIGURE 44–13 (a) Light beam goes straight across an elevator which is not accelerating. (b) The light beam bends (exaggerated) according to an observer in an accelerating elevator whose speed increases in the upward direction.

The principle of equivalence can be used to predict that light ought to be deflected by the gravitational force due to a massive object. Consider another thought experiment, in which an elevator is in free space where virtually no gravity acts. If a light beam is emitted by a flashlight attached to the side of the elevator, the beam travels straight across the elevator and makes a spot on the opposite side if the elevator is at rest or moving at constant velocity (Fig. 44–13a). If instead the elevator is accelerating upward, as in Fig. 44–13b, the light beam still travels straight across in a reference frame at rest. In the upwardly accelerating elevator, however, the beam is observed to curve downward. Why? Because during the time the light travels from one side of the elevator to the other, the elevator is moving upward at a vertical speed that is increasing relative to the light. Next we note that according to the equivalence principle, an upwardly accelerating reference frame is equivalent to a downward gravitational field. Hence, we can picture the curved light path in Fig. 44–13b as being due to the effect of a gravitational field. Thus, from the principle of equivalence, we expect gravity to exert a force on a beam of light and to bend it out of a straight-line path!

That light is affected by gravity is an important prediction of Einstein's general theory of relativity. And it can be tested. The amount a light beam would be deflected from a straight-line path must be small even when passing a massive object. (For example, light near the Earth's surface after traveling 1 km is predicted to drop only about 10^{-10} m, which is equal to the diameter of a small atom and not detectable.) The most massive object near us is the Sun, and it was calculated that light from a distant star would be deflected by 1.75″ of arc (tiny but detectable) as it passed by the edge of the Sun (Fig. 44–14).

FIGURE 44–14 (a) Two stars in the sky observed from Earth. (b) If the light from one of these stars passes very near the Sun, whose gravity bends the rays, the star will appear higher than it actually is (follow the ray backwards). [Not to scale.]



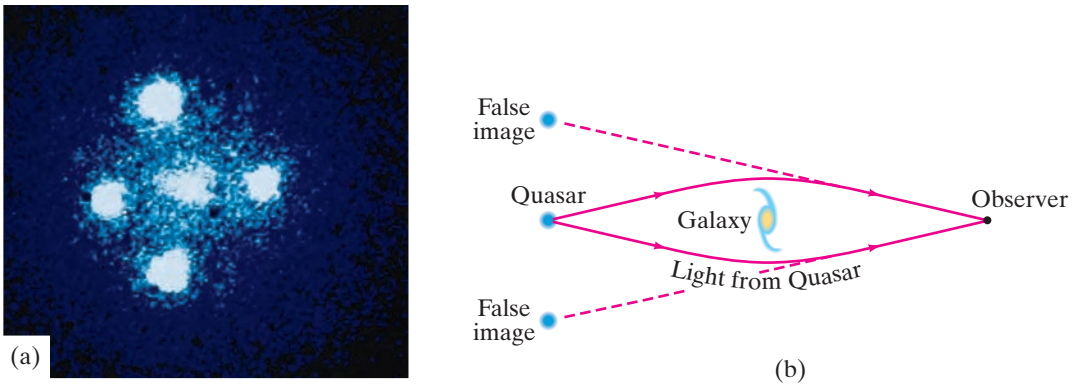


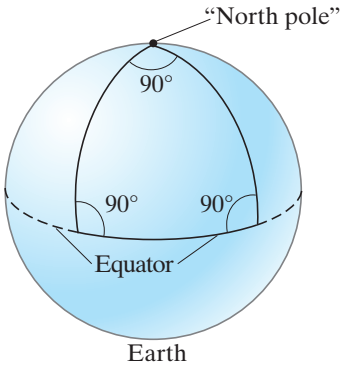
FIGURE 44–15 (a) Hubble Space Telescope photograph of the so-called "Einstein cross", thought to represent "gravitational lensing": the central spot is a relatively nearby galaxy, whereas the four other spots are thought to be images of a single quasar *behind* the galaxy. (b) Diagram showing how the galaxy could bend the light coming from the quasar behind it to produce the four images. See also Fig. 44–14. [If the shape of the nearby galaxy and distant quasar were perfect spheres and perfectly aligned, we would expect the "image" of the distant quasar to be a circular ring or halo instead of the four separate images seen here. Such a ring is called an "Einstein ring."]

However, such a measurement could be made only during a total eclipse of the Sun, so that the Sun's tremendous brightness would not obscure the starlight passing near its edge. An opportune eclipse occurred in 1919, and scientists journeyed to the South Atlantic to observe it. Their photos of stars just behind the Sun revealed shifts in accordance with Einstein's prediction. Another example of gravitational deflection of light is **gravitational lensing**, as described in Fig. 44–15. The very distant galaxies shown in the XDF photo at the start of this Chapter, page 1322, are thought to be visible only because of gravitational lensing (and magnification of their emitted light) by nearer galaxies—as if the nearby galaxies acted as a magnifying glass.

The mathematician Fermat showed in the 1600s that optical phenomena, including reflection, refraction, and effects of lenses, can be derived from a simple principle: that light traveling between two points follows the shortest path in space. Thus if gravity curves the path of light, then gravity must be able to curve space itself. That is, *space itself can be curved*, and it is gravitational mass that causes the curvature. Indeed, the curvature of space—or rather, of four-dimensional space-time—is a basic aspect of Einstein's General Relativity.

What is meant by **curved space**? To understand, recall that our normal method of viewing the world is via Euclidean plane geometry. In Euclidean geometry, there are many axioms and theorems we take for granted, such as that the sum of the angles of any triangle is 180°. Non-Euclidean geometries, which involve curved space, have also been imagined by mathematicians. It is hard enough to imagine three-dimensional curved space, much less curved four-dimensional space-time. So let us try to understand the idea of curved space by using two-dimensional surfaces.

FIGURE 44–16 On a two-dimensional curved surface, the sum of the angles of a triangle may not be 180°.



Consider, for example, the two-dimensional surface of a sphere. It is clearly curved, Fig. 44–16, at least to us who view it from the outside—from our three-dimensional world. But how would hypothetical two-dimensional creatures determine whether their two-dimensional space was flat (a plane) or curved? One way would be to measure the sum of the angles of a triangle. If the surface is a plane, the sum of the angles is 180°, as we learn in plane geometry. But if the space is curved, and a sufficiently large triangle is constructed, the sum of the angles will *not* be 180°. To construct a triangle on a curved surface, say the sphere of Fig. 44–16, we must use the equivalent of a straight line: that is, the shortest distance between two points, which is called a **geodesic**. On a sphere, a geodesic is an arc of a great circle (an arc in a plane passing through the center of the sphere) such as the Earth's equator and the Earth's longitude lines. Consider, for example, the large triangle of Fig. 44–16: its sides are two longitude lines passing from the north pole to the equator, and the third side is a section of the equator as shown. The two longitude lines make 90° angles with the equator (look at a world globe to see this more clearly). They make an angle with each other at the north pole, which could be, say, 90° as shown; the sum of these angles is $90° + 90° + 90° = 270°$. This is clearly *not* a Euclidean space. Note, however, that if the triangle is small in comparison to the radius of the sphere, the angles will add up to nearly 180°, and the triangle (and space) will seem flat.

Another way to test the curvature of space is to measure the radius r and circumference C of a large circle. On a plane surface, $C = 2\pi r$. But on a two-dimensional spherical surface, C is *less* than $2\pi r$, as can be seen in Fig. 44–17. The proportionality between C and r is *less* than 2π. Such a surface is said to have *positive curvature*. On the saddlelike surface of Fig. 44–18, the circumference of a circle is greater than $2\pi r$, and the sum of the angles of a triangle is less than 180°. Such a surface is said to have a *negative curvature*.

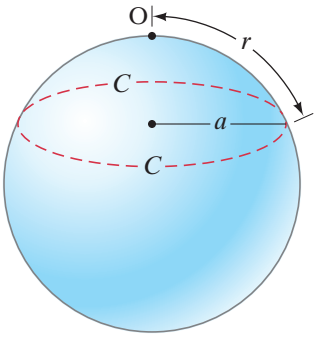


FIGURE 44–17 On a spherical surface (a two-dimensional world) a circle of circumference C is drawn (red) about point O as the center. The radius of the circle (not the sphere) is the distance r along the surface. (Note that in our three-dimensional view, we can tell that $C = 2\pi a$. Since $r > a$, then $C < 2\pi r$.)

Curvature of the Universe

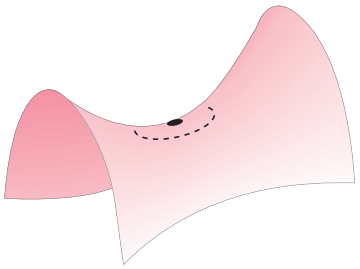
What about our universe? On a large scale (not just near a large mass), what is the overall curvature of the universe? Does it have positive curvature, negative curvature, or is it flat (zero curvature)? We perceive our world as Euclidean (flat), but we can not exclude the possibility that space could have a curvature so slight that we don't normally notice it. This is a crucial question in cosmology, and it can be answered only by precise experimentation.

If the universe had a positive curvature, the universe would be *closed*, or *finite* in volume. This would *not* mean that the stars and galaxies extended out to a certain boundary, beyond which there is empty space. There is no boundary or edge in such a universe. The universe is all there is. If a particle were to move in a straight line in a particular direction, it would eventually return to the starting point—perhaps eons of time later.

On the other hand, if the curvature of space was zero or negative, the universe would be *open*. It could just go on forever. An open universe could be *infinite*; but according to recent research, even that may not necessarily be so.

Today the evidence is very strong that the universe on a large scale is very close to being flat. Indeed, it is so close to being flat that we can't tell if it might have very slightly positive or very slightly negative curvature.

FIGURE 44–18 Example of a two-dimensional surface with negative curvature.



Black Holes

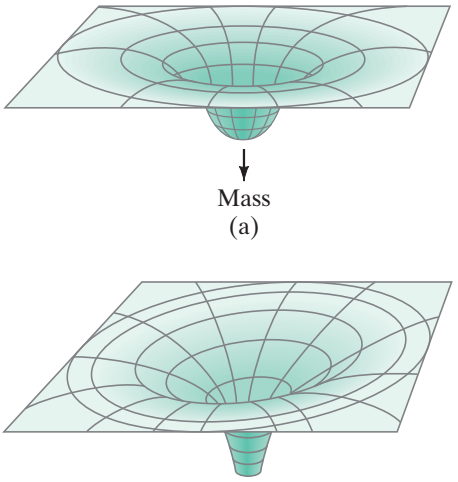
According to Einstein's theory of general relativity (sometimes abbreviated GR), space-time is curved near massive objects. We might think of space as being like a thin rubber sheet: if a heavy weight is placed on the sheet, it sags as shown in Fig. 44–19a. The weight corresponds to a huge mass that causes space (space itself!) to curve. Thus, in the context of general relativity[†] we do not speak of the "force" of gravity acting on objects. Instead we say that objects and light rays move as they do because space-time is curved. An object starting at rest or moving slowly near the great mass of Fig. 44–19a would follow a geodesic (the equivalent of a straight line in plane geometry) toward that great mass.

The extreme curvature of space-time shown in Fig. 44–19b could be produced by a **black hole**. A black hole, as we mentioned in Section 44–2, has such strong gravity that even light cannot escape from it. To become a black hole, an object of mass M must undergo **gravitational collapse**, contracting by gravitational self-attraction to within a radius called the **Schwarzschild radius**,

$$R = \frac{2GM}{c^2},$$

where G is the gravitational constant and c the speed of light. If an object collapses to within this radius, it is predicted by general relativity to collapse to a point at $r = 0$, forming an infinitely dense singularity. This prediction is uncertain, however, because in this realm we need to combine quantum mechanics with gravity, a unification of theories not yet achieved (Section 44–12).

FIGURE 44–19 (a) Rubber-sheet analogy for space-time curved by matter. (b) Same analogy for a black hole, which can "swallow up" objects that pass near.



I EXERCISE C What is the Schwarzschild radius for an object with 10 solar masses?

[†]Alexander Pope (1688–1744) wrote an epitaph for Newton:
"Nature, and Nature's laws lay hid in night:
God said, *Let Newton be!* and all was light."

Sir John Squire (1884–1958), perhaps uncomfortable with Einstein's profound thoughts, added:
"It did not last: the Devil howling '*Ho!*
Let Einstein be!' restored the status quo."

The Schwarzschild radius also represents the event horizon of a black hole. By **event horizon** we mean the surface beyond which no emitted signals can ever reach us, and thus inform us of events that happen beyond that surface. As a star collapses toward a black hole, the light it emits is pulled harder and harder by gravity, but we can still see it. Once the matter passes within the event horizon, the emitted light cannot escape but is pulled back in by gravity (= curvature of space-time).

All we can know about a black hole is its mass, its angular momentum (rotating black holes), and its electric charge. No other information, no details of its structure or the kind of matter it was formed of, can be known because no information can escape.

How might we observe black holes? We cannot see them because no light can escape from them. They would be black objects against a black sky. But they do exert a gravitational force on nearby objects, and also on light rays (or photons) that pass nearby (just like in Fig. 44–15). The black hole believed to be at the center of our Galaxy $\left(M \approx 4 \times 10^6\, M_{\text{Sun}}\right)$ was discovered by examining the motion of matter in its vicinity. Another technique is to examine stars which appear to move as if they were one member of a *binary system* (two stars rotating about their common center of mass), but without a visible companion. If the unseen star is a black hole, it might be expected to pull off gaseous material from its visible companion (as in Fig. 44–10). As this matter approached the black hole, it would be highly accelerated and should emit X-rays of a characteristic type before plunging inside the event horizon. Such X-rays, plus a sufficiently high mass estimate from the rotational motion, can provide evidence for a black hole. One of the many candidates for a black hole is in the binary-star system Cygnus X-1. It is widely believed that the center of most galaxies is occupied by a black hole with a mass 10^6 to 10^9 times the mass of a typical star like our Sun.

EXERCISE D A black hole has radius R. Its mass is proportional to (*a*) R, (*b*) R^2, (*c*) R^3. Justify your answer.

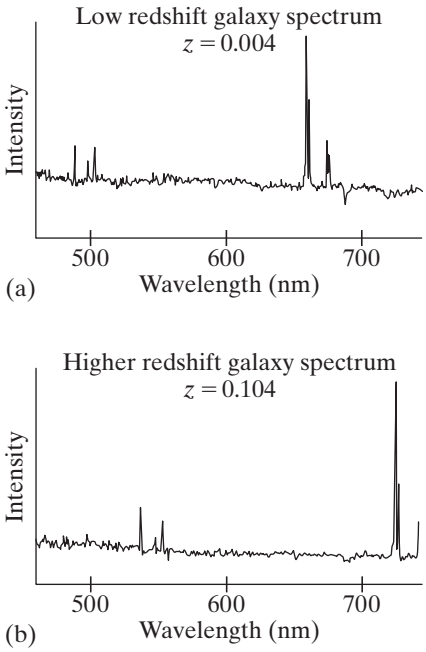


FIGURE 44–20 Atoms and molecules emit and absorb light of particular frequencies depending on the spacing of their energy levels, as we saw in Chapters 37 to 40. (a) The spectrum of light received from a relatively slow-moving galaxy. (b) Spectrum of a galaxy moving away from us at a much higher speed. Note how the peaks (or lines) in the spectrum have moved to longer wavelengths. The redshift is $z = (\lambda_{\text{obs}} - \lambda_{\text{rest}})/\lambda_{\text{rest}}$.

44–5 The Expanding Universe: Redshift and Hubble's Law

We discussed in Section 44–2 how individual stars evolve from their birth to their death as white dwarfs, neutron stars, or black holes. But what about the universe as a whole: is it static, or does it change? One of the most important scientific discoveries of the twentieth century was that distant galaxies are racing away from us, and that the farther they are from us at a given time, the faster they are moving away. How astronomers arrived at this astonishing idea, and what it means for the past history of the universe as well as its future, will occupy us for the remainder of the book.

Evidence that the universe is expanding was first put forth by Georges Lemaître (1894–1966), a Belgian physicist and Catholic priest, in 1927 and also by Edwin Hubble in 1929. This idea was based on distance measurements of galaxies (Section 44–3), and determination of their velocities by the Doppler shift of spectral lines in the light received from them (Fig. 44–20). In Chapter 16 we saw how the frequency of sound is higher and the wavelength shorter if the source and observer move toward each other. If the source moves away from the observer, the frequency is lower and the wavelength longer. The **Doppler effect** occurs also for light, but the formula for light is slightly different than for sound and is given by[†]

$$\lambda_{\text{obs}} = \lambda_{\text{rest}} \sqrt{\frac{1 + v/c}{1 - v/c}}, \qquad \left[\begin{array}{l}\text{source and observer moving}\\ \text{away from each other}\end{array}\right] \qquad \textbf{(44–3)}$$

where λ_{rest} is the emitted wavelength as seen in a reference frame at rest with respect to the source, and λ_{obs} is the wavelength observed in a frame moving with velocity v away from the source along the line of sight. (For relative motion

[†]For light there is no medium and we can make no distinction between motion of the source and motion of the observer (special relativity), as we did for sound which travels in a medium.

PSE5_CH44_v1.6_23Jun-20

toward each other, $v < 0$ in this formula.) When a distant source emits light of a particular wavelength, and the source is moving away from us, the wavelength appears longer to us: the color of the light (if it is visible) is shifted toward the red end of the visible spectrum, an effect known as a **redshift**. (If the source moves toward us, the color shifts toward the blue or shorter wavelength.)

In the spectra of stars in other galaxies, lines are observed that correspond to lines in the known spectra of particular atoms (see Section 37–10 and Figs. 35–22 and 37–20). What Hubble found was that the lines seen in the spectra from distant galaxies were generally *redshifted*, and that the amount of shift seemed to be approximately proportional to the distance of the galaxy from us. That is, the velocity v of a galaxy moving away from us is proportional to its distance d from us:

$$v = H_0 d. \qquad \textbf{(44–4)}$$

HUBBLE'S LAW

This is **Hubble's law**, one of the most fundamental astronomical ideas. It was first suggested, in 1927, by Georges Lemaître, who also first proposed what later came to be called the Big Bang. The constant H_0 is called the **Hubble constant**.

The value of H_0 until recently was uncertain by over 20%, and thought to be between 15 and 25 km/s/Mly. But recent measurements now put its value more precisely at

$$H_0 = 21 \text{ km/s/Mly}$$

(that is, 21 km/s per million light-years of distance). The current uncertainty is close to 1%, or ± 0.2 km/s/Mly. [H_0 can be written in terms of parsecs (Section 44–3) as $H_0 = 68$ km/s/Mpc (that is, 68 km/s per megaparsec of distance) with an uncertainty of about ± 1 km/s/Mpc.]

Redshift Origins

Galaxies very near us seem to be moving randomly relative to us: some move towards us (blueshifted), others away from us (redshifted); their speeds are on the order of $0.001c$. But for more distant galaxies, the velocity of recession is much greater than the velocity of local random motion, and so is dominant and Hubble's law (Eq. 44–4) holds very well. More distant galaxies have higher recession velocity and a larger redshift, and we call their redshift a **cosmological redshift**. We interpret this redshift today as due to the *expansion of space* itself. We can think of the originally emitted wavelength λ_{rest} as being stretched out (becoming longer) along with the expanding space around it, as suggested in Fig. 44–21. Although Hubble thought of the redshift as a Doppler shift, now we prefer to understand it in this sense of expanding space. (But note that atoms in galaxies do not expand as space expands; they keep their regular size.)

There is a third way to produce a redshift, which we mention for completeness: a **gravitational redshift**. Light leaving a massive star is gaining in gravitational potential energy (just like a stone thrown upward from Earth). So the kinetic energy of each photon, hf, must be getting smaller (to conserve energy). A smaller frequency f means a larger (longer) wavelength λ ($= c/f$), which is a redshift.

The amount of a redshift is specified by the **redshift parameter**, z, defined as

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}} = \frac{\Delta\lambda}{\lambda_{\text{rest}}}, \qquad \textbf{(44–5a)}$$

where λ_{rest} is a wavelength as seen by an observer at rest relative to the source, and λ_{obs} is the wavelength measured by a moving observer. Equation 44–5a can be written as

$$z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{rest}}} - 1 \qquad \textbf{(44–5b)}$$

and

$$z + 1 = \frac{\lambda_{\text{obs}}}{\lambda_{\text{rest}}}. \qquad \textbf{(44–5c)}$$

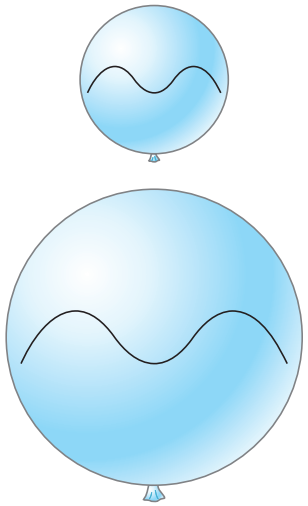


FIGURE 44–21 Simplified model of a 2-dimensional universe, imagined as a balloon. As you blow up the balloon (= expanding universe), the wavelength of a wave on its surface gets longer (redshifted).

For low speeds not close to the speed of light ($v \lesssim 0.1\,c$), the Doppler formula (Eq. 44–3) can be used to show (Problem 36) that z is proportional to the speed of the source toward or away from us:

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}} = \frac{\Delta\lambda}{\lambda_{\text{rest}}} \approx \frac{v}{c}. \qquad [v \ll c] \quad \textbf{(44–6)}$$

But redshifts are not always small, in which case the approximation of Eq. 44–6 is not valid. For high z galaxies, not even Eq. 44–3 applies because the redshift is due to the expansion of space (cosmological redshift), not the Doppler effect. Our Chapter-Opening Photograph, page 1322, shows two very distant high z galaxies, $z = 8.8$ and 11.9, which are also shown enlarged.

Scale Factor

The expansion of space can be described as a scaling of the typical distance between two points or objects in the universe. If two distant galaxies are a distance d_0 apart at some initial time, then a time t later they will be separated by a greater distance $d(t)$. The **scale factor** is the same as for light, expressed in Eq. 44–5a:

$$\frac{d(t) - d_0}{d_0} = \frac{\Delta\lambda}{\lambda} = z$$

or

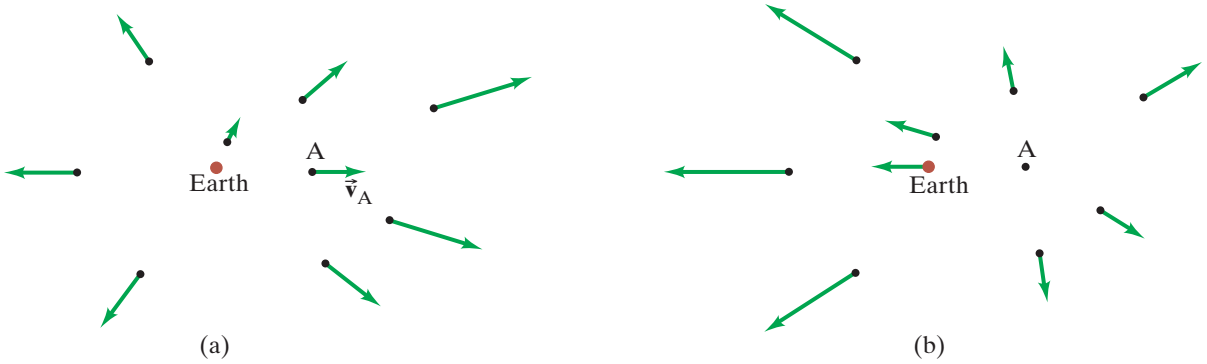
$$\frac{d(t)}{d_0} = 1 + z.$$

Thus, for example, if a galaxy has $z = 3$, then the scale factor is now $(1 + 3) = 4$ times larger than when the light was emitted from that galaxy. That is, the average distance between galaxies has become 4 times larger. Thus the factor by which the wavelength has increased since it was emitted tells us by what factor the universe (or the typical distance between objects) has increased.

Expansion, and the Cosmological Principle

What does it mean that distant galaxies are all moving away from us, and with ever greater speed the farther they are from us? It seems to suggest some kind of explosive expansion that started at some very distant time in the past. And at first sight we seem to be in the middle of it all. But we aren't. The expansion appears the same from any other point in the universe. To understand why, see Fig. 44–22. In Fig. 44–22a we have the view from Earth (or from our Galaxy). The velocities of surrounding galaxies are indicated by arrows, pointing away from us, and the arrows are longer (faster speeds) for galaxies more distant from us. Now, what if we were on the galaxy labeled A in Fig. 44–22a? From Earth, galaxy A appears to be moving to the right at a velocity, call it $\vec{\mathbf{v}}_A$, represented by the arrow pointing to the right. If we were *on* galaxy A, Earth would appear to be moving to the left at velocity $-\vec{\mathbf{v}}_A$.

FIGURE 44–22 Expansion of the universe looks the same from any point in the universe. If you are on Earth as shown in part (a), or you are instead at galaxy A (which is at rest in the reference frame shown in (b)), all other galaxies appear to be racing away from you.



PSE5_CH44_v1.6_23Jun-20

To determine the velocities of other galaxies relative to A, we vectorially add the velocity vector, $-\vec{\mathbf{v}}_A$, to all the velocity arrows shown in Fig. 44–22a. This yields Fig. 44–22b, where we see that the universe is expanding away from galaxy A as well; and the velocities of galaxies receding from A are proportional to their current distance from A. *The universe looks pretty much the same from different points.*

Thus the expansion of the universe can be stated as follows: all galaxies are racing away from *each other* at an average rate of about 21 km/s per million light-years of distance between them. The ramifications of this idea are profound, and we discuss them in a moment.

A basic assumption in cosmology has been that on a large scale, the universe would look the same to observers at different places at the same time. In other words, the universe is both *isotropic* (looks the same in all directions) and *homogeneous* (would look the same if we were located elsewhere, say in another galaxy). This assumption is called the **cosmological principle**. On a local scale, say in our solar system or within our Galaxy, it clearly does not apply (the sky looks different in different directions). But it has long been thought to be valid if we look on a large enough scale, so that the average population density of galaxies and clusters of galaxies ought to be the same in different areas of the sky. This seems to be valid on distances greater than about 700 Mly. The expansion of the universe (Fig. 44–22) is consistent with the cosmological principle; and the near uniformity of the cosmic microwave background radiation (discussed in Section 44–6) supports it. Another way to state the cosmological principle is that *our place in the universe is not special.*

The expansion of the universe, as described by Hubble's law, strongly suggests that galaxies must have been closer together in the past than they are now. This is, in fact, the basis of the *Big Bang* theory of the origin of the universe, which pictures the universe as a relentless expansion starting from a very hot and compressed beginning. We discuss the Big Bang in detail shortly, but first let us see what can be said about the age of the universe.

One way to estimate the age of the universe uses the Hubble constant. With $H_0 \approx 21$ km/s per 10^6 light-years, the time required for the galaxies to arrive at their present separations would be approximately (starting with $v = d/t$ and using Hubble's law, Eq. 44–4),

$$t = \frac{d}{v} = \frac{d}{H_0 d} = \frac{1}{H_0} \approx \frac{(10^6 \text{ ly})(0.95 \times 10^{13} \text{ km/ly})}{(21 \text{ km/s})(3.16 \times 10^7 \text{ s/yr})} \approx 14 \times 10^9 \text{ yr},$$

or 14 billion years. The age of the universe calculated in this way is called the *characteristic expansion time* or "Hubble age." It is a very rough estimate and assumes the rate of expansion of the universe was constant (which today we are quite sure is not true). Today's best measurements give the age of the universe as about 13.8×10^9 yr, in remarkable agreement with the rough Hubble age estimate.

*Steady-State Model

Before discussing the Big Bang in detail, we mention one alternative to the Big Bang—the **steady-state model**—which assumed that the universe is infinitely old and on average looks the same now as it always has. (This assumed uniformity in time as well as space was called the *perfect cosmological principle.*) According to the steady-state model, no large-scale changes have taken place in the universe as a whole, particularly no Big Bang. To maintain this view in the face of the recession of galaxies away from each other, matter would need to be created continuously to maintain the assumption of uniformity. The rate of mass creation required is very small—about one nucleon per cubic meter every 10^9 years.

The steady-state model provided the Big Bang model with healthy competition in the mid-twentieth century. But the discovery of the cosmic microwave background radiation (next Section), as well as other observations of the universe, has made the Big Bang model universally accepted.

44–6 The Big Bang and the Cosmic Microwave Background

The expansion of the universe suggests that typical objects in the universe were once much closer together than they are now. This is the basis for the idea that the universe began about 14 billion years ago as an expansion from a state of very high density and temperature known affectionately as the **Big Bang**.

The birth of the universe was not an explosion, because an explosion blows pieces out into the surrounding space. Instead, the Big Bang was the start of an expansion of space itself. The observable universe was relatively very small at the start and has been expanding, getting ever larger, ever since. The initial tiny universe of extremely dense matter is not to be thought of as a concentrated mass in the midst of a much larger space around it. The initial tiny but dense universe was the *entire universe*. There wouldn't have been anything else. When we say that the universe was once smaller than it is now, we mean that the average separation between objects (such as electrons or galaxies) was less. The universe may have been infinite in extent even then, and it may still be now (only bigger). The **observable universe** (that which we have the possibility of observing because light has had time to reach us) is, however, finite.

A major piece of evidence supporting the Big Bang is the **cosmic microwave background** radiation (or CMB) whose discovery came about as follows.

In 1964, Arno Penzias and Robert Wilson pointed their horn antenna for detecting radio waves (Fig. 44–23) into the sky. With it they detected widespread emission, and became convinced that it was coming from outside our Galaxy. They made precise measurements at a wavelength $\lambda = 7.35$ cm, in the microwave region of the electromagnetic spectrum (Fig. 31–12). The intensity of this radiation was found initially not to vary by day or night or time of year, nor to depend on direction. It came from all directions in the universe with equal intensity, to a precision of better than 1%. It could only be concluded that this radiation came from the universe as a whole.

The intensity of this CMB measured at $\lambda = 7.35$ cm corresponds to blackbody radiation (see Section 37–1) at a temperature of about 3 K. When radiation at other wavelengths was measured by the COBE satellite (COsmic Background Explorer), the intensities were found to fall on a nearly perfect blackbody curve as shown in Fig. 44–24, corresponding to a temperature of 2.725 K (± 0.002 K).

The remarkable uniformity of the CMB was in accordance with the cosmological principle. But theorists felt that there needed to be some small

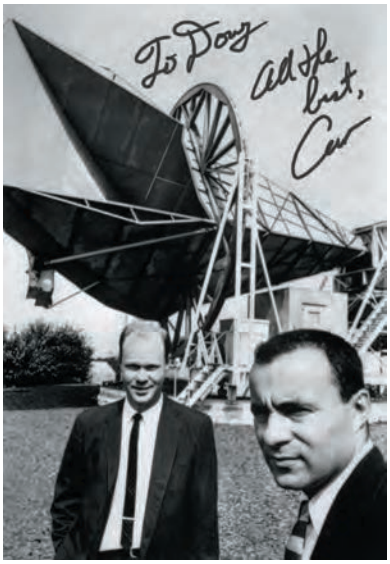


FIGURE 44–23 Photo of Arno Penzias (right, who signed it "Arno") and Robert Wilson. Behind them their "horn antenna."

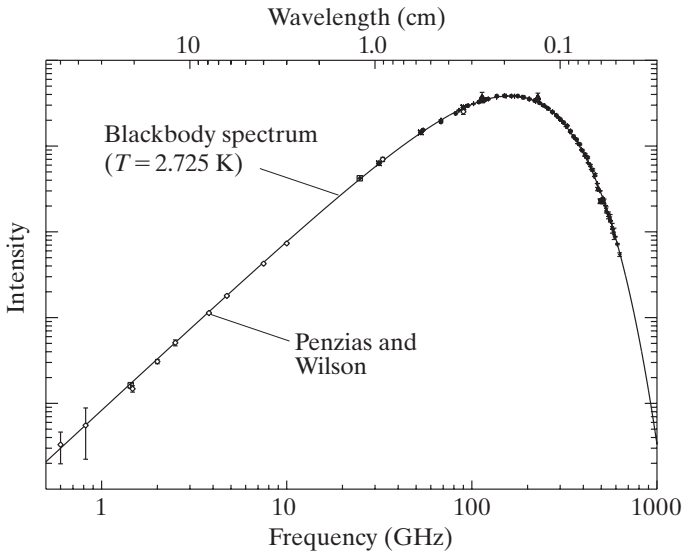


FIGURE 44–24 Spectrum of cosmic microwave background radiation, showing blackbody curve and experimental measurements including at the frequency detected by Penzias and Wilson. (Thanks to G. F. Smoot and D. Scott. The vertical bars represent the most recent experimental uncertainty in a measurement.)

inhomogeneities, or "anisotropies," in the CMB that would have provided "seeds" at which galaxy formation could have started. Small areas of slightly higher density, which could have contracted under gravity to form clusters of galaxies, were indeed found. These tiny inhomogeneities in density and temperature were detected first by the COBE satellite experiment in 1992, led by George Smoot and John Mather (Fig. 44–25).

This discovery of the **anisotropy** of the CMB ranks with the discovery of the CMB itself in the history of cosmology. The blackbody fit and the anisotropy were the culmination of decades of research by pioneers such as Richard Muller, Paul Richards, and David Wilkinson. Subsequent experiments gave us greater detail with the WMAP (Wilkinson Microwave Anisotropy Probe) results from 2003–2012, Fig. 44–26, and even more recently with the European Planck satellite results in 2013.

The CMB provides strong evidence in support of the Big Bang, and gives us information about conditions in the very early universe. In fact, in the late 1940s, George Gamow and his collaborators calculated that a Big Bang origin of the universe should have generated just such a microwave background radiation.

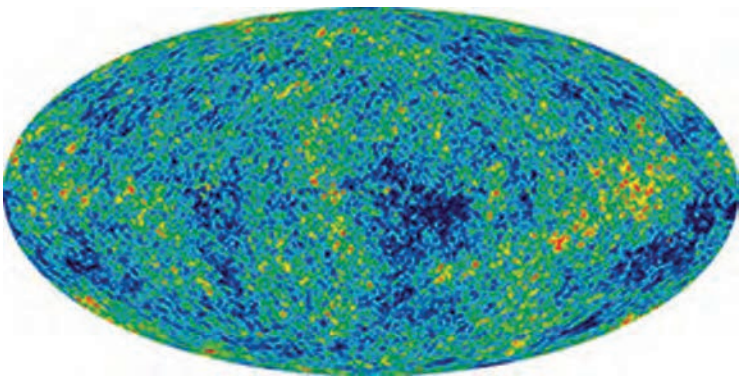
To understand why, let us look at what a Big Bang might have been like. (Today we usually use the term "Big Bang" to refer to the *process*, starting from a moment after the birth of the universe through the subsequent expansion.) The temperature must have been extremely high at the start, so high that there could not have been any atoms in the very early stages of the universe (high energy collisions would have broken atoms apart into nuclei and free electrons). Instead, the universe would have consisted solely of radiation (photons) and a plasma of charged electrons and other elementary particles. The universe would have been opaque—the photons in a sense "trapped," traveling very short distances before being scattered again, primarily by electrons. Indeed, the details of the microwave background radiation provide strong evidence that matter and radiation were once in equilibrium at a very high temperature. As the universe expanded, the energy spread out over an increasingly larger volume and the temperature dropped. When the temperature approached ~3000 K, some 380,000 years later, could nuclei and electrons combine together as stable atoms. With the disappearance of free electrons, as they combined with nuclei to form atoms, the radiation would have been freed—**decoupled** from matter, we say. The universe became *transparent* because photons were now free to travel nearly unimpeded straight through the universe.

It is this radiation, from 380,000 years after the birth of the universe, that we now see as the CMB. As the universe expanded, so too the wavelengths of the radiation lengthened, thus redshifting to longer wavelengths that correspond to lower temperature (recall Wien's law, $\lambda_P T = \text{constant}$, Section 37–1), until they would have reached the 2.7-K background radiation we observe today.



FIGURE 44–25 COBE scientists John Mather (chief scientist and responsible for measuring the blackbody form of the spectrum) and George Smoot (chief investigator for anisotropy experiment) shown here during celebrations for their Dec. 2006 Nobel Prize, given for their discovery of the spectrum and anisotropy of the CMB using the COBE instrument.

FIGURE 44–26 Measurements of the cosmic microwave background radiation over the entire sky, color-coded to represent differences in temperature from the average 2.725 K: the color scale ranges from $+200\,\mu$K (red) to $-200\,\mu$K (dark blue), representing slightly hotter and colder spots (associated with variations in density). Results are from the WMAP satellite in 2012: the angular resolution is 0.2°.



Looking Back Toward the Big Bang—Lookback Time

Figure 44–27 shows our Earth point of view, looking out in all directions back toward the Big Bang and the brief (380,000-year-long) period when radiation was trapped in the early plasma (yellow band). The time it takes light to reach us from an event is called its **lookback time**. The "close-up" insert in Fig. 44–27 shows a photon scattering repeatedly inside that early plasma and then exiting the plasma in a straight line. No matter what direction we look, our view of the very early universe is blocked by this wall of plasma. It is like trying to look into a very thick fog or into the surface of the Sun—we can see only as far as its surface, called the **surface of last scattering**, but not into it. Wavelengths from there are redshifted by $z \approx 1100$. Time $\Delta t'$ in Fig. 44–27 is the lookback time (not real time that goes forward).

Recall that when we view an object far away, we are seeing it as it was then, when the light was emitted, not as it would appear today.

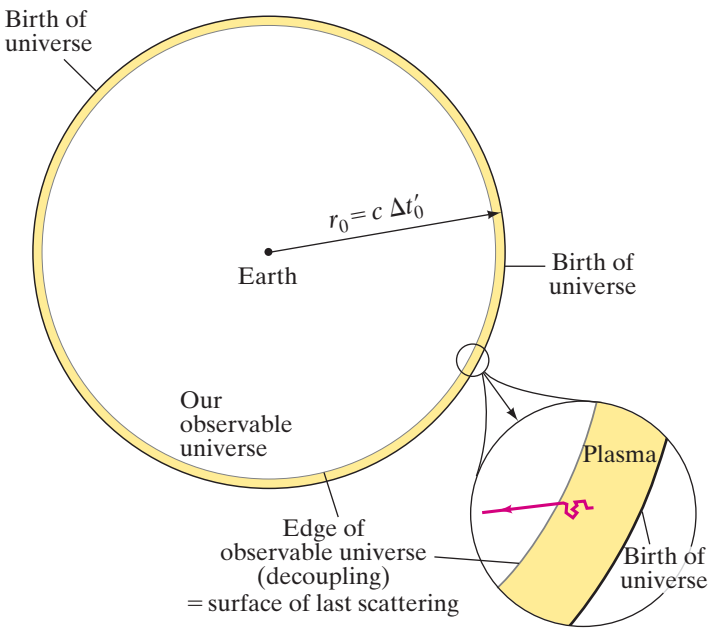


FIGURE 44–27 When we look out from the Earth, we look back in time. Any other observer in the universe would see more or less the same thing. The farther an object is from us, the longer ago the light we see had to have left it. We cannot see quite as far as the Big Bang because our view is blocked; we can see only as far as the "surface of last scattering," which radiated the CMB. The insert on the lower right shows the earliest 380,000 years of the universe when it was opaque: a photon is shown scattering many times and then (at decoupling, 380,000 yr after the birth of the universe) becoming free to travel in a straight line. If this photon wasn't heading our way when "liberated," many others were. Galaxies are not shown, but would be concentrated close to Earth in this diagram because they were created relatively recently. *Note*: This diagram is not a normal map. Maps show a section of the world as might be seen all *at a given time*. This diagram shows space (like a map), but each point is *not* at the same time. The light coming from a point a distance r from Earth took a time $\Delta t' = r/c$ to reach Earth, and thus shows an event that took place long ago, a time $\Delta t' = r/c$ in the past, which we call its "lookback time." The universe began $\Delta t'_0 = 13.8\,\text{Gyr}$ ago.

The Observable Universe

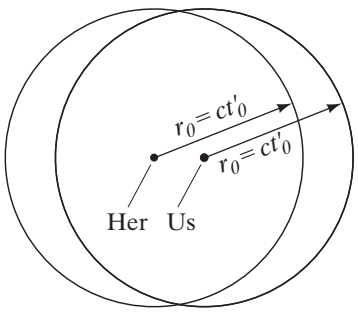
Figure 44–27 can easily be misinterpreted: it is not a picture of the universe at a given instant, but is intended to suggest how we look out in all directions from our observation point (the Earth, or near it). Be careful not to think that the birth of the universe took place in a circle or a sphere surrounding us as if Fig. 44–27 were a photo taken at a given moment. What Fig. 44–27 does show is what we can see, the *observable universe*. Better yet, it shows the *most* we could see.

We would undoubtedly be arrogant to think that we could see the entire universe. Indeed, theories assume that we cannot see everything, that the **entire universe** is greater than the **observable universe**, which is a sphere of radius $r_0 = ct_0$ centered on the observer, with t_0 being the age of the universe. We can never see further back than the time it takes light to reach us.

Consider, for example, an observer in another galaxy, very far from us, located to the left of our observation point in Fig. 44–27. That observer would not yet have seen light coming from the far right of the large circle in Fig. 44–27 that we see—it will take some time for that light to reach her. But she will have already, some time ago, seen the light coming from the left that we are seeing now. In fact, her observable universe, superimposed on ours, is suggested by Fig. 44–28.

The edge of our observable universe is called the **horizon**. We could, in principle, see as far as the horizon, but not beyond it. An observer in another galaxy, far from us, will have a different horizon.

FIGURE 44–28 Two observers, on widely separated galaxies, have different horizons, different observable universes.



PSE5_CH44_v1.6_23Jun-20

44–7 The Standard Cosmological Model: Early History of the Universe

In the last decade or two, a convincing theory of the origin and evolution of the universe has been developed, now called the **Standard Cosmological Model**. Part of this theory is based on recent theoretical and experimental advances in elementary particle physics, and part from observations of the universe including COBE, WMAP, and Planck satellites. Indeed, cosmology and elementary particle physics have cross-fertilized to a surprising extent.

Let us go back to the earliest of times—as close as possible to the Big Bang—and follow a Standard Model theoretical scenario of events as the universe expanded and cooled after the Big Bang. Initially we talk of extremely small time intervals as well as extremely high temperatures, far higher than any temperature in the universe today. Figure 44–29 is a compressed graphical representation of the events, and it may be helpful to consult it as we go along.

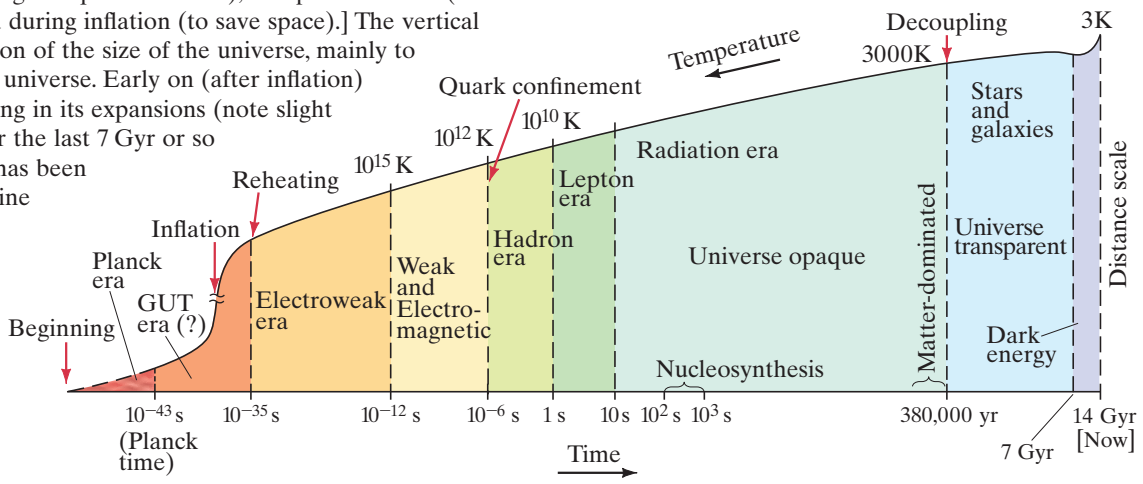


FIGURE 44–29 Compressed graphical representation of the development of the universe after the Big Bang, according to modern cosmology. [The time scale is mostly logarithmic (each factor of 10 in time gets equal treatment), except at the start (there can be no $t = 0$ on a log scale), and during inflation (to save space).] The vertical height is a rough indication of the size of the universe, mainly to suggest expansion of the universe. Early on (after inflation) the universe is decelerating in its expansions (note slight downward curve); but for the last 7 Gyr or so (= thin strip on right) it has been accelerating, so the size line on the top curves upward at upper right.

The History

We begin at a time only a minuscule fraction of a second after the "beginning" of the universe, 10^{-43} s. This time (sometimes referred to as the **Planck time**) is an unimaginably short time, and predictions can be only speculative. Earlier, we can say nothing because we do not have a theory of quantum gravity which would be needed for the incredibly high densities and temperatures during this "Planck era."

The first theories of the Big Bang assumed the universe was extremely hot in the beginning, maybe 10^{32} K, and then gradually cooled down while expanding. In those first moments after 10^{-43} s, the four forces of nature were thought to be united—there was only one force (Chapter 43, Fig. 43–22). Then a kind of "phase transition" would have occurred during which the gravitational force would have "condensed out" as a separate force. This and subsequent phase transitions, as shown in Fig. 43–22, are analogous to phase transitions water undergoes as it cools from a gas condensing into a liquid, and with further cooling freezes into ice.[†] The *symmetry* of the four forces would have been broken leaving the strong, weak, and electromagnetic forces still unified, and the universe would have entered the **grand unified era** (GUT—see Section 43–11).

[†]It may be interesting to point out that this story of origins here bears some resemblance to ancient accounts (nonscientific) that mention the "void," "formless wasteland" (or "darkness over the deep"), "abyss," "divide the waters" (= a phase transition?), not to mention the sudden appearance of light.

This scenario of a *hot* Big Bang is now doubted by some important theorists, such as Andreí Linde, whose theories suggest the universe was much cooler at the Planck time. But what happened next to the universe, though very strange, is accepted by most cosmologists: a brilliant idea, suggested by Linde and Alan Guth in the early 1980s, proposed that the universe underwent an incredible exponential expansion, increasing in size by a factor of 10^{30} or maybe much more, in a tiny fraction of a second, perhaps 10^{-35} s or 10^{-32} s. The usefulness of this **inflationary scenario** is that it solved major problems with earlier Big Bang models, such as explaining why the universe is flat, as well as the thermal equilibrium to provide the nearly uniform CMB, as discussed below.

When inflation ended, whatever energy caused it then ended up being transformed into elementary particles with very high kinetic energy, corresponding to very high temperature $\left(\text{Eq. 18–4, } \overline{K} = \frac{3}{2}kT\right)$. That process is referred to as **reheating**, and the universe was now a "soup" of leptons, quarks, and other particles. We can think of this "soup" as a plasma of particles and antiparticles, as well as photons—all in roughly equal numbers—colliding with one another frequently and exchanging energy.

The temperature of the universe at the end of inflation was much lower than that expected by the hot Big Bang theory. But it would have been high enough so that the weak and electromagnetic forces were unified into a single force, and this stage of the universe is sometimes called the **electroweak era**. Approximately 10^{-12} s after the Big Bang, the temperature dropped to about 10^{15} K corresponding to randomly moving particles with an average kinetic energy K of about 100 GeV (see Eq. 18–4):

$$K \approx kT \approx \frac{(1.4 \times 10^{-23}\,\text{J/K})(10^{15}\,\text{K})}{1.6 \times 10^{-19}\,\text{J/eV}} \approx 10^{11}\,\text{eV} = 100\,\text{GeV}.$$

(We usually ignore the factor $\frac{3}{2}$ in Eq. 18–4, as these are order-of-magnitude quantities. Also, be careful to distinguish K (italic) from K.) At that time, symmetry between weak and electromagnetic forces would have broken down, and the weak force separated from the electromagnetic.

As the universe cooled down to about 10^{12} K ($K \approx 100$ MeV), approximately 10^{-6} s after the Big Bang, quarks stop moving freely and begin to "condense" into more normal particles: nucleons and the other hadrons and their antiparticles. With this **confinement of quarks**, the universe entered the **hadron era**. But it did not last long. Very soon the vast majority of hadrons disappeared. To see why, let us focus on the most familiar hadrons: nucleons and their antiparticles. When the average kinetic energy of particles was somewhat higher than 1 GeV, protons, neutrons, and their antiparticles were continually being created out of the energies of collisions involving photons and other particles, such as

$$\begin{aligned} \text{photons} &\rightarrow \text{p} + \overline{\text{p}} \\ &\rightarrow \text{n} + \overline{\text{n}}. \end{aligned}$$

But just as quickly, particles and antiparticles would annihilate: for example

$$\text{p} + \overline{\text{p}} \rightarrow \text{photons or leptons.}$$

So the processes of creation and annihilation of nucleons were in equilibrium. The numbers of nucleons and antinucleons were high—roughly as many as there were electrons, positrons, or photons. But as the universe expanded and cooled, and the average kinetic energy of particles dropped below about 1 GeV, which is the minimum energy needed in a typical collision to create nucleons and antinucleons (about 940 MeV each), the process of nucleon creation could not continue. Annihilation could continue, however, with antinucleons annihilating nucleons, until almost no nucleons were left. But not quite zero. Somehow we need to explain our present world of matter (nucleons and electrons) with very little antimatter in sight.

PSE5_CH44_v1.6_23Jun-20

To explain our world of matter, we might suppose that earlier in the universe, after the inflationary period, a slight excess of quarks over antiquarks was formed.[†] This would have resulted in a slight excess of nucleons over antinucleons. And it is these "leftover" nucleons that we are made of today. The excess of nucleons over antinucleons was probably about one part in 10^9. During the hadron era, there should have been about as many nucleons as photons. After it ended, the "leftover" nucleons thus numbered only about one nucleon per 10^9 photons, and this ratio has persisted to this day. Protons, neutrons, and all other heavier particles were thus tremendously reduced in number by about 10^{-6} s after the Big Bang. The lightest hadrons, the pions, soon disappeared, about 10^{-4} s after the Big Bang; because they are the lightest mass hadrons (140 MeV), pions were the last hadrons able to be created as the temperature (and average kinetic energy) dropped. Lighter particles, including electrons and neutrinos, were the dominant form of matter, and the universe entered the **lepton era**.

By the time the first full second had passed (clearly the most eventful second in history!), the universe had cooled to about 10 billion degrees, 10^{10} K. The average kinetic energy was about 1 MeV. This was still sufficient energy to create electrons and positrons and balance their annihilation reactions, since their masses correspond to about 0.5 MeV. So there were about as many e^+ and e^- as there were photons. But within a few more seconds, the temperature had dropped sufficiently so that e^+ and e^- could no longer be formed. Annihilation $(e^+ + e^- \rightarrow \text{photons})$ continued. And, like nucleons before them, electrons and positrons all but disappeared from the universe—except for a slight excess of electrons over positrons (later to join with nuclei to form atoms). Thus, about $t = 10$ s after the Big Bang, the universe entered the **radiation era** (Fig. 44–29). Its major constituents were photons and neutrinos. But the neutrinos, partaking only in the weak force, rarely interacted. So the universe, until then experiencing significant amounts of energy in matter and in radiation, now became **radiation-dominated**: much more energy was contained in radiation than in matter, a situation that would last more than 50,000 years.

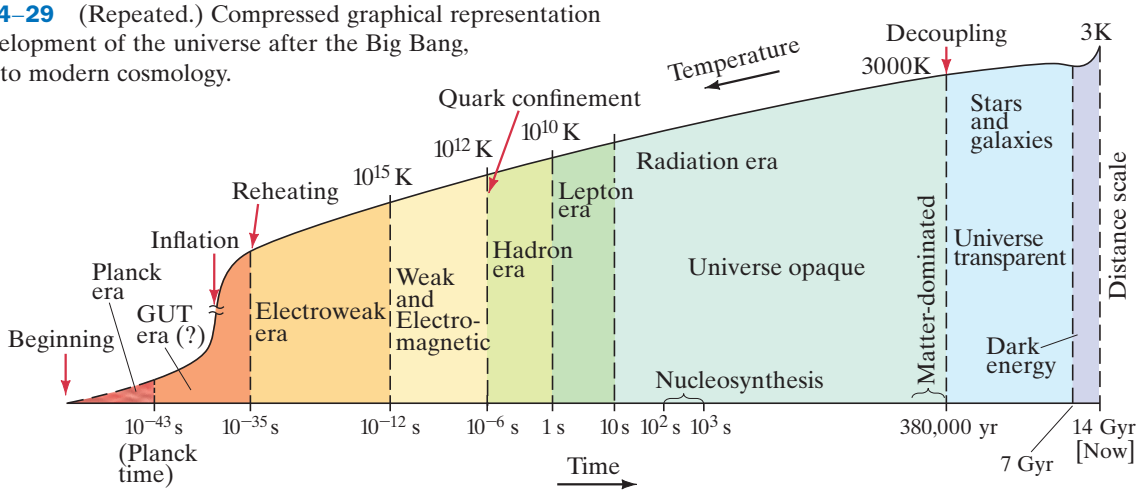


FIGURE 44–29 (Repeated.) Compressed graphical representation of the development of the universe after the Big Bang, according to modern cosmology.

Meanwhile, during the next few minutes, crucial events were taking place. Beginning about 2 or 3 minutes after the Big Bang, nuclear fusion began to occur. The temperature had dropped to about 10^9 K, corresponding to an average kinetic energy $K \approx 100$ keV, where nucleons could strike each other and be able to fuse (Section 42–4), but now cool enough so newly formed nuclei would not be immediately broken apart by subsequent collisions. Deuterium, helium, and very tiny amounts of lithium nuclei were made. But the universe was cooling too quickly, and larger nuclei were not made. After only a few minutes, probably not even a quarter of an hour after the Big Bang, the temperature dropped far enough that nucleosynthesis stopped, not to start again for millions of years (in stars).

[†]Why this could have happened is a question for which we are seeking an answer today.

Thus, after the first quarter hour or so of the universe, matter consisted mainly of bare nuclei of hydrogen (about 75%) and helium (about 25%)[†] as well as electrons. But radiation (photons) continued to dominate.

Our story is almost complete. The next important event is thought to have occurred 380,000 years later. The universe had expanded to about $\frac{1}{1000}$ of its present scale, and the temperature had cooled to about 3000 K. The average kinetic energy of nuclei, electrons, and photons was less than an electron volt. Since ionization energies of atoms are on the order of eV, then as the temperature dropped below this point, electrons could orbit the bare nuclei and remain there (without being ejected by collisions), thus forming atoms. This period is often called the **recombination** epoch (a misnomer since electrons had never before been combined with nuclei to form atoms). With the disappearance of free electrons and the birth of atoms, the photons—which had been continually scattering from the free electrons—now became free to spread throughout the universe. As mentioned in the previous Section, we say that the photons became **decoupled** from matter. Thus *decoupling* occurred at *recombination*. The energy contained in radiation had been decreasing (lengthening in wavelength as the universe expanded); and at about $t = 56{,}000$ yr (even before decoupling) the energy contained in matter became dominant over radiation. The universe was said to have become **matter-dominated** (marked on Fig. 44–29). As the universe continued to expand, the electromagnetic radiation cooled further, to 2.7 K today, forming the cosmic microwave background radiation we detect from everywhere in the universe.

After the birth of atoms, then stars and galaxies could begin to form: by self-gravitation around mass concentrations (inhomogeneities). Stars began to form about 200 million years after the Big Bang, galaxies after almost 10^9 years. The universe continued to evolve until today, some 14 billion years after it started.

* * *

This scenario, like other scientific models, cannot be said to be "proven." Yet this model is remarkably effective in explaining the evolution of the universe we live in and the observations made up to now; it also makes predictions which can be tested against the next generation of observations.

A major event, and something only discovered recently, is that when the universe was about half as old as it is now (about 7 Gyr ago), its expansion began to accelerate. This was a big surprise because it was assumed the expansion of the universe would slow down due to gravitational attraction of all objects toward each other. This acceleration in the expansion of the universe is said to be due to "dark energy," as we discuss in Section 44–9. On the right in Fig. 44–29 is a narrow vertical strip that represents the most recent 7 billion years or so of the universe, during which *dark energy* seems to have dominated.

44–8 Inflation: Explaining Flatness, Uniformity, and Structure

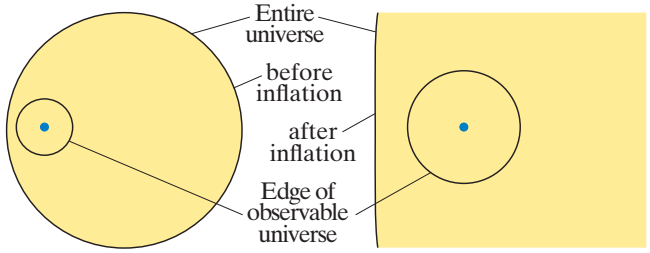
The idea that the universe underwent a period of exponential inflation early in its life, expanding by a factor of 10^{30} or more (previous Section), was first put forth by Alan Guth and Andreí Linde. Many sophisticated models based on this general idea have since been proposed. The energy required for this wild expansion may have been due to fields somewhat like the Higgs field (Section 42–10). So far, the evidence for inflation is indirect; yet it is a feature of most viable cosmological models because it alone is able to provide natural explanations for several remarkable features of our universe.

[†]This Standard Model prediction of a 25% primordial production of helium agrees with what we observe today—the universe *does* contain about 25% He—and it is strong evidence in support of the Standard Big Bang Model. Furthermore, the theory says that 25% He abundance is fully consistent with there being three neutrino types, which is the number we observe. And it sets an upper limit of four to the maximum number of possible neutrino types. This is a striking example of the powerful connection between particle physics and cosmology.

PSE5_CH44_v1.6_23Jun-20

Flatness

First of all, our best measurements suggest that the universe is flat, that it has zero curvature. As scientists, we would like some reason for this remarkable result. To see how inflation explains flatness, consider a simple 2-dimensional model of the universe as we did earlier in Figs. 44–16 and 44–21. A circle in this 2-dimensional universe (= surface of a sphere, Fig. 44–30a) represents the *observable* universe as seen by an observer at the blue dot. A possible hypothesis is that inflation occurred over a time interval that very roughly doubled the age of the universe from, let us say, $t = 1 \times 10^{-35}\,\text{s}$ to $t = 2 \times 10^{-35}\,\text{s}$. The size of the *observable* universe $(r = ct)$ would have increased by a factor of two during inflation, while the radius of curvature of the *entire* universe increased by an enormous factor of 10^{30} or more. Thus the edge of our 2-D sphere representing the entire universe would have seemed flat to a high degree of precision, as shown in Fig. 44–30b. Even if the time of inflation was a factor of 10 or 100 (instead of 2), the expansion factor of 10^{30} or more would have blotted out any possibility of observing anything but a flat universe.



(a) Before inflation (b) After inflation

FIGURE 44–30 (a) Simple 2-D model of the entire universe; the observable universe is suggested by the small circle centered on us (blue dot). (b) Edge of universe is essentially flat after the 10^{30}-fold expansion during inflation.

CMB Uniformity

Inflation also explains why the Cosmic Background Radiation is so uniform. Without inflation, the tiny universe at $10^{-35}\,\text{s}$ would not have been small enough for all parts of it to have been in contact and so reach the same temperature (information cannot travel faster than c). To see this, suppose that the currently observable universe came from a region of space about 1 cm in diameter at $t = 10^{-36}\,\text{s}$, as per original Big Bang theory. In that $10^{-36}\,\text{s}$, light could have traveled $d = ct = (3 \times 10^8\,\text{m/s})(10^{-36}\,\text{s}) = 10^{-27}\,\text{m}$, way too small for the opposite sides of a 1-cm-wide "universe" to have been in communication. But if that region had been 10^{30} times smaller $(= 10^{-32}\,\text{m})$, as proposed by the inflation model, there could have been contact and thermal equilibrium to produce the observed nearly uniform CMB. Inflation, by making the very early universe extremely small, assures that all parts of that region which is today's observable universe could have been in thermal equilibrium. And after inflation the universe could be large enough to give us today's observable universe.

Galaxy Seeds, Fluctuations, Magnetic Monopoles

Inflation also gives us a clue as to how the present structure of the universe (galaxies and clusters of galaxies) came about. We saw earlier that, according to the uncertainty principle, energy might be not conserved by an amount ΔE for a time $\Delta t \approx \hbar/\Delta E$. Forces, whether electromagnetic or other types, can undergo such tiny **quantum fluctuations** according to quantum theory, but they are so tiny they are not detectable unless magnified in some way. That is what inflation might have done: it could have magnified those fluctuations perhaps 10^{30} times in size, which would give us the density irregularities seen in the cosmic microwave background (Planck, Fig. 44–26). That would be very nice, because the density variations we see in the CMB are what we believe were the seeds that later coalesced under gravity into galaxies and galaxy clusters, and our models fit the data extremely well.

Sometimes it is said that the quantum fluctuations occurred in the **vacuum state** or vacuum energy. This could be possible because the vacuum is no longer considered to be empty, as we discussed in Section 43–3 relative to positrons as holes in a negative energy sea of electrons. Indeed, the vacuum is thought to be filled with fields and particles occupying all the possible negative energy states.

Also, the virtual exchange particles that carry the forces, as discussed in Chapter 43, could leave their brief virtual states and actually become real as a result of the 10^{30} magnification of space (according to inflation) and the very short time over which it occurred $(\Delta t = \hbar/\Delta E)$.

Inflation helps us too with the puzzle of why **magnetic monopoles** (Section 27–1) have never been observed, yet isolated magnetic poles may well have been copiously produced at the start. After inflation, they would have been so far apart that we have never stumbled on one.

Inflation may solve outstanding problems, but we may need new physics to understand how inflation occurred. Many predictions of inflationary theory have been confirmed by recent cosmological observations.

44–9 Dark Matter and Dark Energy

According to the Standard Big Bang Model, the universe is evolving and changing. Individual stars are being created, evolving, and then dying to become white dwarfs, neutron stars, or black holes. At the same time, the universe as a whole is expanding. One important question is whether the universe will continue to expand forever. Until the late 1990s, the universe was thought to be dominated by matter which interacts by gravity, and the fate of the universe was connected to the curvature of space-time (Section 44–4). If the universe had *negative* curvature, the expansion of the universe would never stop, although the rate of expansion would decrease due to the gravitational attraction of its parts. Such a universe would be *open* and infinite. If the universe is *flat* (no curvature), it would still be open and infinite but its expansion would slowly approach a zero rate. If the universe had *positive* curvature, it would be *closed* and finite; the effect of gravity would be strong enough that the expansion would eventually stop and the universe would begin to contract, collapsing back onto itself in a **big crunch**.

Critical Density

According to the above scenario (which does not include inflation or the recently discovered acceleration of the universe), the fate of the universe would depend on the average mass–energy density in the universe. For an average mass density greater than a critical value known as the **critical density**, estimated to be about

$$\rho_c \approx 10^{-26}\,\text{kg/m}^3$$

(i.e., a few nucleons/m^3 on average throughout the universe), space-time would have a positive curvature and gravity would prevent expansion from continuing forever. Eventually (if $\rho > \rho_c$) gravity would pull the universe back into a big crunch. If instead the actual density was equal to the critical density, $\rho = \rho_c$, the universe would be flat and open, just barely expanding forever. If the actual density was less than the critical density, $\rho < \rho_c$, the universe would have negative curvature and would easily expand forever. See Fig. 44–31. Today we believe the universe is very close to flat.

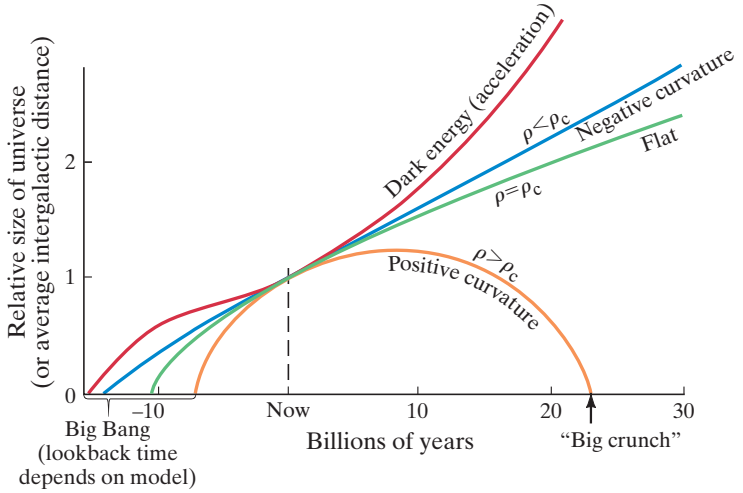


FIGURE 44–31 Three future possibilities for the universe, depending on the density ρ of ordinary matter, plus a fourth possibility that includes dark energy. Note that all curves have been chosen to have the same slope ($= H_0$, the Hubble parameter) right now. Looking back in time, the Big Bang occurs where each curve touches the horizontal (time) axis.

PSE5_CH44_v1.6_23Jun-20

But recent evidence suggests the universe is expanding at an *accelerating* rate, as discussed below.

EXAMPLE 44–7 **ESTIMATE** **Critical density of the universe.** Use energy conservation and escape velocity (Section 8–7) to estimate the critical density of the universe.

APPROACH At the critical density, ρ_c, any given galaxy of mass m will just be able to "escape" away from our Galaxy. As we saw in Section 8–7, escape can just occur if the total energy E of the galaxy satisfies

$$E = K + U = \tfrac{1}{2}mv^2 - G\frac{mM}{R} = 0.$$

Here R is the distance of that galaxy m from us. We approximate the total mass M that pulls inward on m as the total mass within a sphere of radius R around us (Appendix E). If we assume the density of galaxies is roughly constant, then $M = \frac{4}{3}\pi\rho_c R^3$.

SOLUTION Substituting this M into the equation above, and setting $v = HR$ (Hubble's law, Eq. 44–4), we obtain

$$\frac{GM}{R} = \tfrac{1}{2}v^2$$

or

$$\frac{G\left(\frac{4}{3}\pi\rho_c R^3\right)}{R} = \tfrac{1}{2}(HR)^2.$$

We solve for ρ_c:

$$\rho_c = \frac{3H^2}{8\pi G} \approx \frac{3\left[(22\,\text{km/s/Mly})\,(1\,\text{Mly}/10^{19}\,\text{km})\right]^2}{8(3.14)\left(6.67\times10^{-11}\,\text{N}\cdot\text{m}^2/\text{kg}^2\right)} \approx 10^{-26}\,\text{kg/m}^3.$$

EXERCISE E Return to the Chapter-Opening Questions, page 1322, and answer them again. Try to explain why you may have answered differently the first time.

Dark Matter

COBE, WMAP, Planck, and other experiments have convinced scientists that the universe is flat and $\rho = \rho_c$. But this ρ cannot be only normal baryonic matter (atoms are 99.9% baryons—protons and neutrons—by weight). These recent experiments put the amount of normal baryonic matter in the universe at only about 5% of the critical density. What is the other 95%? There is strong evidence for a significant amount of nonluminous matter in the universe referred to as **dark matter**, which acts normally under gravity, but does not absorb or radiate light sufficiently to be visible. For example, observations of the rotation of galaxies suggest that they rotate as if they had considerably more mass than we can see. Recall from Chapter 5, Eq. 5–6, that for a satellite of mass m revolving around Earth (mass M)

$$m\frac{v^2}{r} = G\frac{mM}{r^2}$$

and hence $v = \sqrt{GM/r}$. If we apply this equation to stars in a galaxy, we see that their speed depends on galactic mass. Observations show that stars farther from the galactic center revolve much faster than expected if there is only the pull of visible matter, suggesting a great deal of invisible matter. Similarly, observations of the motion of galaxies within clusters also suggest that they have considerably more mass than can be seen. In fact, dark matter was first proposed in the 1930s, based on such observations.

Furthermore, theory suggests that without dark matter, galaxies and stars probably would not have formed and would not exist. Dark matter seems to hold the universe together.

What might this nonluminous matter in the universe be? We don't know yet. But we hope to find out soon. It cannot be made of ordinary baryonic matter (protons and neutrons), so it must consist of some other sort of elementary particle, perhaps created at a very early time. Perhaps it is made up of previously undetected *weakly interacting massive particles* (**WIMP**s), possibly supersymmetric particles (Section 43–12) such as neutralinos. We are anxiously awaiting the results of intense searches for such particles, looking both at what arrives from far out in the cosmos with underground detectors[†] and satellites, and by producing them in particle colliders (the LHC, Section 43–1).

Dark matter makes up roughly 25% of the mass–energy of the universe, according to the latest observations and models. Thus the total mass–energy is 25% dark matter plus 5% baryons for a total of about 30%, which does not bring ρ up to ρ_c. What is the other 70%? We are not sure about that either, but we have given it a name: "dark energy."

Dark Energy—Cosmic Acceleration



FIGURE 44–32 Saul Perlmutter, center, flanked by Adam G. Riess (left) and Brian P. Schmidt, at the Nobel Prize celebrations, December 2011.

In 1998, just before the turn of the millennium, two groups, one led by Saul Perlmutter and the other by Brian Schmidt and Adam Riess (Fig. 44–32), reported a huge surprise. Gravity was assumed to be the predominant force on a large scale in the universe, and it was thought that the expansion of the universe ought to be slowing down in time because gravity acts as an attractive force between objects. But measurements of Type Ia supernovae (our best standard candles—see Section 44–3) unexpectedly showed that very distant (high z) supernovae were dimmer than expected. That is, given their great distance d as determined from their low brightness, their speed v as determined from the measured z was less than expected according to Hubble's law. This result suggests that nearer galaxies are moving away from us relatively faster than those very distant ones, meaning the expansion of the universe in more recent epochs has sped up. This **acceleration** in the expansion of the universe (in place of the expected deceleration due to gravitational attraction between masses) seems to have begun roughly 7 billion years ago (7 Gyr, which would be about halfway back to what we call the Big Bang).

What could be causing the universe to accelerate in its expansion, against the attractive force of gravity? Does our understanding of gravity need to be revised? We don't yet know the answers to these questions. There are several speculations. But somehow there seems to be a long-range *repulsive* effect on space, like a negative gravity, causing objects to speed away from each other ever faster. Whatever it is, it has been given the name **dark energy**. Many scientists say dark energy is the biggest mystery facing physical science today. The name "dark energy" doesn't tell us much: As the poet[‡] sang a half century ago, "Something is happening but we don't know what it is… ." Here are some thoughts:

One idea is a sort of quantum field given the name **quintessence**. Another possibility suggests an energy latent in space itself (**vacuum energy**) and relates to an aspect of General Relativity known as the **cosmological constant** (symbol Λ). When Einstein developed his equations, he found that they offered no solutions for a static universe. In those days (1917) it was thought the universe was static—unchanging and everlasting. Einstein added an arbitrary constant (Λ) to his equations to provide solutions for a static universe.[§] A decade later, when Hubble showed us an expanding universe, Einstein discarded his cosmological constant as no longer needed ($\Lambda = 0$). But today, measurements are consistent

[†]In deep mines and under mountains to block out most other particles.

[‡]Bob Dylan, "Ballad of a Thin Man." "Dark matter," on the other hand, is a good name because it refers to something we can't see but that acts gravitationally, like regular matter.

[§]It seems strange that Einstein and other scientists believed in a static universe. The ancients, including the Roman Lucretius argued against it: *If there was no birth-time of earth and heaven and they have been from everlasting, why before the* [Trojan] *war …. have not other poets as well sung other themes?* [The reference is to Homer being the oldest known writings.] The ancient Hebrews also argued for a beginning (like our Big Bang): see Genesis.

PSE5_CH44_v1.6_23Jun-20

with dark energy being due to a nonzero cosmological constant, although further measurements are needed to see subtle differences among theories.

There is increasing evidence that the effects of some form of dark energy are very real. Observations of the CMB, supernovae, and large-scale structure (Section 44–10) agree well with theories and computer models when they input dark energy as providing about 70% of the mass–energy in the universe, and when the total mass–energy density equals the critical density ρ_c.

Today's best estimate of how the mass–energy in the universe is distributed is approximately, rounded off (but see also Fig. 44–33):

70% dark energy

30% matter, subject to the known gravitational force.

Of this 30%, about

25% is dark matter

5% is baryons (what atoms are made of); of this 5% only $\frac{1}{10}$ is readily visible matter—stars and galaxies (that is, 0.5% of the total); the other $\frac{9}{10}$ of ordinary matter, which is not visible, is mainly gaseous plasma.

It is remarkable that only 0.5% of all the mass–energy in the universe is visible as stars and galaxies.

The idea that the universe is dominated by completely unknown forms of matter and energy seems bizarre. Nonetheless, the ability of our present model to precisely explain observations of the CMB anisotropy, cosmic expansion, and large-scale structure (next Section) presents a compelling case.

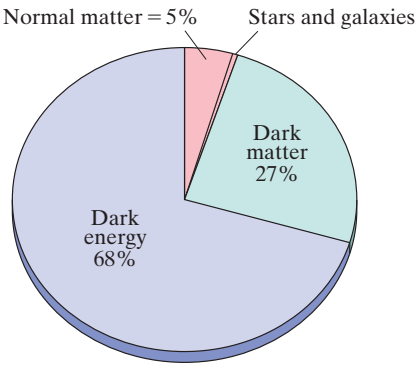


FIGURE 44–33 Portions of total mass–energy in the universe today (approximate).

44–10 Large-Scale Structure of the Universe

The beautiful Planck pictures of the sky (Fig. 44–26) show small but significant inhomogeneities in the temperature of the cosmic microwave background (CMB). These anisotropies reflect compressions and expansions in the primordial plasma just before decoupling (Fig. 44–29), from which galaxies and clusters of galaxies formed. Analyses of the irregularities in the CMB using mammoth computer simulations predict a large-scale distribution of galaxies very similar to what is seen today (Fig. 44–34). These simulations are very successful if they contain dark energy and dark matter; and the dark matter needs to be *cold* (slow speed—think of Eq. 18–4, $\frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$. where T is temperature), rather than "hot" dark matter such as neutrinos which move at or very near the speed of light. Indeed, the modern **cosmological model** is called the **ΛCDM** model, where lambda (**Λ**) stands for the cosmological constant, and CDM is **cold dark matter**.

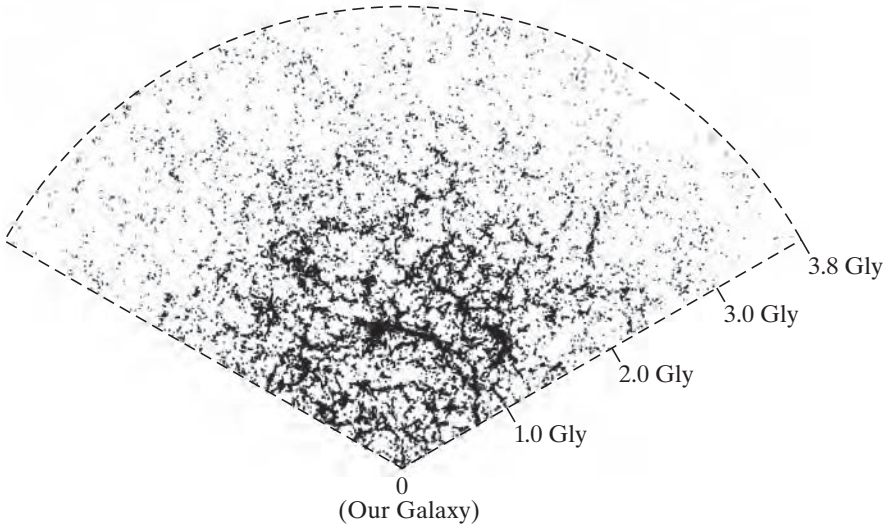


FIGURE 44–34 Distribution of some 50,000 galaxies in a 2.5° slice through almost half of the sky above the equator, as measured by the Sloan Digital Sky Survey (SDSS). Each dot represents a galaxy. The distance from us is obtained from the redshift and Hubble's law, and is given in units of 10^9 light-years (Gly). The point 0 represents us, our observation point. This diagram may seem to put us at the center, but remember that at greater distances, fewer galaxies are bright enough to be detected, thus resulting in an apparent thinning out of galaxies. Note the "walls" and "voids" of galaxies.

Cosmologists have gained substantial confidence in this cosmological model from such a precise fit between observations and theory. They can also extract very precise values for cosmological parameters which previously were only known with low accuracy. The CMB is such an important cosmological observable that every effort is being made to extract all of the information it contains. A new generation of ground, balloon, and satellite experiments is observing the CMB with greater resolution and sensitivity. They may detect interaction of *gravity waves* (produced in the inflationary epoch) with the CMB and thereby provide direct evidence for cosmic inflation, and also provide information about elementary particle physics at energies far beyond the reach of man-made accelerators and colliders.

44–11 Gravitational Waves—LIGO

Einstein, in his General Theory of Relativity (Section 44–4) predicted that there could be **gravitational waves**. That is, space itself could support waves: space itself could be "wavy" as waves moved along. Any wave needs a source, such as a rock thrown into a lake causes circular waves to move outward from the rock's point of entry. See photos in Fig. 15–1, page 428.

Gravitational waves are expected to be very tiny ripples in the fabric of space-time; tiny and difficult to detect. Very sensitive instruments have been developed over the last few decades to detect gravitational waves, and use them as a tool to understand the universe.

In early 2016, their direct detection was announced by LIGO, the Laser Interferometer Gravitational-wave Observatory. How did they do it? They used very large interferometers. Recall the Michelson interferometer with 2 perpendicular arms (Section 34–6, Fig. 34–24) which could sit on a large table-top and was used to detect (unsuccessfully) the "ether" assumed to be the medium to carry light waves. See the Special Theory of Relativity, especially Section 36–2, and Fig. 36–3.

LIGO uses interferometers whose two perpendicular arms are vastly longer, 4 km each. A passing gravitational wave changes the relative lengths of the two arms and causes a phase difference in the light beams (from a laser source) traveling along the two arms. Because LIGO has large interferometers at two distant locations (Louisiana and Washington state), and both detected a $\frac{1}{4}$-second long pulse of gravitational waves pass by on September 14, 2015, they could estimate (by triangulation—see Section 1–6 and Fig. 44–11) that the source of the tiny ripples they detected was 1.3 billion light-years away.

What was the source? Analysis of the frequencies present in the pulse allowed them to judge that 2 black holes orbiting each other had spiraled closer and closer until they collided and merged into one, releasing a tremendous amount of energy $\left(\approx 3M_{\text{Sun}}c^2\right)$. That energy warped space-time, like a rock striking a lake. The change in length of the interferometer arms was only about 10^{-18} m. Astonishing instruments!

To finally make a direct detection of gravitational waves was a triumph. But there is more. For the first time, the possibility that binary black holes exist, and could merge, was confirmed. (Like binary stars, they would orbit each other, slowly moving closer, and possibly merging given enough time.)

Additional large interferometers are already in place near Pisa, Italy (Virgo), and being constructed in India and Japan. With more, and more sensitive, detectors in place, gravitational waves should help us to observe details about the universe that would otherwise be invisible and unknown.

44–12 Finally . . .

When we look up into the night sky, we see stars; and with the best telescopes, we see galaxies and the exotic objects we discussed earlier, including rare supernovae. But even with our best instruments we do not see the processes going on inside stars and supernovae that we hypothesized (and believe to exist).

PSE5_CH44_v1.6_23Jun-20

We are dependent on brilliant theorists who come up with viable theories and verifiable models. We depend on complicated computer models whose parameters are varied until the outputs compare favorably with our observations and analyses of Planck or WMAP and other experiments. And we now have a surprisingly precise idea about some aspects of our universe: it is nearly or completely flat, it is about 14 billion years old, it contains only 5% "normal" baryonic matter (for atoms), and so on.

The questions raised by cosmology are difficult and profound, and may seem removed from everyday "reality." We can always say, "the Sun is shining, it's going to shine on for an unimaginably long time, all is well." Nonetheless, the questions of cosmology are deep ones that fascinate the human intellect. One aspect that is especially intriguing is this: calculations on the formation and evolution of the universe have been performed that deliberately varied the values—just slightly—of certain fundamental physical constants. The result would be a universe in which life as we know it and possibly life of any kind could not exist. [For example, if the difference in mass between a proton and a neutron were zero, or less than the mass of the electron, $0.511\,\text{MeV}/c^2$, there would be no atoms: electrons would be captured by protons to make neutrons.] Such results have contributed to a philosophical idea called the **anthropic principle**, which says that if the universe were even a little different than it is, we could not be here. We physicists are trying to find out if there are some undiscovered fundamental laws that determined those conditions that allowed us to exist. A poet might say that the universe is exquisitely tuned, almost as if to accommodate us.

Summary

The night sky contains myriads of stars including those in the Milky Way, which is a "side view" of our **Galaxy** looking along the plane of the disk. Our Galaxy includes over 10^{11} stars. Beyond our Galaxy are billions of other galaxies.

Astronomical distances are measured in **light-years** $(1\,\text{ly} \approx 10^{13}\,\text{km})$. The nearest star is about 4 ly away and the nearest large galaxy is 2 million ly away. Our Galactic disk has a diameter of about 100,000 ly. [Distances are sometimes specified in **parsecs**, where 1 parsec = 3.26 ly.]

Stars are believed to begin life as collapsing masses of gas (protostars), largely hydrogen. As they contract, they heat up (potential energy is transformed to kinetic energy). When the temperature reaches about 10 million degrees, nuclear fusion begins and forms heavier elements (**nucleosynthesis**), mainly helium at first. The energy released during these reactions heats the gas so its outward pressure balances the inward gravitational force, and the young star stabilizes as a **main-sequence** star. The tremendous luminosity of stars comes from the energy released during these thermonuclear reactions. After billions of years, as helium is collected in the core and hydrogen is used up, the core contracts and heats further. The outer envelope expands and cools, and the star becomes a **red giant** (larger diameter, redder color).

The next stage of stellar evolution depends on the mass of the star, which may have lost much of its original mass as its outer envelope escaped into space. Stars of residual mass less than about 1.4 solar masses cool further and become **white dwarfs**, eventually fading and going out altogether. Heavier stars contract further due to their greater gravity: the density approaches nuclear density, the huge pressure forces electrons to combine with protons to form neutrons, and the star becomes essentially a huge nucleus of neutrons. This is a **neutron star**, and the energy released during its final core collapse is believed to produce **supernova** explosions. If the star is very massive, it may contract even further and form a **black hole**, which is so dense that no matter or light can escape from it.

In the **general theory of relativity**, the **equivalence principle** states that an observer cannot distinguish acceleration from a gravitational field. Said another way, gravitational and inertial masses are the same. The theory predicts gravitational bending of light rays to a degree consistent with experiment. Gravity is treated as a curvature in space and time, the curvature being greater near massive objects. The universe as a whole may be curved. With sufficient mass (and no dark energy), the curvature of the universe would be positive, and the universe is *closed* and *finite*; otherwise, it would be *open* and *infinite*. Today we believe the universe is **flat**.

Distant galaxies display a **redshift** in their spectral lines, originally interpreted as a Doppler shift. The universe is observed to be **expanding**, its galaxies racing away from each other at speeds (v) proportional to the distance (d) between them:

$$v = H_0 d, \qquad \textbf{(44–4)}$$

which is known as **Hubble's law** $(H_0$ is the **Hubble constant**$)$. This expansion of the universe suggests an explosive origin, the **Big Bang**, which occurred about 13.8 billion years ago. It is not like an ordinary explosion, but rather an expansion of space itself.

The **cosmological principle** assumes that the universe, on a large scale, is homogeneous and isotropic.

Important evidence for the Big Bang model of the universe was the discovery of the **cosmic microwave background** radiation (CMB), which conforms to a blackbody radiation curve at a temperature of 2.725 K.

PSE5_CH44_v1.6_23Jun-20

The **Standard Model** of the Big Bang provides a possible scenario as to how the universe developed as it expanded and cooled after the Big Bang. Starting at 10^{-43} seconds after the Big Bang, according to this model, the universe underwent a brief but rapid exponential expansion, referred to as **inflation**. Shortly thereafter, quarks were **confined** into hadrons (the **hadron era**). About 10^{-4} s after the Big Bang, the majority of hadrons disappeared, having combined with anti-hadrons, producing photons, leptons, and energy, leaving mainly photons and leptons to freely move, thus introducing the **lepton era**. By the time the universe was about 10 s old, the electrons too had mostly disappeared, having combined with their antiparticles; the universe was **radiation-dominated**. A couple of minutes later, nucleosynthesis began, but lasted only a few minutes. It then took almost four hundred thousand years before the universe was cool enough for electrons to combine with nuclei to form atoms (**recombination**). Photons, up to then continually being scattered off of free electrons, could now move freely—they were **decoupled** from matter and the universe became transparent. The background radiation had expanded and cooled so much that its total energy became less than the energy in matter, and **matter dominated** increasingly over radiation. Then stars and galaxies formed, producing a universe not much different than it is today—some 14 billion years later.

Recent observations indicate that the universe is essentially flat, that it contains an as-yet unknown type of **dark matter**, and that it is dominated by a mysterious **dark energy** which exerts a sort of negative gravity causing the expansion of the universe to accelerate. The total contributions of baryonic (normal) matter, dark matter, and dark energy sum up to the **critical density**.

Questions

1. The Milky Way was once thought to be "murky" or "milky" but is now considered to be made up of point sources. Explain.
2. A star is in equilibrium when it radiates at its surface all the energy generated in its core. What happens when it begins to generate more energy than it radiates? Less energy? Explain.
3. Describe a red giant star. List some of its properties.
4. Does the H–R diagram directly reveal anything about the core of a star?
5. Why do some stars end up as white dwarfs, and others as neutron stars or black holes?
6. If you were measuring star parallaxes from the Moon instead of Earth, what corrections would you have to make? What changes would occur if you were measuring parallaxes from Mars?
7. *Cepheid variable* stars change in luminosity with a typical period of several days. The period has been found to have a definite relationship with the average intrinsic luminosity of the star. How could these stars be used to measure the distance to galaxies?
8. What is a geodesic? What is its role in General Relativity?
9. If it were discovered that the redshift of spectral lines of galaxies was due to something other than expansion, how might our view of the universe change? Would there be conflicting evidence? Discuss.
10. Almost all galaxies appear to be moving away from us. Are we therefore at the center of the universe? Explain.
11. If you were located in a galaxy near the boundary of our observable universe, would galaxies in the direction of the Milky Way appear to be approaching you or receding from you? Explain.
12. Compare an explosion on Earth to the Big Bang. Consider such questions as: Would the debris spread at a higher speed for more distant particles, as in the Big Bang? Would the debris come to rest? What type of universe would this correspond to, open or closed?
13. If nothing, not even light, escapes from a black hole, then how can we tell if one is there?
14. The Earth's age is often given as about 4.6 billion years. Find that time on Fig. 44–29. Modern humans have lived on Earth on the order of 200,000 years. Where is that on Fig. 44–29?
15. Why were atoms, as opposed to bare nuclei, unable to exist until hundreds of thousands of years after the Big Bang?
16. (*a*) Why are Type Ia supernovae so useful for determining the distances of galaxies? (*b*) How are their distances actually measured?
17. Under what circumstances would the universe eventually collapse in on itself?
18. (*a*) Why did astronomers expect that the expansion rate of the universe would be decreasing (decelerating) with time? (*b*) How, in principle, could astronomers hope to determine whether the universe used to expand faster than it does now?
19. Describe how we can estimate the distance from us to other stars. Which methods can we use for nearby stars, and which can we use for very distant stars? Which method gives the most accurate distance measurements for the most distant stars?
20. (*a*) Describe some of the evidence that the universe began with a "Big Bang." (*b*) How does the curvature of the universe affect its future destiny? (*c*) How does dark energy affect the possible future of the universe?
21. Explain what the 2.7-K cosmic microwave background radiation is. Where does it come from? Why is its temperature now so low?

PSE5_CH44_v1.6_23Jun-20

MisConceptual Questions

1. Which one of the following is *not* expected to occur on an H–R diagram during the lifetime of a single star?
   (*a*) The star will move off the main sequence toward the upper right of the diagram.
   (*b*) Low-mass stars will become white dwarfs and end up toward the lower left of the diagram.
   (*c*) The star will move along the main sequence from one place to another.
   (*d*) All of the above.
2. When can parallax be used to determine the approximate distance from the Earth to a star?
   (*a*) Only during January and July.
   (*b*) Only when the star's distance is relatively small.
   (*c*) Only when the star's distance is relatively large.
   (*d*) Only when the star appears to move directly toward or away from the Earth.
   (*e*) Only when the star is the Sun.
   (*f*) Always.
   (*g*) Never.
3. Observations show that all galaxies tend to move away from Earth, and that more distant galaxies move away from Earth at faster velocities than do galaxies closer to the Earth. These observations imply that
   (*a*) the Earth is the center of the universe.
   (*b*) the universe is expanding.
   (*c*) the expansion of the universe will eventually stop.
   (*d*) All of the above.
4. Which process results in a tremendous amount of energy being emitted by the Sun?
   (*a*) Hydrogen atoms burn in the presence of oxygen—that is, hydrogen atoms oxidize.
   (*b*) The Sun contracts, decreasing its gravitational potential energy.
   (*c*) Protons in hydrogen atoms fuse, forming helium nuclei.
   (*d*) Radioactive atoms such as uranium, plutonium, and cesium emit gamma rays with high energy.
   (*e*) None of the above.
5. Which of the following methods can be used to find the distance from us to a star outside our galaxy? Choose all that apply.
   (*a*) Parallax.
   (*b*) Using luminosity and temperature from the H–R diagram and measuring the apparent brightness.
   (*c*) Using supernova explosions as a "standard candle."
   (*d*) Redshift in the line spectra of elements and compounds.
6. The history of the universe can be determined by observing astronomical objects at various (large) distances from the Earth. This method of discovery works because
   (*a*) time proceeds at different rates in different regions of the universe.
   (*b*) light travels at a finite speed.
   (*c*) matter warps space.
   (*d*) older galaxies are farther from the Earth than are younger galaxies.
7. Where did the Big Bang occur?
   (*a*) Near the Earth.
   (*b*) Near the center of the Milky Way Galaxy.
   (*c*) Several billion light-years away.
   (*d*) Throughout all space.
   (*e*) Near the Andromeda Galaxy.
8. When and how were virtually all of the elements of the Periodic Table formed?
   (*a*) In the very early universe a few seconds after the Big Bang.
   (*b*) At the centers of stars during their main-sequence phases.
   (*c*) At the centers of stars during novae.
   (*d*) At the centers of stars during supernovae.
   (*e*) On the surfaces of planets as they cooled and hardened.
9. We know that there must be dark matter in the universe because
   (*a*) we see dark dust clouds.
   (*b*) we see that the universe is expanding.
   (*c*) we see that stars far from the galactic center are moving faster than can be explained by visible matter.
   (*d*) we see that the expansion of the universe is accelerating.
10. Acceleration of the universe's expansion rate is due to
    (*a*) the repulsive effect of dark energy.
    (*b*) the attractive effect of dark matter.
    (*c*) the attractive effect of gravity.
    (*d*) the thermal expansion of stellar cores.

Problems

44–1 to 44–3 Stars, Galaxies, Stellar Evolution, Distances

1. (I) The parallax angle of a star is 0.00035°. How far away is the star?
2. (I) A star exhibits a parallax of 0.23 seconds of arc. How far away is it?
3. (I) If one star is twice as far away from us as a second star, will the parallax angle of the farther star be greater or less than that of the nearer star? By what factor?
4. (II) What is the relative brightness of the Sun as seen from Jupiter, as compared to its brightness from Earth? (Jupiter is 5.2 times farther from the Sun than the Earth is.)
5. (II) When our Sun becomes a red giant, what will be its average density if it expands out to the orbit of Mercury (6×10^{10} m from the Sun)?

6. (II) We saw earlier (Chapter 19) that the rate energy reaches the Earth from the Sun (the "solar constant") is about $1.3 \times 10^3\ \text{W/m}^2$. What is (*a*) the apparent brightness *b* of the Sun, and (*b*) the intrinsic luminosity *L* of the Sun?

7. (II) Estimate the angular width that our Galaxy would subtend if observed from the nearest galaxy to us (Table 44–1). Compare to the angular width of the Moon from Earth.

8. (II) Assuming our Galaxy represents a good average for all other galaxies, how many stars are in the observable universe?

9. (II) Calculate the density of a white dwarf whose mass is equal to the Sun's and whose radius is equal to the Earth's. How many times larger than Earth's density is this?

10. (II) A neutron star whose mass is 1.5 solar masses has a radius of about 11 km. Calculate its average density and compare to that for a white dwarf (Problem 9) and to that of nuclear matter.

*11. (II) A star is 48 pc away. What is its parallax angle? State (*a*) in seconds of arc, and (*b*) in degrees.

*12. (II) What is the parallax angle for a star that is 74 ly away? How many parsecs is this?

*13. (II) A star is 77 pc away. How long does it take for its light to reach us?

14. (II) Calculate the *Q*-values for the He burning reactions of Eq. 44–2. (The mass of the very unstable ^8_4Be is 8.005305 u.)

15. (II) When our Sun becomes a while dwarf, it is expected to be about the size of the Moon. What angular width would it subtend from the present distance to Earth?

16. (II) Suppose that three main-sequence stars could undergo the three changes represented by the three arrows, A, B, and C, in the H–R diagram of Fig. 44–35. For each case, describe the changes in temperature, intrinsic luminosity, and size.

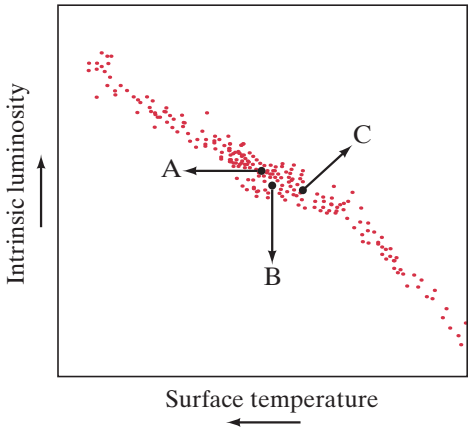


FIGURE 44–35 Problem 16.

17. (III) Suppose two stars of the same apparent brightness *b* are also believed to be the same size. The spectrum of one star peaks at 750 nm whereas that of the other peaks at 450 nm. Use Wien's law and the Stefan-Boltzmann equation (Eq. 19–17) to estimate their relative distances from us. [*Hint:* See Examples 44–4 and 44–5.]

18. (III) Stars located in a certain cluster are assumed to be about the same distance from us. Two such stars have spectra that peak at $\lambda_1 = 470\ \text{nm}$ and $\lambda_2 = 720\ \text{nm}$, and the ratio of their apparent brightness is $b_1/b_2 = 0.091$. Estimate their relative sizes (give ratio of their diameters) using Wien's law and the Stefan-Boltzmann equation, Eq. 19–17.

44–4 General Relativity, Gravity and Curved Space

19. (I) Show that the Schwarzschild radius for Earth is 8.9 mm.

20. (II) What is the Schwarzschild radius for a typical galaxy (like ours)?

21. (II) What mass will give a Schwarzschild radius equal to that of the hydrogen atom in its ground state?

22. (II) What is the maximum sum-of-the-angles for a triangle on a sphere?

23. (II) Describe a triangle, drawn on the surface of a sphere, for which the sum of the angles is (*a*) 359°, and (*b*) 179°.

24. (II) Calculate the escape velocity, using Newtonian mechanics, from an object that has collapsed to its Schwarzschild radius.

25. (II) What is the apparent deflection of a light beam in an elevator (Fig. 44–13) which is 3.2 m wide if the elevator is accelerating downward at $9.8\ \text{m/s}^2$?

44–5 Redshift, Hubble's Law

26. (I) The redshift of a galaxy indicates a recession velocity of 2250 km/s. How far away is it?

27. (I) If a galaxy is traveling away from us at 2.2% of the speed of light, roughly how far away is it?

28. (II) A galaxy is moving away from Earth. The "blue" hydrogen line at 434 nm emitted from the galaxy is measured on Earth to be 455 nm. (*a*) How fast is the galaxy moving? (*b*) How far is it from Earth based on Hubble's law?

29. (II) Estimate the wavelength shift for the 656.3-nm line in the Balmer series of hydrogen emitted from a galaxy whose distance from us is (*a*) 6.0×10^6 ly, (*b*) 6.0×10^7 ly.

30. (II) If an absorption line of calcium is normally found at a wavelength of 393.4 nm in a laboratory gas, and you measure it to be at 419.7 nm in the spectrum of a galaxy, what is the approximate distance to the galaxy?

31. (II) What is the speed of a galaxy with $z = 0.068$?

32. (II) What would be the redshift parameter *z* for a galaxy traveling away from us at $v = 0.068c$?

33. (II) Estimate the distance *d* from the Earth to a galaxy whose redshift parameter $z = 1$.

34. (II) Estimate the speed of a galaxy, and its distance from us, if the wavelength for the hydrogen line at 434 nm is measured on Earth as being 625 nm.

35. (II) Radiotelescopes are designed to observe 21-cm waves emitted by atomic hydrogen gas. A signal from a distant radio-emitting galaxy is found to have a wavelength that is 0.10 cm longer than the normal 21-cm wavelength. Estimate the distance to this galaxy.

36. (III) Starting from Eq. 44–3, show that the Doppler shift in wavelength is $\Delta\lambda/\lambda_{\text{rest}} \approx v/c$ (Eq. 44–6) for $v \ll c$. [*Hint:* Use the binomial expansion.]

44–6 to 44–8 The Big Bang, CMB, Universe Expansion

37. (I) Calculate the wavelength at the peak of the blackbody radiation distribution at 2.7 K using Wien's law.

38. (II) (*a*) What temperature would correspond to 13-TeV collisions at the LHC? (*b*) Estimate the time after the Big Bang probed by this energy. (*c*) To what era in cosmological history does this correspond? [*Hint*: See Fig. 44–29.]

39. (II) Calculate the peak wavelength of the CMB at 1.0 s after the birth of the universe. In what part of the EM spectrum is this radiation?

40. (II) The critical density for closure of the universe is $\rho_c \approx 10^{-26}\,\text{kg/m}^3$ (see Example 44–7). State ρ_c in terms of the average number of nucleons per cubic meter.

41. (II) The scale factor of the universe (average distance between galaxies) at any given time is believed to have been inversely proportional to the absolute temperature. Estimate the size of the universe, compared to today, at (*a*) $t = 10^6$ yr, (*b*) $t = 1$ s, (*c*) $t = 10^{-6}$ s, and (*d*) $t = 10^{-35}$ s.

42. (II) At approximately what time had the universe cooled below the threshold temperature for producing (*a*) kaons $(M \approx 500\,\text{MeV}/c^2)$, (*b*) Υ $(M \approx 9500\,\text{MeV}/c^2)$, and (*c*) muons $(M \approx 100\,\text{MeV}/c^2)$?

44–9 Dark Matter, Dark Energy

43. (II) Only about 5% of the energy in the universe is composed of baryonic matter. (*a*) Estimate the average density of baryonic matter in the observable universe with a radius of 14 billion light-years that contains 10^{11} galaxies, each with about 10^{11} stars like our Sun. (*b*) Estimate the density of dark matter in the universe.

44–10 Structure of Universe

44. (II) Two identical black holes form a binary system and are orbiting one another. Assume they are a distance apart which is twice the Schwartzchild radius in each. Then, assuming Newton mechanics is still valid, how fast are they moving with respect to the center of mass?

General Problems

45. Use conservation of angular momentum to estimate the angular velocity of a neutron star which has collapsed to a diameter of 16 km, from a star whose core radius was equal to that of Earth $(6 \times 10^6\,\text{m})$. Assume its mass is 1.5 times that of the Sun, and that it rotated (like our Sun) about once a month.

46. By what factor does the rotational kinetic energy change when the star in Problem 45 collapses to a neutron star?

47. Assume that the nearest stars to us have an intrinsic luminosity about the same as the Sun's. Their apparent brightness, however, is about 10^{11} times fainter than the Sun. From this, estimate the distance to the nearest stars.

48. A certain pulsar, believed to be a neutron star of mass 1.5 times that of the Sun, with diameter 16 km, is observed to have a rotation speed of 1.0 rev/s. If it loses rotational kinetic energy at the rate of 1 part in 10^9 per day, which is all transformed into radiation, what is the power output of the star?

49. The nearest large galaxy to our Galaxy is about 2×10^6 ly away. If both galaxies have a mass of 4×10^{41} kg, with what gravitational force does each galaxy attract the other? Ignore dark matter.

50. How large would the Sun be if it had its current mass but had a density equal to the critical density of the universe, $\rho_c \approx 10^{-26}\,\text{kg/m}^3$? Express your answer in light-years and compare with the Earth–Sun distance and the diameter of our Galaxy.

51. Two stars, whose spectra peak at 660 nm and 480 nm, respectively, both lie on the main sequence. Use Wien's law, the Stefan-Boltzmann equation, and the H–R diagram (Fig. 44–6) to estimate the ratio of their diameters.

52. (*a*) In order to measure distances with parallax at 100 ly, what minimum angular resolution (in degrees) is needed? (*b*) What diameter mirror or lens would be needed?

53. In the later stages of stellar evolution, a star (if massive enough) will begin fusing carbon nuclei to form, for example, magnesium:

$$^{12}_{6}\text{C} + {}^{12}_{6}\text{C} \rightarrow {}^{24}_{12}\text{Mg} + \gamma.$$

(*a*) How much energy is released in this reaction (see Appendix G)? (*b*) How much kinetic energy must each carbon nucleus have (assume equal) in a head-on collision if they are just to "touch" (use Eq. 41–1) so that the strong force can come into play? (*c*) What temperature does this kinetic energy correspond to?

54. Consider the reaction

$$^{16}_{8}\text{O} + {}^{16}_{8}\text{O} \rightarrow {}^{28}_{14}\text{Si} + {}^{4}_{2}\text{He},$$

and answer the same questions as in Problem 53.

55. Use *dimensional analysis* with the fundamental constants c, G, and $\hbar$ to estimate the value of the so-called *Planck time*. It is thought that physics as we know it can say nothing about the universe before this time.

56. Estimate the mass of our observable universe using the following assumptions: Our universe is spherical in shape, it has been expanding at the speed of light since the Big Bang, and its density is the critical density.

57. Estimate what neutrino mass $(\text{in eV}/c^2)$ would provide the critical density to close the universe. Assume the neutrino density is, like photons, about 10^9 times that of nucleons, and that nucleons make up only (*a*) 2% of the mass needed, or (*b*) 5% of the mass needed.

58. The evolution of stars, as discussed in Section 44–2, can lead to a white dwarf, a neutron star, or even a black hole, depending on the mass. (*a*) Referring to Sections 44–2 and 44–4, give the radius of (i) a white dwarf of 1 solar mass, (ii) a neutron star of 1.5 solar masses, and (iii) a black hole of 3 solar masses. (*b*) Express these three radii as ratios $(r_i : r_{ii} : r_{iii})$.

59. When stable nuclei first formed, about 3 minutes after the Big Bang, there were about 7 times more protons than neutrons. Explain how this leads to a ratio of the mass of hydrogen to the mass of helium of 3:1. This is about the actual ratio observed in the universe.

60. Astronomers use an **apparent magnitude** (m) scale to describe apparent brightness. It uses a logarithmic scale, where a higher number corresponds to a less bright star. (For example, the Sun has magnitude -27, whereas most stars have positive magnitudes.) On this scale, a change in apparent magnitude by $+5$ corresponds to a decrease in apparent brightness by a factor of 100. If Venus has an apparent magnitude of -4.4, whereas Sirius has an apparent magnitude of -1.4, which is brighter? What is the ratio of the apparent brightness of these two objects?

61. Astronomers have measured the rotation of gas around a possible supermassive black hole of about 2 billion solar masses at the center of a galaxy. If the radius from the galactic center to the gas clouds is 68 light-years, estimate the value of z.

62. Calculate the Schwarzschild radius using a semi-classical (Newtonian) gravitational theory, by calculating the minimum radius R for a sphere of mass M such that a photon can escape from the surface. (General Relativity gives $R = 2GM/c^2$.)

63. (*a*) Use special relativity and Newton's law of gravitation to show that a photon of mass $m = E/c^2$ just grazing the Sun will be deflected by an angle $\Delta\theta$ given by

$$\Delta\theta = \frac{2GM}{c^2R}$$

where G is the gravitational constant, R and M are the radius and mass of the Sun, and c is the speed of light. (*b*) Put in values and show $\Delta\theta = 0.87''$. (General Relativity predicts an angle twice as large, $1.74''$.)

64. We cannot use Hubble's law to measure the distances to nearby galaxies, because their random motions are larger than the overall expansion. Indeed, the closest galaxy to us, the Andromeda Galaxy, 2.5 million light-years away, is approaching us at a speed of about 130 km/s. (*a*) What is the shift in wavelength of the 656-nm line of hydrogen emitted from the Andromeda Galaxy, as seen by us? (*b*) Is this a redshift or a blueshift? (*c*) Ignoring the expansion, how soon will it and the Milky Way Galaxy collide?

65. Estimate the radius of a white dwarf whose mass is equal to that of the Sun by the following method, assuming there are N nucleons and $\frac{1}{2}N$ electrons (why $\frac{1}{2}$?): (*a*) Use Fermi-Dirac statistics (Section 40–6) to show that the total energy of all the electrons is

$$E_e = \frac{3}{5}\left(\frac{1}{2}N\right)\frac{h^2}{8m_e}\left(\frac{3}{\pi}\frac{N}{2V}\right)^{\frac{2}{3}}.$$

[*Hint:* See Eqs. 40–12 and 40–13; we assume electrons fill energy levels from 0 up to the Fermi energy.] (*b*) The nucleons contribute to the total energy mainly via the gravitational force (note that the Fermi energy for nucleons is negligible compared to that for electrons—why?). Use a gravitational form of Gauss's law to show that the total gravitational potential energy of a uniform sphere of radius R is

$$-\frac{3}{5}\frac{GM^2}{R},$$

by considering the potential energy of a spherical shell of radius r due only to the mass inside it (why?) and integrate from $r = 0$ to $r = R$. (See also Appendix E.) (*c*) Write the total energy as a sum of these two terms, and set $dE/dR = 0$ to find the equilibrium radius, and evaluate it for a mass equal to the Sun's $(2.0 \times 10^{30}\,\text{kg})$.

66. Determine the radius of a neutron star using the same argument as in Problem 65 but for N neutrons only. Show that the radius of a neutron star, of 1.5 solar masses, is about 11 km.

ANSWERS TO EXERCISES

A: Our Earth and ourselves, 2 years ago.

B: 600 ly (estimating L from Fig. 44–6 as $L \approx 8 \times 10^{26}\,\text{W}$; note that on a log scale, 6000 K is closer to 7000 K than it is to 5000 K).

C: 30 km.

D: (*a*); not the usual R^3, but R: see formula for the Schwarzschild radius.

E: (*c*); (*d*).

APPENDIX

A Mathematical Formulas

A–1 Quadratic Formula

If $ax^2 + bx + c = 0$

then $$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

A–2 Binomial Expansion†

$$(1 \pm x)^n = 1 \pm nx + \frac{n(n-1)}{2!}x^2 \pm \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$$

$$\approx 1 \pm nx \qquad (\text{if } x \ll 1) \qquad \text{[Binomial approximation]}$$

$$(x + y)^n = x^n\left(1 + \frac{y}{x}\right)^n = x^n\left(1 + n\frac{y}{x} + \frac{n(n-1)}{2!}\frac{y^2}{x^2} + \cdots\right)$$

A–3 Other Expansions†

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots \qquad [\theta \text{ in radians}]$$

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots \qquad [\theta \text{ in radians}]$$

$$\tan\theta = \theta + \frac{\theta^3}{3} + \frac{2}{15}\theta^5 + \cdots \qquad \left[|\theta| < \frac{\pi}{2}\text{ rad}\right]$$

In general: $$f(x) = f(0) + \left(\frac{df}{dx}\right)_0 x + \left(\frac{d^2f}{dx^2}\right)_0 \frac{x^2}{2!} + \cdots$$

A–4 Exponents

$$(a^n)(a^m) = a^{n+m} \qquad \frac{1}{a^n} = a^{-n}$$

$$(a^n)(b^n) = (ab)^n \qquad a^n a^{-n} = a^0 = 1$$

$$(a^n)^m = a^{nm} \qquad a^{\frac{1}{2}} = \sqrt{a}$$

[Any number ($\neq 0$) raised to the 0^{th} power is 1.]

A–5 Areas and Volumes

Object	Surface area	Volume
Circle, radius r	πr^2	—
Sphere, radius r	$4\pi r^2$	$\frac{4}{3}\pi r^3$
Right circular cylinder, radius r, height h	$2\pi r^2 + 2\pi rh$	$\pi r^2 h$
Right circular cone, radius r, height h	$\pi r^2 + \pi r\sqrt{r^2 + h^2}$	$\frac{1}{3}\pi r^2 h$

†The exclamation point used after a number, such as 4!, is called a **factorial** number, and means you multiply that number by all the integers smaller than it. That is, $4! = 4 \times 3 \times 2 \times 1 = 24$ and $3! = 3 \times 2 \times 1 = 6$. The quantity x in all these expansions should be unitless, just like the 1.

A–6 Plane Geometry

1. *Equal angles*:

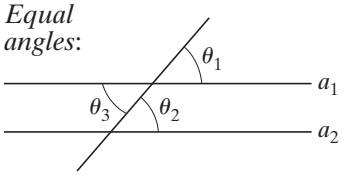


FIGURE A–1 If line a_1 is parallel to line a_2, then $\theta_1 = \theta_2 = \theta_3$.

2. *Equal angles*:

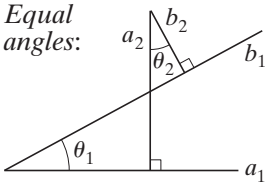


FIGURE A–2 If $a_1 \perp a_2$ and $b_1 \perp b_2$, then $\theta_1 = \theta_2$.

3. The sum of the angles in any plane triangle is 180°.

4. *Pythagorean theorem*:

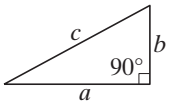


FIGURE A–3

In any right triangle (one angle = 90°) of sides a, b, and c:

$$a^2 + b^2 = c^2$$

where c is the length of the hypotenuse (opposite the 90° angle).

5. *Similar triangles*: Two triangles are said to be similar if all three of their angles are equal (in Fig. A–4, $\theta_1 = \phi_1$, $\theta_2 = \phi_2$, and $\theta_3 = \phi_3$). Similar triangles can have different sizes and different orientations.

(a) Two triangles are similar if any two of their angles are equal. (This follows because the third angles must also be equal since the sum of the angles of a triangle is 180°.)

(b) The ratios of corresponding sides of two similar triangles are equal (Fig. A–4):

$$\frac{a_1}{b_1} = \frac{a_2}{b_2} = \frac{a_3}{b_3}.$$

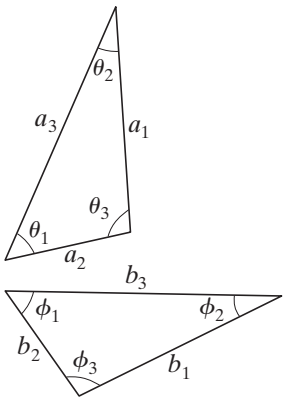


FIGURE A–4

6. *Congruent triangles*: Two triangles are congruent if one can be placed precisely on top of the other. That is, they are similar triangles and they have the same size. Two triangles are congruent if any of the following holds:

(a) The three corresponding sides are equal.

(b) Two sides and the enclosed angle are equal ("side-angle-side").

(c) Two angles and the enclosed side are equal ("angle-side-angle").

A–7 Logarithms

Logarithms are defined in the following way:

$$\text{if } \quad y = A^x, \quad \text{then} \quad x = \log_A y.$$

That is, the logarithm of a number y to the base A is that number which, as the exponent of A, gives back the number y. For **common logarithms**, the base is 10, so

$$\text{if } \quad y = 10^x, \quad \text{then} \quad x = \log y.$$

The subscript 10 on $\log_{10}$ is usually omitted when dealing with common logs. Another important base is the exponential base $e = 2.718\ldots$, a natural number (originating in the integral $\int \frac{1}{x}dx = \ln x$, Section B–4). Such logarithms are called **natural logarithms** and are written ln. Thus,

$$\text{if } \quad y = e^x, \quad \text{then} \quad x = \ln y.$$

For either type, we can take the logarithm of only positive numbers,[†] $y > 0$. For any number y, the two types of logarithm are related by

$$\ln y = 2.3026 \log y.$$

Some simple rules for logarithms are as follows:

$$\log(ab) = \log a + \log b, \tag{i}$$

which is true because if $a = 10^n$ and $b = 10^m$, then $ab = 10^{n+m}$. From the

[†]If $y < 0$, no value of x satisfies $y = 10^x$ (try it) nor $y = e^x$. And if $y = 0$, then x must be $-\infty$, which is hardly useful. Thus we can take $\log y$ only if $y > 0$. Note also that both y and x should in principle be *dimensionless*, without units, just like θ in $\sin\theta$ or $\cos\theta$. Consider the expansions in Appendix A–3 with terms in x, x^2, x^3, etc., whose units would be inconsistent and nonsense.

definition of logarithm, $\log a = n$, $\log b = m$, and $\log(ab) = n + m$; hence, $\log(ab) = n + m = \log a + \log b$. In a similar way, we can show that

$$\log a^n = n \log a \tag{ii}$$

and

$$\log\left(\frac{a}{b}\right) = \log a - \log b \tag{iii}$$

and also, since $\log(b/a) = \log b - \log a$, then

$$\log\left(\frac{b}{a}\right) = -\log\left(\frac{a}{b}\right). \tag{iv}$$

These four rules apply to any kind of logarithm.

If you do not have a calculator that calculates logs, you can easily use a **log table**, such as the small one shown here (Table A–1): the number N whose log we want is given to two digits. The first digit is in the vertical column to the left, and the second digit is in the horizontal row across the top. For example, Table A–1 tells us that $\log 1.0 = 0.000$, $\log 1.1 = 0.041$, and $\log 4.1 = 0.613$. The Table gives logs for numbers between 1.0 and 9.9. For larger or smaller numbers, we use rule (i) above, $\log(ab) = \log a + \log b$. For example, $\log(380) = \log(3.8 \times 10^2) = \log(3.8) + \log(10^2)$. From the Table, $\log 3.8 = 0.580$; and from rule (ii) above $\log(10^2) = 2\log(10) = 2$, since $\log(10) = 1$ by the definition of the logarithm: if $10 = 10^1$, then $1 = \log(10)$. Thus,

$$\begin{aligned}\log(380) &= \log(3.8) + \log(10^2)\\ &= 0.580 + 2\\ &= 2.580.\end{aligned}$$

Similarly,

$$\begin{aligned}\log(0.081) &= \log(8.1) + \log(10^{-2})\\ &= 0.908 - 2 = -1.092.\end{aligned}$$

The reverse process of finding the number N whose log is, say, 2.670, is called "taking the **antilogarithm**." To do so, we separate our number 2.670 into two parts, making the separation at the decimal point:

$$\begin{aligned}\log N = 2.670 &= 2 + 0.670\\ &= \log(10^2) + 0.670.\end{aligned}$$

We now look at Table A–1 to see what number has its log equal to 0.670; none does, so we must **interpolate**: we see that $\log 4.6 = 0.663$ and $\log 4.7 = 0.672$. So the number we want is between 4.6 and 4.7, and closer to the latter by $\frac{7}{9}$, or 4.68. Approximately we can say that $\log 4.68 = 0.670$. Thus

$$\begin{aligned}\log N &= 2 + 0.670\\ &= \log(10^2) + \log(4.68) = \log(4.68 \times 10^2),\end{aligned}$$

so $N = 4.68 \times 10^2 = 468$.

If the given logarithm is negative, say, -2.180, we proceed as follows:

$$\begin{aligned}\log N = -2.180 &= -3 + 0.820\\ &= \log(10^{-3}) + \log(6.6) = \log(6.6 \times 10^{-3}),\end{aligned}$$

so $N = 6.6 \times 10^{-3}$. Notice that we added to our given logarithm the next largest integer (3 in this case) so that we have an integer, plus a decimal number between 0 and 1.0 whose antilogarithm can be looked up in the Table.

TABLE A–1 Short Table of Common Logarithms

N	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.000	0.041	0.079	0.114	0.146	0.176	0.204	0.230	0.255	0.279
2	0.301	0.322	0.342	0.362	0.380	0.398	0.415	0.431	0.447	0.462
3	0.477	0.491	0.505	0.519	0.531	0.544	0.556	0.568	0.580	0.591
4	0.602	0.613	0.623	0.633	0.643	0.653	0.663	0.672	0.681	0.690
5	0.699	0.708	0.716	0.724	0.732	0.740	0.748	0.756	0.763	0.771
6	0.778	0.785	0.792	0.799	0.806	0.813	0.820	0.826	0.833	0.839
7	0.845	0.851	0.857	0.863	0.869	0.875	0.881	0.886	0.892	0.898
8	0.903	0.908	0.914	0.919	0.924	0.929	0.934	0.940	0.944	0.949
9	0.954	0.959	0.964	0.968	0.973	0.978	0.982	0.987	0.991	0.996

A–8 Vectors: see Sections 3–2 to 3–5, 7–2, 11–2.

A–9 Trigonometric Functions and Identities

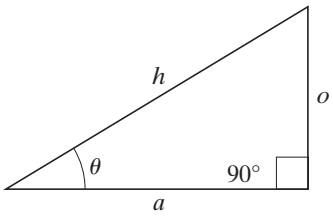


FIGURE A–5

The trigonometric functions are defined as follows (see Fig. A–5, o = side opposite, a = side adjacent, h = hypotenuse. Values are given in Table A–2):

$$\sin\theta = \frac{o}{h} \qquad \csc\theta = \frac{1}{\sin\theta} = \frac{h}{o}$$

$$\cos\theta = \frac{a}{h} \qquad \sec\theta = \frac{1}{\cos\theta} = \frac{h}{a}$$

$$\tan\theta = \frac{o}{a} = \frac{\sin\theta}{\cos\theta} \qquad \cot\theta = \frac{1}{\tan\theta} = \frac{a}{o}$$

and recall that

$$a^2 + o^2 = h^2 \qquad \text{[\textbf{Pythagorean theorem}].}$$

FIGURE A–6

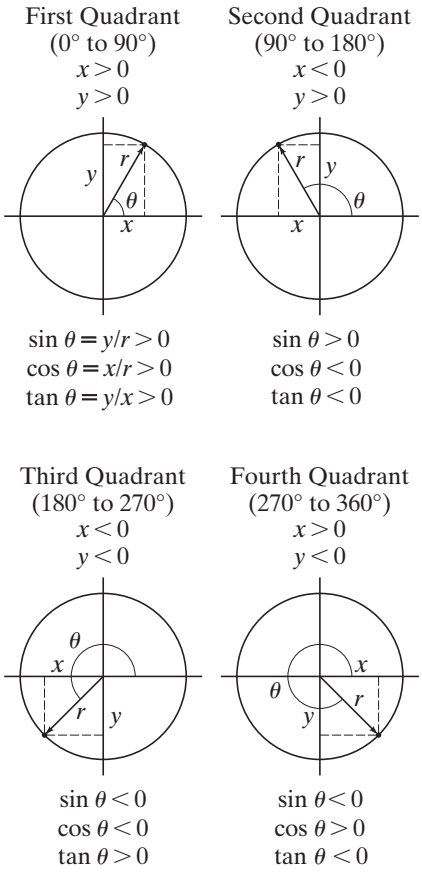


Figure A–6 shows the signs (+ or −) that cosine, sine, and tangent take on for angles θ in the four quadrants (0° to 360°, or 0 to 2π radians). Note that angles are measured counterclockwise from the $+x$ axis as shown; negative angles are measured clockwise from *below* the $+x$ axis: for example, $-30° = +330°$, and so on.

The following are some useful identities among the trigonometric functions:

$$\sin^2\theta + \cos^2\theta = 1$$

$$\sec^2\theta - \tan^2\theta = 1, \qquad \csc^2\theta - \cot^2\theta = 1$$

$$\sin 2\theta = 2\sin\theta\cos\theta$$

$$\cos 2\theta = \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$$

$$\tan 2\theta = \frac{2\tan\theta}{1 - \tan^2\theta}$$

$$\sin(A \pm B) = \sin A\cos B \pm \cos A\sin B$$

$$\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B$$

$$\tan(A \pm B) = \frac{\tan A \pm \tan B}{1 \mp \tan A\tan B}$$

$$\sin(180° - \theta) = \sin\theta \qquad [\text{in radians, } \sin(\pi - \theta) = \sin\theta]$$

$$\cos(180° - \theta) = -\cos\theta$$

$$\sin(90° - \theta) = \cos\theta$$

$$\cos(90° - \theta) = \sin\theta$$

$$\sin(-\theta) = -\sin\theta$$

$$\cos(-\theta) = \cos\theta$$

$$\tan(-\theta) = -\tan\theta$$

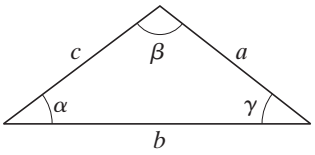
$$\sin\tfrac{1}{2}\theta = \sqrt{\frac{1 - \cos\theta}{2}}, \quad \cos\tfrac{1}{2}\theta = \sqrt{\frac{1 + \cos\theta}{2}}, \quad \tan\tfrac{1}{2}\theta = \sqrt{\frac{1 - \cos\theta}{1 + \cos\theta}}$$

$$\sin A \pm \sin B = 2\sin\left(\frac{A \pm B}{2}\right)\cos\left(\frac{A \mp B}{2}\right).$$

$$\cos A + \cos B = 2\cos\left(\frac{A + B}{2}\right)\cos\left(\frac{A - B}{2}\right)$$

$$\cos A - \cos B = -2\sin\left(\frac{A + B}{2}\right)\sin\left(\frac{A - B}{2}\right)$$

FIGURE A–7



For any triangle (see Fig. A–7):

$$\frac{\sin\alpha}{a} = \frac{\sin\beta}{b} = \frac{\sin\gamma}{c} \qquad \text{[\textbf{Law of sines}]}$$

$$c^2 = a^2 + b^2 - 2ab\cos\gamma. \qquad \text{[\textbf{Law of cosines}]}$$

Values of sine, cosine, tangent are given in Table A–2.

TABLE A–2 Trigonometric Table: Numerical Values of Sin, Cos, Tan

Angle in Degrees	Angle in Radians	Sine	Cosine	Tangent	Angle in Degrees	Angle in Radians	Sine	Cosine	Tangent
0°	0.000	0.000	1.000	0.000					
1°	0.017	0.017	1.000	0.017	46°	0.803	0.719	0.695	1.036
2°	0.035	0.035	0.999	0.035	47°	0.820	0.731	0.682	1.072
3°	0.052	0.052	0.999	0.052	48°	0.838	0.743	0.669	1.111
4°	0.070	0.070	0.998	0.070	49°	0.855	0.755	0.656	1.150
5°	0.087	0.087	0.996	0.087	50°	0.873	0.766	0.643	1.192
6°	0.105	0.105	0.995	0.105	51°	0.890	0.777	0.629	1.235
7°	0.122	0.122	0.993	0.123	52°	0.908	0.788	0.616	1.280
8°	0.140	0.139	0.990	0.141	53°	0.925	0.799	0.602	1.327
9°	0.157	0.156	0.988	0.158	54°	0.942	0.809	0.588	1.376
10°	0.175	0.174	0.985	0.176	55°	0.960	0.819	0.574	1.428
11°	0.192	0.191	0.982	0.194	56°	0.977	0.829	0.559	1.483
12°	0.209	0.208	0.978	0.213	57°	0.995	0.839	0.545	1.540
13°	0.227	0.225	0.974	0.231	58°	1.012	0.848	0.530	1.600
14°	0.244	0.242	0.970	0.249	59°	1.030	0.857	0.515	1.664
15°	0.262	0.259	0.966	0.268	60°	1.047	0.866	0.500	1.732
16°	0.279	0.276	0.961	0.287	61°	1.065	0.875	0.485	1.804
17°	0.297	0.292	0.956	0.306	62°	1.082	0.883	0.469	1.881
18°	0.314	0.309	0.951	0.325	63°	1.100	0.891	0.454	1.963
19°	0.332	0.326	0.946	0.344	64°	1.117	0.899	0.438	2.050
20°	0.349	0.342	0.940	0.364	65°	1.134	0.906	0.423	2.145
21°	0.367	0.358	0.934	0.384	66°	1.152	0.914	0.407	2.246
22°	0.384	0.375	0.927	0.404	67°	1.169	0.921	0.391	2.356
23°	0.401	0.391	0.921	0.424	68°	1.187	0.927	0.375	2.475
24°	0.419	0.407	0.914	0.445	69°	1.204	0.934	0.358	2.605
25°	0.436	0.423	0.906	0.466	70°	1.222	0.940	0.342	2.747
26°	0.454	0.438	0.899	0.488	71°	1.239	0.946	0.326	2.904
27°	0.471	0.454	0.891	0.510	72°	1.257	0.951	0.309	3.078
28°	0.489	0.469	0.883	0.532	73°	1.274	0.956	0.292	3.271
29°	0.506	0.485	0.875	0.554	74°	1.292	0.961	0.276	3.487
30°	0.524	0.500	0.866	0.577	75°	1.309	0.966	0.259	3.732
31°	0.541	0.515	0.857	0.601	76°	1.326	0.970	0.242	4.011
32°	0.559	0.530	0.848	0.625	77°	1.344	0.974	0.225	4.331
33°	0.576	0.545	0.839	0.649	78°	1.361	0.978	0.208	4.705
34°	0.593	0.559	0.829	0.675	79°	1.379	0.982	0.191	5.145
35°	0.611	0.574	0.819	0.700	80°	1.396	0.985	0.174	5.671
36°	0.628	0.588	0.809	0.727	81°	1.414	0.988	0.156	6.314
37°	0.646	0.602	0.799	0.754	82°	1.431	0.990	0.139	7.115
38°	0.663	0.616	0.788	0.781	83°	1.449	0.993	0.122	8.144
39°	0.681	0.629	0.777	0.810	84°	1.466	0.995	0.105	9.514
40°	0.698	0.643	0.766	0.839	85°	1.484	0.996	0.087	11.430
41°	0.716	0.656	0.755	0.869	86°	1.501	0.998	0.070	14.301
42°	0.733	0.669	0.743	0.900	87°	1.518	0.999	0.052	19.081
43°	0.750	0.682	0.731	0.933	88°	1.536	0.999	0.035	28.636
44°	0.768	0.695	0.719	0.966	89°	1.553	1.000	0.017	57.290
45°	0.785	0.707	0.707	1.000	90°	1.571	1.000	0.000	∞

APPENDIX

B Derivatives and Integrals

B–1 Derivatives: General Rules

(See also Section 2–3.)

$$\frac{dx}{dx} = 1$$

$$\frac{d}{dx}[af(x)] = a\frac{df}{dx} \qquad [a = \text{constant}]$$

$$\frac{d}{dx}[f(x) + g(x)] = \frac{df}{dx} + \frac{dg}{dx}$$

$$\frac{d}{dx}[f(x)g(x)] = \frac{df}{dx}g + f\frac{dg}{dx}$$

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g\frac{df}{dx} - f\frac{dg}{dx}}{g^2}$$

$$\frac{d}{dx}\left[\frac{1}{g(x)}\right] = -\frac{1}{g^2}\frac{dg}{dx}$$

$$\frac{d}{dx}[f(y)] = \frac{df}{dy}\frac{dy}{dx} \qquad \textbf{[chain rule]}$$

$$\frac{dx}{dy} = \frac{1}{\left(\frac{dy}{dx}\right)} \qquad \left[\text{if } \frac{dy}{dx} \neq 0.\right]$$

B–2 Derivatives: Particular Functions

$$\frac{da}{dx} = 0 \qquad [a = \text{constant}]$$

$$\frac{d}{dx}x^n = nx^{n-1}$$

$$\frac{d}{dx}\sin ax = a\cos ax$$

$$\frac{d}{dx}\cos ax = -a\sin ax$$

$$\frac{d}{dx}\tan ax = a\sec^2 ax = \frac{a}{\cos^2 ax}$$

$$\frac{d}{dx}\ln ax = \frac{1}{x}$$

$$\frac{d}{dx}e^{ax} = ae^{ax}$$

B–3 Indefinite Integrals: General Rules

(See also Section 7–3.)

$$\int dx = x$$

$$\int af(x)\,dx = a\int f(x)\,dx \qquad [a = \text{constant}]$$

$$\int [f(x) + g(x)]dx = \int f(x)\,dx + \int g(x)\,dx$$

$$\int u\,dv = uv - \int v\,du \qquad \textbf{[integration by parts}: \text{see also Section B–6]}$$

B–4 Indefinite Integrals: Particular Functions

(An arbitrary constant can be added to the right side of each equation.)

$$\int a\,dx = ax \qquad [a = \text{constant}]$$

$$\int x^m\,dx = \frac{1}{m+1}x^{m+1} \qquad [m \neq -1]$$

$$\int \sin ax\,dx = -\frac{1}{a}\cos ax$$

$$\int \cos ax\,dx = \frac{1}{a}\sin ax$$

$$\int \tan ax\,dx = -\frac{1}{a}\ln|\cos ax|$$

$$\int \sin^2 ax\,dx = \frac{x}{2} - \frac{\sin 2ax}{4a}$$

$$\int \frac{1}{x}\,dx = \ln x \qquad [x > 0]^{\dagger}$$

$$\int \frac{1}{c+ax}\,dx = \frac{1}{a}\ln(c+ax) \qquad [(c+ax) > 0]$$

$$\int e^{ax}\,dx = \frac{1}{a}e^{ax}$$

$$\int \frac{dx}{x^2+a^2} = \frac{1}{a}\tan^{-1}\frac{x}{a} \qquad [a \neq 0]$$

$$\int \frac{dx}{x^2-a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| \qquad [a \neq 0,\ x \neq \pm a]$$

$$\int \frac{dx}{\sqrt{x^2 \pm a^2}} = \ln\left(x + \sqrt{x^2 \pm a^2}\right)$$

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \sin^{-1}\left(\frac{x}{a}\right) \qquad [\text{if } x^2 \leq a^2]$$

$$\int \frac{dx}{(x^2 \pm a^2)^{\frac{3}{2}}} = \frac{\pm x}{a^2\sqrt{x^2 \pm a^2}}$$

$$\int \frac{x\,dx}{(x^2 \pm a^2)^{\frac{3}{2}}} = \frac{-1}{\sqrt{x^2 \pm a^2}}$$

$$\int xe^{-ax}\,dx = -\frac{e^{-ax}}{a^2}(ax+1)$$

$$\int x^2e^{-ax}\,dx = -\frac{e^{-ax}}{a^3}(a^2x^2 + 2ax + 2)$$

[†]The natural logarithm, and thus also the natural number $e = 2.718\ldots$, originates from this integral.

B–5 A Few Definite Integrals $[a > 0]$

$$\int_0^\infty x^n e^{-ax}\,dx = \frac{n!}{a^{n+1}}$$

$$\int_0^\infty e^{-ax^2}\,dx = \sqrt{\frac{\pi}{4a}}$$

$$\int_0^\infty xe^{-ax^2}\,dx = \frac{1}{2a}$$

$$\int_0^\infty x^2e^{-ax^2}\,dx = \sqrt{\frac{\pi}{16a^3}}$$

$$\int_0^\infty x^3e^{-ax^2}\,dx = \frac{1}{2a^2}$$

$$\int_0^\infty x^{2n}e^{-ax^2}\,dx = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^{n+1}a^n}\sqrt{\frac{\pi}{a}} \qquad [n = 1, 2, 3, \cdots]$$

B–6 Integration by Parts

Sometimes a difficult integral can be simplified by carefully choosing the functions u and v in the identity:

$$\int u\,dv = uv - \int v\,du. \qquad \text{[Integration by parts]}$$

This identity follows from the property of derivatives

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

or as differentials: $d(uv) = u\,dv + v\,du$.

For example $\int xe^{-x}\,dx$ can be integrated by choosing $u = x$ and $dv = e^{-x}\,dx$ in the "integration by parts" equation above:

$$\int xe^{-x}\,dx = (x)(-e^{-x}) + \int e^{-x}\,dx$$
$$= -xe^{-x} - e^{-x} = -(x+1)e^{-x}.$$

APPENDIX C

Numerical Integration

This Appendix discusses how to solve certain Problems numerically, often needing a computer to do the sums. Some of this material is also covered in Chapter 7, Section 7–3. Here we discuss its use in kinematics (Chapter 2).

If we are given the velocity v of an object as a function of time t, we can obtain the displacement, x. Suppose the velocity as a function of time, $v(t)$, is given as a graph (rather than as an equation that could be integrated as discussed in Section 2–8), as shown in Fig. C–1a. If we are interested in the time interval from t_1 to t_2, as shown, we divide the time axis into many small subintervals, $\Delta t_1, \Delta t_2, \Delta t_3, \ldots$, which are indicated by the dashed vertical lines. For each subinterval, a horizontal dashed line is drawn to indicate the average velocity during that time interval. The displacement during any subinterval is given by Δx_i, where the subscript i represents the particular subinterval $(i = 1, 2, 3, \ldots)$. From the definition of average velocity (Eq. 2–2) we have

$$\Delta x_i = \bar{v}_i \, \Delta t_i.$$

Thus the displacement during each subinterval equals the product of $\bar{v}_i$ and Δt_i, and equals the area of the dark rectangle in Fig. C–1a for that subinterval. The total displacement between times t_1 and t_2 is the sum of the displacements over all the subintervals:

$$x_2 - x_1 = \sum_{t_1}^{t_2} \bar{v}_i \, \Delta t_i, \qquad \textbf{(C–1a)}$$

where x_1 is the position at t_1 and x_2 is the position at t_2. This sum equals the area of all the rectangles shown.

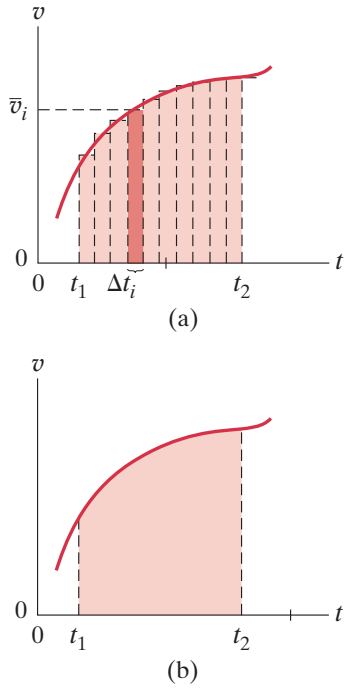


FIGURE C–1 Graph of v vs. t for a particular motion of a particle. In (a), the time axis is broken into subintervals of width Δt_i, the average velocity during each Δt_i is $\bar{v}_i$, and the area of all the rectangles, $\Sigma \bar{v}_i \, \Delta t_i$, closely approximates the total displacement $(x_2 - x_1)$ during the total time $(t_2 - t_1)$. In (b), $\Delta t_i \to 0$ and the total area under the curve between t_1 and t_2 is equal to $(x_2 - x_1)$.

It is often difficult to estimate $\bar{v}_i$ with precision for each subinterval from the graph. We can get greater accuracy in our calculation of $x_2 - x_1$ by breaking the time interval $t_2 - t_1$ into more, but narrower, subintervals. Ideally, we can let each Δt_i approach zero, so we approach (in principle) an infinite number of subintervals. In this limit, the area of all these infinitesimally thin rectangles becomes exactly equal to the area under the curve (Fig. C–1b). Thus *the total displacement between any two times is equal to the area between the velocity curve and the t axis between the two times* t_1 *and* t_2. This limit can be written

$$x_2 - x_1 = \lim_{\Delta t \to 0} \sum_{t_1}^{t_2} \bar{v}_i \, \Delta t_i$$

or, using standard calculus notation,

$$x_2 - x_1 = \int_{t_1}^{t_2} v(t) \, dt. \qquad \textbf{(C–1b)}$$

We have let $\Delta t \to 0$ and renamed it dt to indicate that it is now infinitesimally small. The average velocity, $\bar{v}$, over an infinitesimal time dt is the instantaneous velocity at that instant, which we have written $v(t)$ to remind us that v is a function of t. The symbol $\int$ is an elongated S and indicates a sum over an infinite number of infinitesimal subintervals. We say that we are taking the **integral** of $v(t)$ over dt from time t_1 to time t_2, and this is equal to the area between the $v(t)$ curve and the t axis between the times t_1 and t_2 (Fig. C–1b). The integral in Eq. C–1b is a **definite integral**, since the limits t_1 and t_2 are specified.

Similarly, if we know the acceleration as a function of time, we can obtain the velocity by the same process. We use the definition of average acceleration (Eq. 2–5) and solve for Δv:

$$\Delta v = \bar{a} \, \Delta t.$$

If a is known as a function of t over some time interval t_1 to t_2, we can subdivide this time interval into many subintervals, Δt_i, just as we did in Fig. C–1a. The change in velocity during each subinterval is $\Delta v_i = \bar{a}_i \, \Delta t_i$. The total change in velocity from time t_1 until time t_2 is

$$v_2 - v_1 = \sum_{t_1}^{t_2} \bar{a}_i \, \Delta t_i, \qquad \textbf{(C–2a)}$$

where v_2 represents the velocity at t_2 and v_1 the velocity at t_1. This relation can be written as an integral by letting $\Delta t \to 0$ (the number of intervals then approaches infinity)

$$v_2 - v_1 = \lim_{\Delta t \to 0} \sum_{t_1}^{t_2} \bar{a}_i \, \Delta t_i$$

or

$$v_2 - v_1 = \int_{t_1}^{t_2} a(t) \, dt. \qquad \textbf{(C–2b)}$$

Equations C–2 will allow us to determine the velocity v_2 at some time t_2 if the velocity is known at t_1 and a is known as a function of time.

If the acceleration or velocity is known at discrete intervals of time, we can use the summation forms of the above equations, Eqs. C–1a and C–2a, to estimate velocity or displacement. This technique is known as **numerical integration**. We now take an Example that can also be evaluated analytically, so we can compare the results.

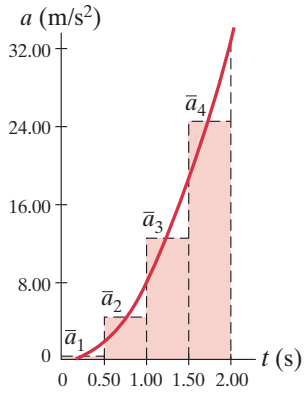


FIGURE C–2 Example C–1.

EXAMPLE C–1 Numerical integration. An object starts from rest at $t = 0$ and accelerates at a rate $a(t) = (8.00\,\text{m/s}^4)t^2$. Determine its velocity after 2.00 s using numerical methods.

APPROACH Let us first divide up the interval $t = 0.00\,\text{s}$ to $t = 2.00\,\text{s}$ into four subintervals each of duration $\Delta t_i = 0.50\,\text{s}$ (Fig. C–2). We use Eq. C–2a with $v_2 = v$, $v_1 = 0$, $t_2 = 2.00\,\text{s}$, and $t_1 = 0$. For each of the subintervals we need to estimate $\bar{a}_i$. There are various ways to do this and we use the simple method of choosing $\bar{a}_i$ to be the acceleration $a(t)$ at the midpoint of each interval (an even simpler but usually less accurate procedure would be to use the value of a at the start of the subinterval). That is, we evaluate $a(t) = (8.00\,\text{m/s}^4)t^2$ at $t = 0.25\,\text{s}$ (which is midway between 0.00 s and 0.50 s), 0.75 s, 1.25 s, and 1.75 s.

SOLUTION The results are as follows:

i	1	2	3	4
$\bar{a}_i\,(\text{m/s}^2)$	0.50	4.50	12.50	24.50

Now we use Eq. C–2a, and note that all Δt_i equal 0.50 s (so they can be factored out):

$$v(t = 2.00\,\text{s}) = \sum_{t=0}^{t=2.00\,\text{s}} \bar{a}_i\,\Delta t_i$$

$$= (0.50\,\text{m/s}^2 + 4.50\,\text{m/s}^2 + 12.50\,\text{m/s}^2 + 24.50\,\text{m/s}^2)(0.50\,\text{s})$$

$$= 21.0\,\text{m/s}.$$

We can compare this result to the analytic solution given by Eq. C–2b since the functional form for a is integrable analytically:

$$v = \int_0^{2.00\,\text{s}} (8.00\,\text{m/s}^4)t^2\,dt = \frac{8.00\,\text{m/s}^4}{3}t^3\Big|_0^{2.00\,\text{s}}$$

$$= \frac{8.00\,\text{m/s}^4}{3}[(2.00\,\text{s})^3 - (0)^3] = 21.33\,\text{m/s}$$

or 21.3 m/s to the proper number of significant figures. This analytic solution is precise, and we see that our numerical estimate is not far off even though we only used four Δt intervals. It may not be close enough for purposes requiring high accuracy. If we use more and smaller subintervals, we will get a more accurate result. If we use 10 subintervals, each with $\Delta t = 2.00\,\text{s}/10 = 0.20\,\text{s}$, we have to evaluate $a(t)$ at $t = 0.10\,\text{s}$, $0.30\,\text{s}, \ldots,$ $1.90\,\text{s}$ to get the $\bar{a}_i$, and these are as follows:

i	1	2	3	4	5	6	7	8	9	10
$\bar{a}_i\,(\text{m/s}^2)$	0.08	0.72	2.00	3.92	6.48	9.68	13.52	18.00	23.12	28.88

Then, from Eq. C–2a we obtain

$$v(t = 2.00\,\text{s}) = \sum \bar{a}_i\,\Delta t_i = (\sum \bar{a}_i)(0.200\,\text{s})$$

$$= (106.4\,\text{m/s}^2)(0.200\,\text{s}) = 21.28\,\text{m/s},$$

where we have kept an extra significant figure to show that this result is much closer to the (precise) analytic one but still is not quite identical to it. The percentage difference has dropped from 1.4% $(0.3\,\text{m/s}^2/21.3\,\text{m/s}^2)$ for the four-subinterval computation to only 0.2% (0.05/21.3) for the 10-subinterval one.

In the Example above we were given an analytic function that was integrable, so we could compare the accuracy of the numerical calculation to the known precise one. But what do we do if the function is not integrable, so we can't compare our numerical result to an analytic one? That is, how do we know if we've taken enough subintervals so that we can trust our calculated estimate to be accurate to within some desired uncertainty, say 1%? What we can do is compare two successive numerical calculations: the first done with n subintervals and the second with, say, twice as many subintervals ($2n$). If the two results are within the desired uncertainty (say 1%), we can usually assume that the calculation with more subintervals is within the desired uncertainty of the true value. If the two calculations are not that close, then a third calculation, with more subintervals (maybe double, maybe 10 times as many, depending on how good the previous approximation was) must be done, and compared to the previous one.

PROBLEM SOLVING
How many subintervals give sufficient accuracy?

The procedure is not hard to automate using a computer spreadsheet application or algorithm.

If we wanted to also obtain the displacement x at some time, we would have to do a second numerical integration over v, which means we would first need to calculate v for many different times. Programmable calculators and computers are very helpful for doing the long sums.

APPENDIX D

More on Dimensional Analysis

An important use of dimensional analysis (Section 1–7) is to obtain the *form* of an equation: how one quantity depends on others. To take a concrete example, let us try to find an expression for the period T of a simple pendulum: a small "bob" of mass m that can swing back and forth on the end of a thin cord (Section 14–5). First, we try to figure out what T could depend on, and make a list of these variables. It might depend on the length ℓ of the cord, on the mass m of the bob, on the angle of swing θ, and on the acceleration due to gravity, g. It might also depend on air resistance (we would use the viscosity of air), the gravitational pull of the Moon, and so on; but everyday experience suggests that the Earth's gravity is the major force involved, so we ignore the other possible forces. So let us assume that T is a function of ℓ, m, θ, and g, and that each of these factors is present to some power:

$$T = C\ell^w m^x \theta^y g^z.$$

C is a dimensionless constant, and w, x, y, and z are exponents we want to solve for. We now write down the dimensional equation (Section 1–7) for this relationship:

$$[T] = [L]^w[M]^x[L/T^2]^z.$$

Because θ has no dimensions (a radian is a length divided by a length—see Eq. 10–1a), it does not appear. We simplify and obtain

$$[T] = [L]^{w+z}[M]^x[T]^{-2z}.$$

To have dimensional consistency, we must have

$$\begin{aligned} 1 &= -2z \\ 0 &= w + z \\ 0 &= x. \end{aligned}$$

We solve these equations and find that $z = -\frac{1}{2}$, $w = \frac{1}{2}$, and $x = 0$. Thus our desired equation must be

$$T = C\sqrt{\ell/g}\, f(\theta), \qquad \textbf{(D–1)}$$

where $f(\theta)$ is some function of θ that we cannot determine using this technique. Nor can we determine in this way the dimensionless constant C. (To obtain C and f, we would have to do an analysis such as that in Chapter 14 using Newton's laws, which reveals that $C = 2\pi$ and $f \approx 1$ for small θ.) But look what we *have* found, using only dimensional consistency. We obtained the form of the expression that relates the period of a simple pendulum to the major variables of the situation, ℓ and g (see Eq. 14–12c), and saw that it does not depend on the mass m.

How did we do it? And how useful is this technique? Basically, we had to use our intuition as to which variables were important and which were not. This is not always easy, and often requires a lot of insight. As to usefulness, the final result in our example could have been obtained from Newton's laws, as in Chapter 14. But in many physical situations, such a derivation from other laws cannot be done. In those situations, dimensional analysis can be a powerful tool.

In the end, any expression derived by the use of dimensional analysis (or by any other means, for that matter) must be checked against experiment. For example, in our derivation of Eq. D–1, we can compare the periods of two pendulums of different lengths, ℓ_1 and ℓ_2, whose amplitudes (θ) are the same. Using Eq. D–1, we would have

$$\frac{T_1}{T_2} = \frac{C\sqrt{\ell_1/g}\, f(\theta)}{C\sqrt{\ell_2/g}\, f(\theta)} = \sqrt{\frac{\ell_1}{\ell_2}}.$$

Because C and $f(\theta)$ are the same for both pendula, they cancel out, so we can experimentally determine if the ratio of the periods varies as the ratio of the square roots of the lengths. This comparison to experiment checks our derivation, at least in part; C and $f(\theta)$ could be determined by further experiments.

APPENDIX E

Gravitational Force Due to a Spherical Mass Distribution

In Chapter 6 we stated that the gravitational force exerted by or on a uniform sphere acts as if all the mass of the sphere were concentrated at its center, if the other object (exerting or feeling the force) is outside the sphere. In other words, the gravitational force that a uniform sphere exerts on a particle outside it is

$$F = G\frac{mM}{r^2}, \qquad [m \text{ outside sphere of mass } M]$$

where m is the mass of the particle, M the mass of the sphere, and r the distance of m from the center of the sphere. Now we will derive this result. We will use the concepts of infinitesimally small quantities and integration.

First we consider a very thin, uniform spherical shell (like a thin-walled basketball) of mass M whose thickness t is small compared to its radius R (Fig. E–1). The force on a particle of mass m at a distance r from the center of the shell can be calculated as the vector sum of the forces due to all the particles of the shell. We imagine the shell divided up into thin (infinitesimal) circular strips so that all points on a strip are equidistant from our particle m. One of these circular strips, labeled AB, is shown in Fig. E–1. It is $R\,d\theta$ wide, t thick, and has a radius $R\sin\theta$. The force on our particle m due to a tiny piece of the strip at point A is represented by the vector $\vec{\mathbf{F}}_A$ shown; and the force due to a tiny piece of the strip at point B, which is diametrically opposite A, is the force $\vec{\mathbf{F}}_B$. We take the two pieces at A and B to be of equal mass, so $F_A = F_B$. The horizontal components of $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ are each equal to

$$F_A\cos\phi$$

and point toward the center of the shell. The vertical components of $\vec{\mathbf{F}}_A$ and $\vec{\mathbf{F}}_B$ are of equal magnitude and point in opposite directions, and so cancel. Since for every point on the strip there is a corresponding point diametrically opposite (as with A and B), we see that the net force due to the entire strip points toward the center of the shell. Its magnitude will be

$$dF = G\frac{m\,dM}{\ell^2}\cos\phi,$$

where dM is the mass of the entire circular strip and ℓ is the distance from all points on the strip to m, as shown. We write dM in terms of the density ρ; by density we mean the mass per unit volume (Section 13–2). Hence, $dM = \rho\,dV$, where dV is the volume of the strip and equals $(2\pi R\sin\theta)(t)(R\,d\theta)$. Then the force dF due to the circular strip shown is

$$dF = G\frac{m\rho 2\pi R^2 t\sin\theta\,d\theta}{\ell^2}\cos\phi. \qquad \textbf{(E–1)}$$

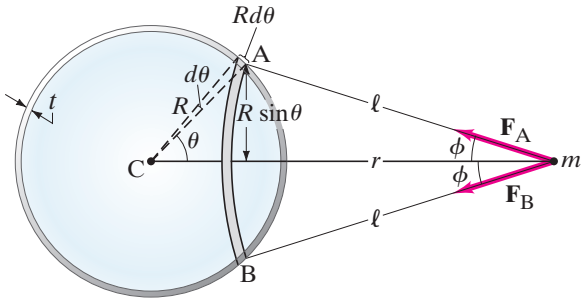


FIGURE E–1 Calculating the gravitational force on a particle of mass m due to a uniform spherical shell of radius R and mass M.

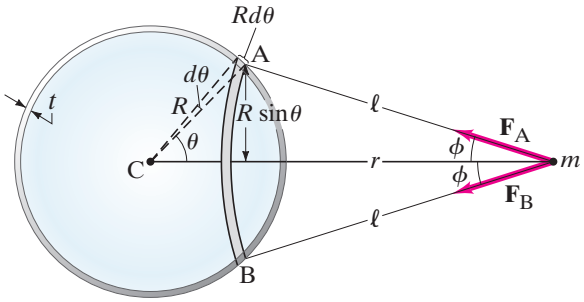


FIGURE E–1 (Repeated.) Calculating the gravitational force on a particle of mass m due to a uniform spherical shell of radius R and mass M.

To get the total force F that the entire shell exerts on the particle m, we must integrate over all the circular strips: that is, we integrate

$$dF = G\frac{m\rho 2\pi R^2 t \sin\theta\, d\theta}{\ell^2}\cos\phi \qquad \textbf{(E–1)}$$

from $\theta = 0°$ to $\theta = 180°$. But our expression for dF contains ℓ and ϕ, which are functions of θ. From Fig. E–1 we can see that

$$\ell\cos\phi = r - R\cos\theta.$$

Furthermore, we can write the law of cosines (Appendix A–9, page A-4) for triangle CmA:

$$\cos\theta = \frac{r^2 + R^2 - \ell^2}{2rR}. \qquad \textbf{(E–2)}$$

With these two expressions we can reduce our three variables (ℓ, θ, ϕ) to only one, which we take to be ℓ (it is simplest, as we shall see).We do two things with Eq. E–2: (1) We put it into the equation for $\ell\cos\phi$ above:

$$\cos\phi = \frac{1}{\ell}(r - R\cos\theta) = \frac{r^2 + \ell^2 - R^2}{2r\ell}$$

and (2) we take the differential of both sides of Eq. E–2 (because $\sin\theta\, d\theta$ appears in the expression for dF, Eq. E–1), considering r and R to be constants when summing over the strips:

$$-\sin\theta\, d\theta = -\frac{2\ell\, d\ell}{2rR} \quad \text{or} \quad \sin\theta\, d\theta = \frac{\ell\, d\ell}{rR}.$$

We insert these into Eq. E–1 for dF and find

$$dF = Gm\rho\pi t\frac{R}{r^2}\left(1 + \frac{r^2 - R^2}{\ell^2}\right)d\ell.$$

Now we integrate to get the net force on our thin shell of radius R. To integrate over all the strips ($\theta = 0°$ to 180°), we must go from $\ell = r - R$ to $\ell = r + R$ (see Fig. E–1). Thus,

$$F = Gm\rho\pi t\frac{R}{r^2}\left[\int d\ell + (r^2 - R^2)\int\frac{d\ell}{\ell^2}\right]_{\ell=r-R}^{\ell=r+R}$$

or

$$F = Gm\rho\pi t\frac{R}{r^2}\left[\ell - \frac{r^2 - R^2}{\ell}\right]_{\ell=r-R}^{\ell=r+R}$$

$$= Gm\rho\pi t\frac{R}{r^2}(4R).$$

The volume V of the spherical shell is its area $(4\pi R^2)$ times the thickness t. Hence the shell's mass is $M = \rho V = \rho 4\pi R^2 t$, and finally

$$F = G\frac{mM}{r^2}. \qquad \left[\begin{array}{c}\text{particle of mass } m \text{ outside a}\\ \text{thin uniform spherical shell of mass } M\end{array}\right]$$

This result gives us the force a thin shell exerts on a particle of mass m a distance r

from the center of the shell, and *outside* the shell. We see that the force is the same as that between m and a particle of mass M at the center of the shell. In other words, for purposes of calculating the gravitational force exerted on or by a uniform spherical shell, we can consider all its mass concentrated at its center.

What we have derived for a shell holds also for a solid sphere, since a solid sphere can be considered as made up of many concentric shells, from $R = 0$ to $R = R_0$, where R_0 is the radius of the solid sphere. Why? Because if each shell has mass dM, we write for each shell, $dF = Gm\,dM/r^2$, where r is the distance from the center C to mass m and is the same for all shells. Then the total force equals the sum or integral over dM, which gives the total mass M. Thus the result

$$F = G\frac{mM}{r^2} \qquad \left[\begin{array}{c}\text{particle of mass } m \text{ outside}\\ \text{solid sphere of mass } M\end{array}\right] \qquad \textbf{(E–3)}$$

is valid for a solid sphere of mass M with uniform density, and even for a sphere where the density varies with distance from the center, $\rho = \rho(R)$. (It is not valid if the density varies within each shell—that is, depends not only on R.) Thus the gravitational force exerted on or by spherical objects, including nearly spherical objects like the Earth, Sun, and Moon, can be considered to act as if the objects were point particles.

PROBLEM SOLVING
If particle m is inside *the sphere*

This result, Eq. E–3, is true only if the mass m is outside the sphere. Let us next consider a point mass m that is located *inside* the spherical shell of Fig. E–1. Here, r would be less than R, and the integration over ℓ would be from $\ell = R - r$ to $\ell = R + r$, so

$$\left[\ell - \frac{r^2 - R^2}{\ell}\right]_{R-r}^{R+r} = 0.$$

Thus the force on any mass inside the shell would be *zero*. This result has particular importance for the electrostatic force, which is also an *inverse square law*. For the gravitational situation, we see that at points within a solid sphere, say 1000 km below the Earth's surface, only the mass up to that radius contributes to the net force. The outer shells beyond the point we are considering contribute zero net gravitational effect.

The results we have obtained here can also be reached using the gravitational analog of Gauss's law for electrostatics (Chapter 22, page 648).

APPENDIX F

Differential Form of Maxwell's Equations

Maxwell's equations can be written in another form that is often more convenient than Eqs. 31–5. This material is usually covered in more advanced courses, and is included here simply for completeness.

We quote here two theorems, without proof, that are derived in vector analysis textbooks. The first is called **Gauss's theorem** or the **divergence theorem**. It relates the integral over a surface of any vector function $\vec{\mathbf{F}}$ to a volume integral over the volume enclosed by that surface:

$$\oint_{\text{Area } A} \vec{\mathbf{F}} \cdot d\vec{\mathbf{A}} = \int_{\text{Volume } V} \vec{\nabla} \cdot \vec{\mathbf{F}}\, dV. \qquad \text{[Gauss's theorem]} \qquad \textbf{(F–1)}$$

The operator $\vec{\nabla}$ is the **del operator**, defined in Cartesian coordinates as

$$\vec{\nabla} = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}.$$

The quantity

$$\vec{\nabla} \cdot \vec{\mathbf{F}} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

is called the **divergence** of $\vec{\mathbf{F}}$. Terms like $\partial F_y/\partial y$ mean that we take the derivative of the y component of F with respect only to y, treating x and z as constant.

The second theorem is **Stokes's theorem**, and relates a line integral around a closed path to a surface integral over any surface enclosed by that path:

$$\oint_{\text{Line}} \vec{\mathbf{F}} \cdot d\vec{\boldsymbol{\ell}} = \int_{\text{Area } A} \vec{\nabla} \times \vec{\mathbf{F}} \cdot d\vec{\mathbf{A}}. \qquad \text{[Stokes's theorem]} \qquad \textbf{(F–2)}$$

The quantity $\vec{\nabla} \times \vec{\mathbf{F}}$ is called the **curl** of $\vec{\mathbf{F}}$. (See Section 11–2 on the vector product.)

We now use these two theorems to obtain the differential form of Maxwell's equations in free space. We apply Gauss's theorem (Eq. F–1) to Eq. 31–5a (Gauss's law):

$$\frac{Q}{\epsilon_0} = \oint_A \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \int \vec{\nabla} \cdot \vec{\mathbf{E}}\, dV.$$

The charge Q can be written as a volume integral over the **charge density**, or *charge* per unit volume, ρ: $Q = \int \rho\, dV$. Then

$$\int \vec{\nabla} \cdot \vec{\mathbf{E}}\, dV = \frac{1}{\epsilon_0}\int \rho\, dV.$$

Both sides now contain volume integrals over the same volume, and for this to be true over *any* volume, whatever its size or shape, the integrands must be equal:

$$\vec{\nabla} \cdot \vec{\mathbf{E}} = \frac{\rho}{\epsilon_0}. \qquad \textbf{(F–3)}$$

This is the differential form of Gauss's law.

The second of Maxwell's equations, $\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0$, Eq. 31–5b, is treated in the same way, and we obtain

$$\vec{\nabla} \cdot \vec{\mathbf{B}} = 0, \qquad \textbf{(F–4)}$$

which is the second of Maxwell's equations in differential form.

Next, we apply Stokes's theorem (Eq. F–2) to the third of Maxwell's equations, Eq. 31–5c:

$$-\frac{d\Phi_B}{dt} = \oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = \int \vec{\boldsymbol{\nabla}} \times \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}.$$

Since the magnetic flux $\Phi_B = \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}$, we have

$$\int \vec{\boldsymbol{\nabla}} \times \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = -\frac{\partial}{\partial t} \int \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}}$$

where we use the partial derivative, $\partial\vec{\mathbf{B}}/\partial t$, since B may also depend on position. These are surface integrals over the same area, and to be true over any area, even a very small one, we must have

$$\vec{\boldsymbol{\nabla}} \times \vec{\mathbf{E}} = -\frac{\partial \vec{\mathbf{B}}}{\partial t}. \qquad \textbf{(F–5)}$$

This is the third of Maxwell's equations in differential form. Finally, we look at the last of Maxwell's equations, Eq. 31–5d:

$$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}.$$

We apply Stokes's theorem and write the electric flux as $\Phi_E = \int \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}$:

$$\int \vec{\boldsymbol{\nabla}} \times \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = \mu_0 I + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}.$$

The conduction current I can be written in terms of the current density $\vec{\mathbf{j}}$, using Eq. 25–12:

$$I = \int \vec{\mathbf{j}} \cdot d\vec{\mathbf{A}}.$$

Then Maxwell's fourth equation becomes:

$$\int \vec{\boldsymbol{\nabla}} \times \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = \mu_0 \int \vec{\mathbf{j}} \cdot d\vec{\mathbf{A}} + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}}.$$

For this to be true over any area A, whatever its size or shape, the integrands on each side of the equation must be equal:

$$\vec{\boldsymbol{\nabla}} \times \vec{\mathbf{B}} = \mu_0 \vec{\mathbf{j}} + \mu_0 \epsilon_0 \frac{\partial \vec{\mathbf{E}}}{\partial t}, \qquad \textbf{(F–6)}$$

which is the differential form of Maxwell's fourth equation.

Equations F–3, 4, 5, and 6 are Maxwell's equations in differential form for free space. They are summarized in Table F–1.

TABLE F–1 Maxwell's Equations in Free Space[†]

Integral form	Differential form
$\oint \vec{\mathbf{E}} \cdot d\vec{\mathbf{A}} = \frac{Q}{\epsilon_0}$	$\vec{\boldsymbol{\nabla}} \cdot \vec{\mathbf{E}} = \frac{\rho}{\epsilon_0}$
$\oint \vec{\mathbf{B}} \cdot d\vec{\mathbf{A}} = 0$	$\vec{\boldsymbol{\nabla}} \cdot \vec{\mathbf{B}} = 0$
$\oint \vec{\mathbf{E}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{d\Phi_B}{dt}$	$\vec{\boldsymbol{\nabla}} \times \vec{\mathbf{E}} = -\frac{\partial \vec{\mathbf{B}}}{\partial t}$
$\oint \vec{\mathbf{B}} \cdot d\vec{\boldsymbol{\ell}} = \mu_0 I + \mu_0 \epsilon_0 \frac{d\Phi_E}{dt}$	$\vec{\boldsymbol{\nabla}} \times \vec{\mathbf{B}} = \mu_0 \vec{\mathbf{j}} + \mu_0 \epsilon_0 \frac{\partial \vec{\mathbf{E}}}{\partial t}$

[†] $\vec{\boldsymbol{\nabla}}$ stands for the *del operator* $\vec{\boldsymbol{\nabla}} = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}$ in Cartesian coordinates.

APPENDIX G

Selected Isotopes

(1) Atomic Number Z	(2) Element	(3) Symbol	(4) Mass Number A	(5) Atomic Mass[†]	(6) % Abundance (or Radioactive Decay[‡] Mode)	(7) Half-life (if radioactive)
0	(Neutron)	n	1	1.008665	β^-	10.17 min
1	Hydrogen	H	1	1.007825	99.9866%	
	[proton	p	1	1.007276]		
	Deuterium	${}_1^2\mathrm{H}$	2	2.014102	0.0134%	
	[deuteron	d or D	2	2.013553]		
	Tritium	${}_1^3\mathrm{H}$	3	3.016049	β^-	12.32 yr
	[triton	t or T	3	3.015501]		
2	Helium	He	3	3.016029	0.0001343%	
			4	4.002603	99.9998657%	
3	Lithium	Li	6	6.015123	7.49%	
			7	7.016003	92.51%	
4	Beryllium	Be	7	7.016929	EC, γ	53.22 days
			9	9.012183	100%	
5	Boron	B	10	10.012937	19.6%	
			11	11.009305	80.4%	
6	Carbon	C	11	11.011433	β^+, EC	20.364 min
			12	12.000000	98.91%	
			13	13.003355	1.09%	
			14	14.003242	β^-	5730 yr
7	Nitrogen	N	13	13.005739	β^+, EC	9.965 min
			14	14.003074	99.631%	
			15	15.000109	0.369%	
8	Oxygen	O	15	15.003066	β^+, EC	122.24 s
			16	15.994915	99.759%	
			18	17.999160	0.202%	
9	Fluorine	F	19	18.998403	100%	
10	Neon	Ne	20	19.992440	90.48%	
			22	21.991385	9.25%	
11	Sodium	Na	22	21.994437	β^+, EC, γ	2.6018 yr
			23	22.989769	100%	
			24	23.990963	β^-, γ	14.957 h
12	Magnesium	Mg	24	23.985042	78.99%	
13	Aluminum	Al	27	26.981538	100%	
14	Silicon	Si	28	27.976927	92.223%	
			31	30.975363	β^-, γ	157.36 min
15	Phosphorus	P	31	30.973762	100%	
			32	31.973908	β^-	14.268 days

[†]The masses (atomic mass units) given in column (5) are those for the neutral atom, including the Z electrons (except for the special cases of proton, deuteron, triton, whose masses are given as nuclei without an electron).

[‡]Chapter 41; EC = electron capture.

(1) Atomic Number Z	(2) Element	(3) Symbol	(4) Mass Number A	(5) Atomic Mass	(6) % Abundance (or Radioactive Decay Mode)	(7) Half-life (if radioactive)
16	Sulfur	S	32	31.972071	94.99%	
			35	34.969032	β^-	87.37 days
17	Chlorine	Cl	35	34.968853	75.76%	
			37	36.965903	24.24%	
18	Argon	Ar	40	39.962383	99.6035%	
19	Potassium	K	39	38.963706	93.2581%	
			40	39.963998	0.0117%	
					β^-, EC, γ, β^+	1.248×10^9 yr
20	Calcium	Ca	40	39.962591	96.941%	
21	Scandium	Sc	45	44.955908	100%	
22	Titanium	Ti	48	47.947941	73.72%	
23	Vanadium	V	51	50.943957	99.750%	
24	Chromium	Cr	52	51.940505	83.789%	
25	Manganese	Mn	55	54.938043	100%	
26	Iron	Fe	56	55.934936	91.754%	
27	Cobalt	Co	59	58.933194	100%	
			60	59.933816	β^-, γ	5.2712 yr
28	Nickel	Ni	58	57.935342	68.0769%	
			60	59.930785	26.2231%	
29	Copper	Cu	63	62.929597	69.15%	
			65	64.927790	30.85%	
30	Zinc	Zn	64	63.929142	49.17%	
			66	65.926034	27.73%	
31	Gallium	Ga	69	68.925574	60.108%	
32	Germanium	Ge	72	71.922076	27.45%	
			74	73.921178	36.52%	
33	Arsenic	As	75	74.921595	100%	
34	Selenium	Se	80	79.916522	49.80%	
35	Bromine	Br	79	78.918338	50.66%	
36	Krypton	Kr	84	83.911498	56.987%	
37	Rubidium	Rb	85	84.911790	72.17%	
38	Strontium	Sr	86	85.909261	9.86%	
			88	87.905612	82.58%	
			90	89.907731	β^-	28.79 yr
39	Yttrium	Y	89	88.905841	100%	
40	Zirconium	Zr	90	89.904699	51.45%	
41	Niobium	Nb	93	92.906373	100%	
42	Molybdenum	Mo	98	97.905404	24.29%	
43	Technetium	Tc	98	97.907211	β^-, γ	4.2×10^6 yr
44	Ruthenium	Ru	102	101.904340	31.55%	
45	Rhodium	Rh	103	102.905494	100%	
46	Palladium	Pd	106	105.903480	27.33%	
47	Silver	Ag	107	106.905092	51.839%	
			109	108.904756	48.161%	
48	Cadmium	Cd	114	113.903365	28.754%	
49	Indium	In	115	114.903879	95.719%; β^-	4.41×10^{14} yr
50	Tin	Sn	120	119.902202	32.58%	
51	Antimony	Sb	121	120.903810	57.21%	

(1) Atomic Number Z	(2) Element	(3) Symbol	(4) Mass Number A	(5) Atomic Mass	(6) % Abundance (or Radioactive Decay Mode)	(7) Half-life (if radioactive)
52	Tellurium	Te	130	129.906223	34.08%; $\beta^-\,\beta^-$	$>2.7 \times 10^{24}$ yr
53	Iodine	I	127	126.904472	100%	
			131	130.906126	β^-, γ	8.0252 days
54	Xenon	Xe	132	131.904155	26.909%	
			136	135.907214	8.857%; $\beta^-\,\beta^-$	2.19×10^{21} yr
55	Cesium	Cs	133	132.905452	100%	
56	Barium	Ba	137	136.905827	11.23%	
			138	137.905247	71.70%	
57	Lanthanum	La	139	138.906359	99.91119%	
58	Cerium	Ce	140	139.905446	88.449%	
59	Praseodymium	Pr	141	140.907658	100%	
60	Neodymium	Nd	142	141.907729	27.153%	
61	Promethium	Pm	145	144.912756	EC, α	17.7 yr
62	Samarium	Sm	152	151.919739	26.74%	
63	Europium	Eu	153	152.921237	52.19%	
64	Gadolinium	Gd	158	157.924112	24.84%	
65	Terbium	Tb	159	158.925354	100%	
66	Dysprosium	Dy	164	163.929181	28.260%	
67	Holmium	Ho	165	164.930328	100%	
68	Erbium	Er	166	165.930299	33.503%	
69	Thulium	Tm	169	168.934218	100%	
70	Ytterbium	Yb	174	173.938868	32.025%	
71	Lutetium	Lu	175	174.940777	97.401%	
72	Hafnium	Hf	180	179.946560	35.08%	
73	Tantalum	Ta	181	180.947999	99.98799%	
74	Tungsten (wolfram)	W	184	183.950933	30.64%; α	$>8.9 \times 10^{21}$ yr
75	Rhenium	Re	187	186.955752	62.60%; β^-	4.33×10^{10} yr
76	Osmium	Os	191	190.960928	β^-, γ	14.99 days
			192	191.961479	40.78%	
77	Iridium	Ir	191	190.960592	37.23%	
			193	192.962924	62.77%	
78	Platinum	Pt	195	194.964794	33.775%	
79	Gold	Au	197	196.966570	100%	
80	Mercury	Hg	199	198.968281	16.94%	
			202	201.970644	29.74%	
81	Thallium	Tl	205	204.974427	70.46%	
82	Lead	Pb	206	205.974465	24.1%	
			207	206.975897	22.1%	
			208	207.976652	52.4%	
			210	209.984188	β^-, γ, α	22.20 yr
			211	210.988735	β^-, γ	36.164 min
			212	211.991896	β^-, γ	10.64 h
			214	213.999804	β^-, γ	27.06 min
83	Bismuth	Bi	209	208.980399	100%	
			211	210.987269	α, γ, β^-	2.14 min
84	Polonium	Po	209	208.982430	α, β^+, γ	124 yr
			210	209.982874	α, γ, EC	138.376 days
85	Astatine	At	210	209.987147	α, β^+, γ	8.1 h

(1) Atomic Number Z	(2) Element	(3) Symbol	(4) Mass Number A	(5) Atomic Mass	(6) % Abundance (or Radioactive Decay Mode)	(7) Half-life (if radioactive)
86	Radon	Rn	222	222.017576	α, γ	3.8215 days
87	Francium	Fr	223	223.019734	β^-, γ, α	22.00 min
88	Radium	Ra	226	226.025409	α, γ	1600 yr
89	Actinium	Ac	227	227.027751	β^-, γ, α	21.772 yr
90	Thorium	Th	228	228.028740	α, γ	1.9124 yr
			232	232.038054	99.98%; α, γ	1.40×10^{10} yr
91	Protactinium	Pa	231	231.035883	100%; α, γ	3.276×10^4 yr
92	Uranium	U	232	232.037155	α, γ	68.9 yr
			233	233.039634	α, γ	1.592×10^5 yr
			235	235.043928	0.7204%; α, γ	7.04×10^8 yr
			236	236.045566	α, γ	2.342×10^7 yr
			238	238.050787	99.2742%; α, γ	4.468×10^9 yr
			239	239.054292	β^-, γ	23.45 min
93	Neptunium	Np	237	237.048172	α, γ	2.144×10^6 yr
			239	239.052938	β^-, γ	2.356 days
94	Plutonium	Pu	239	239.052162	α, γ	24,110 yr
			244	244.064204	α	8.00×10^7 yr
95	Americium	Am	243	243.061380	α, γ	7364 yr
96	Curium	Cm	247	247.070353	α, γ	1.56×10^7 yr
97	Berkelium	Bk	247	247.070306	α, γ	1380 yr
98	Californium	Cf	251	251.079587	α, γ	900 yr
99	Einsteinium	Es	252	252.082980	α, EC, γ	471.7 days
100	Fermium	Fm	257	257.095105	α, γ	100.5 days
101	Mendelevium	Md	258	258.098430	α, γ	51.5 days
102	Nobelium	No	259	259.100998	α, EC	58 min
103	Lawrencium	Lr	262	262.109610	α, EC, fission	$\approx$4 h
104	Rutherfordium	Rf	267	267.121790	fission	2.5 h
105	Dubnium	Db	268	268.125670	fission	29 h
106	Seaborgium	Sg	269	269.182570	α	5 min
107	Bohrium	Bh	271	271.135180	α	10 min
108	Hassium	Hs	277	277.151900	fission	110 s
109	Meitnerium	Mt	276	276.151710	α	10 s
110	Darmstadtium	Ds	281	281.164720	α, fission	14 s
111	Roentgenium	Rg	282	282.169410	α	1.6 min
112	Copernicium	Cn	285	285.177320	α	32 s
113	Nihonium	Nh	286	286.182520	α	7 s
114	Flerovium	Fl	289	289.190620	α	2.4 s
115	Moscovium	Mc	290	290.196350	α	410 ms
116	Livermorium	Lv	293	293.204690	α	80 ms
117	Tennessine	Ts	294	294.210970	α	70 ms
118	Oganesson	Og	294	294.214130	α	1.15 ms

This page intentionally left blank

Answers to Odd-Numbered Problems

Chapter 1

1. (*a*) 3;
(*b*) 4;
(*c*) 3;
(*d*) 1;
(*e*) 2;
(*f*) 4;
(*g*) 2.

3. (*a*) 869,000;
(*b*) 9100;
(*c*) 0.25;
(*d*) 476;
(*e*) 0.0000362.

5. (*a*) 4%;
(*b*) 0.4%;
(*c*) 0.07%.

7. 2.3 m^2.

9. 0.24 rad, 12°.

11. 3×10^{-5}%.

13. 20%.

15. (*a*) 3 megavolts;
(*b*) 2 micrometers;
(*c*) 5 kilodays;
(*d*) 18 hectobucks;
(*e*) 900 nanoseconds.

17. (*a*) 13.4;
(*b*) 49.3.

19. (*a*) 1.4×10^{10} years;
(*b*) 4.4×10^{17} s.

21. 3.95 m.

23. (*a*) $\dfrac{0.621 \text{ mi/h}}{1 \text{ km/h}}$;
(*b*) $\dfrac{3.28 \text{ ft/s}}{1 \text{ m/s}}$;
(*c*) $\dfrac{0.278 \text{ m/s}}{1 \text{ km/h}}$.

25. (*a*) 9.46×10^{15} m;
(*b*) 6.31×10^4 AU.

27. 100.

29. 0.20 mi/gal.

31. (*a*) 3.16×10^7 s;
(*b*) 3.16×10^{16} ns;
(*c*) 3.17×10^{-8} yr.

33. (*a*) 7.478×10^{13} m^2;
(*b*) 6.84.

35. (*a*) 10^3;
(*b*) 10^5;
(*c*) 10^{-1};
(*d*) 10^9.

37. 500 hr.

39. 10^{14} cells.

41. 10^{11} gal/y.

43. 2×10^9 kg/yr.

45. 11 days.

47. 500 jelly beans.

49. 6×10^4 m.

51. 4×10^8 kg/year.

53. 6.0 s.

55. (*a*) L/T^4, L/T^2;
(*b*) m/s^4, m/s^2.

57. (*b*) 10^{-43} s.

59. 1.18×10^9 atoms/m^2.

61. 56,000 years.

63. 75 min.

65. 2×10^{-4} m.

67. 4×10^5 metric tons, 9×10^7 gal.

69. 150 m long, 25 m wide, 15 m high, 6×10^4 m^3.

71. 1.15 mi.

73. 120 m.

75. 200 plumbers.

77. L/m, L/year, L.

79. 4×10^{51} kg.

Chapter 2

1. 47 m.

3. 0.61 cm/s, no, actual distance traveled given.

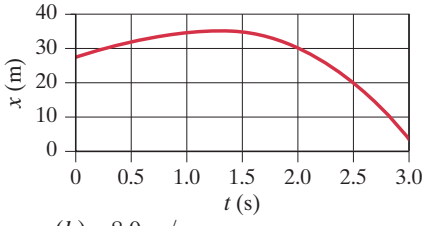
5. (*a*) 360 km;
(*b*) 80 km/h.

7. (*a*) 3.68 m/s;
(*b*) 0 m/s.

9. 56 s.

11. 1.8 min.

13. (*a*)



(*b*) −8.0 m/s;
(*c*) 1.3 s.

15. 25 km/h.

17. (*a*) $t = 0$ s to $t \approx 18$ s;
(*b*) $t \approx 27$ s;
(*c*) $t \approx 38$ s;
(*d*) both directions.

19. 6.11 m/s.

21. (*a*) 6.08 m/s^2;
(*b*) 7.88×10^4 km/h^2.

23. 6.5 m/s^2, 0.66 *g*'s.

25. (*a*) 21.2 m/s;
(*b*) 2.00 m/s^2.

27. (*a*) m/s, m/s^2;
(*b*) $2B$ m/s^2;
(*c*) $(A + 12B)$m/s, $2B$ m/s^2;
(*d*) $A - 3Bt^{-4}$.

29. 13 s later.

31. (*a*) 2.5 m/s^2;
(*b*) 0.6 m/s^2.

33. −3.8 m/s^2.

35. 1.4 m/s^2, 110 m.

37. 1.20×10^2 m.

39. 260 m/s^2.

41. (*a*) 560 m;
(*b*) 47 s;
(*c*) 23 m, 21 m.

43. No, 27.2 m past the intersection.

45. 22 m/s.

47. 6.3 s.

49. 10.5 s.

51. 1.0 m/s^2, 7.8 m/s.

53. (*a*) 8.8 s;
(*b*) 86 m/s.

55. 13 m/s, 8.3 m.

57. 1.09 s.

59. (*a*) 10.4 m/s;
(*b*) 0.774 s, 2.90 s;
(*c*) There are two times at which the object reaches that height—once when moving upwards, once when moving downwards.

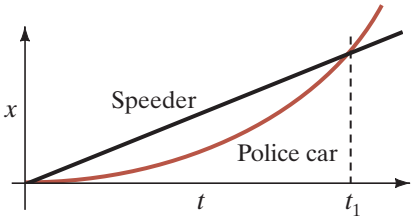
61. 26.1 s, 256 m/s (922 km/h).

63. (*a*) 5.80 s;
(*b*) 41.4 m/s;
(*c*) 99.5 m.

65. 12 m/s.

67. (*a*) 14 m/s;
(*b*) 25 m.

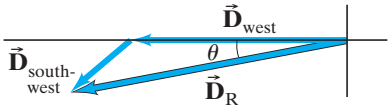
69. 2.1 m.
71. 29.3 m.
73. (*a*) $7.5\ \text{m/s} + \frac{2}{3}(3.0\ \text{m/s}^{5/2})t^{3/2}$;
(*b*) $(7.5\ \text{m/s})t + \frac{4}{15}(3.0\ \text{m/s}^{5/2})t^{5/2}$;
(*c*) $6.7\ \text{m/s}^2$, 3.0×10^1 m/s, 82 m.
75. 6 ×.
77. 1.2 m.
79. (*a*) The longer any particular bolt falls, the higher its speed.
(*b*) decrease;
(*c*) $0, d_1, 3d_1, 5d_1$, and so on.
81. (*a*) 14.4 s, no;
(*b*) no, 4.6 s.
83. 29.0 m.
85. (*a*)



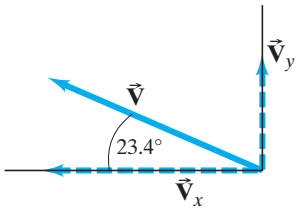
(*b*) 22 s;
(*c*) $3.6\ \text{m/s}^2$;
(*d*) 78 m/s;
(*e*) constant acceleration of police car.
87. (*a*) 0.82 m;
(*b*) 4.9 m/s;
(*c*) 1.0 s.
89. Stop.
91. 74 m.
93. (*a*) 1.60 s;
(*b*) 11.5 m;
(*c*) They won't collide unless the launch point is more than 91.3 m above the ground.
95. (*a*) 12.5 min;
(*b*) 11.1 min.
97. (*a*) 560 m;
(*b*) 28 s.
99. 64 m.
101. (*a*) 4.3×10^6 bits/s;
(*b*) 67%.
103. For no collision, $x > \dfrac{(v_M - v_A)^2}{-2a}$.

Chapter 3

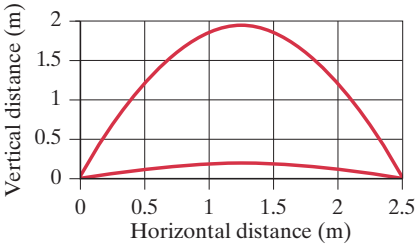
1. 339 km, 14.3° south of west.



3. 11.6 units, −35.9°.
5. (*a*)



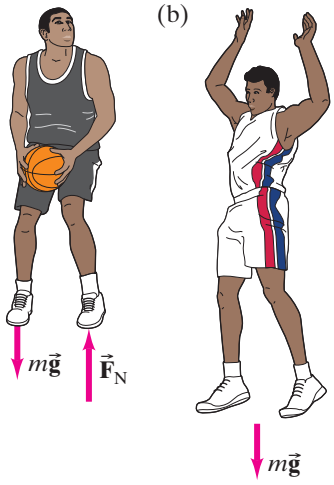
(*b*) −20.0 units, 8.66 units;
(*c*) 21.8 units, 23.4° above the −*x* axis.
7. (*a*) 1.3 units in the *x*-direction;
(*b*) 12.3 units in the *x*-direction;
(*c*) 12.3 units in the negative *x*-direction.
9. −2840 m, 3590 m, 2450 m, 5190 m.
11. (*a*) $-53.7\hat{\mathbf{i}} + 4.9\hat{\mathbf{j}}$, 53.9, 174.8°;
(*b*) $53.7\hat{\mathbf{i}} - 4.9\hat{\mathbf{j}}$, 53.9, 354.8°;
(*c*) They are opposite of each other.
13. (*a*) $-127.9\hat{\mathbf{i}} - 34.5\hat{\mathbf{j}}$, 132.4, 195.1°;
(*b*) $124.0\hat{\mathbf{i}} - 96.4\hat{\mathbf{j}}$, 157.1, 322.1°;
(*c*) $-20.5\hat{\mathbf{i}} - 75.3\hat{\mathbf{j}}$, 78.1, 254.8°.
15. (*a*) 4.2, 315°;
(*b*) $1.0\hat{\mathbf{i}} - 5.0\hat{\mathbf{j}}$, 5.1, 280°.
17. (*a*) parallel;
(*b*) perpendicular;
(*c*) $\vec{\mathbf{V}}_2 = 0$.
19. $(9.60\hat{\mathbf{i}} - 3.00t\hat{\mathbf{k}})\text{m/s}$, $-3.00\ \hat{\mathbf{k}}\ \text{m/s}^2$.
21. (*a*) cannot be determined;
(*b*) $3.78\ \text{m/s}^2$, 31.9°;
(*c*) cannot be determined.
23. (*a*) $0.900\ \text{m/s}^2$, downward;
(*b*) 16.7 s.
25. (*a*) $(2.48\hat{\mathbf{i}} + 2.75\hat{\mathbf{j}})\text{cm/s}$;
(*b*) 3.70 cm/s, 48.0°.
27. 410 m, −65°.
29. 85 km/h, 71° north of east.
31. $7.5\ \text{m/s}\hat{\mathbf{j}}$, $(4.2\hat{\mathbf{i}} + 6.3\hat{\mathbf{j}})\text{m}$.
33. 6.0×10^1 m, 8.8 m.
35. 18.5 m.
37. 18°, 72°.



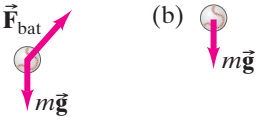
39. (*a*) 34.7 m;
(*b*) 5.32 s;
(*c*) 153 m;
(*d*) 30.9 m/s.
41. 23° or 67°.
43. (*a*) 9.72 m/s;
(*b*) 0.81 m.
45. (*a*) 4.0×10^1 m/s;
(*b*) 22 m/s.
47. 255 m.
49. (*a*) 55 m;
(*b*) 61 m;
(*c*) 35 m/s, 77° below the horizontal.
51. $\theta(t) = \tan^{-1}(gt/v_0)$.
53. (*a*) 8.9 m/s;
(*b*) She will land exactly on the edge of the opposite bank, neither long nor short.
55. (*a*) 30.3°, 59.7°;
(*b*) 2.92 m, 8.56 m;
(*c*) 1.54 s, 2.64 s.
57. 0.88 s, 0.95 m.
59. No, it is 0.81 m to low; 4.6 m to 34.7 m.
61. $R = \dfrac{v_0 \cos\theta_0}{g}\left[v_0 \sin\theta_0 + \sqrt{v_0^2 \sin^2\theta_0 - 2gh}\right]$.
63. 1.66 m/s, 25° relative to river.
65. (*a*) 1.71 m/s;
(*b*) 3.94 m/s.
67. (*a*) 631 km/h, 5.47° east of south;
(*b*) 15.6 km.
69. $v_T/\tan\theta$.
71. (*a*) 46 m;
(*b*) 92 s.
73. $(-55\hat{\mathbf{i}} + 35\hat{\mathbf{j}})\text{km/h}$,
$(55\hat{\mathbf{i}} - 35\hat{\mathbf{j}})\text{km/h}$.
75. (*a*) 55.4° upstream;
(*b*) 13.8 km/h;
(*c*) 0.142 h.
77. $1.7\ \text{m/s}^2$.
79. 33 m/s.
83. 2.0 m/s.
85. 11.1 m/s to 11.3 m/s.
87. 0.91%.
89. (*a*) $(2.3\hat{\mathbf{i}} + 3.7\hat{\mathbf{j}})\text{m/s}$, 4.3 m/s, 57° above the horizontal;
(*b*) 4.2 m;
(*c*) $(2.3\hat{\mathbf{i}} - 9.1\hat{\mathbf{j}})\text{m/s}$, 9.4 m/s, −76° (below the horizontal).
91. (*a*) 15.6 m;
(*b*) 27.6°.
93. 3.1 s.
95. (*a*) 18 m/s, 16 m/s;
(*b*) 6.5 m below the top.
97. 6.8 m/s, 61°.
99. $(2.5 - 3.0t)\text{m}\hat{\mathbf{i}} + (-3.1 + 2.25t^2)\text{m}\hat{\mathbf{j}}$, $4.5\ \text{m/s}^2\ \hat{\mathbf{j}}$, parabolic.
101. 1.9 m/s.

Chapter 4

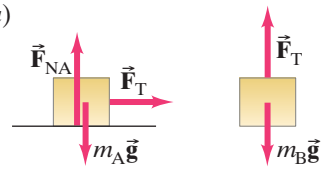
1. 63 N.
3. 1630 N.
5. -3100 N.
7. -1.3×10^6 N, 39%, 1.3×10^6 N.
9. 210 N.
11. >1.5 kg.
13. (*a*) 196 N, 196 N;
(*b*) 98.0 N, 294 N.
15. -0.4 m/s^2, downward.
17. -2.5 m/s^2, downward.
19. 0.705 m/s^2.
21. (*a*) 4.69 m/s^2;
(*b*) 38 m/s;
(*c*) 64 s.
23. 5.06×10^4 N, 4.45×10^4 N.
25. (*a*) 7.5 s, 13 s, 7.5 s;
(*b*) 12%, 0%, -12%;
(*c*) 55%.
27. (*a*) 7.4 m/s;
(*b*) 2100 N.
29. (*a*) (b)



31. (*a*) (b)



33. (*a*) 380 N;
(*b*) 1.5 m/s^2.
35. 36 N, 38 N, 74 N.
37. 2.
39. 12°.
41. $\frac{5}{2}\frac{F_0}{m}t_0^2$.
43. (*a*) 12 N;
(*b*) 260 N.
45. (*a*) 3.67 m/s^2;
(*b*) 9.39 m/s.
47. 30°.
49. (*a*) 1.4 m;
(*b*) 12 kN, no.
51. (*a*) 0.6 m/s^2;
(*b*) 1.5×10^5 N.
53. 4.0×10^2 m.
55. (*a*)



(*b*) $g\dfrac{m_B}{m_A+m_B}$, $g\dfrac{m_A m_B}{m_A+m_B}$.

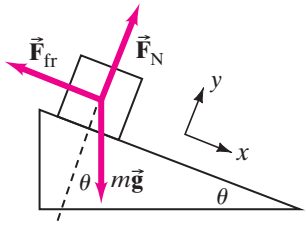
57. $g\dfrac{m_B+\dfrac{\ell_B}{\ell_A+\ell_B}m_C}{m_A+m_B+m_C}$.

59. 1.74 m/s^2, $F_{TB} = 20.9$ N,
$F_{TA} = 22.6$ N; $a_{\underset{59}{\text{Prob}}} < a_{\underset{4\text{-}12}{\text{Ex.}}}$;
$(F_{TB})_{\underset{59}{\text{Prob}}} < (F_{TB})_{\underset{4\text{-}12}{\text{Ex.}}}$;
$(F_{TA})_{\underset{59}{\text{Prob}}} > (F_{TA})_{\underset{4\text{-}12}{\text{Ex.}}}$.

61. 1.9 m/s^2.
63. (*a*) $\dfrac{3m_Bm_C - m_Am_C - 4m_Am_B}{4m_Am_B + m_Am_C + m_Bm_C}g$,
$\dfrac{m_Bm_C - 3m_Am_C + 4m_Am_B}{4m_Am_B + m_Am_C + m_Bm_C}g$,
$\dfrac{m_Am_C + m_Bm_C - 4m_Am_B}{4m_Am_B + m_Am_C + m_Bm_C}g$;
(*b*) $\dfrac{4m_Am_Bm_C}{4m_Am_B + m_Am_C + m_Bm_C}g$,
$\dfrac{8m_Am_Bm_C}{4m_Am_B + m_Am_C + m_Bm_C}g$.
65. 4.1 m.
67. 0.86 m/s^2.
69. 8.5°.
71. Vertical ropes, 127 N;
30° ropes, 147 N;
60° ropes, 255 N.
73. 26 N.
75. 92 m/s.
77. (*a*) -1.9 m/s^2;
(*b*) 0.616 kg;
(*c*) 7.9 N, 6.03 N.
79. (*a*) $Mg/2$;
(*b*) $Mg/2$, $Mg/2$, 3 $Mg/2$, Mg.
81. 870 N, 72° above the horizontal.
83. 2.01×10^4 N.
85. 1.6×10^7 s.
87. (*a*) 37 $\dfrac{\text{N}}{\text{m/s}}$;
(*b*) 120 N.
89. 16 m/s.
91. (*a*) 2.0×10^1 N;
(*b*) 3.3 N.
93. 2.0 m.
95. (*a*) 2000 N;
(*b*) 6;
(*c*) 30°.

Chapter 5

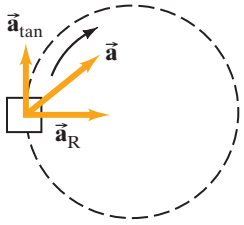
1. 76 N, 0 N.
3. 0.20.
5. (*a*)



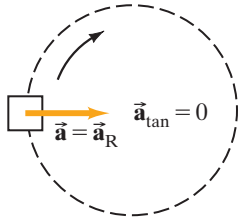
$\vec{\mathbf{F}}_{fr}$ is a static frictional force.
(*b*) Same diagram, but $\vec{\mathbf{F}}_{fr}$ is a kinetic frictional force.
(*c*) The kinetic frictional force would point in the opposite direction.

7. -5.3 m/s^2.
11. (*b*) 55 m.
13. 312 m.
15. (*a*) 43 N;
(*b*) 38 N.
17. (*a*) 0.45;
(*b*) 5.1 m/s;
(*c*) 5.8 m;
(*d*) 15 m/s.
19. (*a*) 1.7 m/s^2;
(*b*) 430 N;
(*c*) 1.7 m/s^2, 220 N.
21. (*a*) 0.80 m;
(*b*) 1.3 s.
23. (*a*) 5.0 kg;
(*b*) 6.7 kg.
25. (*a*) 0.22 s;
(*b*) 0.16 m.
27. 0.40.
29. (*a*) 3.1 m/s^2;
(*b*) 1.7 N.
31. 1.1 m/s^2.
33. (*a*) 0.53;
(*b*) 2.6 m/s^2;
(*c*) -2.6 m/s^2;
(*d*) 83 N, 73 N.
35. 24 m/s, yes.
37. 12 m/s.
39. 0.84.
41. 280 N.
43. 13 m.
45. 30 m.
47. 8.6 m/s.
49. 35 m/s.

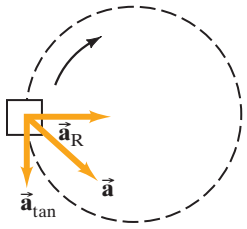
51. 7.0×10^3 N down the plane.
53. 77°.
55. 8.9×10^{-2} m, slide outward away from the center.
57. (*a*) 7380 N;
(*b*) 545 N;
(*c*) 29.4 m/s.
59. 28.3 m/s, 0.410 rev/s.
61. 1100 m.
63. 0.33.
65. 3.0×10^3 N, 550 N.
67. (*a*)



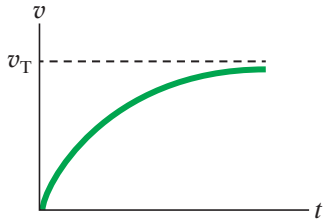
(*b*)



(*c*)



69. (*a*) 11 m/s^2;
(*b*) 17 m/s^2.
71. $m(b + ct^2)$, $\frac{m}{r}(v_0 + bt + \frac{1}{3}ct^3)^2$.
73. (*a*) 3×10^{-5} kg/s;
(*b*) 1 s.
75. (*a*) $\sqrt{mg/c}$;
(*b*) 0.2 kg/m;
(*c*) above.



77. (*a*) 15 kg/m;
(*b*) 630 N.
79. 10 m.
81. 0.57.
83. 89.5 N, 0.598.
85. 0.18, no outward force.
87. 2800 N, 1900 N.
89. 35°.
91. (*a*) 5.95×10^{-3} m/s^2;
(*b*) 3.56×10^{22} N, the Sun.
93. (*a*) 3.0 m/s;
(*b*) 9.6 N, 19 N.
95. (*a*) 1.0×10^1 kg;
(*b*) 0.6 m/s^2.
97. (*a*) 58 s;

103. $r_{\max} = \dfrac{g(\cos\phi + \mu_s \sin\phi)}{4\pi^2 f^2(\sin\phi - \mu_s \cos\phi)}$,

$r_{\min} = \dfrac{g(\cos\phi - \mu_s \sin\phi)}{4\pi^2 f^2(\sin\phi + \mu_s \cos\phi)}$.

105. 8.5 m/s^2, 45 m/s^2, 5.0×10^1 N.

Chapter 6

1. 1.80×10^3 N.
3. 1.62 m/s^2.
5. 27.4 m/s^2.
7. 3.46×10^8 m.
9. 0.05%.
11. (*a*) Less;
(*b*) 0.23%.
13. Multiplied by 1.26.
15. 2.0 m/s^2.
17. 11%.
19. 2.0×10^2 kg.
21. 5×10^{12} m/s^2.
25. (*b*) g decreases as r increases;
(*c*) 9.42 m/s^2 vs. 9.43 m/s^2.
27. 9.78 m/s^2, 0.099° below radially inward.
29. 7460 m/s.
31. (*a*) 568 N;
(*b*) 568 N;
(*c*) 671 N;
(*d*) 470 N;
(*e*) 0 N.
33. 5.2×10^{39} kg, 2.6×10^9 solar masses.
35. (*a*) The close satellite is moving about 1.3 × faster;
(*b*) The close satellite has a period about 0.51 × shorter.
37. 4.8×10^4 s.
39. (*a*) 0 N, so no direction;
(*b*) 88 N, towards the Moon.
41. 9.0 s.
43. $\sqrt{GM/\ell}$.
45. 160 years.
47. 5070 s or 84.5 min.
49. (*a*) 1.90×10^{27} kg;
(*b*) Europa: 1.90×10^{27} kg,
Ganymede: 1.89×10^{27} kg,
Callisto: 1.90×10^{27} kg, yes they are consistent.
51. 2.0×10^{11} m.
53. 5.4×10^{12} m, in the solar system, Pluto.
55. (*b*) 2.92×10^4 m/s, 3.02×10^4 m/s.
57. (*a*) 180 AU;
(*b*) 360 AU;
(*c*) 360:1.
59. 1.07×10^{-2} m/s^2, towards the center of the Earth.
61. (*a*) $-\dfrac{2Gmy}{(x_0^2 + y^2)^{3/2}}\hat{\mathbf{j}}$;

(*b*) $y = \pm\dfrac{x_0}{\sqrt{2}}$, $g = \dfrac{4Gm}{3\sqrt{3}x_0^2}$.

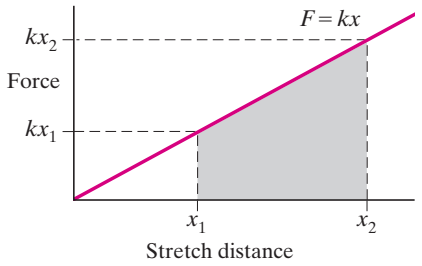
63. (*b*) 7.5×10^{29} kg.
65. 5070 s or 84.5 min.
67. $T_{\text{inner}} = 2.0 \times 10^4$ s,
$T_{\text{outer}} = 7.1 \times 10^4$ s,
$T_{\text{inner}} \approx \frac{1}{2}T_{\text{Sat}}$,
$T_{\text{outer}} \approx 2T_{\text{Sat}}$.
69. (*a*) 7000 m;
(*b*) 5×10^{-3} m/s^2;
(*c*) 4×10^4 s or 11 h.
71. 7.4×10^{36} kg, $3.7 \times 10^6\, m_{\text{Sun}}$.
73. (*a*) 0.716;
(*b*) 1.284.
75. 0.948, 3.98.
77. 2.3 *g*'s.
79. $-\sqrt{Gm_E/r_E}$.
81. 8.99 d.
83. 1.21×10^6 m.
85. (*a*) 2.5×10^4 m;
(*b*) 2.0 h.

Chapter 7

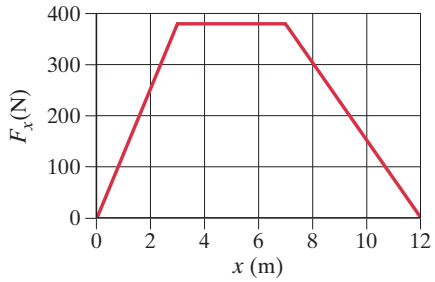
1. 1.0×10^4 J.
3. 1.51×10^4 J.
5. 1900 J.
7. (*a*) 1.10 *Mg*;
(*b*) 1.10 *Mgh*.
9. 6000 J.
13. 18 J.
15. −240 J, 0 J, 240 J.
17. 2 m.
23. $b\hat{\mathbf{i}} - \frac{5}{3}b\hat{\mathbf{j}}$, $b \neq 0$.
25. (*a*) 7;
(*b*) −230;
(*c*) 28.
29. 55.9°, 43.2°, 113°.

35. 62.7°.

37. 0.15 J.



39. 3.0×10^3 J.



41. (*a*) 2800 J;
(*b*) 2100 J.

43. $\frac{1}{5}bd^5$.

45. 6.0 J.

47. 40.0 J.

49. 102 J.

51. $\dfrac{m^2g^2}{2k}$.

53. 96 J.

55. (*a*) $\sqrt{3}$;
(*b*) 1/4.

57. (*a*) -2.6×10^5 J;
(*b*) 2.6×10^5 J;
(*c*) 0 J.

59. 340 N.

61. 2.25.

63. 2.2×10^4 N/m.

65. 0.16.

67. (*a*) 640 J;
(*b*) −470 J;
(*c*) 0 J;
(*d*) 2.5 m/s.

69. (*a*) $\frac{1}{2}mv_2^2\left(1 + 2\dfrac{v_1}{v_2}\right)$;
(*b*) $\frac{1}{2}mv_2^2$;
(*c*) $W_{\text{Earth}} = \frac{1}{2}mv_2^2\left(1 + 2\dfrac{v_1}{v_2}\right)$,
$W_{\text{train}} = \frac{1}{2}mv_2^2$.

71. 5.7 m/s, 11.3 m/s.

73. (*a*) 2.15×10^5 J;
(*b*) 21.0 m/s;
(*c*) 2.44 m.

75. 16 m/s.

77. 15 m/s.

79. $\dfrac{A}{k}e^{-0.10k}$.

81. 86 kJ, 42°.

83. (*a*) 370 N;
(*b*) 370 N.

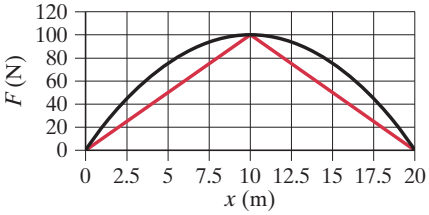
87. (*a*) -8.4×10^4 J;
(*b*) 7.6×10^4 N;
(*c*) -2.1×10^5 J.

89. (*a*) $\dfrac{v}{\mu_k g}$;
(*b*) $\dfrac{v^2}{2\mu_k g}$;
(*c*) mv^2;
(*d*) $\frac{1}{2}mv^2$, $\frac{1}{2}mv^2$.

91. 5800 N/m.

93. (*a*) 130 N (28 lbs), yes;
(*b*) 470 N, (110 lbs), no.

95. (*a*) 1000 J;



(*b*) 1330 J.

97. 8.3×10^4 N/m.

99. (*b*) 0.83, 6.8.

Chapter 8

1. 2300 J.

3. 1.07 m.

5. 6.0×10^1 J.

7. $-\dfrac{k}{2x^2} + \dfrac{k}{8\ \text{m}^2}$.

9. (*a*) Yes;
(*b*) $\frac{1}{2}kx^2 - \frac{1}{4}ax^4 - \frac{1}{5}bx^5 + C$.

11. 47.5 m/s.

13. 4.69 m/s.

15. 71 cm.

17. 1.6×10^5 N/m.

19. (*a*) 10.4 m/s;
(*b*) 5.69 m.

21. 2*d*, no.

23. 0.26 m.

25. $\dfrac{9.6\ Mg}{h}$.

27. 4.0×10^7 J.

29. (*a*) 25 m/s;
(*b*) 370 m.

31. 21 m/s.

33. 173 m/s.

35. 27 m.

37. 0.36.

39. (*a*) 0.16 m;
(*b*) 1.0;
(*c*) 0.80 m/s.

41. 1/4.

45. (*a*) -3.4×10^{10} J, -2.0×10^{10} J;
(*b*) 1.7×10^{10} J, 1.0×10^{10} J;
(*c*) 7×10^9 J.

47. (*a*) 10,700 m/s;
(*b*) 11,600 m/s;
(*c*) 11,200 m/s.

49. (*a*) $-\sqrt{\dfrac{GM_E}{2r^3}}$;
(*b*) 1.09×10^4 m/s.

51. $\dfrac{GMm}{12r_E}$.

53. (*b*) −1/2.

55. 2.41×10^4 W.

57. (*a*) 1100 J;
(*b*) 1100 W.

59. (*b*) 0.10 hp;
(*c*) 130 kW.

61. 480 W.

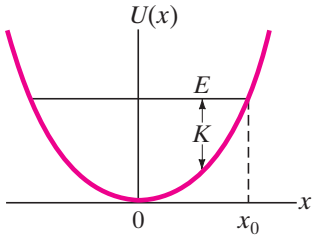
63. 5.3×10^4 W.

65. 15.4 W.

67. 3.0×10^1 hp.

69. (*a*) −250 W;
(*b*) 1900 W;
(*c*) −81 W, 440 W.

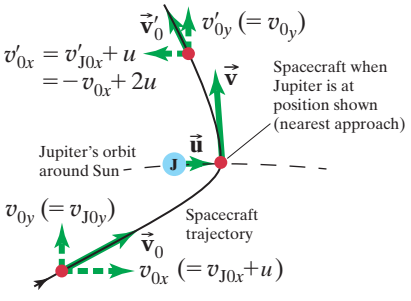
71.



73. (*a*) $0 \le x < 3$ m, 6 m $< x < 9$ m;
(*b*) $x = 3$ m, 6 m, 9 m;
(*c*) $x \approx 4$ m.

75. $a^2/4b$.

77.



79. (*a*) 42 m/s;
(*b*) 1.5×10^5 W.

81. (*a*) 28.2 m/s;
(*b*) 116 m.

85. 4.5×10^4 N.
87. (*a*) 7.0×10^5 J;
(*b*) 42 W, 5.7×10^{-2} hp;
(*c*) 280 W, 0.38 hp.
89. (*a*) 2.5 *r*;
(*b*) 11 *mg*;
(*c*) 5 *mg*;
(*d*) *mg*.
91. 8.0 m/s.
93. (*a*) 35°;
(*b*) 450 N;
(*c*) 750 N.
95. 1.7×10^6 J.
97. 8.1×10^6 J.
99. (*a*) 5220 m/s;
(*b*) 3190 m/s.
101. $x = \sqrt{a/b}$.
103. $U_{\text{launch}}/U_{\text{land}} = 16.2$.

Chapter 9

1. 0.27 kg · m/s.
3. 0.4 m/s.
5. $(9.6t\,\hat{\mathbf{i}} - 9.4\vec{\mathbf{k}})$N.
7. 2500 m/s.
9. 1:2.
11. 25 N at 35° above horizontal.
13. $\frac{3}{2}v_0\hat{\mathbf{i}} - v_0\hat{\mathbf{j}}$.
15. $(4.0\hat{\mathbf{i}} + 3.3\hat{\mathbf{j}} - 3.3\hat{\mathbf{k}})$m/s.
17. (*a*) $(116\text{ m/s})\hat{\mathbf{i}} + (58.0\text{ m/s})\hat{\mathbf{j}}$, or 130 m/s, 26.6° above the horizontal;
(*b*) 4.37×10^5 J.
19. (*a*) 1.7 kg · m/s;
(*b*) 490 N.
21. (*a*) 0.14 m/s;
(*b*) 521 N;
(*c*) $K_{\text{astronaut}} = 391$ J, $K_{\text{capsule}} = 22$ J.
23. (*a*) 5 N · s;
(*b*) 80 m/s.
25. (*a*) $2\dfrac{mv}{\Delta t}$;
(*b*) $2\dfrac{mv}{t}$.
27. (*a*) $0.964\text{ N} + (1.39\text{ N/s})t$;
(*b*) 17.4 N;
(*c*) $0.629\text{ N}\sqrt{(2.35 - (0.0768/\text{s})t)} + (1.39\text{ N/s})t$, *17.2 N*.
29. $v'_A = -3.60$ m/s, $v'_B = 2.00$ m/s (they exchange velocities).
31. $m_2 = 3m_1$.
33. (*a*) 1.00;
(*b*) 0.890;
(*c*) 0.286;
(*d*) 0.0192.
35. 14 or 0.072.
37. $v_2 = \sqrt{2}\,v_1$.
39. $K_{\text{heavy}} = 1800$ J , $K_{\text{light}} = 2800$ J.
41. 42%.
43. $v'_{\text{light}} = 0.20$ m/s in opposite direction, $v'_{\text{heavy}} = 2.5$ m/s in original direction.
45. (*a*) 920 m/s;
(*b*) 0.999.
47. (*b*) 1.39 m/s, 48.1°.
49. (*a*) 30°;
(*b*) $\dfrac{v}{\sqrt{3}}, \dfrac{v}{\sqrt{3}}$;
(*c*) 2/3.
53. $v'_1 = 3.8$ m/s, 30° above the incident direction,
$v'_2 = 3.8$ m/s, 30° below the incident direction,
$v'_3 = 1.0$ m/s, opposite the incident direction.
55. 0.438 m.
57. $3.8\,\ell_0$.
59. From the back left corner: $x_{\text{CM}} = 1.2\ell$, $y_{\text{CM}} = 0.9\ell$.
61. (0.89 m, −1.1 m).
63. $(0\hat{\mathbf{i}} + 0\hat{\mathbf{j}} + \frac{3}{4}h\hat{\mathbf{k}})$.
65. (*a*) 4.66×10^6 m from the center of the Earth.
67. (*a*) 5.0 m;
(*b*) 2.6 m;
(*c*) 3.6 m.
69. $-2.3\hat{\mathbf{i}}$ m/s $- 3.5\hat{\mathbf{j}}$ m/s.
71. 41 m.
73. −82 kg/s.
75. 0.93 hp.
77. 22 m.
79. (*a*) $v'_A = 3.65$ m/s, $v'_B = 4.45$ m/s;
(*b*) $\Delta p_A = -370$ kg · m/s, $\Delta p_B = 370$ kg · m/s.
81. $4.35\text{ kg} \cdot \text{m/s}(\hat{\mathbf{j}} - \hat{\mathbf{i}})$.
83. (*a*) -1.3×10^5 N;
(*b*) 8.5 *g*'s.
85. (*a*) 0.3 m/s;
(*b*) −5.9 m/s, 4.6 m/s;
(*c*) 0.5 m/s;
(*d*) 0.7 m/s;
(*e*) 3.2 m/s;
(*f*) The results for (*c*) and (*e*) are reasonable. The result for (*d*) is not reasonable.
87. (*a*) $v'_m = -4.67$ m/s, $v'_M = 3.73$ m/s;
(*b*) 2.23 m.
89. (*b*) 8.0×10^8 m;
(*c*) 35 ly;
(*d*) 5400 stars.
91. (*a*) 13 m;
(*b*) 49 m.
93. (*a*) 1.5×10^{21} J;
(*b*) 38,000 bombs.
95. (*a*) No;
(*b*) m_B/m_A;
(*c*) m_B/m_A;
(*d*) stays at rest.
97. $0\hat{\mathbf{i}} + \dfrac{4R}{3\pi}\hat{\mathbf{j}}$.
99. (*a*) $\dfrac{M_0 v_0}{M_0 + \dfrac{dM}{dt}t}$;
(*b*) 6.4 m/s, yes.
101. (*b*) $F_{\text{on person}} = 1000$ N, $F_{\text{friction}} = 700$ N.
103. (*a*) 6.0 m/s;
(*b*) 14 m/s;
(*c*) 4.9°.

Chapter 10

1. (*a*) 0.785 rad, $\pi/4$ rad;
(*b*) 1.05 rad, $\pi/3$ rad;
(*c*) 1.57 rad, $\pi/2$ rad;
(*d*) 6.283 rad, 2π rad;
(*e*) 7.77 rad, $89\pi/36$ rad.
3. 5300 m.
5. (*a*) 750 rad/sec;
(*b*) 23 m/s;
(*c*) 4.5×10^7 bit/s.
7. (*a*) 250 rad/sec;
(*b*) 44 m/s, 1.1×10^4 m/s^2.
9. (*a*) $\dfrac{\pi}{30}$ rad/sec, 1.05×10^{-1} rad/sec;
(*b*) $\dfrac{\pi}{1800}$ rad/sec, 1.75×10^{-3} rad/sec;
(*c*) $\dfrac{\pi}{21{,}600}$ rad/sec, 1.45×10^{-4} rad/sec;
(*d*) 0.
11. 9.30×10^{-3} rad, 9.06×10^{-3} rad.
13. (*a*) 1.99×10^{-7} rad/s;
(*b*) 7.27×10^{-5} rad/s.
15. (*a*) 3.4 rad/s^2;
(*b*) 110 m/s^2, 1.0 m/s^2.
17. R_2/R_1.
19. 2.8×10^4 rev.
21. (*a*) $-0.42\dfrac{\text{rad}}{\text{s}^2}$;
(*b*) 190 s.
23. (*a*) −3.1 rad/s^2;
(*b*) 12 s;
(*c*) 95 m.

25. (*a*) $9.5 - 26.0t + 6.4t^3$;
(*b*) $-26.0 + 19.2t^2$;
(*c*) 1.0×10^2 rad/s, 150 rad/s^2;
(*d*) 49 rad/s;
(*e*) 96 rad/s^2.
27. (*a*) 97 m · N.
29. (*a*) 36 m · N;
(*b*) 31 m · N.
31. (*a*) 340 N;
(*b*) 1800 N.
33. (*a*) 14 m · N;
(*b*) −13 m · N.
35. 1.6×10^4 m · N.
37. (*a*) 1.37×10^{-3} kg · m^2;
(*b*) 6.00×10^{-2} m · N.
39. (*a*) 72 rad/s^2;
(*b*) 620 N.
41. (*a*) 0.69 kg · m^2;
(*b*) 2.8×10^{-2} m · N.
43. (*a*) 0.081 m · N;
(*b*) 14 s.
45. (*a*) 1.90×10^3 kg · m^2.
(*b*) 9700 m · N.
47. (*a*) $14M\ell^2$;
(*b*) $\frac{14}{3}M\ell\alpha$;
(*c*) Perpendicular to both the rod and the axis of rotation.
49. (*a*) R_0;
(*b*) $\sqrt{\frac{1}{2}R_0^2 + \frac{1}{12}w^2}$;
(*c*) $\sqrt{\frac{1}{2}}R_0$;
(*d*) $\sqrt{\frac{1}{2}(R_1^2 + R_2^2)}$;
(*e*) $\sqrt{\frac{2}{5}}r_0$;
(*f*) $\sqrt{\frac{1}{12}}\ell$;
(*g*) $\sqrt{\frac{1}{3}}\ell$;
(*h*) $\sqrt{\frac{1}{12}(\ell^2 + w^2)}$.
51. (*a*) 0.69 m/s^2;
(*b*) 2%.
55. (*a*) $5.3Mr_0^2$;
(*b*) −15%.
57. (*a*) 0.042 m from the center, on a line to the small mass;
(*b*) 0.48 kg · m^2.
59. (*b*) $I_y = \frac{1}{12}M\ell^2$, $I_x = \frac{1}{12}Mw^2$.
61. 108 hp.
63. (*a*) 2.6×10^{29} J;
(*b*) 2.7×10^{33} J.
65. 1.40×10^4 J.
67. 1.9 m/s.
69. 9.22 m/s.
71. (*a*) 8.37 m/s, 34.2 rad/s;
(*b*) 5/2;
(*c*) Only ω_{bottom} depends on the radius; nothing depends on the mass.
73. (*a*) $-\frac{5}{7}a$;
(*b*) $\frac{2}{7}a$.
75. (*a*) $\dfrac{12v_0^2}{49\mu_k g}$;
(*b*) $\frac{5}{7}v_0$, $\frac{5}{7}v_0/r_0$.
77. (*a*) 1.6 m/s;
(*b*) 0.48 m.
79. (*a*) 4.1×10^5 J;
(*b*) 0.18;
(*c*) 1.3 m/s^2;
(*d*) 6%.
81. 0.74 m · N.
83. (*a*) 1.1 m/s;
(*b*) 94%.
85. 2.4 m · N, arm muscles.
87. (*a*) $\omega_R = \omega_F R_F / R_R$;
(*b*) 4.0;
(*c*) 1.5.
89. (*a*) 3.1×10^{-47} kg · m^2;
(*b*) 1.9×10^{-47} kg · m^2.
91. (*a*) $Mg\sqrt{\dfrac{h}{2R - h}}$;
(*b*) $\dfrac{Mg\sqrt{2Rh - h^2}}{R - h}$.
93. 660 m · N.
95. (*a*) 1.1 m;
(*b*) 0.81 m.
97. (*a*) $h = 2.7R$;
(*b*) $h = 2.7(R - r)$.
99. $\frac{7}{5}r\left(1 - \dfrac{\mu_s Mg}{F}\right) \le h \le \frac{7}{5}r\left(1 + \dfrac{\mu_s Mg}{F}\right)$.
101. (*a*) $\dfrac{x}{y}g$;
(*b*) small, large, upwards and towards the rear of the bicycle;
(*c*) 3.6 m/s^2.
103. (*a*) $\sqrt{\dfrac{3g}{\ell}(1 - \sin\phi)}$;
(*b*) $\sqrt{3g\ell}$.
105. (*a*) $\sqrt{\frac{4}{3}gh}$;
(*b*) $\sqrt{\frac{4}{3}gh}$.
107. (*a*) 1.7×10^8 J;
(*b*) 2100 rad/s;
(*c*) 25 min.

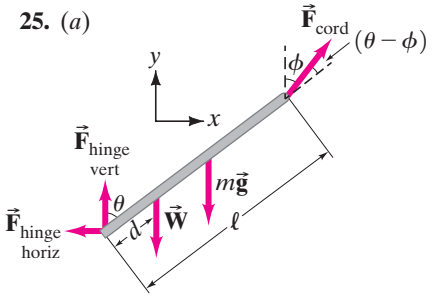
Chapter 11

1. 3.41 kg · m^2/s.
3. $\frac{1}{2}\omega$.
5. 1.3 kg · m^2, pulling her limbs in toward the center of her body.
7. (*a*) 7.08×10^{33} kg · m^2/s;
(*b*) 2.67×10^{40} kg · m^2/s.
9. 2.5 rev/s.
11. (*a*) $-\omega_W I_W / I_P$;
(*b*) $-\dfrac{I_W}{2I_P}\omega_W$;
(*c*) $-\omega_W I_W / I_P$;
(*d*) 0.
13. (*a*) 0.47 rad/s;
(*b*) 0.80 rad/s.
15. $\dfrac{\omega_{\text{out}}}{\omega_{\text{in}}} = 0.3$.
17. $(\frac{1}{12}M + \frac{1}{2}m)\ell^2\omega$.
19. (*a*) 3.7×10^{16} J;
(*b*) 1.9×10^{20} kg · m^2/s.
21. −0.35 rad/s.
25. 45°.
31. (*a*) $-7.0\hat{\mathbf{i}} - 15\hat{\mathbf{j}} + 4\,\hat{\mathbf{k}}$;
(*b*) 170°.
33. Yes, no.
39. $(-63\hat{\mathbf{i}} - 51\hat{\mathbf{j}} + 55\hat{\mathbf{k}})$kg · m^2/s.
41. (*a*) $(\frac{7}{9}m + \frac{1}{6}M)\ell^2\omega^2$;
(*b*) $(\frac{14}{9}m + \frac{1}{3}M)\ell^2\omega$.
43. (*a*) $\left((M_A + M_B)R_0 + \dfrac{I}{R_0}\right)v$;
(*b*) $\dfrac{M_B g}{M_A + M_B + \dfrac{I}{R_0^2}}$.
45. (*a*) $\dfrac{(m_A r_A^2 + m_B r_B^2)(\omega^2 \sin^2\phi)}{2d\tan\phi}$;
(*b*) 24 N.
47. $\dfrac{m^2v^2}{g(m + M)(\frac{4}{3}M + m)}$.
49. (*a*) 4.4 rad/s;
(*b*) 0;
(*c*) 18 rad/s;
(*d*) 2500 J.
51. (*a*) 14 m/s;
(*b*) 5.0 rad/s.
53. (*b*) $\dfrac{5v_0}{2R}$;
(*c*) $\frac{1}{14}v_0$;
(*d*) $-\frac{1}{14}v_0$.
55. (*a*) 2.6 s;
(*b*) 5.2 s.
57. 16 rad/s.
59. An axis perpendicular to the plane of the orbit; no.
61. (*a*) 9.80 m/s^2, inward along a radial line;
(*b*) 9.78 m/s^2, 0.0988° south from an inward radial line;
(*c*) 9.77 m/s^2, inward along a radial line.
63. Due north or due south.
65. $(m\omega^2 r - F_{fr})\,\hat{\mathbf{i}} + (F_{\text{spoke}} - 2m\omega v)\,\hat{\mathbf{j}} + (F_N - mg)\,\hat{\mathbf{k}}$;
67. (*a*) $(-36\hat{\mathbf{i}} + 42\hat{\mathbf{j}} - 21\hat{\mathbf{k}})$kg · m^2/s;
(*b*) $(16\hat{\mathbf{j}} - 8.0\hat{\mathbf{k}})$m · N.

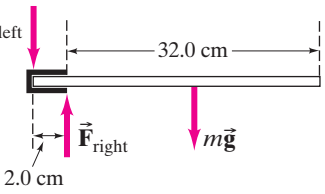
69. (*b*) 0.750.
71. $-0.22\ \text{kg}\cdot\text{m}^2/\text{s}\ \hat{\mathbf{k}}$.
73. $v\left[-\sin(\omega t)\,\hat{\mathbf{i}}+\cos(\omega t)\,\hat{\mathbf{j}}\right],\left(\frac{v}{R}\right)\hat{\mathbf{k}}$.
75. (*a*) The tire will turn to the right;
(*b*) $\Delta L = 0.19L_0$.
77. (*a*) $820\ \text{kg}\cdot\text{m}^2/\text{s}^2$;
(*b*) $820\ \text{m}\cdot\text{N}$;
(*c*) 850 W.
79. $\vec{\mathbf{a}}_{\text{tan}} = -\alpha R\sin\theta\,\hat{\mathbf{i}} + \alpha R\cos\theta\,\hat{\mathbf{j}}$
(*a*) $mR^2\alpha\hat{\mathbf{k}}$;
(*b*) $mR^2\alpha\hat{\mathbf{k}}$.
81. 0.965.
83. (*a*) 14,000 rev/s;
(*b*) 4300 rev/s.
85. 0.64 m.

Chapter 12

1. 558 N, 117°.
3. (*a*) 2.3 m from tower;
(*b*) 4100 kg.
5. (*a*) $F_A = 1700$ N, downward;
$F_B = 2300$ N, upward;
(*b*) $F_A = 2.0\times10^3$ N, downward;
$F_B = 2800$ N, upward.
7. $F_{\text{near}} = 2900$ N downward,
$F_{\text{far}} = 1300$ N downward.
9. (*a*) 2.3 m from the adult;
(*b*) 2.6 m from the adult.
11. $F_{\text{left}} = 260$ N, $F_{\text{right}} = 190$ N.
13. 20 N to 50 N.
15. 0.64 m.
17. (*a*) 410 N;
(*b*) 328 N, 410 N.
19. 1.5×10^4 N, 7.0×10^3 N.
21. (*a*) 230 N;
(*b*) 1.0×10^2 N.
23. (*a*) 323 N;
(*b*) 5.0×10^2 N.
25. (*a*)



(*b*) −9 N, 51 N;
(*c*) 2.4 m.
27. 0.50.
29. $5.2\ \text{m/s}^2$.
31. 970 N.
33. (*a*) 270 N;
(*b*) 270 N.
35. 1400 N, 2100 N.
37. 29°.
39. 4.01×10^{-2} m.
41. (*a*) $1.4\times10^6\ \text{N/m}^2$;
(*b*) 6.9×10^{-6};
(*c*) 6.6×10^{-5} m.
43. $9.0\times10^7\ \text{N/m}^2$, 900 atmospheres.
45. 33 kg.
47. 3.2 mm.
49. (*a*) No;
(*b*) 1.3×10^{-3} m.
51. $A_1 = 1.6\times10^{-3}\ \text{m}^2$,
$A_2 = 9.1\times10^{-3}\ \text{m}^2$.
53. 2.7 cm.
55. (*a*) $F_T = 140$ kN,
$\vec{\mathbf{F}}_A = 160$ kN, 23° above AC;
(*b*) $F_{AB} = F_{BC} = F_{DC} = F_{DE} =$ 72 kN compression,
$F_{CA} = 107$ kN compression,
$F_{DB} = 72$ kN tension,
$F_{CE} = 36$ kN compression.
57. $F_{AB} = F_{DB} = F_{DE} = 7.5\times10^4$ N, compression,
$F_{ACx} = F_{CE} = 3.7\times10^4$ N, tension
$F_{BC} = F_{DC} = 7.5\times10^4$ N, tension.
59. 12 m.
61. (*a*) $\mu_s < \ell/2h$;
(*b*) $\mu_s > \ell/2h$.
63. (*a*)



(*b*) $F_{\text{Right}} = 550$ N, $F_{\text{Left}} = 490$ N,
$mg = 65$ N;
(*c*) $11\ \text{m}\cdot\text{N}$.
65. (*a*) 0.78 N;
(*b*) 0.98 N.
67. 3.5×10^{-4}m.
69. The painter can walk to within 5 cm of the right edge, and to within 17 cm of the left edge. Both ends are dangerous.
71. 2.45 m.
73. 2.0×10^3 N.
75. (*a*) $(5.4\times10^{-6})\%$;
(*b*) 1.1×10^{-17} m.
77. 150 N, 0.83 m.
79. 5500 N, no.
81. (*a*) 0.27 *mg*;
(*b*) 0.57 *mg*;
(*c*) horizontal at the lowest point, 62° below the horizontal at the attachment points.
83. (*a*) $\dfrac{g(2mx+M\ell)}{2\ell\sin\theta}$;
(*b*) $F_{\substack{\text{hinge}\\\text{horiz}}} = \dfrac{g(2mx+M\ell)}{2\ell\tan\theta}$,
$F_{\substack{\text{hinge}\\\text{vert}}} = mg\left(1-\dfrac{x}{\ell}\right)+\frac{1}{2}Mg$.
85. 0.257.
87. 3.8.
89. (*b*) $\theta_{\mu_W=0.40} = 46°$, $\theta_{\mu_W=0} = 51°$, 11%.
91. $F_A(t) = -(8.0\times10^4\ \text{N/s})t + 1.0\times10^6$ N,
$F_B(t) = (8.0\times10^4\ \text{N/s})t + 1.1\times10^5$ N.
93. (*a*) 7.7×10^4 N;
(*b*) $F_{Hx} = 1.2\times10^5$ N,
$F_{Hy} = 4.2\times10^5$ N.
95. (*a*) 7.7 m/s;
(*b*) 4500 N;
(*c*) 1.9×10^5 N;
(*d*) $7.4\times10^6\ \text{N/m}^2$, won't break;
(*e*) $3.2\times10^8\ \text{N/m}^2$, will break.

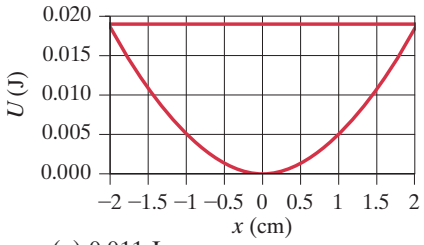
Chapter 13

1. 2×10^{11} kg.
3. 710 kg.
5. 0.8612.
7. (*a*) $6.1\times10^6\ \text{N/m}^2$;
(*b*) $1.5\times10^5\ \text{N/m}^2$.
9. (*a*) 4.5×10^5 N;
(*b*) 4.5×10^5 N.
11. 13 m.
13. 140 cm^2.
15. (*a*) 2.4×10^7 N, $1.19\times10^5\ \text{N/m}^2$;
(*b*) $1.19\times10^5\ \text{N/m}^2$.
17. $683\ \text{kg/m}^3$.
19. $4.3\times10^5\ \text{N/m}^2$.
21. (*a*) $1.29\times10^5\ \text{N/m}^2$;
(*b*) $9.4\times10^4\ \text{N/m}^2$.
23. (*c*) $t \geq 0.38\ h$, no.
27. $3240\ \text{kg/m}^3$.
29. 810 kg.
31. 10.5%.
33. 32 bottles.
35. (*a*) 7.40×10^5 N;
(*b*) 1.0×10^4 N.
39. 2700 balloons.
41. 9 N downward, 21 N upward.
43. 4.11 kg.
45. 4.0 m/s.
47. $4.12\times10^{-3}\ \text{m}^3/\text{s}$.
49. 1.2×10^5 N.
51. 9.7×10^4 Pa.
55. 2.5 m/s, 2.2 atm.

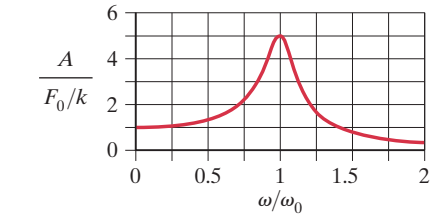
57. (*a*) $\sqrt{\frac{2P_2}{\rho} + 2g(y_2 - y_1)}$;
(*b*) 15 m/s.
61. 1400 N.
63. 0.43 m/s.
65. 7700 Pa.
67. 0.098 m.
69. (*a*) 2200, turbulent;
(*b*) 4400, turbulent.
71. 1.0 m.
73. 1.1×10^{-2} N.
75. 1.5 mm.
79. (*a*) 0.71 m;
(*b*) 0.68 m;
(*c*) 1.0 m.
81. 0.047 atm change.
83. 1.18×10^5 Pa.
85. 0.28 N.
87. 0.6 atm.
89. (*a*) 88 Pa/s
(*b*) 50 s.
91. 5.3 km.
93. 1/2.
95. $d_1 = d_0\left(\frac{v_0^2}{v_0^2 + 2gy}\right)^{\frac{1}{4}}$.
97. (*b*) Above.
99. 5.29×10^{18} kg.
101. 1.5×10^4 N.
103. (*a*) 7.2 m/s;
(*b*) 12 L/min;
(*c*) 0.72 m/s.
105. 6.2 s.

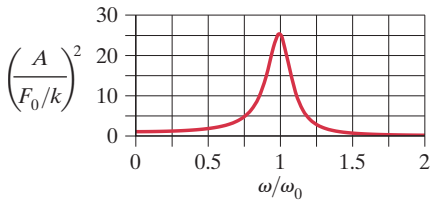
Chapter 14

1. 1.08 m.
3. (*a*) $(7.8\text{ cm})\cos(9.5\,t)$;
(*b*) −1.1 cm.
5. 490 N/m.
7. (*a*) 0.18 N/m;
(*b*) 3.1 Hz.
9. 0.91 kg.
11. 0.248 s.
13. (*a*) $A_A = 2.5$ m, $A_B = 3.5$ m;
(*b*) $f_A = 0.25$ Hz, $f_B = 0.50$ Hz;
(*c*) $T_A = 4.0$ s, $T_B = 2.0$ s;
(*d*) $x_A = (2.5\text{ m})\sin(\frac{1}{2}\pi t)$,
$x_B = (3.5\text{ m})\cos(\pi t)$.
15. (*a*) $y(t) =$
$(0.280\text{ m})\sin[(27.5\text{ rad/s})t]$;
(*b*) $t_{\text{longest}} =$
$5.71 \times 10^{-2}\text{ s} + n(0.228\text{ s})$,
$n = 0, 1, 2, \cdots$,
$t_{\text{shortest}} =$
$1.71 \times 10^{-1}\text{ s} + n(0.228\text{ s})$,
$n = 0, 1, 2, \cdots$.
17. (*a*) 5.0 m/s;
(*b*) 1.4×10^4 m/s^2.
19. (*a*) $y = (0.16\text{ m})\cos(9.2\,t)$;
(*b*) 0.17 s;
(*c*) 1.5 m/s;
(*d*) 14 m/s^2, at the release point.
21. $\frac{1}{2\pi}\sqrt{\frac{2k}{m}}$.
23. $\frac{1}{2\pi}\sqrt{\frac{3k}{M}}$.
27. (*a*) 0.650 m;
(*b*) 1.18 Hz;
(*c*) 14.8 J;
(*d*) 12.4 J, 2.37 J.
29. $\sqrt{3}$.
31.



(*a*) 0.011 J;
(*b*) 0.008 J;
(*c*) 0.5 m/s.
33. (*a*) 52 N/m;
(*b*) 0.052 J.
35. (*a*) 380 N/m;
(*b*) 4.1 kg.
37. (*a*) 0.436 s, 2.29 Hz;
(*b*) 0.147 m;
(*c*) 30.6 m/s^2;
(*d*) $x(t) = (0.147\text{ m})\sin(4.58\pi t)$;
(*e*) 1.99 J;
(*f*) 1.67 J.
39. 91 N/m.
41. 3.5 s.
43. 0.330 m.
45. (*a*) −6.2°;
(*b*) 9.7°;
(*c*) −14°.
47. Shorten it by 0.4 mm.
49. (*a*) −0.15 rad;
(*b*) $\pi/15$ rad;
(*c*) $\pi/15$ rad.
51. 1.6 g.
53. (*a*) $2\pi\sqrt{\frac{(M + \frac{1}{3}m)\ell}{(M + \frac{1}{2}m)g}}$;
(*b*) $1 - \sqrt{\frac{(M + \frac{1}{2}m)}{(M + \frac{1}{3}m)}}$;
(*c*) 2.
55. (*a*) 1.6 s,
57. 1.08 s.
59. Decreases by a factor of ~6.
61. (*a*) -2.2×10^{-3}%;
(*b*) 24.
63. 0.030 kg/s.
67. 2.9 m/s.
71.

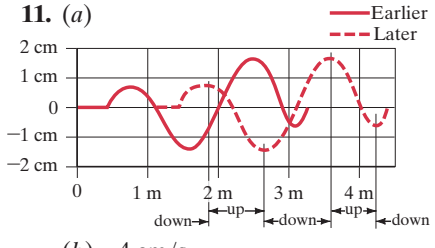




75. 0.11 m.
77. (*a*) 1.16 *f*;
(*b*) 0.81 *f*.
79. $x(t) =$
$(0.118\text{ m})\cos[(6.0\pi\text{ rad/s})t + 2.2]$.
81. (*a*) 0.44 s;
(*b*) 3 mm.
83. 0.13 m/s, 0.12 m/s^2, 1.2%.
85. 0.9922 m, 1.6 mm, 0.164 m.
87. 2.8 cm.
89. (*a*) 98 N/m;
(*b*) 8.8 cm.
91. 5.4 mm.
93. (*a*) $\pm 0.866x_0$;
(*b*) $\pm\frac{1}{2}x_0$.
95. 84.5 min.
97. 1.36 Hz.

Chapter 15

1. 2.5 m/s.
3. (*a*) 1400 m/s;
(*b*) 4100 m/s;
(*c*) 5100 m/s.
5. 0.58 m.
7. 4.8 N.
9. (*a*) 84 m/s;
(*b*) 9600 N.
11. (*a*)



(*b*) −4 cm/s.
13. 0.01 s.
15. The intensity at 12 m is 4 times greater.
17. (*a*) $I_{15\text{ km}}/I_{55\text{ km}} = 13$;
(*b*) $A_{15\text{ km}}/A_{55\text{ km}} = 3.7$.

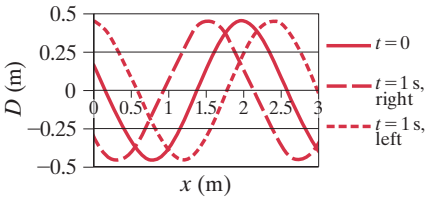
19. (*a*) 0.24 W;
(*b*) 0.20 cm.

21. (*b*) 6.0×10^2 W.

23. (*a*) $D(x, t) = 0.015 \sin(35x + 1200t)$;
(*b*) 34 m/s.

25. The traveling wave is given by $D = 0.22 \sin(5.6x + 34t)$.
(*a*) 1.1 m;
(*b*) 5.4 Hz;
(*c*) 0.22 m;
(*d*) 6.1 m/s, negative x direction;
(*e*) 7.5 m/s, 0.

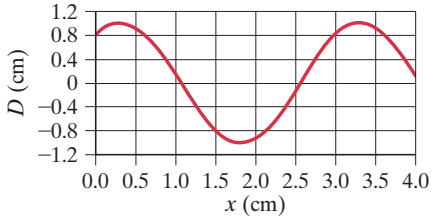
27. (*a*)



(*b*) $D(x, t) = (0.45\text{ m})\cos[2.6(x - 2.0t) + 1.2]$;
(*d*) $D(x, t) = (0.45\text{ m})\cos[2.6(x + 2.0t) + 1.2]$.

29. (*a*) 0.603 m;
(*b*) 4.37×10^{-4} s;
(*c*) 26°.

31. (*a*)



(*b*) $D(x, t) = (1.00\text{ cm})\sin\left[(2.09\text{ cm}^{-1})x + (1540\text{ rad/s})t + 0.93\right]$, x in cm, t in s.

39. $\pi/50$.

41. (*a*) 0.84 m;
(*b*) 0.26 N;
(*c*) 0.59 m.

43. (*c*) It is all kinetic energy.

45. 662 Hz.

47. 80 Hz: fundamental or first harmonic, 160 Hz: first overtone or second harmonic, 240 Hz: second overtone or third harmonic, 400 Hz: fourth overtone or fifth harmonic.

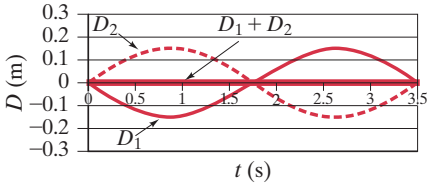
49. 80 Hz.

53. (*a*) 1.3 kg;
(*b*) 0.32 kg;
(*c*) 5.2×10^{-2} kg.

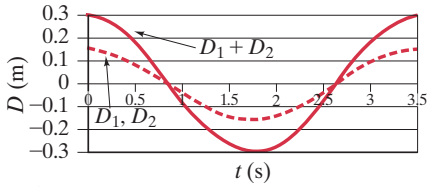
55. (*a*) 5.2 c1,2 pt, 6.7 Hz, 7.0×10^1 cm/s.
(*c*) 74 cm/s.

57. 325 Hz.

59. (*a*)



(*b*)



61. 4, 8, 12.

63. 0.77 m/s.

65. 5.2 km/s.

67. 26°.

69. 41°.

71. (*a*) 0.042 m;
(*b*) 0.55 rad.

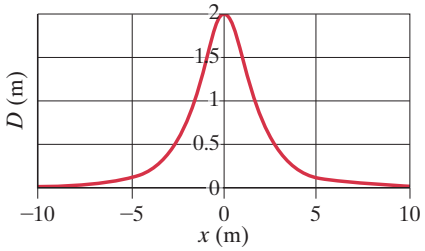
73. (*a*) $t = \dfrac{2}{v}\sqrt{D^2 + \left(\frac{1}{2}x\right)^2}$.

75. (*a*) 0.05 m;
(*b*) 9.

77. 0.69 m.

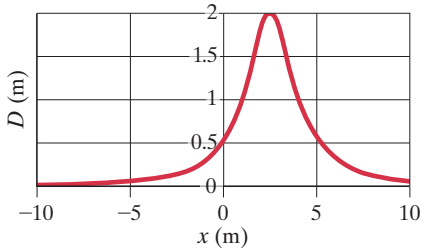
79. The less dense material, 1.4.

81. (*a*) $t = 0$ sec;

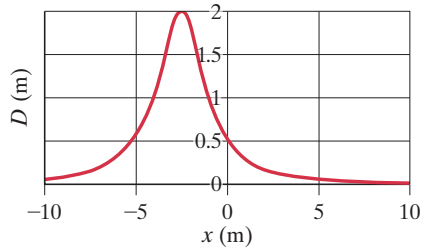


(*b*) $D(x,t) = \dfrac{4.0\text{ m}^3}{[x - (2.4\text{ m/s})t]^2 + 2.0\text{ m}^2}$;

(*c*) $t = 1.0$ sec, moving right;



(*d*) $D(x,t) = \dfrac{4.0\text{ m}^3}{[x + (2.4\text{ m/s})t]^2 + 2.0\text{ m}^2}$,
$t = 1.0$ sec, moving left.



83. $f_{\text{light}}/f_{\text{heavy}} = \sqrt{2}$.

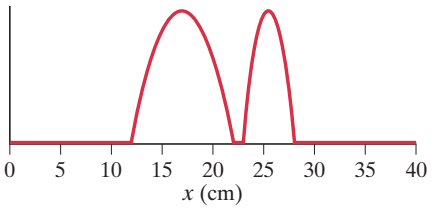
85. (*a*) G: 784 Hz, 1180 Hz; B: 998 Hz, 1480 Hz;
(*b*) 1.59;
(*c*) 1.26;
(*d*) 0.630.

87. 6.36 m from the end where the first pulse originated.

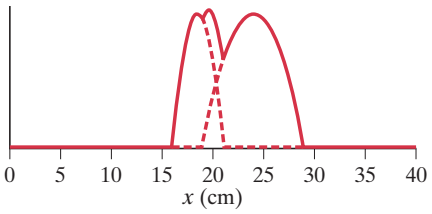
91. Yes.

93. 18 W.

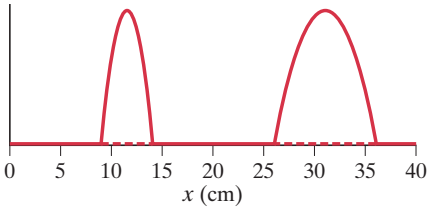
95. $t = 1.0$ s



$t = 2.0$ s



$t = 3.0$ s



Chapter 16

1. 500 m.

3. −18°C.

5. (*a*) 8%;
(*b*) 4%.

7. (*a*) 4.5×10^{-9} m;
(*b*) 4.5×10^{-11} m.

9. (*a*) $-(6.0 \times 10^{-5}\text{ Pa}) \times \cos\left[(0.45\pi\text{ m}^{-1})x - (150\pi\text{ rad/s})t\right]$;
(*b*) $-(6.0 \times 10^{-3}\text{ Pa}) \times \cos\left[(45\pi\text{ m}^{-1})x - (1.5 \times 10^4\pi\text{ rad/s})t\right]$.

11. 65 dB.

13. Lowest: 70 – 80 Hz;
Highest: 20,000 Hz.

15. (*a*) 10^9;
(*b*) 10^{12}.

17. 134 dB.

19. 7 dB.

21. (*a*) 2.4×10^{-5} m;
(*b*) 3.0×10^1 Pa.

23. (*a*) Higher frequency, 3.5;
(*b*) 12.

25. (*a*) 9.4×10^{-6} W;
(*b*) 6 million people.
26. 140 dB.
27. (*a*) $\beta_{280} = 120$ dB, $\beta_{45} = 110$ dB;
(*b*) 450 W.
29. (*a*) 79.4 Hz, 238 Hz, 397 Hz, 556 Hz;
(*b*) 159 Hz, 318 Hz, 476 Hz, 635 Hz.
31. 10.
33. 260 Hz.
35. (*a*) 0.652 m;
(*b*) 1.30 m, 262 Hz;
(*c*) 1.30 m, 262 Hz.
37. −2.3%.
39. 3.9 m, open.
41. (*a*) 253;
(*b*) 253.
43. 19 Hz, 38 Hz.
45. $x_1 = 4.2$ cm, $x_2 = 8.2$ cm,
$x_3 = 11.9$ cm, $x_4 = 15.5$ cm,
$x_5 = 18.8$ cm, $x_6 = 22.0$ cm.
47. ±0.50 Hz.
49. 347 Hz.
51. 3.7 Hz.
53. 9 beats per second.
55. (*a*) 343 Hz;
(*b*) 1030 Hz, 1720 Hz.
57. (*a*) 1770 Hz;
(*b*) 1530 Hz.
59. 3 Hz.
61. (*a*) 2.8 s;
(*b*) 7.3 s.
63. (*a*) 1750 Hz, 1440 Hz;
(*b*) 1880 Hz, 1340 Hz;
(*c*) 1640 Hz, 1530 Hz.
65. (*a*) 770 Hz;
(*b*) 770 Hz;
(*c*) 770 Hz;
(*d*) 770 Hz;
(*e*) 796 Hz;
(*f*) 797 Hz.
67. 20 km/h.
69. (*a*) 76;
(*b*) 0.76°.
73. 0.12 s.
75. (*a*) 0.17 m;
(*b*) 11 m;
(*c*) 0.5%.
77. 14 W.
79. (*a*) 57 Hz, 69 Hz, 86 Hz, 110 Hz, 170 Hz.
81. (*a*) −6 dB, not audible;
(*b*) 20 dB, audible.
83. 114 m/s.
85. 635 Hz.
87. 0.44, 0.20, 0.088.
89. (*a*) 280 m/s, 57 N;
(*b*) 0.19 m;
(*c*) string: 880 Hz; pipe: 1320 Hz.
91. 2.9×10^{-9} J.
93. 2.62 m.
95. 10.6 m.
97. (*a*) 1.3 m;
(*b*) 3, 4, 5, and 6.
99. 50 Hz, 6th overtone.
101. (*a*) 3200 Hz;
(*b*) 1.6 m;
(*c*) 0.11 m.
103. 0.17 m/s.

Chapter 17

1. 3.2×10^{22}.
3. (*a*) 20° C;
(*b*) 3500° F.
5. 101.3° F.
7. (*a*) 57.9° C;
(*b*) 17.6° C.
9. 0.08 m.
11. 981 kg/m^3.
13. −65° C.
15. 0.83%.
17. 6.59 mL.
19. (*a*) 5.0×10^{-5}/C°;
(*b*) Copper.
23. (*a*) 2.7 cm;
(*b*) 0.3 cm.
25. 45 min lost.
27. 4.6×10^7 N/m^2.
29. (*a*) 27° C;
(*b*) 5500 N.
31. −459.67° F.
33. (*a*) 4300 K, 15×10^6 K;
(*b*) 6.4%, 1.8×10^{-3}%.
35. 1.42 m^3.
37. 1.25 kg/m^3.
39. (*a*) 21.2 m^3;
(*b*) 2.22 atm.
41. 1.57×10^8 Pa.
43. 333° C.
45. 4.23 atm.
47. −130° C.
49. (*a*) 3 m^3 or 3000 L;
(*b*) 0.01 kg/m^3;
(*c*) 1/100.
51. 7.0 min.
53. 181° C.
55. 4.29 cm^3.
57. 55.5 mol, 3.34×10^{25} molecules.
59. 4×10^{-17} Pa.
61. 300 molecules/cm^3.
63. (*a*) 71.2 torr;
(*b*) 218° C.
65. 223 K.
67. (*a*) Low;
(*b*) 2.8×10^{-2}%.
69. 21%.
71. 9.1 L, not advisable.
73. (*a*) 1500 kg;
(*b*) 200 kg enters the house.
75. 20 air molecules.
77. (*a*) Lower;
(*b*) 0.36%.
79. 14%.
81. 3.34×10^{-9} m.
83. 950 min.
85. (*a*) 660 kg/m^3;
(*b*) −3%.
87. ±0.11 C°.
89. 2300 m.
91. 76.3 cm.
93. Slope: 0.0492 mL/C°;
effective $\beta = 4.92 \times 10^{-4}$/C°;
$\beta_{\text{liquid}} = 5.01 \times 10^{-4}$/C°; glycerin.
95. Thermal expansion method: 260 L, 260 L; Ideal gas method: 430 L, 270 L.
97. 0.03 atm.

Chapter 18

1. 6×10^3 m/s.
3. (*a*) 4.5;
(*b*) 5.2.
5. 2450°C.
7. (*a*) 3600 J;
(*b*) 11 m/s.
9. $\sqrt{3}$.
13. (*b*) 5.6%.
15. 50.0°C.
17. (*a*) 461 m/s;
(*b*) 22 round trips per second.
19. (*c*) 3×10^{28} collisions per second.
21. Triple the temperature.
25. (*a*) Solid or vapor;
(*b*) 5.11 atm $\leq P \leq$ 73 atm and
$-56.6°\text{C} \leq T \leq 31°\text{C}$.
29. 1500 Pa or 0.015 atm.
31. 94°C.
33. 7.0×10^1 g.
35. 2.90 atm.
37. (*a*) Greater;
(*b*) 2.0×10^{-4}%;
(*c*) 0.603%.
39. (*a*) 89.8°C;
(*b*) 6090 m.

41. 4.7×10^{-10} m.

43. $P_{\text{Van der Waals}} = 0.95 P_{\text{Ideal}}$ for $V = 4.16 \times 10^{-5}$ m^3 and $V = 2.16 \times 10^{-4}$ m^3; $P_{\text{Van der Waals}} = 1.05\ P_{\text{Ideal}}$ for $V = 3.46 \times 10^{-5}$ m^3.

45. 0.010 Pa.

47. (*a*) 3.9×10^{-10} m; (*b*) 1.8×10^{-10} m.

49. 5×10^9 collisions per second.

51. 7.3×10^{-3} Pa.

53. 3.5 h, convection is more important that diffusion.

55. (*b*) 4×10^{-11} mol/s; (*c*) 0.6 s.

57. 260 m/s, 3.7×10^{-22} atm.

59. (*a*) 290 m/s; (*b*) 9.5 m/s.

61. (*a*) 1.61×10^5 K; (*b*) 2.01×10^4 K.

63. (*a*) 5.8×10^{23}; (*b*) 1:1 ; (*c*) $(v_{\text{rms}})_{\text{He}}/(v_{\text{rms}})_{\text{O}_2} = \sqrt{8}$.

65. (*a*) $r_{\substack{\text{inter-}\\\text{molecular}}} = 2.07 \times 10^{-9}$ m, $d_{\substack{\text{inter-}\\\text{molecular}}}/d_{\substack{\text{actual}\\\text{molecule}}} \approx 14$; (*b*) 55 cm; (*c*) $r_{\substack{\text{inter-}\\\text{molecular}}} = 1.44 \times 10^{-9}$ m, $d_{\substack{\text{inter-}\\\text{molecular}}}/d_{\substack{\text{actual}\\\text{molecule}}} \approx 9.6$; (*d*) At 1 atm: 3.8×10^{-2}, at 3 atm: 0.11%.

67. $U/K = 1.27 \times 10^{-4}$, yes.

69. 1.5×10^5 K.

71. 2210 m/s, 3.47.

73. (*a*) 330 m/s; (*b*) 3.9×10^7 collisions/s; (*c*) 3.5×10^{26} collisions/s.

75. 2×10^{13} m.

77. (*b*) $a = 0.365 \dfrac{\text{N} \cdot \text{m}^4}{\text{mol}^2}$, $b = 4.28 \times 10^{-5}$ m^3/mol.

Chapter 19

1. 10.5°C.

3. (*a*) 1.0×10^7 J; (*b*) 2.9 kWh; (*c*) \$0.35 per day, no.

5. 4.2×10^5 J, 1.0×10^2 kcal.

7. 170 s.

9. 2100 J/kg · C°.

11. (*a*) 4.0×10^2 s (*b*) 83 min.

13. 18.5°C.

15. 330 Cal.

17. 43 C°.

19. 0.39 C°.

21. 8.0×10^6 J.

23. 0.316 kg.

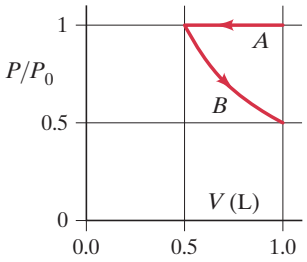
25. (*a*) 5.2×10^5 J; (*b*) 1.5×10^5 J.

27. 6600 kcal.

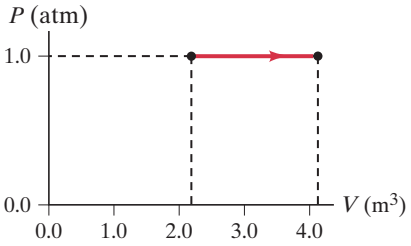
29. 360 m/s.

31. 9.8×10^{-3} kg.

33.



35. (*a*) 1.9×10^5 J; (*b*) 4.9×10^5 J; (*c*)



37. 3.5×10^{-3} kg.

39. (*a*) 480 J; (*b*) 0 J; (*c*) 480 J into the gas.

41. (*a*) 0 J; (*b*) 2630 J; (*c*) rise.

43. (*a*) 1.72×10^5 J; (*b*) 2.09×10^6 J.

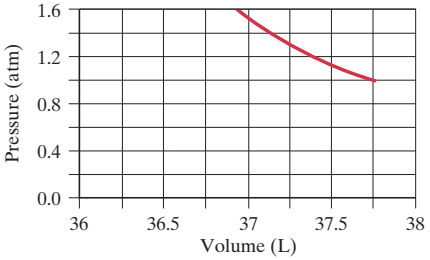
45. $RT \ln \dfrac{(V_2 - b)}{(V_1 - b)} + a\left(\dfrac{1}{V_2} - \dfrac{1}{V_1}\right)$.

47. (*a*) 30 J; (*b*) 68 J; (*c*) −84 J; (*d*) −114 J; (*e*) −15 J.

49. 3.85×10^5 J.

55. (*a*) 0.241 kcal/kg · K; (*b*) triatomic.

57. (*a*)



(*b*) 5.40×10^3 J; (*c*) −104 J; (*d*) 5.30×10^3 J.

61. 0.15 m^3.

63. (*d*) 35°C.

65. 1.6×10^4 W.

67. 7.5×10^4 s.

69. 20 bulbs.

71. 130°C.

73. 350 Btu/h, 1.0×10^2 W.

75. (*a*) 3.2×10^{26} W; (*b*) 1.1×10^3 W/m^2.

77. 100°C, 1/3 of the mass as steam, 2/3 of the mass as liquid water.

79. (*a*) 33 W; (*b*) 5500 W.

81. (*a*) 1.8×10^5 W; (*b*) 3.0×10^8 J; (*c*) \$770.

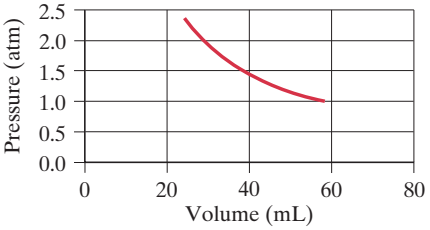
83. 0.15 C°.

85. (*a*) 8.0×10^2 W; (*b*) 5.3 g.

87. 1.4×10^5 s ≈ 1.6 d.

89. 290 J.

91. (*a*) 4.79 cm in diameter; (*b*)



(*c*) $W = 4.99$ J, $\Delta E_{\text{int}} = 0$, $Q = 4.99$ J added.

93. 110°C.

95. 305 J.

97. (*a*) 1.73×10^{17} W; (*b*) 278 K.

99. (*a*) The ice warms from −20°C to −15°C;
(*b*) 3.00 g of ice will melt;
(*c*) The water will warm to 12°C;
(*d*) 0.442 g of water will be converted to steam;
(*e*) The steam warms from 110°C to 115°C.

101. (*a*) 3.4 W;
(*b*) 2.3 C°/s;
(*c*) no;
(*d*) 86°C;
(*e*) Conduction, convection, evaporation.

103. 15%.

Chapter 20

1. 0.25 or 25%.

3. 1400 MW.

5. (*a*) 1.8×10^4 J/s;
(*b*) 8.2×10^4 J/s;
(*c*) 33 km/gal, 20 mi/gal, 8.6 km/L, 5.4 mi/L.

7. 0.230 or 23.0%.

9. 250°C.

13. 1.7×10^{13} J/h.

15. 1400 m.

17. 86°C.

19. (*a*) $P_a = 4.1 \times 10^5$ Pa, $P_b = 2.1 \times 10^5$ Pa;
(*b*) $V_c = 34.4$ L, $V_d = 17$ L;
(*c*) 2100 J;
(*d*) 1500 J;
(*e*) 600 J;
(*f*) 0.28.

21. 8.55.

23. 6.5.

25. −15°C.

27. (*a*) 17;
(*b*) 7.1×10^7 J.

29. (*a*) −6°C;
(*c*) 25%.

31. 78 L.

33. \$1300, 11 y, \$13,000.

35. 1300 J/K.

37. 0.19 J/K.

39. 4×10^4 J/K.

41. $4.77 \times 10^{-2} \dfrac{\text{J/K}}{\text{s}}$.

43. 8.7 J/K.

45. (*a*) $e_{\text{actual}}/e_{\text{ideal}} = 0.67$;
(*b*) 0.42 J/K;
(*c*) 0.

47. (*a*) $(93\ m)$J/K, yes;
(*b*) $(-93\ m)$J/K, no.

49. (*a*) 1100 J/K;
(*b*) 1130 J/K;
(*c*) −990 J/K.

51. (*a*) Adiabatic;
(*b*) $\Delta S_{\text{adiabatic}} = 0$, $\Delta S_{\text{isothermal}} = -nR\ln 2$;
(*c*) $\Delta S_{\text{environment}} = 0$, $\Delta S_{\text{isothermal}} = nR\ln 2$.

55. −17 mJ/K.

59. $S_{4\text{ heads}} = 0$,
$S_{3\text{ heads, }1\text{ tail}} = 1.91 \times 10^{-23}$ J/K,
$S_{2\text{ heads, }2\text{ tails}} = 2.47 \times 10^{-23}$ J/K,
$S_{1\text{ head, }1\text{ tails}} = 1.91 \times 10^{-23}$ J/K,
$S_{4\text{ tails}} = 0$.

61. (*a*) 2.47×10^{-23} J/K;
(*b*) -9.2×10^{-22} J/K;
(*c*) $\Delta S_{\substack{\text{coin}\\\text{examples}}} \ll \Delta S_{\substack{\text{thermo}\\\text{examples}}}$.

63. (*a*) 7.63×10^{-23} J/K, yes;
(*b*) -7.63×10^{-23} J/K, no.

65. (*a*) 1.03×10^6 kWh;
(*b*) 5.5×10^4 kW.

67. Yes.

69. (*a*) 0.41 mol;
(*b*) 4.0×10^2 K;
(*c*) 810 J;
(*d*) -7.0×10^2 J;
(*e*) 810 J;
(*f*) 0.13;
(*g*) 0.24.

71. (*a*) 1.0×10^2 kg/s;
(*b*) 9.8×10^7 gal/h.

73. (*a*) 0.19;
(*b*) $e_{\text{rectangle}}/e_{\text{Carnot}} = 0.23$.

75. (*a*) 0.281;
(*b*) 1.08×10^5 W, 2.2×10^9 J, 5.3×10^5 kcal.

77. (*a*) 4.8 C°;
(*b*) 69.4 J/kg · K.

79. (*b*) 0.55.

81. 98 W.

83. $e_{\text{Sterling}} = \left(\dfrac{T_H - T_L}{T_H}\right)\left[\dfrac{\ln\dfrac{V_b}{V_a}}{\ln\dfrac{V_b}{V_a} + \dfrac{3}{2}\left(\dfrac{T_H - T_L}{T_H}\right)}\right]$,
$e_{\text{Sterling}} < e_{\text{Carnot}}$.

85. −19°C.

87. 13 J/K.

89. (*a*) See Table below;†
(*b*) 1/27;
(*c*) 1/9.

91. Photovoltaic panel: 15 min, mirror: 3.3 min.

93. (*a*) 300 kg;
(*b*) 2000 m^2.

†**TABLE for Problem 89a in Chapter 20**

Macrostate	Microstates (r = red, o = orange, g = green)			Number of Microstates
3 red, 0 orange, 0 green	r r r			1
2 red, 1 orange, 0 green	r r o	r o r	o r r	3
2 red, 0 orange, 1 green	r r g	r g r	g r r	3
1 red, 2 orange, 0 green	r o o	o r o	o o r	3
1 red, 0 orange, 2 green	r g g	g r g	g g r	3
1 red, 1 orange, 1 green	r o g o g r	r g o g r o	o r g g o r	6
0 red, 3 orange, 0 green	o o o			1
0 red, 2 orange, 1 green	g o o	o g o	o o g	3
0 red, 1 orange, 2 green	o g g	g o g	g g o	3
0 red, 0 orange, 3 green	g g g			1

Chapter 21

1. 2.7×10^{-3} N.

3. 1.21×10^{4} N.

5. 1.5×10^{-13}%.

7. 3.59 cm.

9. -4.6×10^{8} C, 0.

11. (*a*) $q_1 = q_2 = \frac{1}{2}Q_T$;
(*b*) $q_1 = 0$, $q_2 = Q_T$.

13. Top charge: 2.00×10^{2} N, *y*-direction; Lower left charge: 2.00×10^{2} N, 30° below the $-x$ axis; Lower right charge: 2.00×10^{2} N, 30° below the $+x$ axis.

15. Each force is 7.16×10^{7} N, and lies along the line from the center of the square outwards from the charge.

17. $10.1\dfrac{kQ^2}{\ell^2}$, 61° above the positive *x*-axis.

19. 1.8×10^{12} electrons.

21. (*a*) $T = 2\pi\sqrt{\dfrac{m\pi\varepsilon_0 d^3}{Qq}}$;
(*b*) 0.2 ps.

23. 1.05×10^{5} N/C, south.

25. 195 N/C $\hat{\mathbf{j}}$.

27. 1.33×10^{14} m/s^2, opposite to the field.

29. 0.25 N/C.

31. 1.16×10^{-10} C.

33. 1.17×10^{6} N/C, towards the $-33.8\ \mu$C charge.

35. 65 cm.

37. (*a*) $\dfrac{\sqrt{3}kQ}{\ell^2}$ at 240°;
(*b*) $\dfrac{kQ}{\ell^2}$ at 330°.

39. $\dfrac{Q}{4\pi\varepsilon_0}\left\{\dfrac{(x-\frac{1}{2}\ell)}{\left[(x-\frac{1}{2}\ell)^2+R^2\right]^{3/2}} + \dfrac{(x+\frac{1}{2}\ell)}{\left[(x+\frac{1}{2}\ell)^2+R^2\right]^{3/2}}\right\}\hat{\mathbf{i}}$.

41. $\vec{\mathbf{E}}_A = 3.0 \times 10^{6}$ N/C at 90°, $\vec{\mathbf{E}}_B = 7.8 \times 10^{6}$ N/C at 56°, yes.

43. $x_M = \pm a/\sqrt{2}$.

47. $\dfrac{8\lambda\ell z}{\pi\varepsilon_0(4z^2+\ell^2)(4z^2+2\ell^2)^{1/2}}\hat{\mathbf{k}}$.

49. (*a*) $E_x = \dfrac{\lambda\ell}{4\pi\varepsilon_0 x\sqrt{x^2+\ell^2}}$,
$E_y = \dfrac{\lambda}{4\pi\varepsilon_0 x\sqrt{x^2+\ell^2}}(x-\sqrt{x^2+\ell^2})$.
Note that $E_y < 0$, and so the electric field points to the right and down.

51. $E(x) = \dfrac{Q}{4\pi\varepsilon_0 x(x+\ell)}$.

53. (*a*) 5.70×10^{-3} m;
(*b*) 5.48×10^{-9} s.

55. (*a*) To the right;
(*b*) 32 N/C.

57. $-10\ \text{m}\hat{\mathbf{i}} + 9.5\ \text{m}\hat{\mathbf{j}}$.

59. (*a*) 1.1×10^{-28} C · m;
(*b*) 2.4×10^{-24} m · N;
(*c*) 1.7×10^{-24} m · N;
(*d*) 4.8×10^{-24} J.

61. (a) $\sin\theta \approx \theta$;
(*b*) $\dfrac{1}{2\pi}\sqrt{\dfrac{pE}{I}}$.

65. (a) 4×10^{-10} N;
(*b*) 7×10^{-10} N;
(*c*) 6×10^{-5} N.

67. 3.9×10^{9} C.

69. 2.4×10^{-10} N.

71. 5.2×10^{-11} m.

73. $\dfrac{Q^2}{2\varepsilon_0 A}$, 8×10^{-7} C.

75. 5×10^{-9} C.

77. 8.0×10^{-9} C.

79. 16°.

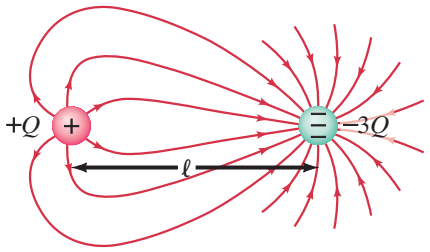
81. $E_A = 3.4 \times 10^{4}$ N/C, to the right,
$E_B = 2.3 \times 10^{4}$ N/C, to the left,
$E_C = 5.6 \times 10^{3}$ N/C, to the right,
$E_D = 3.4 \times 10^{3}$ N/C, to the left.

83. -5.6×10^{-6} C; unstable.

85. (*a*) 1.11×10^{6} N/C, downward;
(*b*) -1.96×10^{-5} C/m^2.

87. (*a*) 3.61×10^{24} electrons;
(*b*) 1.79×10^{-18}.

89.



91. $\vec{\mathbf{E}} = \dfrac{Qx}{4\pi\varepsilon_0(x^2+a^2)^{3/2}}\hat{\mathbf{i}} - \dfrac{2Qa}{4\pi^2\varepsilon_0(x^2+a^2)^{3/2}}\hat{\mathbf{j}}$.

Chapter 22

1. (*a*) 34 N · m^2/C;
(*b*) 24 N · m^2/C;
(*c*) 0.

3. (*a*) 0;
(*b*) $\Phi_{\substack{\text{thru faces}\\ \text{parallel to } E}} = 0$,
$\Phi_{\substack{\text{thru face with}\\ E \text{ entering}}} = -E_0\ell^2$,
$\Phi_{\substack{\text{thru face with}\\ E \text{ leaving}}} = E_0\ell^2$.

5. 1.64×10^{-8} C.

7. (*a*) -1.1×10^{5} N · m^2/C;
(*b*) 0.

9. -1.5×10^{-6} C.

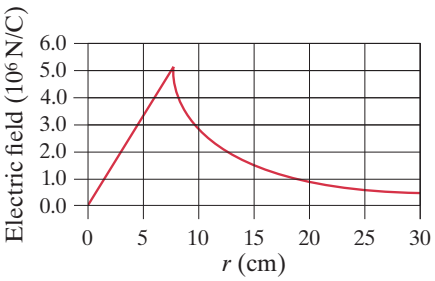
11. 2.6×10^{-5} C/m.

13. -7.97×10^{-11} C.

15. (*a*) 3.0×10^{4} N/C, towards the wire;
(*b*) 1.0×10^{5} N/C, towards the wire;

17. (*a*) $\dfrac{1}{4\pi\varepsilon_0}\dfrac{Q}{r^2}$;
(*b*) 0;
(*c*) $\dfrac{1}{4\pi\varepsilon_0}\dfrac{Q}{r^2}$;
(*d*) Makes $E = 0$ inside the boundaries of the shell;
(*e*) Polarizes the shell: negative on the inside boundary, positive on the outside boundary.

19.



21. (*a*) 6.49×10^{7} N/C, radially outward;
(*b*) 0;
(*c*) 6.49×10^{5} N/C, radially outward.

23. (*a*) $-8.00\ \mu$C;
(*b*) $+1.60\ \mu$C.

25. (*a*) 0;
(*b*) σ/ε_0;
(*c*) The fields will not change.

27. (*a*) 0;
(*b*) $\dfrac{\sigma_1 r_1^2}{\varepsilon_0 r^2}$;
(*c*) $\dfrac{\sigma_1 r_1^2 + \sigma_2 r_2^2}{\varepsilon_0 r^2}$;
(*d*) $\sigma_1 r_1^2 + \sigma_2 r_2^2 = 0$;
(*e*) $\sigma_1 = 0$.

29. (*a*) 0;
(*b*) $\dfrac{Q}{4\pi\varepsilon_0 r^2}\dfrac{(r^3 - r_1^3)}{(r_0^3 - r_1^3)}$;
(*c*) $\dfrac{Q}{4\pi\varepsilon_0 r^2}$.

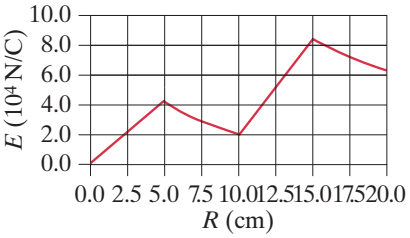
31. (*a*) $-q$;
(*b*) $Q + q$;
(*c*) $\dfrac{q}{4\pi\varepsilon_0 r^2}$;
(*d*) 0;
(*e*) $\dfrac{q+Q}{4\pi\varepsilon_0 r^2}$.

33. (*a*) $\dfrac{\sigma R_0}{\varepsilon_0 R}$, radially outward;
(*b*) 0;
(*c*) They are the same if $\lambda = 2\pi R_0\sigma$.

35. (*a*) 0;
(*b*) $\dfrac{Q}{2\pi\varepsilon_0 R\ell}$, radially outward;
(*c*) 0;
(*d*) $\dfrac{eQ}{4\pi\varepsilon_0\ell}$.

37. (*a*) 2.8×10^7 m/s;
(*b*) 5.5×10^5 m/s.

39. (*a*) $\dfrac{\rho_E R}{2\varepsilon_0}$;
(*b*) $\dfrac{\rho_E R_1^2}{2\varepsilon_0 R}$;
(*c*) $\dfrac{\rho_E(R_1^2 + R^2 - R_2^2)}{2\varepsilon_0 R}$;
(*d*) $\dfrac{\rho_E(R_1^2 + R_3^2 - R_2^2)}{2\varepsilon_0 R}$;
(*e*)



41. (*a*) 0;
(*b*) $\dfrac{1}{4\pi\varepsilon_0}\dfrac{Q}{(75R_0)^2}$.

43. (*a*) $\dfrac{\rho_E d}{2\varepsilon_0}$, perpendicularly away from the slab.

45. (*a*) 35 N, towards the other plate;
(*b*) 0.18 J.

47. (*a*) 0;
(*b*) $-\dfrac{\rho_0(d-x)}{\varepsilon_0}\hat{\mathbf{i}}$;
(*c*) $-\dfrac{\rho_0(d+x)}{\varepsilon_0}\hat{\mathbf{i}}$.

49. $\dfrac{Q}{4\pi\varepsilon_0}\dfrac{r^2}{r_0^4}$, radially outward.

51. 362 N · m^2/C for both spheres.

53. $\oint \vec{\mathbf{g}} \cdot d\vec{\mathbf{A}} = -4\pi G M_{\text{encl}}$.

55. $\varepsilon_0 a\ell^3$.

57. (*a*) 0;
(*b*) $\dfrac{1}{25\pi\varepsilon_0}\dfrac{Q}{r_0^2} \le E_{\substack{\text{sphere}\\\text{surface}}} \le \dfrac{1}{\pi\varepsilon_0}\dfrac{Q}{r_0^2}$;
(*c*) No;
(*d*) No.

59. (*a*) 3.4×10^{11} C/m^3;
(*b*) 1.1×10^{-19} C;
(*c*) 3.5×10^{11} N/C.

61. (*a*) $\dfrac{\rho_E r_0}{6\varepsilon_0}$, towards point C;
(*b*) $\dfrac{17\rho_E r_0}{54\varepsilon_0}$, towards the left.

63. (*a*) 0;
(*b*) 5.65×10^5 N/C, to the right;
(*c*) 5.65×10^5 N/C, to the right;
(*d*) -5.00×10^{-6} C/m^2;
(*e*) 5.00×10^{-6} C/m^2.

65. (*a*) On the inner surface of the shell;
(*b*) $E_{\text{inside}} = \dfrac{2.7 \times 10^4 \text{ N} \cdot \text{m}^2/\text{C}}{r^2}$,
$E_{\text{outside}} = 0$.

Chapter 23

1. 6.3×10^{-4} J.

3. -0.77 V.

5. (*a*) 1.9×10^{-7} m;
(*b*) 1.1 mV increase.

7. -138 V.

9. 2.8×10^{-2} m.

11. 1.1×10^{-6} C.

13. 2×10^{-5} m.

15. (*a*) -0.96 GV;
(*b*) 0.96 GV.

17. (*a*) They have the same potential;
(*b*) $Q\dfrac{r_2}{(r_1 + r_2)}$.

19. (*a*) 15,000 V;
(*b*) 2.7×10^{-5} C/m^2.

21. (*a*) 0;
(*b*) -44.1 V;
(*c*) -44.1 V.

23. (*a*) $V_0 - \dfrac{\sigma R_0}{\varepsilon_0}\ln\dfrac{R}{R_0}$;
(*b*) V_0;
(*c*) No.

25. 2.8×10^{-9} C.

27. (*a*) 5.8×10^5 V;
(*b*) 9.2×10^{-14} J.

29. 0.28 J.

31. 1.8×10^3 m/s.

33. 1.07×10^6 m/s.

35. $\dfrac{\sigma}{2\varepsilon_0}\left(\sqrt{x^2 + R_2^2} - \sqrt{x^2 + R_1^2}\right)$.

37. 32 m/s.

39. $\dfrac{Q}{8\pi\varepsilon_0\ell}\left[\ln\left(\dfrac{x+\ell}{x-\ell}\right)\right], x > \ell$.

41. $\dfrac{a}{6\varepsilon_0}\left[(R_0^2 - 2x^2)(x^2 + R_0^2)^{1/2} + 2x^3\right]$, $x > 0$.

43. 1 mm.

45. (*a*) 2.6×10^{-3} V;
(*b*) 1.8×10^{-3} V;
(*c*) -1.8×10^{-3} V.

49. Plate at origin:
$\sigma = -7.5 \times 10^{-11}$ C/m^2,
other plate: $\sigma = 7.5 \times 10^{-11}$ C/m^2.

51. $\vec{\mathbf{E}} = (-by + cyz)\hat{\mathbf{i}} + (-2ay - bx + cxz)\hat{\mathbf{j}} + (cxy)\hat{\mathbf{k}}$.

53. (*a*) Field on y-axis:
$\vec{\mathbf{E}} = \dfrac{1}{4\pi\varepsilon_0}\dfrac{Q}{y\sqrt{\ell^2 + y^2}}\hat{\mathbf{j}}$;
(*b*) Field on x-axis:
$\vec{\mathbf{E}} = \dfrac{1}{4\pi\varepsilon_0}\left(\dfrac{Q}{x^2 - \ell^2}\right)\hat{\mathbf{i}}$.

55. -82.5 kV.

57. 1.3 eV.

59. (*a*) 4.8 keV;
(*b*) $v_e/v_p = 42.8$.

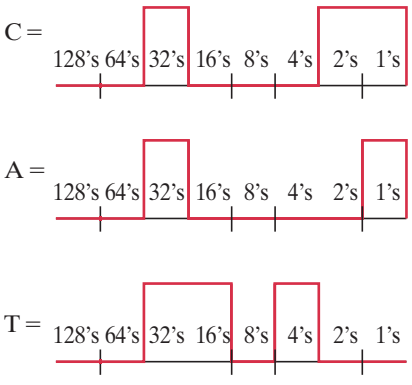
61. $\dfrac{3Q^2}{20\pi\varepsilon_0 r_0}$.

63. 111000001101.

65. 43,690.

67. (*a*) 65,536 different possible voltages;
(*b*) 16,777,216 different possible voltages;
(*c*) 16,777,216 different colors.

69. (*a*) C = 00100011, T = 00110100;
(*b*) B = 34, H = 40;
(*c*)



71. (*a*) 2.29 V;
(*b*) 10110001.

73. 000; 000; 000; 000; 000; 010; 011; 001; 100; 100; 010; 010; 010; 000; 000; 000.

75. (*a*) 2850 Mb/s;
(*b*) ≈ 150 times present day-rate, compression;
(*c*) 0.7%;
(*d*) 0.40 frames/s.

77. 1.7×10^6 V.

79. (*a*) $\frac{16}{\sqrt{3}} \frac{1}{4\pi\varepsilon_0} \frac{Q}{\ell}$;

(*b*) $\left(3 + \frac{3}{\sqrt{2}} + \frac{1}{\sqrt{3}}\right) \frac{1}{4\pi\varepsilon_0} \frac{Q}{\ell}$;

(*c*) $4\left(3 + \frac{3}{\sqrt{2}} + \frac{1}{\sqrt{3}}\right) \frac{1}{4\pi\varepsilon_0} \frac{Q^2}{\ell}$.

81. $K_\alpha = 2K_p$.

83. (*a*) 2.8×10^3 m/s;
(*b*) 2.1×10^3 m/s.

85. 1.03×10^6 m/s.

87. $V_{\text{left}} = \frac{Q}{\pi\varepsilon_0 \ell}\left(\frac{\sqrt{3}}{6} - 2\right)$,

$V_{\text{top}} = -\frac{\sqrt{3}Q}{2\pi\varepsilon_0 \ell}$,

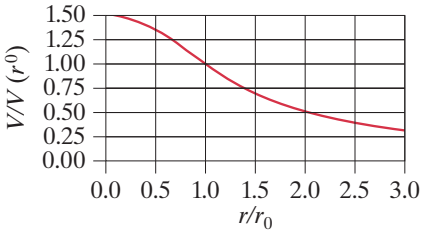
$V_{\text{right}} = -\frac{Q}{\pi\varepsilon_0 \ell}\left(1 + \frac{\sqrt{3}}{6}\right)$.

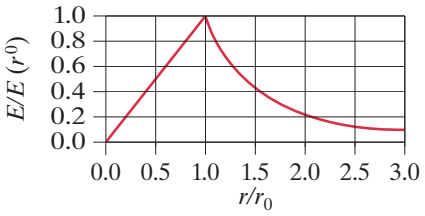
89. (*a*) 1.3×10^6 V;
(*b*) 2.0 kg.

91. (*a*) $\frac{Q}{4\pi\varepsilon_0 r}$;

(*b*) $\frac{Q}{8\pi\varepsilon_0 r_0}\left(3 - \frac{r^2}{r_0^2}\right)$;

(*c*)





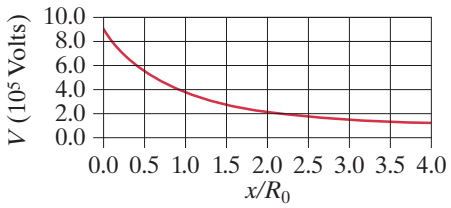
93. Binary values: 1110 ; 101101 ; 1100111.

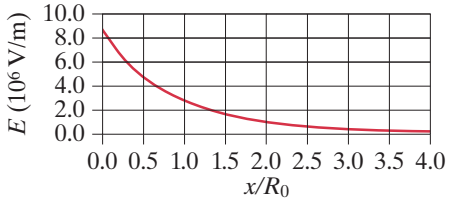
95. (*a*) 19 kV;
(*b*) $\frac{4Bx\hat{\mathbf{i}}}{(x^2 + R^2)^3}$;
(*c*) 1.9×10^5 V/m$\hat{\mathbf{i}}$.

97. 3000 V, little actual charge transfer.

99. (*a*) 27 V;
(*b*) 2.2×10^{-18} J; 14 eV;
(*c*) -2.2×10^{-18} J ; −14 eV;
(*d*) 2.2×10^{-18}; 14 eV.

101.





103. (*a*) $\frac{\rho_E}{3\varepsilon_0}\frac{(r_2^3 - r_1^3)}{r}$;

(*b*) $\frac{\rho_E}{\varepsilon_0}\left(\frac{1}{2}r_2^2 - \frac{1}{6}r^2 - \frac{1}{3}\frac{r_1^3}{r}\right)$;

(*c*) $\frac{\rho_E}{2\varepsilon_0}(r_2^2 - r_1^2)$, yes.

105. (*a*) 6×10^5 V;
(*b*) 1×10^{-5} C.

Chapter 24

1. 3.6×10^{-6} F.

3. 9.00×10^{-5} C.

5. $Q_1 = Q_0 \frac{C_1}{C_1 + C_2}$,

$Q_2 = Q_0\left(\frac{C_2}{C_1 + C_2}\right)$,

$V_1 = V_2 = \frac{Q_0}{C_1 + C_2}$.

7. 1.2 C.

9. 150 m^2.

11. 7.10×10^{-4} F.

13. 9×10^{-16} m, no.

15. 2.86×10^{-8} C.

17. 5.3×10^4 V/m.

19. (*a*) 0.22 μm $\leq x \leq$ 220 μm;

(*b*) $\Delta x = \frac{x^2 \Delta C}{\varepsilon_0 A}$;

(*c*) $x_{\min}$: 0.01%, $x_{\max}$: 10%.

21. 3.0×10^3 pF, yes.

25. (*a*) parallel;

(*b*) $\varepsilon_0 A\left(\frac{1}{d_1} + \frac{1}{d_2}\right)$;

(*c*) $C_{\min} = \frac{4\varepsilon_0 A}{d_1 + d_2}$, $C_{\max} = \infty$.

27. $C_{\max} = 1.89 \times 10^{-8}$ F, parallel;
$C_{\min} = 1.68 \times 10^{-9}$ F, series.

29. $Q_1 = Q_2 = 14.8\ \mu$C, $Q_3 = 29.6\ \mu$C,
$Q_4 = 44.4\ \mu$C;
$V_1 = V_2 = 0.925$ V, $V_3 = 1.850$ V,
$V_4 = 1.56$ V, $V_{ab} = 3.41$ V.

31. $Q_1 = \frac{C_1 C_2}{C_1 + C_2} V_0$, $Q_2 = \frac{C_2^2}{C_1 + C_2} V_0$.

33. 1000 pF, in parallel.

35. 2.7 μF.

37. $\frac{\varepsilon_0 A}{d}\left(1 - \frac{\theta\sqrt{A}}{2d}\right)$.

39. 1.0×10^{-7} J/m^3.

41. 2.3×10^3 J.

43. (*a*) 1.9×10^{-4} J;
(*b*) 4.0×10^{-5} J;
(*c*) $Q_{\text{parallel}} = 1.6 \times 10^{-5}$ C,
$Q_{\text{series}} = 3.3 \times 10^{-6}$ C.

45. (*a*) 4.74×10^{-4} J;
(*b*) 2.01×10^{-4} J;
(*c*) -2.72×10^{-4} J.

47. 3.7.

51. 5.6×10^{-6} J/m.

53. 2.1×10^{-10} F.

55. 1.1×10^{-4} J.

57. 3.8×10^{-7} C.

59. $\frac{2\varepsilon_0 A}{d}\frac{K_1 K_2}{K_1 + K_2}$.

61. (*a*) $Q_{1\ (\text{no dielectric})} = 0.43\ Q_0$,
$Q_{2\ (\text{dielectric})} = 1.57\ Q_0$;
(*b*) $V_1 = V_2 = 0.43 V_0$.

63. 4.76×10^{-9} F, 0.280 m^2.

65. (*a*) $F = \frac{h}{\ell} = \left[\frac{1}{(K_{\text{liq}} - K_V)} \times \left[\frac{C \ln(R_a/R_b)}{2\pi\varepsilon_0 \ell} - K_V\right]\right]$;
(*b*) Full: 1500 pF, empty: 1100 pF.

69. $\frac{\varepsilon_0 A}{d - \ell + \frac{\ell}{K}}$.

71. (*a*) 6×10^{-12} F;
(*b*) 5×10^{-11} C;
(*c*) 150 V/m;
(*d*) 2×10^{-10} J;
(*e*) Capacitance, charge, and work.

73. 610 V.

75. (*a*) Multiplied by 4;
(*b*) multiplied by 4;
(*c*) divided by 2.

77. 5.5×10^{-6} C.

79. 83 days.

81. 1.2.

83. (*a*) 25 J;
(*b*) 940 kW.

85. (*a*) 4.2×10^{-11} C;
(*b*) 4.2×10^{-11} C;
(*c*) 20 V;
(*d*) 2×10^{-10} J.

87. (*a*) Parallel;
(*b*) 7.7 pF to 35 pF.

89. (*a*) $\dfrac{Q^2x}{2A\varepsilon_0}$.

91. (*a*) 2.7×10^{-7} C, 15,000 V/m, 5.9×10^{-9} F, 6.0×10^{-6} J;
(*b*) 8.5×10^{-7} C, 15,000 V/m, 1.9×10^{-8} F, 1.9×10^{-5} J.

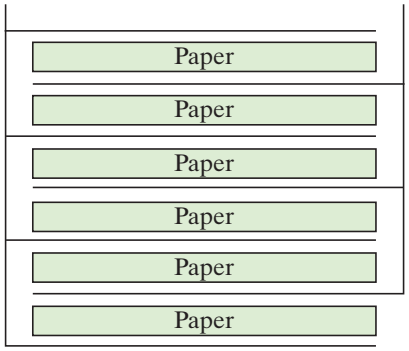
93. (*c*) 3.7;
(*d*) 0.23.

95. (*b*) 2×10^{-16} F;
(*c*) 0.8 μm.

97. (*a*) 1.1 μm^2;
(*b*) 17 Mbytes.

99. (*a*) 1.0×10^2 hp;
(*b*) 1.1×10^9 J;
(*c*) 3.0×10^3 F;
(*d*) 2×10^6 g;
(*e*) No.

101. (*a*) 8.8 nF;
(*b*) 14 μC;
(*c*)



(*d*) 12.62 mm;
(*e*) 1700 V.

Chapter 25

1. 1400 V.

3. 1.09×10^{19} electrons/s.

5. 6.9×10^{-11} A.

7. (*a*) 18 A;
(*b*) 1.7×10^4 C.

9. (*a*) 8.9 Ω;
(*b*) 6.5×10^3 C.

11. (*a*) 4.5 A;
(*b*) 4.8 A.

13. 4.8×10^{-2} Ω.

15. $d_{\text{Tungsten}} = 3.3$ mm.

17. 24 C°.

19. 150 Ω.

21. (*a*) 3.8×10^{-4} Ω;
(*b*) 1.5×10^{-3} Ω;
(*c*) 6.0×10^{-3} Ω.

23. 59.7° C.

25. $\dfrac{1}{4\pi\sigma}\left(\dfrac{1}{r_1} - \dfrac{1}{r_2}\right)$.

27. (*b*) 5.8×10^{-4} Ω;
(*c*) 0.51 Ω.

29. 19 Ω.

31. 130 A.

33. (*a*) 1250 W;
(*b*) 11.5 Ω;
(*c*) 7.68 Ω.

35. 1/4 as bright.

37. $26.

39. (*a*) 22 Ω, 0.41 W;
(*b*) 4.

41. 11 kW.

43. (*a*) 33%;
(*b*) 75 W.

45. 0.11 mm.

47. 0.51 A, 0.36 A.

49. 410 V.

51. (*a*) $V_0/\sqrt{2}$;
(*b*) $V_0/\sqrt{2}$.

53. (*a*) 7.1×10^{-10} m/s;
(*b*) 9.6 A/m^2;
(*c*) 1.6×10^{-7} V/m.

55. 2.5 A/m^2, north.

57. 35 m/s.

59. 7.4 h.

61. 0.13 S.

63. (*a*) $40.7 per month;
(*b*) 1400 kg of coal.

65. 0.22 kg/s.

67. Left end: 5.1×10^5 A/m^2, right end: 2.0×10^5 A/m^2.

69. 3.7×10^5 kg.

71. 4.2×10^{-3} m.

73. 1.56 Ω.

75. (*b*) As large as possible.

77. (*a*) 7.4 hp;
(*b*) 220 km.

79. (*a*) 33 Hz;
(*b*) 0.990 A;
(*c*) $V = (30.1 \sin 210\,t)$ V.

81. $140 million.

83. (*a*) 4.0 kW;
(*b*) 18 W;
(*c*) 12 W;
(*d*) $0.30/month.

85. Divided by 4.

87. (*a*) $55/kWh;
(*b*) $280/kWh;
(*c*) 5.0×10^2 times greater, 2.5×10^3 times greater.

89. 1.60×10^{-4} Ω.

91. 1400 W.

93. (*a*) 0.634;
(*b*) 1.01 Ω;
(*c*) 1.59 Ω;
(*d*) 330 W in Cu, 520 W in Al;
(*e*) 2.05 mm;
(*f*) 0.481 Ω, yes.

95. 1.5×10^{-6} m.

Chapter 26

1. (*a*) 5.91 V;
(*b*) 5.99 V.

3. 0.060 Ω.

5. 8.6 V.

7. 2.

9. 140 Ω.

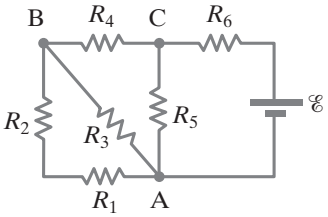
11. 0.4 Ω.

13. 560 Ω, 2.0%.

15. 21 Ω.

17. (*a*) 588 Ω;
(*b*) 8.51×10^{-2} A, 6.24×10^{-2} A, 8.51×10^{-2} A.

19.



$I_1 = I_2 = 0.217$ mA, $I_3 = 0.434$ mA, $I_4 = 0.652$ mA, $I_5 = 1.09$ mA, $I_6 = 1.74$ mA, $V_{AB} = 1.85$ V.

21. Connect 18 of the resistors in series.

23. 77 V.

25. (*a*) V_1 and V_2 increase, V_3 and V_4 decrease;
(*b*) I_1 and I_2 increase, I_3 and I_4 decrease;
(*c*) It increases;
(*d*) Before: $I_1 = 0.109$ A, $I_2 = 0$, $I_3 = I_4 = 0.054$ A;
After $I_1 = 0.122$ A, $I_2 = I_3 = I_4 = 0.041$ A.

27. 0.33 A.

29. 0.

31. (*a*) −27 V;
(*b*) 85 V battery: 83 V,
45 V battery: 43 V.

33. $I_1 = 0.60$ A, left; $I_2 = 0.27$ A, left.

37. 0.48 A.

39. 0.17 A.

41. (*a*) $\dfrac{R(5R' + 3R)}{8(R + R')}$;
(*b*) $\dfrac{1}{2}R$.

43. 2.40 V.

45. (*a*) 1.60×10^{-9} F;
(*b*) 1.29×10^{-5} s.

47. 5.9×10^{-3} s.

49. 0.16 s.

51. (*a*) 6.8 V, 15 μC;
(*b*) 5.3×10^{-5} s.

53. (*a*) 8.0 V;
(*b*) 14 V;
(*c*) 8.0 V;
(*d*) 4.8 μC.

55. 2.9×10^{-5} A.

57. (*a*) Put 4.4×10^{-4} Ω in parallel;
(*b*) Put 45 kΩ in series.

59. 76 kΩ.

61. 18 V, 11 V, 16% too low in each case.

63. 9.38×10^{-4} A, 5.0 V, current: 17% higher, voltage: 17% lower.

65. 11 V.

67. 140 kΩ.

69. 1.8 V, 118 V.

71. (*a*) R_2R_3/R_1;
(*b*) 112 Ω.

75. 600 cells, 0.54 m^2, 4 banks of 150 cells in series per bank, with the banks in parallel, always have the panel pointed directly at the sun.

77. 45 V or −21 V.

79. 100 Ω.

81. (*a*) 0.141 Ω;
(*b*) 0.53 A;
(*c*) Cu: 29 mV, Al: 46 mV.

83. 8 mA.

85. 9.0 Ω.

87. 1.2×10^5 Ω.

93. (*a*) 5.8×10^{-5} A, upward;
(*b*) −17 V.

97. $I_{5\Omega} = 1.57$ A, $I_{6\Omega} = 1.52$ A, $I_{12\Omega} = 0.762$ A, $I_{6.8\Omega} = 0.713$ A.

99. 2.00×10^2 MΩ.

101.

Series	Parallel
$R_1 + R_2$	$\dfrac{R_1R_2}{R_1 + R_2}$
$\dfrac{V_B - V_A}{R_1 + R_2}$	$(V_B - V_A)\dfrac{R_1 + R_2}{R_1R_2}$
$V_B - V_A$	$V_B - V_A$
$V_B - V_A$	$V_B - V_A$
$V_1 = (V_B - V_A)\dfrac{R_1}{R_1 + R_2}$ $V_2 = (V_B - V_A)\dfrac{R_2}{R_1 + R_2}$	$V_1 = V_B - V_A$ $V_2 = V_B - V_A$
$\dfrac{V_AR_1 + V_BR_2}{R_1 + R_2}$ (R_1 closest to point B)	Not applicable
$I_1 = \dfrac{V_B - V_A}{R_1 + R_2}$ $I_2 = \dfrac{V_B - V_A}{R_1 + R_2}$	$I_1 = \dfrac{V_B - V_A}{R_1}$ $I_2 = \dfrac{V_B - V_A}{R_2}$

Chapter 27

1. 0.68 km.

3. (*a*) 6.7 N/m;
(*b*) 3.8 N/m.

5. 1.7 N.

7. 33°.

9. 1400 A, no.

11. 1.91 N, 41.8° below the negative *y*-axis.

13. $(0.34\,\hat{\mathbf{i}} + 0.17\,\hat{\mathbf{k}})$T.

17. 8.1×10^{-14} N, north.

19. 3×10^7 m/s, no.

21. (*a*) no force;
(*b*) downward;
(*c*) upward;
(*d*) into the page;
(*e*) to the left;
(*f*) to the left.

23. 0.61 m.

25. $[(0.82\hat{\mathbf{i}} - 1.1\hat{\mathbf{j}} + 0.1\hat{\mathbf{k}})] \times 10^{-15}$ N.

27. $L_{\text{final}} = \dfrac{1}{2}L_{\text{initial}}$.

29. (*a*) 7.1×10^5 m/s;
(*b*) 4.3×10^{-2} m;
(*c*) 3.8×10^{-7} s.

31. $r_d = r_\alpha = \sqrt{2}\,r_p$.

33. (*a*) negative;
(*b*) $\dfrac{qB_0(d^2 + \ell^2)}{2d}$.

35. 1.37×10^{-6} m.

37. (*a*) $\sqrt{3}\dfrac{mv_0}{qB_0}$;
(*b*) $\dfrac{4\pi m}{3qB_0}$.

39. 0.359 m · N.

41. (*a*) 7.01×10^{-5} m · N;
(*b*) North.

43. (*a*) $-6.9\,\hat{\mathbf{k}}$ A · m^2;
(*b*) $(4.2\hat{\mathbf{i}} - 3.8\hat{\mathbf{j}})$m · N;
(*c*) −4.5 J.

45. Drops by 18%.

47. 21 μA.

49. (*a*) 5.85×10^6 m/s;
(*b*) 1.28×10^{-2} m.

51. (*a*) 0.53 T;
(*b*) 0.61 T.

53. (*a*) 2.2×10^{-4} V/m;
(*b*) 2.7×10^{-4} m/s;
(*c*) 3.2×10^{28} electrons/m^3.

55. (*a*) It determines the polarity of the emf;
(*b*) 0.56 m/s.

59. ${}^2_1$H(deuteron),
${}^4_2$He(alpha particle).

61. 2.5 m.

63. 3.8×10^{-4} m · N.

65. 0.17 T at 27° westward from the vertical.

67. $\dfrac{6.0\,mg}{\ell B}$, toward the left.

69. (*c*) 48 MeV.

71. 1×10^9 A.

73. 33 A, down.

75. −0.13 eV.

77. 0.25 N, to the north and 68° above the horizontal.

79. $\dfrac{a^2IB_0}{b}\,\hat{\mathbf{i}}$.

81. $\pi\sqrt{\dfrac{mb(3a + b)}{3(a + b)NIaB}}$.

83. 7.0×10^{-6} N.

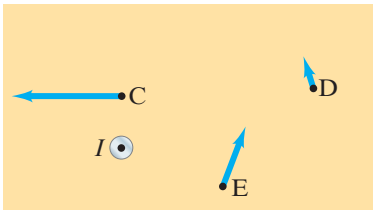
85. (*a*) Downward from the pivot;
(*b*) 0.028 T;
(*c*) 0.12 T.

Chapter 28

1. 2.6×10^{-4} T, $B_{\text{cable}}/B_{\text{Earth}} = 5.2$.

3. 0.21 N, attractive.

5.

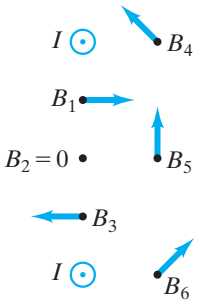


7. 8.5×10^{-5} T @ 82° or $(1.1 \times 10^{-5}\ \text{T})\hat{\mathbf{i}} + (8.4 \times 10^{-5}\ \text{T})\hat{\mathbf{j}}$.

9. 4.0×10^{-5} T, 9.7° below the horizontal.

11. 4.4×10^{-17} T.

13.



15. Middle wire: 0.063 N/m, attracting; left wire: 0.031 N/m, repelling.

17. 33 A, downward.

19. $\dfrac{\mu_0 I}{2\pi}\left(\dfrac{d-2x}{x(d-x)}\right)\hat{\mathbf{j}}$.

21. 4.33×10^{-5} T.

23. 5.91×10^{-5} T @ 5.83° below the negative x axis.

25. $\dfrac{\mu_0 I\, I'}{\pi}\left[\dfrac{a}{2d} - \dfrac{\sqrt{3}}{3}\ln\left(1 + \dfrac{\sqrt{3}a}{2d}\right)\right]$.

27. 0.182 A.

29. (*a*) 4.5 mT;
(*b*) 2.7 mT;
(*c*) 1.5 mT.

31. (*a*) 0.554 m;
(*b*) 9.42 mT.

33. 0.14 N to the south.

35. (*a*) $\dfrac{\mu_0 I_0 R^2}{2\pi R_1^3}$;
(*b*) $\dfrac{\mu_0 I_0}{2\pi R}$;
(*c*) $\dfrac{\mu_0 I_0}{2\pi R}\dfrac{(R_3^3 - R^3)}{(R_3^3 - R_2^3)}$;
(*d*) 0.

37. $\dfrac{\mu_0 I\theta}{4\pi}\left(\dfrac{R_2 - R_1}{R_1 R_2}\right)\hat{\mathbf{k}}$
(out of the plane of the page).

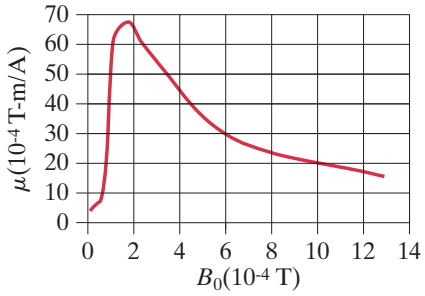
39. 9.2×10^{-8} m · N, in a direction to cause the inner loop to rotate into the same plane as the outer loop with the currents parallel. We assumed that the magnetic field from the larger loop is constant across the surface of the smaller loop.

45. $\dfrac{\mu_0 I}{4\pi d}\left(\sqrt{2} - \dfrac{\sqrt{5}}{2}\right)\hat{\mathbf{k}}$
(out of the plane of the page).

49. (*a*) $\dfrac{2\sqrt{2}d^2\mu_0 I\,\hat{\mathbf{i}}}{\pi(d^2 + 4x^2)(d^2 + 2x^2)^{1/2}}$;
(*b*) Yes, $\vec{\boldsymbol{\mu}} = d^2 I\hat{\mathbf{i}}$.

51. 0.5 T.

53.



55. M : 8.9×10^{-4} N/m at 90°,
N : 5.2×10^{-4} N/m at 300°,
P : 5.2×10^{-4} N/m at 240°.

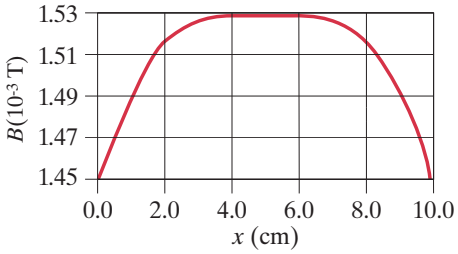
57. 2.0×10^2 A.

59. (*a*) 1.2×10^{-5} N toward the wire;
(*b*) 0.

61. (*a*) 410 A, to the right;
(*b*) No;
(*c*) 410 A, to the left, yes.

63. (*a*) $1.1 \times 10^{-6}\ \text{T} \cdot \text{m}^3 \times$
$\left[\dfrac{1}{[(0.100\ \text{m})^2 + x^2]^{3/2}} + \dfrac{1}{[(0.100\ \text{m})^2 + (x - 0.100\ \text{m})^2]^{3/2}}\right]$;

(*b*)



(*c*) -1.1×10^{-2} %.

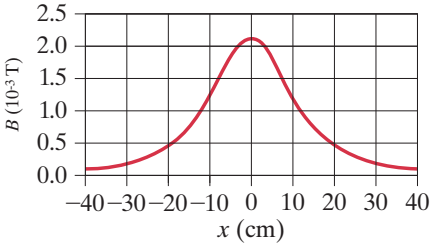
65. 1.9×10^{-6} *g*'s.

67. $\dfrac{\mu_0 I\sqrt{5}}{2\pi a}$, into the page.

71. 9.3×10^{-5} T, into the page.

73. 0.03 N.

75. $B = \dfrac{7.0686 \times 10^{-6}\ \text{T} \cdot \text{m}^3}{[(0.15\ \text{m})^2 + x^2]^{3/2}}$.



77. 7.1 A.

79. (*a*) $\dfrac{\mu_0 I}{2\pi R}(1 + \pi)\hat{\mathbf{k}}$;
(*b*) $\dfrac{\mu_0 I}{2\pi R}\sqrt{1 + \pi^2}$, 177° from the $+x$ axis, towards the $+z$ axis.

Chapter 29

3. 0.16 V.

5. (*a*) 1.3×10^{-2} Wb;
(*b*) 55°;
(*c*) 7.5×10^{-3} Wb.

7. Counterclockwise.

9. (*a*) Clockwise;
(*b*) 0.034 V;
(*c*) 0.014 A.

11. (*a*) Zero;
(*b*) clockwise;
(*c*) counterclockwise;
(*d*) clockwise.

13. (*a*) $(-7.2 + 1.1t^2)$V;
(*b*) −6.1 V, 11 V.

15. 0.012 V, 0.012 V.

17. 5.94 C.

19. (*a*) Zero;
(*b*) 0.99A, counterclockwise.

21. (*a*) $\dfrac{\mu_0 Ia}{2\pi}\ln\left(1 + \dfrac{a}{b}\right)$;
(*b*) $\dfrac{\mu_0 Ia^2 v}{2\pi b(b + a)}$;
(*c*) clockwise;
(*d*) $\dfrac{\mu_0^2 I^2 a^4 v}{4\pi^2 R b^2 (b + a)^2}$.

23. 0.85 m/s.

25. (*a*) 0.17 V;
(*b*) 6.3 mA;
(*c*) 6.7×10^{-4} N.

27. 1.3 T.

29. (*b*) $v_0 e^{-\frac{B^2\ell^2}{mR}t}$.

31. (*a*) $\dfrac{\mu_0 Iv}{2\pi}\ln\left(\dfrac{b + a}{b}\right)$, point toward the wire;
(*b*) same magnitude, away from the wire.

33. 26.2 V.

35. (*a*) 530 V;
(*b*) double it.

37. 57 V.

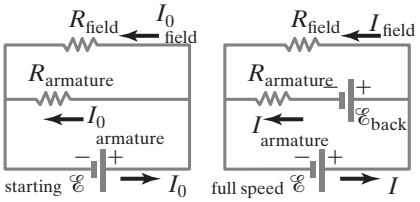
39. 130 V.

41. 13 A.

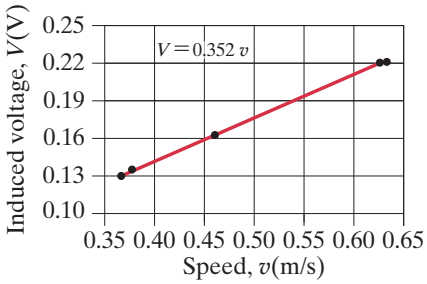
43. (*a*) 2800 V;
(*b*) 160 V.

45. 410 V, 51 A.

47. 50, 4.8 V.
49. (*a*) 5.0×10^4 V(rms);
(*b*) 0.10 (10%).
51. 7×10^9 m.
53. 4.8-cm diameter wire.
55. 5.0×10^{-9} N, in the $-\hat{\mathbf{j}}$ direction.
59. 8.61×10^{-3} J.
61. 0.62 nC.
63. (*a*) 41 kV;
(*b*) 2.9×10^7 W;
(*c*) 7.4×10^5 W;
(*d*) 2.8×10^7 W.
65. (*a*) Step-down;
(*b*) 2.9 A;
(*c*) 0.29 A;
(*d*) 4.1 Ω.
67. 4.9×10^{-2} A, left to right.
69. 1.6×10^{17} electrons.
71. (*a*) 40 A;
(*b*) 114 V;
(*c*) 12 W;
(*d*) 4800 W.
73. (*a*) Clockwise;
(*b*) no induced current;
(*c*) counterclockwise;
(*d*) no effect.
75. $0.20\,\text{J}(1 - e^{-20t})$.
77. (*a*)



(*b*) 41.9A;
(*c*) 6.9 A.
79. $\frac{1}{2} B\omega R^2$.
81. $\frac{V_0}{\pi r^2}$, decreasing.
83. $2R^2\mu_0 n I_0 \omega$.
85. (*a*)



(*b*) −0.065%;
(*c*)

Speed (m/s)	Induced Voltage (V)	Theoretical Induced Voltage (V)	% diff.
0.367	0.128	0.130	−1.32%
0.379	0.135	0.134	0.78%
0.465	0.164	0.164	−0.21%
0.623	0.221	0.220	0.37%
0.630	0.222	0.223	−0.30%

Chapter 30

1. $\mu_0 n_1 n_2 \pi r_2^2$.
3. (*a*) 3.65×10^{-2} H;
(*b*) 4.47 V.
5. 7.8×10^{-2} H.
7. 0.498 H.
9. 11.1 V.
11. 2.2.
15. (*a*) $u_E = 4.4 \times 10^{-4}$ J/m^3,
$u_B = 1.6 \times 10^6$ J/m^3, 2.8×10^{-10};
(*b*) 6.0×10^8 V/m.
17. $u_B = 2.3$ J/m^3,
$u_E = 8.1 \times 10^{-15}$ J/m^3.
19. $\frac{\mu_0 I^2}{4\pi} \ln\left(\frac{r_2}{r_1}\right)$.
21. (*a*) 2.1 ms;
(*b*) 15 Ω.
23. (*a*) $\frac{LV_0^2}{2R^2}(1 - e^{-t/\tau})^2$;
(*b*) 7.6τ.
25. (*b*) 6500 V.
27. (*a*) 2.25;
(*b*) 1.5.
29. (*a*) 17.5 kHz;
(*b*) 7.03 mA;
(*c*) 4.33×10^{-6} J.
31. 1.7 Ω.
33. 2.1 kΩ, decreased.
35. (*a*) 320 Ω;
(*b*) 1.9×10^{-2} Ω.
37. 5.9×10^{-2} H.
39. 7.7×10^{-2} H.
41. (*a*) 0.070;
(*b*) 1.00.
43. (*a*) 1.00×10^4 Ω;
(*b*) 1.54×10^4 Ω.
45. 2000 Ω.
47. 1100 Hz.
49. 8.72 kΩ, −3.52°, 71.7 mA.
51. (*a*) 0.40 A;
(*b*) −89°;
(*c*) 0.32 W.
53. 0.88 H.
55. (*a*) 2.8 A;
(*b*) −8.0°;
(*c*) 340 V, −8.0°;
(*d*) 290 V, 98.0°;
(*e*) 340 V, −82.0°.
57. 5.4×10^5 Hz.
59. (*a*) 4.23×10^{-9} F;
(*b*) 35.8 mA.
61. (*a*) 21 kHz;
(*b*) 420 V;
(*c*) 210.
63. 140 loops.
65. 75:1.
67. 7.50×10^{-5} J, 1.2 s.
69. (*c*) 83 μH.
71. 2.5×10^{15} J.
73. 46 m, 21 km, 0.70 kΩ.
75. (*a*) 41 mV;
(*b*) 130 mV.
79. Inductor, 47 mH, resistor, 12 Ω.
81. 64 mH, 2.0×10^1 Ω.
85. Put a 35-mH inductor in series with the device.
87. (*a*) $\frac{V_{20}}{\omega L - 1/\omega C} \sin(\omega t - \pi/2)$;
(*c*) $\frac{1}{\omega^2 LC}$;
(*d*) $V_{1,\text{out}} = V_1$.
89. (*a*) $\frac{V_0}{R} \sin \omega t$;
(*b*) $\frac{V_0}{X_L} \sin\left(\omega t - \frac{\pi}{2}\right)$;
(*c*) $\frac{V_0}{X_C} \sin\left(\omega t + \frac{\pi}{2}\right)$;
(*d*) $I(t) = \left[\frac{V_0}{R}\sqrt{1 + \left(R\omega C - \frac{R}{\omega L}\right)^2} \times \sin(\omega t + \phi)\right]$,
$\tan^{-1}\left(R\omega C - \frac{R}{\omega L}\right)$.
(*e*) $Z = \frac{R}{\sqrt{1 + \left(R\omega C - \frac{R}{\omega L}\right)^2}}$;
(*f*) $\frac{1}{\sqrt{1 + \left(R\omega C - \frac{R}{\omega L}\right)^2}}$.

93. (*a*) 2.2 pF;
(*b*) 6.0 MΩ;
(*c*) 0.42 mA, not dangerous;
(*d*) 35 mA, dangerous.

95. $I_0 = \dfrac{V_0}{\sqrt{R^2 + \omega^2 L^2}}$, $\phi = \tan^{-1}\dfrac{\omega L}{R}$.

97. (*b*) As $f \to 0$, $A \to 0$; as $f \to \infty$, $A \to 1$.

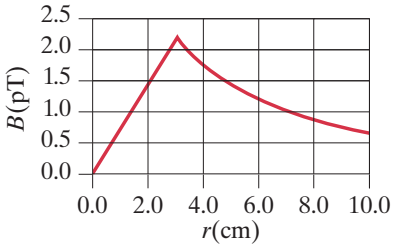
99. (*a*) $f_0 = 7.1$ kHz, $(V_R)_{\text{rms}} = \boxed{V_{\text{rms}}}$;
(*b*) $(V_R)_{\text{rms}} = 0.86\, V_{\text{rms}}$.

Chapter 31

1. $1.1 \times 10^5 \dfrac{\text{V/m}}{\text{s}}$.

3. $1.1 \times 10^{15} \dfrac{\text{V}}{\text{m} \cdot \text{s}}$.

7. (*b*) $r \le r_0 : B_0(r) = (7.5 \times 10^{-11}\ \text{T/m})r$,
$r > r_0 : B_0(r) = (6.8 \times 10^{-14}\ \text{T} \cdot \text{m})\dfrac{1}{r}$;
(*c*)



9. 4.95 V/m.

11. 8.33 min.

13. (*a*) $-\hat{\mathbf{k}}$;
(*b*) $B_0 = E_0 k/\omega$, $-\hat{\mathbf{j}}$.

15. (*a*) 8.386×10^{-3} m;
(*b*) 2.5×10^{18} Hz.

17. (*a*) 3.00×10^5 m;
(*b*) 0.341 m;
(*c*) no.

19. 4.0×10^{16} m.

21. 9.0×10^3 wavelengths, 3.542×10^{-15} s.

23. 3400 rad/s, or 32,000 rev/min.

25. 7.33×10^{-7} W/m^2.

27. 6.36×10^{-7} J/h.

29. (*a*) 54 V/m;
(*b*) 3.9 W/m^2.

31. 1240 V/m, 4.14×10^{-6} T.

33. (*a*) 150 J;
(*b*) 2.4×10^9 V/m.

35. (*a*) 2×10^8 ly;
(*b*) 2000 time larger.

37. (*a*) 4000 Pa, 3.3×10^{-9} N;
(*b*) 8×10^6 m/s^2.

39. 400 m^2.

41. 0.16 m.

43. 5.56 m, 0.434 m.

45. 9.4×10^{-13} F.

47. (*a*) 1.2×10^{-6} H;
(*b*) 1.1×10^{-10} F.

49. 7.14×10^{-4} V/m,
1.35×10^{-9} W/m^2.

51. 10 ns.

53. (*a*) 1.35 s;
(*b*) inside.

55. 8.33 min.

57. 6.9 V/m, 2.3×10^{-8} T.

59. (*a*) 9.2×10^{-11} J;
(*b*) 8.7×10^{-6} V/m, 2.9×10^{-14} T.

61. 61 km.

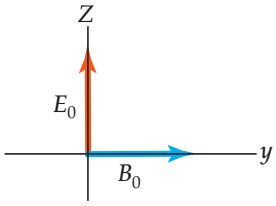
65. (*a*) 2.0×10^{-8} N;
(*b*) 9.2×10^{-7} N;
(*c*) gravity, by a factor of 50.

67. 1.1×10^{-4}%.

69. 2.25×10^8 m/s, 75.2%.

71. (*b*) 1.5×10^{-16} A;
(*c*) 3.0×10^{-20} T,
$\dfrac{B_{\text{D}}}{B_{\substack{\text{steady}\\\text{current}}}} = 1.5 \times 10^{-16}$.

73. (*d*)



both $\vec{\mathbf{E}}$ and $\vec{\mathbf{B}}$ rotate counterclockwise.

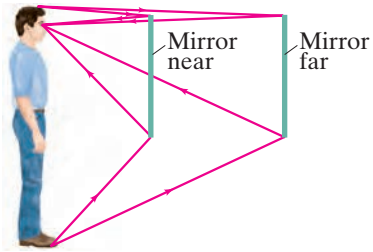
75. (*a*) 2.0×10^{-3} J/s;
(*b*) 0.87 V/m;
(*c*) 0.87 V;
(*d*) 1.7×10^{-2} V.

77. 19 kW.

79. (*a*) $+x$;
(*b*) $\beta = \alpha c$;
(*c*) $B_z = \dfrac{E_0}{c} e^{-(\alpha x - \beta t)^2}$.

Chapter 32

1.



3. 13°.

7. 4.0×10^{-6} m^2.

11. 15.0 cm.

13. (*a*) 12 cm;
(*b*) 17 cm;
(*c*) inverted.

15. 6.3 m.

17. (*a*) Convex;
(*b*) 20 cm behind the mirror;
(*c*) −83 cm;
(*d*) −170 cm.

19. (*b*) −6.0 cm;
(*c*) 1.3 mm.

25. (*a*) Center of curvature;
(*b*) real;
(*c*) inverted;
(*d*) −1.

29. 0.80 m.

35. (*a*) 2.21×10^8 m/s;
(*b*) 2.00×10^8 m/s;
(*c*) 1.97×10^8 m/s.

37. 46.1°.

39. 35.0°.

41. 71.8°.

43. (*a*) 27.1°;
(*b*) 31.2°;
(*c*) 31.2°.

45. 26°.

47. 60.5°.

51. 0.20°.

53. (*b*) 0;
(*c*) 0.5°.

55. 1.26.

57. 6.19×10^{-4} m.

59. (*a*) 2.3×10^{-6} s;
(*b*) 1.7×10^{-8} s.

61. $n \ge 1.64$.

63. 17.5 cm.

67. 1.58, light flint glass.

69. 51.1°.

71. 4.83 m.

73. 36 cm for real image, 18 cm for virtual image.

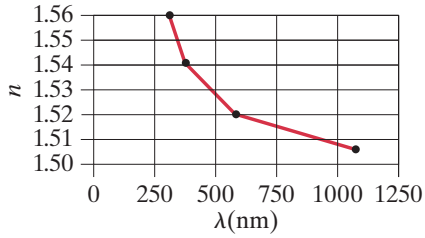
77. 430 nm.

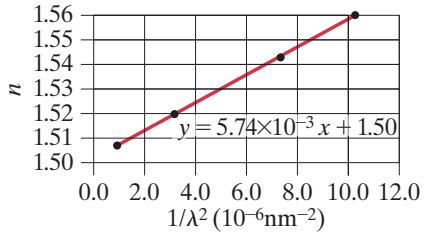
83. 38.9°.

85. 78.4°.

87. 11 cm, 14 cm.

89.





$A = 1.50$, $B = 5740\ \text{nm}^2$.

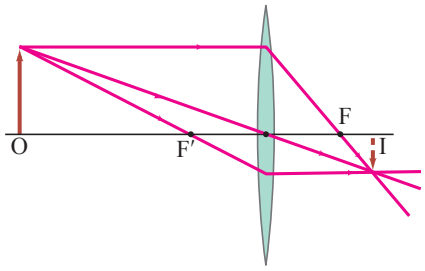
91. 0.18 m toward the viewer from the coin and 0.53 m above the coin.

93. $n = 2$.

95. (*a*) Yes;
(*b*) left: concave, middle: plane, right: convex;
(*c*) concave: same side of mirror as focal point and center of curvature; greater than the center of curvature; plane: anywhere; convex: right: opposite side of mirror as focal point and center of curvature.

Chapter 33

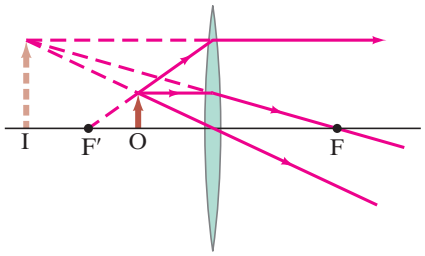
1. (*a*) ≈ 630 mm;
(*b*) 635 mm.



3. (*a*) 3.77 D, converging;
(*b*) −0.160 m, diverging.

5. (*a*) 106 mm;
(*b*) 109 mm;
(*c*) 117 mm;
(*d*) 0.513 m.

7. (*a*) Virtual, upright, magnified;
(*b*) converging;
(*c*) +7.4 D.



9. (*a*) 0.04 m away from the lens;
(*b*) 0.01 m toward the lens.

11. (*a*) −45.0 mm;
(*b*) −105 mm.

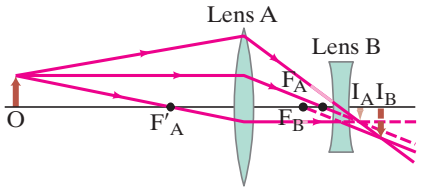
13. 21.3 cm, 64.7 cm.

17. (*b*) 94.9 cm;
(*c*) 94.9 cm.

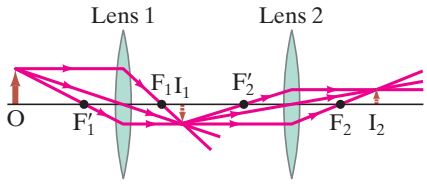
19. (*a*) 22.8 cm beyond the second lens;
(*b*) −0.738×.

21. −14.0 cm.

23. (*a*) 0.10 m beyond the diverging lens;
(*b*) −1.0×;
(*c*)



25. (*a*) Image ≈ 30 cm beyond second lens, half the size of the object;



(*b*) 29 cm beyond the second lens, 0.46× the size of the object.

27. 1.597.

29. 0.34 m.

33. $\frac{1}{125}$ s.

37. (*a*) 49 mm;
(*b*) 14 mm.

39. (*a*) 0.26 mm;
(*b*) 0.47 mm;
(*c*) 1.3 mm;
(*d*) 0.56×, 2.7×.

41. $5.1\ \text{m} \leq d_o < \infty$.

43. $f/2.5$.

45. 0.91 D.

47. (*a*) Nearsighted;
(*b*) 30.6 cm.

49. −9.1 D, −7.7 D.

51. 18.4 cm, 64.3 cm.

53. 0.3 cm.

55. 1.6×.

57. (*a*) 3.7×;
(*b*) 13 mm;
(*c*) 6.7 cm.

59. (*a*) 6.8×;
(*b*) 2.0×.

61. −31×, 89 cm.

63. 24 cm.

65. −30×.

67. −110×.

69. −9×.

71. 7.0×.

73. 1.7 cm.

75. (*a*) 868×;
(*b*) 1.9 cm, 0.289 cm;
(*c*) 0.293 cm.

77. (*a*) 14 cm;
(*b*) 130×.

79. (*a*) 17 cm;
(*b*) 5.4×;
(*c*) 78%.

81. (*a*) Converging;
(*b*) 230 cm.

83. 46.8 cm, converging.

85. Left lens is a converging, right lens is diverging.

87. 2 cm from the original object, towards the lens.

89. 50.0 cm.

91. 45 m.

93. Infinity.

95. −0.578 D (upper part), +2.0 D (lower part).

97. (*a*) $M = -2.3\times$;
(*b*) 4.5 D.

99. (*a*) $\frac{f^2 v_o}{(d_o - f)^2}$;
(*b*) away;
(*c*) 2 *f*.

101. (*b*) $m = +2$ for both.

103. (*a*) 27.0 cm;
(*b*) 1.6 m;
(*c*) 14 cm.

105. -7.8 cm.
107. -10 cm; $d_{ia} = -10$ cm, $d_{ib} = 30$ cm.
109. $d_i = -d_o$.
111. 88 pixels/mm, 11 pixels/mm.
113. (*c*) 125 mm, 156 mm, 94 mm smaller.

Chapter 34

3. 4.3×10^{-6} m.
5. 5.8×10^{-7} m, yellow.
7. 0.2 mm.
9. 1.2×10^{-4} m.
11. 1.2×10^{-5} m.
15. 1.9×10^{-3} m.
17. 3.56×10^{-3} m.
19. ¼.
23. $I_\theta = I_0\left[\dfrac{3 + 2\sqrt{2}\cos\left(\dfrac{2\pi d \sin\theta}{\lambda}\right)}{3 + 2\sqrt{2}}\right]$.
25. 650 nm.
27. (*a*) 180 nm;
(*b*) 360 nm, 540 nm.
29. (*b*) 290 nm.
31. 9.8 μm.
33. (*a*) 120 nm;
(*b*) 240 nm.
35. 33 dark bands.
41. 9 lenses.
43. 2.15×10^{-4} m.
45. 8.31×10^{-5} m.
47. 2.9×10^{-4} m.
49. 57.3°, 32.7°.
51. 44.5%.
53. 37°, 53°.
55. 45°.
57. 9.4%.
59. $I_0 \cos^2\theta \sin^2\theta$ or $\frac{1}{4} I_0 \sin^2 2\theta$, 45°.
61. 3×10^{28} lm, 2×10^{27} cd.
63. 37 polarizers.
65. 0.5 cm.
67. (*a*) 1.3×10^{-4} m;
(*b*) 3.9×10^{-7} m.
69. 460 nm.
71. 4.7 W.
73. 530 nm.
75. $d_1 = \dfrac{\lambda}{2n_1}$, $d_2 = \dfrac{\lambda}{2n_2}$.
77. $\sin\theta_m = \dfrac{m\lambda}{d} \pm \sin\theta_i$,
$m = 0, 1, 2, \cdots$.
79. 330 nm, 160 nm.
81. $\theta_{\text{destructive}} = 0°, 180°$;
$\theta_{\text{constructive}} = 90°, 270°$.
83. 2.40×10^2 nm.
85. (*a*) 0;
(*b*) 0.10;
(*c*) 0.
87. 2.0×10^2 m.
89. (*a*) 60°;
(*b*) 72°;
(*c*) 84°.
91. 126 nm.
93. 1 lens: 92% transmitted, 8% reflected; 6 lenses: 61% transmitted, 39% reflected; 6 lenses with coating: 89% transmitted, 11% reflected.

Chapter 35

1. 2.0°.
3. 2.46×10^{-5} m.
5. (*a*) 16°, 34°, and 56°;
(*b*) 34°;
(*c*) no minima.
7. 4.8 cm.
9. 1.0×10^{-6} m.
11. (*a*) 54°;
(*b*) 26°.
13. $I/I_0 = 0.068$.
15. $d = 5.5D$.
17. 7.
21. (*a*) 3;
(*b*) 9;
(*c*) 15;
(*d*) 31.
23. $(1.54 \times 10^{-5})°$.
25. 1.7×10^{11} m.
27. 1.0×10^4 m.
29. 730 lines/mm, 88 lines/mm.
31. (*a*) 120 m;
(*b*) no.
33. 4.4°.
35. 5.0×10^{-7} m.
37. 1600 slits/cm.
39. 30°.
41. 600 nm − 700 nm 2nd order overlaps with 400 nm − 467 nm 3rd order.
43. 614 mm, 899 mm.
45. 572 nm.
47. $\ell_1 = 8$ cm, $\ell_2 = 45$ cm, second order.
49. (*c*) −0.65°, −30°.
51. (*a*) $R_1 = 1.5 \times 10^4$, $R_2 = 3.0 \times 10^4$;
(*b*) 14 pm.
53. 12.7°.
57. 630 nm.
59. 0.7 m.
61. 13°.
63. 48 km.
65. 9.9×10^{-4} m.
67. Yes.
69. Second order.
71. 0.77 arc min, they can be resolved.
73. 12,500 slits/cm.
75. 2.62×10^{-10} m.
77. 110 m.
79. 1 cm.

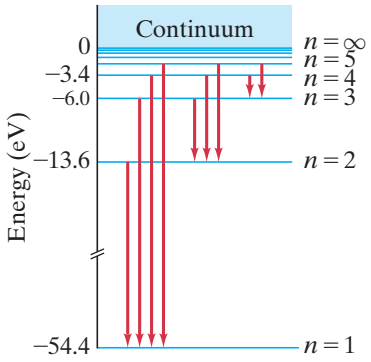
Chapter 36

1. 2.25×10^{-6} s.
3. 29 ly.
5. 0.76 *c*.
7. (*a*) 0.141 *c*;
(*b*) 0.140 *c*.
9. (*a*) 6.47 m, 1.35 m;
(*b*) 14.8 s;
(*c*) 0.670 *c*;
(*d*) 14.8 s.
11. (*a*) 3.0 y;
(*b*) 8.3 y.
13. 6.97×10^{-8} %.
15. 0.9633 *c*.
17. 0.894 *c*.
19. (*a*) (130 m, 20 m, 0);
(*b*) (325 m, 20 m, 0).
21. 2.52×10^8 m/s @ 20.8° above the *x* axis.
23. (*a*) 0.972 *c*;
(*b*) 0.49 *c*.
25. 0.94 *c*, 36°.
27. (*a*) $\ell_0\sqrt{1 - (v\cos\theta/c)^2}$;
(*b*) $\tan^{-1}(\gamma \tan\theta)$.
29. (*a*) A;
(*b*) 2.14×10^{-14} s;
(*c*) B.
31. $-\gamma \dfrac{v\ell}{c^2}$, bulb B.
33. 5.2×10^{-19} kg · m/s.
35. 0.56 *c*.
37. (*a*) 160 %;
(*b*) 270 %.
39. 4×10^{-28} kg.
41. 2.67×10^{-30} kg.
43. 5.6.
45. 3.7.
47. 1.6 GeV/*c*.
49. 0.9655 *c*.
51. 140 MeV/c^2.

53. $K = 34$ Mev, 5% too low, $p = 254$ MeV/c, 3% too low.
55. 237.0483 u.
57. 240 MeV.
59. 2×10^7 kg.
61. 0.1066 c.
65. (*b*) 2.7×10^{-5}.
67. 9.3×10^{-8} s.
69. 0.8465 c.
71. (*a*) $\tan^{-1}\sqrt{\dfrac{c^2}{v^2} - 1}$;
(*c*) $\tan^{-1}\dfrac{c}{v}$, $\sqrt{v^2 + c^2}$.
73. (*a*) 0.77 m/s;
(*b*) 0.21 m.
75. 1.022 MeV or 1.637×10^{-13} J.
79. (*a*) 4×10^9 kg/s;
(*b*) 4×10^7 y;
(*c*) 1×10^{13} y.
81. 28.32 MeV.
83. (*a*) 2.71 GeV/c or 1.45×10^{-18} kg · m/s;
(*b*) 0;
(*c*) 16.6 GeV/c or 8.85×10^{-18} kg · m/s.
87. (*a*) 0.0814 c;
(*b*) 0.085 c.
89. 2.1×10^{22} J.
93. 1.0 m × 0.86 m.
95. (*a*) 23 m;
(*b*) 38 min.
97. 2.9×10^{16} J.
99. (*a*) 0.986 c;
(*b*) $c(1 - 5.0 \times 10^{-7})$.
103. 3.3 T.

Chapter 37

1. (*a*) 10.6 μm, far infrared;
(*b*) 829 nm, infrared;
(*c*) 0.69 mm, microwave;
(*d*) 1.06 mm, microwave.
3. 5.4×10^{-20} J, 0.34 eV.
5. (*a*) 133 J;
(*b*) 267 J;
(*c*) 4.00×10^2 J;
(*d*) $(133n)$ J;
(*e*) −533 J.
9. 2.7×10^{-19} J to 4.9×10^{-19} J, 1.7 eV to 3.0 eV.
11. 3.50×10^{-7} m.
13. 2×10^{13} Hz, 1×10^{-5} m.
15. 3 photons.
17. (*a*) 2.3 eV;
(*b*) 0.3 V.
19. 2.80 eV.
21. 3.7 eV.
23. (*a*) 1.8 eV;
(*b*) 3.0 eV.
25. (*a*) 1.7 eV;
(*b*) 3.0 eV.
27. 0.004.
29. (*a*) 2.43×10^{-12} m;
(*b*) 1.32×10^{-15} m.
31. (*a*) 8.8×10^{-6};
(*b*) 0.049.
33. (*a*) 297 eV;
(*b*) 0.145 nm.
35. 1.83 MeV.
37. 212 MeV, 5.86×10^{-15} m.
39. 1.712 MeV, 7.26×10^{-13} m.
41. 2.5×10^{-12} m.
43. 1840.
45. 1000 m/s.
47. (*a*) 3×10^{-10} m;
(*b*) 9×10^{-11} m;
(*c*) 2.7×10^{-11} m.
49. 690 m/s.
51. 15.3 pm.
53. 1.51 eV.
55. 122 eV.
57. 91.2 mm.
59. 39.9 nm.
61.



63. 972, -1.4×10^{-5} eV.
65. 97.23 nm, 102.6 nm, 121.5 nm, 486.2 nm, 656.3 nm, 1875 nm.
69. 3.28×10^{15} Hz.
71. 5.1×10^{26} photons/s.
73. 3.3×10^{52} photons/y.
75. 4.0×10^{-14} m.
77. 4.40×10^{-40}.
79. 1200 m/s.
81. (*a*) 0.76 V;
(*b*) 5.2×10^5 m/s;
(*c*) 1.4 nm.
85. 9.3×10^{-9} N.
89. (*a*) 1.7 nm.
91. 7.88×10^4 K.
93. (*c*) 150 eV;
(*d*) 1.3×10^{-66} J, 4.3×10^{-33} m/s;
(*e*) 0.14 nm.
95. 13.6 eV.
97. 1.1×10^{-3} s.
99. 3.8×10^{-7} W/m^2, 1.7×10^{-2} V/m.

Index

Note: The abbreviation *defn* means the page cited gives the definition of the term; *ff* means "also the following pages;" *fn* means the reference is in a footnote; *pr* means it is found in a Problem or Question.

4K, 680
A (atomic mass number), 1227
Aberration:
  chromatic, 982 *fn*, 985, 1037
  of lenses, 984–85, 1033, 1036
  spherical, 932, 947, 984, 985, 1037
ABS, 131
Absolute pressure, 371
Absolute scale of temperature, 496, 503, 507–8
Absolute space, 1057, 1061–62
Absolute time, 1057
Absolute zero, 598
  kinetic energy near, 520
  of temperature, 502–3
Absorbed dose, 1273
Absorption lines, 1040–41, 1111, 1197, 1200–1201
Absorption spectrum, 1040–41, 1111, 1200
Absorption wavelength, 1118
Abundances, natural, 1227
Ac circuits, 730–32, 747 *fn*, 875–89
Ac generator, 846–47
Ac motor, 796
Accademia del Cimento, Florence, Italy, 495
Accelerating reference frames, 88, 90, 163–64, 318–21
Acceleration, 14, 27–44, 63–65
  angular, 267–72, 274–79
  average, 27–28
  centripetal, 124 *ff*
  constant, 30–43, 65
  constant angular, 270–71
  Coriolis, 320–21
  cosmic, 1352–53
  due to gravity, 37–43, 89 *fn*, 94, 149–51, 328 *pr*
  in *g*'s, 41
  instantaneous, 29–30, 63–64
  mass and, 103, 121
  of the Moon, 125, 146
  motion at constant, 30–43, 65–73
  radial, 124 *ff*, 133
  related to force, 88–91
  and slope, 29
  tangential, 133–34, 267–68
  uniform, 30–43, 65–73
  variable, 43–44
  velocity vs., 28, 40
Accelerators, particle, 1290–96
Accelerometer, 102
Acceptor level, 1212
Accommodation of eye, 976
Accuracy, 3–6
  precision vs., 6
Achilles tendon, 358 *pr*
Achromatic doublet, 985
Achromatic lens, 985
Acre (*defn*), 10
Actinides, 1168
Action at a distance, 162, 618
Action potential, 737
Action–reaction (Newton's third law), 91–94
Activation, of matrix, 681
Activation energy, 521, 1190, 1192
Active galactic nuclei (AGN), 1326
Active matrix, 681–82
Active solar heating, 599
Activity, 1241
  and half-life, 1244
  source, 1272
Actual-sized images, 952 *pr*
ADC, 676
Addition:
  significant figures in, 5
  of vectors, 55–61
  of velocities, 73–75
  relativistic, 1076
Addressing pixels, 680–81
Adenine (A), 631
Adenosine triphosphate, 216
Adhesion, 387–88
Adiabatic lapse rate, 573 *pr*
Adiabatic processes, 552–53
  expansion of a gas, 559–60
ADP, 216, 1191–92
AFM, 1110, 1151
AGN, 1326
Agrigento, Sicily, 344
A · h, 721
Air:
  atmospheric, pressure variation with elevation, 370
  composition, 525 *fn*
  mass of in a room, 505
  speed of air molecules, 519
  underground circulation, 384
Air bags, 32, 254 *pr*
Air cleaner, electrostatic, 710 *pr*
Air columns, vibrations of, 470–72
Air conditioners, 585–86
Air parcel, 573 *pr*
Air pollution, 598
Air pressure, 370–71, 373, 396 *pr*
  variation with elevation, 370
Air resistance, 37–38, 134–35, 196
Airplane, 383
  banking on turns, 137 *pr*, 144 *pr*
  roar, 466
  runway, 32
Airplane wing, dynamic lift, 383
Airy disk, 1033
Alcor, 1054 *pr*
Alkali metals, 1168
Allowed transitions, 1162–63, 1196–97, 1199, 1200
Alpenhorn, 491 *pr*
Alpha decay, 1234–37, 1240
  and tunneling, 1151, 1236
Alpha particle (or ray), 1151, 1234–37
Alternating current (ac), 730–32, 747 *fn*, 881–88
Alternators, 848
Altitude:
  air pressure variation, 370
  cooking time for food, 526–27
  water boiling point change, 526–27
Alveoli, 19 *pr*
AM radio, 917
Amino acids, 1194
Ammeter, 767–70, 795
  digital, 767, 770
AMOLED, 1217
Amorphous solids, 1202
Ampère, André, 720, 815
Ampere (A) (unit), 720, 814–15
  operational definition of, 815
Ampere-hour (A·h), 721
Ampère's law, 815–21, 899–903
Amplification, 894 *pr*
Amplifiers, 1218
Amplitude, 401, 430, 437
  and energy of simple harmonic oscillator, 408
  intensity related to, 436, 466
  pressure variation, 463
  of vibration, 401, 404
  of wave, 401, 430, 436, 437 *ff*, 1130
Amplitude modulation (AM), 918
Analog information, 856
Analog (mechanical) watch, 424 *pr*
Analog meters, 767–70, 795
Analog signal:
  voltage, 676
  with and without noise, 678
Analog-to-digital converter (ADC), 676
Analyzer, Polaroid as, 1011
Anderson, Carl, 1299
Andromeda, 1325
Aneroid barometer, 373
Aneroid gauge, 372–73
Angle, 8 *fn*, 265
  attack, 383
  critical, 943
  of dip, 784
  of incidence, 444, 449, 927, 939
  phase, 403, 438, 885–86
  polarizing, 1013, 1014
  radian measure of, 265
  of reflection, 444, 927
  of refraction, 449, 939
  solid, 8 *fn*, 1016 *fn*
Angstrom (Å) (unit), 18 *pr*, 941 *fn*
Angular acceleration, 267–72, 274–79
  constant, 270–71
Angular displacement, 266, 413
Angular frequency, 404
Angular magnification, 935, 936–38, 979, 981, 983
Angular momentum, 302–18, 1113
  in atoms, 1113, 1160–63, 1171–74
  conservation, law of, 303–5, 315–17, 1240
  direction of, 306–7, 314–15, 317–18
  nuclear, 1229
  of a particle, 309–10
  quantized in atoms, 1160–61
  quantized in molecules, 1196–97
  relation between torque and, 310–15
  total, 1173
  and uncertainty principle, 1134
  vector, 306, 309
Angular position, 265, 1134
Angular quantities, 265 *ff*
  vector nature, 270

Angular velocity, 266–71
of precession, 317–18
vector, 325 *pr*
Animals, underground, 384
Anisotropy of CMB, 1343, 1349
Annihilation (e^-e^+, particle–antiparticle), 1105, 1301, 1346
Anode, 679, 719
Antenna, 898, 903, 911, 919, 1004
Anthropic principle, 1355
Anticodon, 1195
Anticyclone (anticyclonic behavior), 321
Antilock brakes (ABS), 131
Antilogarithm, A-3
Antimatter, 1299, 1316, 1318 *pr* (*see also* Antiparticle)
Antineutrino, 1238–39, 1305
Antineutron, 1299
Antinodes, 446–47, 468–72
Antiparticle, 1104, 1239, 1299–1302, 1305 (*see also* Antimatter)
Antiproton, 1299–1300
Antiquark, 1305, 1309
Aorta, 380, 384, 388
Apollo, 82 *pr*
Apparent brightness, 1326–27
Apparent loudness, 465
Apparent magnitude, 1360 *pr*
Apparent weight, 154–55, 376, 394 *pr*
Apparent weightlessness, 154–55
Apple, 67, 146
Newton's, 146
Approximations, 6, 11–13
Aquarium window, 369
Arago, F., 1026
Arch of Titus, 350
Arches, 350–51, 358 *pr*
Archimedes, 375
Archimedes' principle, 375–78
and geology, 378
Area, 10, 14, A-1, inside back cover
of contact, 290
under a curve or graph, 178–80
Arecibo, 1036
Aristotle, 2, 86
Ark, Noah's, 19 *pr*
Arm, 273
Armature, 795, 846
Arrays, in TV and computer monitors, 681
Arteries, 380, 384, 386, 388, 389
hardening of, 384, 386
Arterioles, 380
Arteriosclerosis, 384, 386
Artificial radioactivity, 1233
Artificial satellites, 152–54
ASA number, 972
Associative property, 57
Asteroid belt, 168 *pr*
Asteroids, 168 *pr*, 171 *pr*, 223 *pr*, 262 *pr*, 324 *pr*, 329 *pr*
doomsday, 114 *pr*
Astigmatism, 977, 984 *fn*, 985
Astronauts, 82 *pr*, 126, 293 *pr*
on Moon, 82 *pr*
Astronomical telescope, 980–82
Astronomical unit (AU), 17 *pr*, 169 *pr*
Astronomy, 566
Astrophysics, 19 *pr*, 1322–55
Asymptote, 760
Asymptotic freedom, 1312
Athletes, making, 339
ATLAS, 1295, 1314
Atmosphere, light scattered by, 1015–16
Atmosphere (atm) (unit), 371, 373
Atmospheric pressure, 370–71, 373
decrease with altitude, 370
variation with elevation, 370
Atom trap, 1124 *pr*, 1127 *pr*
magneto-optical, 837 *pr*
Atomic bomb, 1266, 1269
Atomic emission spectra, 1041, 1111
Atomic force microscope (AFM), 1110, 1151
Atomic mass, 493, 1136–38
Atomic mass number, 1227
Atomic mass unit, 8, 493
unified, 1229
Atomic number, 1166, 1169–71, 1227
Atomic spectra, 1111–12, 1116–18
Atomic structure, 207
Bohr model of, 1113–19, 1128, 1158–60
early models of, 1110–11
of hydrogen atoms, 1159–65
of multielectron atoms, 1166–68
nuclear model of, 1110
planetary model of, 1110
quantum mechanics of, 1158–79
shells and subshells in, 1167–68
Atomic theory of matter, 493–94, 609
Atomic weight, 493
Atomizer, 383
Atoms, 506–7, 516–22, 527–31, 1110–20
angular momentum in, 1113, 1160–63, 1171–74
arrangement of, 188 *pr*
binding energy in, 1116
Bohr model of, 1113–19
as cloud, 1159
crystal lattice of, 1202
and de Broglie's hypothesis, 1119–20
definition, 493
diameter, 494
distance between, 494
electric charge in, 611
energy levels in, 1113–19, 1160–61, 1166–67, 1169
hydrogen, 1111–20, 1159–65
ionization energy in, 1116–18
multielectron, 1166–68
neutral, 1229
probability distributions in, 1159, 1163–65
quantum mechanics of, 1158–79
shells and subshells in, 1167–68
size, 494
vector model of, 1184 *pr* (*see also* Atomic structure; Kinetic theory)
ATP, 216, 1191–92
Atrium, heart, 388–89
Attack angle, 383
Attractive forces, 1189–90, 1296
Atwood machine, 101, 296 *pr*, 313
AU, 169 *pr*
Audible range, 461
Audio, 490 *pr*
Audio compact disc, 298 *pr*
Aurora borealis, 792
Autofocusing camera, 462
Automobile spring, 399, 401, 404, 445
Autoradiography, 1277
Average acceleration, 27–28
Average acceleration vector, 63
Average angular acceleration, 267
Average angular velocity, 266
Average position, 1146
Average speed, 22–23, 520, 522
Average velocity, 22–24, 63
vector, 63
Avogadro, Amedeo, 506
Avogadro's hypothesis, 506
Avogadro's number, 504, 506–7
Avogadro's Principle, 506
Axel jumps, 330 *pr*
Axial vector, 270, 307 *fn*
Axis:
instantaneous, 284
of lens, 959
of rotation (*defn*), 265
for solving problems, 61, 98, 99
for statics, 333 *ff*
for torque on objects in equilibrium, 333
Axon, 736–37

Back, forces on, 340
Back emf, 848–50
Back muscles, 340–41
Back pain, 340–41
Background radiation, cosmic microwave, 1342–44, 1348, 1349, 1353
Backlight, 680
Backpack, work done on, 175
Bag, potato chip, 513 *pr*
Bainbridge-type mass spectrometer, 800
Balance (balancing), 341–42
of car wheel, 314–15
for humans, 342
on seesaw, 335
Ball curve, 384
"Ballad of a Thin Man," 1352 *fn*
Ballistic galvanometer, 867 *pr*
Ballistic pendulum, 240–41
Balloons, 398 *pr*
helium, 505
hot air, 492, 515 *pr*
Balmer, J. J., 1112
Balmer formula, 1112, 1116
Balmer series, 1112, 1116–17
Band gap, 1208–9
Band spectra, 1196, 1200–1201
Band theory of solids, 1208–10
and doped semiconductors, 1212
Bandwidth, 678
Banking of curves, 130–32
Banking on turns:
airplanes, 137 *pr*, 144 *pr*
cars, 130–32
Bar codes, 1177
Bar (unit), 371, 373
Bardeen, J., 1218
Barn (bn) (unit), 1261
Barometer, 373, 535 *pr*
Barrel distortion, 985
Barrier, Coulomb, 1151, 1236, 1329
Barrier penetration, 1149–51, 1236
Barrier tunneling, 1149–51, 1236
Baryon, 1305–6, 1309–11, 1352
and quark theory, 1309–11
Baryon number, 1301, 1305–6, 1308–10, 1315
conservation of, 1301
Base, nucleotide, 631, 1194
Base, of transistor, 1218
Base bias voltage, 1218
Base quantities, 8, 9
Base semiconductor, 1218
Baseball, 76 *pr*, 77 *pr*, 80 *pr*, 82 *pr*, 83 *pr*, 84 *pr*, 108 *pr*, 109 *pr*, 172, 234, 322 *pr*, 330 *pr*, 384, 1134
work done on, 172, 182
Baseball bat, 330 *pr*, 412–13
Baseball curve, 384
Basilar artery, 384
Basketball, 50 *pr*, 83 *pr*, 109 *pr*
Bats, baseball, 330 *pr*, 412–13
Bats, Doppler use, 479

Battery, 663, 718–21, 724, 748
  automobile, charging, 748 *fn*, 757–59
Battery jumper cables, 758–9
Battery terminal, 719, 748, 758, 759
  corrosion of, 780 *pr*
Bayer mosaic, 971
Beam splitter, 1010
Beams, 344
  sag of, 346
Bear sling, 106 *pr*, 355 *pr*
Beat frequency, 475–76, 479
Beats, 474–76, 479
  in Doppler, 479
Becquerel, Henri, 1233
Becquerel (Bq) (unit), 1272
Bel (unit), 464
Bell, Alexander Graham, 464
Bending, body, 340
BER, 679
Bernoulli, Daniel, 380, 381
Bernoulli's equation, 381–85
Bernoulli's principle, 380–85
Beta decay, 1234, 1237–40, 1245, 1313
  inverse, 1331
  neutrinoless double, 1321 *pr*
Beta particle (or ray), 1234, 1237 (*see also* Electron)
Betatron, 866 *pr*
Betelgeuse, 566
Bethe, Hans, 1268
Biasing and bias voltage, 1213, 1218
Biceps muscle, 339, 363 *pr*
Bicycle (bike), 112 *pr*, 192 *pr*, 298 *pr*, 299 *pr*, 300 *pr*, 301 *pr*, 307, 314, 327 *pr*
  gears, 299 *pr*
  sprockets, 192 *pr*
  wheel, 302, 306, 314
  spinning, 324 *pr*
Bicyclist, 192 *pr*, 225 *pr*, 301 *pr*
Big Bang theory, 15, 480, 1316, 1322, 1341–55
Big crunch, 1349, 1350
Billiards, 327 *pr*
Bimetallic strip, 495, 509 *pr*, 515 *pr*
Binary code, 856
Binary numbers, 676–79
Binary system, 1331, 1338
Binding energy:
  in atoms, 1116
  in molecules, 224 *pr*, 1188, 1190, 1192–93
  of nuclei, 1230–32
  in solids, 1202
  total, 1093 *pr*, 1231
Binding energy per nucleon (*defn*), 1231
Binoculars, 944, 982
Binomial expansion, A-1, inside back cover
Biological damage, 1271–72
  by radiation, 1271–72
Biological development, entropy and, 594
Biot, Jean Baptiste, 821
Biot-Savart law, 821–24
Bipolar junction transistor, 1218
Birds:
  eyes of, 265
  of prey, 265
Bismuth-strontium-calcium-copper oxide (BSCCO), 736
Bit deletion, 678
Bit depth, 677, 678
Bit-error-rate (BER), 679
Bit-flips, 678
Bit insertion, 678
Bit-rate, 676, 678
Bit slip, 678 *fn*
Bits, 298 *pr*, 856
Black holes, 164, 167 *pr*, 170 *pr*, 1326, 1331, 1337–38, 1350, 1359 *pr*
Blackbody, 1096
Blackbody radiation, 1096–97, 1327, 1342
Blood, 380, 384, 386, 388–89, 483
  circulation, 532
  convection by, 563
  lack to brain, 384
  velocity of, 384
Blood flow, 380, 384, 386, 388, 395 *pr*, 491 *pr*
  electromagnetic measurement, 491 *pr*, 845
Blood pressure, 386
  measuring, 389
Blood transfusion, 395 *pr*
Blu-ray discs, 1178
Blueshift, 1339
Boat, bow wave, 480
Body, bending, 340
Body fat, 393 *pr*
Body heat:
  and convection by blood, 563
  cooling by radiation, 564
  radiative loss of, 564
Body temperature, 549
Bohr, Niels, 1106, 1113, 1120, 1128, 1135–36, 1238
Bohr magneton, 1172, 1230
Bohr model of atom, 1113–19, 1128, 1158–60
Bohr radius, 1114, 1158, 1159, 1163–64
Bohr theory, 1128, 1158–60
Boiling, 525
  water, 203, 526–27
  altitude vs., 526–27
  (*see also* Phase, changes of)
Boiling point, 496 *fn*, 525, 545
  and altitude, 526–27
Boltzmann, Ludwig, 595
Boltzmann constant, 507, 518, 596
Boltzmann distribution, 1175
Boltzmann factor, 1175, 1205
Bomb:
  atomic, 1266, 1269
  fission, 1266
  fusion, 1269
  hydrogen, 1269
Bomb calorimeter, 545, 570 *pr*
Bond (*defn*), 1187–88
  covalent, 1187–89, 1202, 1203
  dipole–dipole, 1193
  dipole–induced dipole, 1193
  hydrogen, 1192–96
  ionic, 1188, 1190, 1202, 1203
  metallic, 1202
  molecular, 1186–89
  partially ionic and covalent, 1188–89
  in solids, 1202–3
  strong, 1187–89, 1192–93, 1202–3
  van der Waals, 1192–96, 1202–3
  weak, 1192–96, 1203
Bond energy, 1187–88, 1192–93
Bond length, 1192, 1222 *pr*
Bonding:
  in molecules, 1186–89
  in solids, 1202–3
Bone fracture, 346, 359 *pr*, 364 *pr*
Born, Max, 1128, 1130
Bose, Satyendranath, 1167
Bose–Einstein statistics, 1204 *fn*
Bosons, 1167, 1204 *fn*, 1304, 1305, 1309–12
Bottomness and bottom quark, 1305 *fn*, 1308–10
Bound charge, 706
Bound state, 1147
Boundary conditions, 1142, 1147–48
Boundary layer, 384
Bow wave, of a boat, 480
Bowling, 263 *pr*
Box, rigid, 1142–46
Boyle, Robert, 502
Boyle's law, 502, 503, 517
Bragg, W. H., 1044
Bragg, W. L., 1044, 1128
Bragg equation, 1044
Bragg peak, 1277
Bragg scattering of X-rays, 1179
Brahe, Tycho, 156
Brake pads, 372
Brakes:
  antilock, 131
  brake system of a car, 372
  hydraulic, 372
Braking a car, 28, 35, 183, 288–89
Branes, 1317
Brattain, W., 1218
Brayton cycle, 607 *pr*
Break-even (fusion), 1271
Breakdown voltage, 666
Breaking point, 342
Breaking the sound barrier, 481
Breath, molecules in, 507, 514 *pr*
Breeder reactor, 1265
Bremsstrahlung, 1170
Brewster, David, 1013
Brewster's angle and law, 1013
Bridge circuit, 778 *pr*
Bridge-type full-wave rectifier, 1220 *pr*
Bridges, 347–49, 363 *pr*
  collapsed, 418
  expansion, 498
Brightness, apparent, 1326–27
British engineering system of units, 8
British thermal unit (unit), 539, 570 *pr*
Broglie, Louis de, 1106, 1120
Bronchoscope, 945
Brown, Robert, 493
Brownian motion, 493
Brunelleschi, Filippo, 352
Brushes, 796, 846
BSCCO, 736
Btu (unit), 539, 570 *pr*
Bubble chamber, 1249, 1299
Buckling, highway, 501
Bugle, 490 *pr*
Building materials, *R*-values of, 562
Built environment, 331
Bulk modulus, 343, 345
Bumpers, car, 191 *pr*
Bungee cords, 221 *pr*
Bungee jumper, 422 *pr*
Buoyancy, 374–78
  center of, 397 *pr*
Buoyant force, 374–78
Burglar alarms, 1101
Burning (= fusion), 1329 *fn*
Burrows, 384
Bus rollover, 328 *pr*
Buttressing, 350

Cable television, 920
Calculations:
  for adding vectors, 61
  in problem solving, 33, 34
  rough, 104
Calculators, 1015
  errors, 5
Caloric, 539
Calorie (unit), 539
  content, 545, 570 *pr*
  relation to joule, 539–40
  working off, 540
Calorimeter, 544, 545, 570 *pr*, 1248, 1249
Calorimetry, 542–45

Camera:
autofocusing, 462
digital and film, 970–75
gamma, 1278
Camera flash unit, 701
Campfire, jar into, 503
Cancer, 1272, 1276–77, 1291
Candela (cd) (unit), 1016
Cantilever, 336
Capacitance, 693–707
of axon, 738
Capacitance bridge, 711 *pr*
Capacitance, variable, 715 *pr*
Capacitive reactance, 884
Capacitor discharge, 762–63
Capacitor microphone, 695
Capacitors, 692–707, 1219
charging of, 759–61, 899–902
in circuits, 698–700, 759–64, 883–84
discharging, 761–63
energy stored in, 700–703
as filters, 883–84
with *R* or *L*, 759–64, 878 *ff*
reactance of, 884
in series and in parallel, 698–700
uses of, 884
Capacity, 694–707, 738
Capillaries, 380, 386, 388
Capillarity, 387–88
Capillary action, 388
Capture, electron, 1239
Car:
battery charging, 757–59
bumpers, 191 *pr*
efficiency, 215, 580
electric, 598, 720, 742 *pr*, 744 *pr*, 854, 869
flywheel, 301 *pr*
power needs, 214–15
rattle, 420 *pr*
skidding, 130–32, 190 *pr*
springs, 399, 401, 404, 415
stopping, 28, 35, 183, 288–89
stuck in mud, 115 *pr*
wheel balancing, 314–15
work on a, 183
work to stop, 183
Carbon (CNO) cycle, 1268, 1286 *pr*
Carbon dating, 1246–48
Carbon dioxide, trees offsetting, 608 *pr*
Carbon footprint, 598
Card stripe reader, 858
Cardiac defibrillator, 703, 712 *pr*, 764 *fn*, 780 *pr*
Carnot, N. L. Sadi, 580
Carnot cycle, 580–83, 588
Carnot efficiency, 581–83
Carnot engine, 580–84, 603 *pr*
Carnot refrigerator, 606 *pr*
Carnot's theorem, 582
Carousel, 264, 267, 268
Carrier frequency, 917
Carrier of force, 1296–98, 1312
Cartesian coordinates, 21
Caruso, Enrico, 417
Cassegrainian focus, 982
CAT scan, 1045
Catalysts, 1192
Catapult, 189 *pr*
Cathedral:
in Florence, 331, 352
at Pisa, 412
Cathode, 679, 719
Cathode ray tube (CRT), 679–82, 799, 919
Cathode rays, 679, 797–98 (*see also* Electron)
Causal laws, 159
Causality, 159
Cavendish, Henry, 147, 150
CCD, 970, 974
CD player, 1177
CDM model of universe, 1353–54
CDMA cell phone, 920
CDs, 18 *pr*, 48 *pr*, 53 *pr*, 298 *pr*, 1024 *pr*, 1177
Celestial sphere, 155
Cell (biological):
energy in, 1192
radiation taken up by, 1273
Cell (electric), 719, 748
Cell phone, 451, 852, 898, 911, 920, 1015
Cells (biological):
energy in, 216
human, molecules in, 519
pressure on, 371
Cellular phones, 451, 852, 898, 911, 920, 1015
Celsius temperature scale, 496–97
Center of buoyancy, 397 *pr*
Center of gravity (CG), 246, 248
and balance, 342
Center of mass (CM), 244–50
and angular momentum, 311
and moment of inertia, 275, 280, 284–87
and statics, 334
and translational motion, 248–50, 284–85
Centi- (prefix), 8
Centigrade temperature scale, 496–97
Centiliter (cL) (unit), 8
Centimeter (cm) (unit), 8
Centipoise (cP) (unit), 385
Centrifugal (pseudo) force, 127, 319–21
Centrifugal pump, 388
Centrifugation, uranium enrichment by, 126
Centrifuge, 294 *pr*
angular acceleration of, 271
Centripetal acceleration, 124 *ff*, 176, 185
Centripetal force, 126–29
Cepheid variables, 1333, 1356 *pr*
CERN, 1293, 1294, 1314
CG, 246, 248, 342
Cgs system of units, 8
Chadwick, James, 1227, 1287 *pr*
Chain reaction, 1262–64, 1266
Chain rule, A-6
Chain sliding off edge of flat surface, 190 *pr*
Chamberlain, Owen, 1299
Chandrasekhar limit, 1330
Change:
of phase, 522–24, 545, 547
in potential energy, 201
in a quantity, 22
Characteristic expansion time, 1341
Characteristic X-rays, 1169
Charge:
at center of ring, 623
on electron (*e*), 8, 9
free and bound, 706
magnetic field due to single moving, 824
(*see also* Electric charge)
Charge-coupled device (CCD), 970
Charge density, 623, 648, A-16
Charge per unit length, 623
Charges, accelerating and EM waves, 905
Charging a battery, 748 *fn*, 757–59
Charging by induction, 612–13
Charles, Jacques, 502, 503
Charles's law, 503, 532 *pr*
Charm, 1305 *fn*, 1308–11
Charmed quark, 1308
Chart of the Nucleotides, 1226
Checking solutions, 33, 244
Chemical bonds, 1187–96
Chemical lasers, 1177
Chemical reactions, temperature dependence, 521
Chemical shift, 1280
Chernobyl, 1265
Chicago storm, 263 *pr*
Child, pushing, on swing, 417
Chimney, 384
Chip, computer, 18 *pr*, 1186, 1212, 1219
Cholesterol buildup, 386
Chord, 25, 266
Chromatic aberration, 982 *fn*, 985, 1037
Chromatic scale, 468, 487 *pr*
Chromatography, 531
Chromodynamics, quantum (QCD), 1298, 1311–14
Chromosome, 631, 632
Cigarette, 567
Circle of confusion, 973
Circuit, digital, 1218
Circuit, electric (*see* Electric circuits)
Circuit breaker, 729–30, 766, 826, 859
Circular apertures, 1033–35
Circular motion, 123–34
nonuniform, 133–34
uniform, 123–26
Circular polarization, 925 *pr*
Circulating pump, 388
Circulation, smoking and, 567
City planning, 112 *pr*
Clarinet, 473
Classical physics (*defn*), 2, 1056, 1129
Clausius, R. J. E., 577, 587
Clausius equation of state, 527
Clausius statement of second law of thermodynamics, 577, 585
Climber's fall, 193 *pr*
Climber's traverse, 110 *pr*
Climbing rope, 115 *pr*
Clock, quartz oscillator, 487 *pr*
Closed system (*defn*), 542
Closed tube, 470–72
Clothes:
dark vs. light, 564
insulating properties of, 562
Cloud, electron, 1159, 1165, 1187–89
Cloud chamber, 1249
Cloud formation, 526
Clusters:
of galaxies, 1325, 1349, 1353
of stars, 1325
Clutch, 305
CM (*see* Center of mass)
CMB, 1342–44, 1348, 1349, 1353–54
CMB anisotropy, 1343, 1349, 1353–54
CMB uniformity, 1349
CMOS, 970, 974
CNO cycle, 1268, 1286 *pr*
CO molecule, 1198
Coal, energy in, vs. uranium, 1265
Coating of lenses, optical, 1008–9
Coaxial cable, 714 *pr*, 818, 874, 911
COBE, 1342
Coefficient:
of kinetic friction, 117–18
of linear expansion, 498
of performance (COP), 585–87, 603 *pr*
of restitution, 259 *pr*
of static friction, 117–18
of viscosity, 385
of volume expansion, 498–501
Coherence, 1001–2

Coherent light, 1001, 1175, 1178
Cohesion, 387–88
Coil (*see* Inductor)
Coin tosses, 595–96
Cold dark matter (CDM) model of universe, 1353–54
Collapse of freeway, 418
Collector (of transistor), 1218
Collider Detector at Fermilab (CDF), 1249
Colliding beams, 1294–96
Collimated beam, 1278 *fn*
Collimated gamma-ray detector, 1278
Collisions, 234–44
  completely inelastic, 240
  conservation of energy and momentum in, 230–33, 236–44
  elastic, 235–39, 241–44
  and impulse, 234–35
  inelastic, 236, 239–41, 252
  nuclear, 239, 243
  in 1-D, 236–41
  in 2- or 3-D, 241–44
Colloids, 366
Colonoscope, 945
Color:
  in digital camera, 971
  of light related to frequency and wavelength, 941–43, 998, 1001, 1007
  of quarks, 1311–13
  of star, 1096, 1328
Color charge, 1311–13
Color-corrected lens, 985
Color force, 1312–13, 1315
Color screens, 680
Coma, 984
Comets, 168 *pr*
Common logarithms, A-2–A-3
Commutative property, 56, 176, 308
Commutator, 796
Compact disc (CD) player, 1177
Compact disc, 18 *pr*, 48 *pr*, 53 *pr*, 298 *pr*, 1024 *pr*, 1177
Compass, magnetic, 782–83, 784
Complementarity, principle of, 1105–6
Complementary metal oxide semiconductor (CMOS), 970
Complete circuit, 720
Completely inelastic collisions, 240
Complex quantities, 1130 *fn*, 1137 *fn*, 1140
Complex wave, 442, 472
Components of vector, 58–62
Composite particles, 1304, 1305, 1309–10
Composite wave, 442, 472
Composition resistors, 723
Compound lenses, 985
Compound microscope, 983–84
Compound nucleus, 1261–62
Compound semiconductors, 1215
Compounds, 493
Compression, 344–45, 431, 462
  in digital-to-analog conversions, 677
  in longitudinal wave, 431, 434–35
  work done in, 552–53
Compression ratio, 584
Compression waves, 435
Compressive strength, 346
Compressive stress, 344–45
Compton, A. H., 1102–3, 1128, 1263
Compton effect, 1102–4, 1271
Compton shift, 1103
  derivation of, 1104
Compton wavelength, 1103
Computed tomography (CT), 1045–47, 1278–79
Computer:
  and digital information, 856
  disks, 856
  hard drive, 856
  keyboard, 695
  LCD for, 1015
  memory, 715 *pr*
  monitor, 679–82
  optical drive, 857
  printers, 619
Computer-assisted tomography (CAT), 1045
Computer chips, 18 *pr*, 1186, 1212, 1219
Computerized axial tomography (CAT), 1045
Concave mirror, 931, 935–37, 982
Concentration gradient, 530, 561
Concert hall, temperature rise, 572 *pr*
Concordance model, 1345
Concrete, prestressed and reinforced, 346
Condensation, 523–26
Condensed-matter physics, 1202–19
Condenser microphone, 695
Condition, boundary, 1142, 1147–48
Conductance, 744 *pr*
Conduction:
  charging by, 612–13
  electrical, 611, 717–70
  of heat, 561–62, 573 *pr*
  in nervous system, 736–37
Conduction band, 1208–9
Conduction current (*defn*), 902
Conduction electrons, 611
Conductive layer, 1217
Conductivity:
  electrical, 725, 735
  thermal, 561
Conductors:
  charge of, 1212
  electric, 611, 627, 720 *ff*
  heat, 561
  quantum theory of, 1208–9
Cone, moment of inertia, 297 *pr*
Cones (retinal), 976
Configuration, electron, 1167–68
Confinement:
  in fusion, 1270–71
  of quarks, 1312, 1345–47
Congruent triangles, A-2
Conical pendulum, 129
Conservation laws, 172, 201
  of angular momentum, 303–5, 315–17
  apparent violation of, in beta decay, 1238
  of baryon number, 1301, 1315
  and collisions, 230–33, 236–44
  of electric charge, 610, 1240, 1301
  in elementary particle interactions, 1297, 1301–2
  of energy, 200–212, 549–50, 1137, 1235, 1238, 1240, 1302
  of lepton number, 1301–2, 1315
  of mechanical energy, 201–6
  of momentum, 230–44, 1301–2
  in nuclear and particle physics, 1240, 1301
  in nuclear processes, 1238
  of nucleon number, 1240, 1301–2, 1306
  of strangeness, 1307
Conservation of angular momentum, 302–7, 315–17
Conservation of energy, 195–217, 235–44, 543, 547, 549, 1137, 1235, 1238, 1240, 1302
  in collisions, 236–39
  kinetic energy, 244
Conservation of linear momentum, 227, 230–44
Conservation of mechanical energy, 201–6
Conservative field, 856
Conservative forces, 195–201
  definition of, 195
  and nonconservative forces, 195
Conserved quantity, 172, 201
Constant acceleration, 30–43, 65
  angular, 270–71
  for freely falling objects, 37–43
  kinematic equations for, 31, 65
  problem-solving strategy, 33–36
Constant-volume gas thermometer, 497, 508
Constants:
  normalization, 1144
  values of: inside front cover
Constructive interference, 444–45, 473–74, 999 *ff*, 1009, 1187
Contact force, 86, 94, 97, 162
Contact lens, 978
Continent, model of, 397 *pr*
Continental drift, 378
Continuity, equation of, 379–82
Continuous laser, 1177
Continuous spectrum, 1040, 1096
Continuous wave, 430
Contraction:
  holes, 499
  length, 1070–71, 1075
Contrast, 1048
Control rods, 1264
Convection, 563
  by blood, 563
  forced, 563
  natural, 563
Convection currents, 563
Convective cooling, 574 *pr*
Conventional current (*defn*), 721
Conventions, sign (geometric optics), 935, 963
Converging lens, 958 *ff*
Conversion factors, 9, inside front cover
Converting units, 9–10
  temperature, 496
Convex mirror, 931, 938
Conveyor belt, 253, 261 *pr*
Cooksey, Donald, 1291
Cooling:
  by convection, 574 *pr*
  by evaporation, 524, 548, 549
  by radiation, 564–65
Coordinate axes, 21
Coordinate system, 67, 68, 104, 130, 244, 334
COP, 585–87
Copenhagen interpretation of quantum mechanics, 1136
Copernicus, Nicolaus, 156
Corbeled arch, 359 *pr*
Cord, tension in, 99–102
Core:
  of Earth, 392 *pr*
  of reactor, 1264
Coriolis acceleration, 320–21
  forgotten in battle, 322 *pr*
Coriolis force, 319–21, 328 *pr*
Corkscrew, 356 *pr*
Cornea, 976
Corona discharge, 666, 710 *pr*
Corrective lenses, 976–78
Correspondence principle, 1086, 1119, 1129

Cosine(s), 59
law of, 188 *pr*
Cosmic acceleration, 1352–53
Cosmic Background Explorer (COBE), 1342
Cosmic microwave background radiation (CMB), 1342–44, 1348, 1349, 1353–54
anisotropy of, 1343, 1349, 1353
uniformity of, 1343, 1349
Cosmic rays, 1290
Cosmological constant, 1352, 1353
Cosmological model, 1345–54
Cosmological principle, 1341
perfect, 1341
Cosmological redshift, 1339
Cosmology, 1316, 1322–55
Coulomb, Charles, 613
Coulomb barrier, 1151, 1236, 1329
Coulomb (C) (unit), 614, 815
operational definition of, 815
Coulomb force (*defn*), 615
Coulomb potential (*defn*), 667
Coulomb (unit), 8
Coulomb's apparatus, 613
Coulomb's law, 613–18, 645–46, 652, 903, 1191
vector form of, 616, 618
Counter emf, 848–50
Counter torque, 849
Counters, 1248–49
Counterweight, 101
Couple, force, 333
Covalent bond, 1187–89, 1202, 1203
Cowan, Clyde, 1303 *fn*
Crane, 112 *pr*
Crashworthiness, of car, 261 *pr*
Crest, 430
Crick, F., 1044
Critical angle, 943
Critical damping, 415
Critical density, of universe, 1350–51
Critical mass, 1263–66
Critical point, 523
Critical reaction, 1263–66
Critical temperature, 523, 735
Cross-country skis, 136 *pr*
Cross product, vector, 307–9
Cross product proof, 325 *pr*
Cross section, 1260–63
Crosstalk, 678
CRT, 679–82, 799, 919
Crystal lattice, 494, 1202
Crystallography, 1044
CT scan, 1045–47, 1278–79
Cue ball, 300 *pr*, 327 *pr*
Curie, Marie, 1128, 1233
Curie, Pierre, 829, 1233
Curie (Ci) (unit), 1272
Curie temperature, 825, 829
Curie's law, 829
Curl, A-16
Current:
electric (*see* Electric current)
induced, 838–59, 869 *ff*
Current density, 732–35
Current gain, 1218
Current sensitivity, 767
Curvature:
of field, 985
of space, 163–64, 1336–38, 1349–51
of universe (space-time), 1336–38, 1349–50
Curves, banking of, 130–32
Cutoff wavelength, 1170–71
Cycle (*defn*), 401
Cyclones, 321
Cyclotron, 808 *pr*, 1291–92
Cyclotron frequency, 791, 1292
Cygnus X-1, 1338
Cytosine (C), 631

Da Vinci, Leonardo, 118
DAC, 677, 780 *pr*
Dam, 393 *pr*
Damage, done by radiation, 1271–72
Damped harmonic motion, 414–17
Dampers, 415
Dante Alighieri, 452 *pr*
Dark energy, 1300, 1348, 1350–53
Dark matter, 1317, 1348, 1350–53
Dashes and dots, 676 *fn*
Data lines, 682
Data stream, 678, 681
Dating:
geological, 1247–48
radioactive, 784, 1246–48
Daughter nucleus (*defn*), 1234
Davisson, C. J., 1107
dB (unit), 464–67
Dc circuits, 747–70
Dc (*defn*), 730
Dc generator, 847, 848
Dc motor, 796
de Broglie, Louis, 1106, 1120, 1128, 1129
de Broglie wavelength, 1106–7, 1119–20, 1130, 1137, 1291
applied to atoms, 1119–20
Dead Sea, 527
Debye equation, 575 *pr*, 608 *pr*
Debye (unit), 672
Decay, 1233
alpha, 1151, 1234–37, 1240
beta, 1234, 1237–39, 1240, 1245, 1313, 1321 *pr*, 1331
of elementary particles, 1300–1314
exponential, 759–62, 876, 1241–42
gamma, 1234, 1239–40
proton, 1305, 1315–16
radioactive, 1233–49
rate of, 1240–44
types of radioactive, 1234, 1240
Z-particle, 1298
Decay constant, 1240–41
Decay series, 1245–46
Deceleration, 28–29
Decibels (unit), 464–67
Decimal numbers, 676
Declination, magnetic, 784
Decommissioning nuclear power plant, 1265
Decoupled photons, 1343, 1348
Dee, 1292
Defects of the eye, 976–78, 985
Defibrillator, heart, 703, 712 *pr*, 780 *pr*, 764 *fn*
Defined values, 8, 9
Definite integrals, A-7, A-9
Definite proportions, law of, 493 *fn*
Deformation, 332
Degeneracy:
electron, 1330
neutron, 1331
Degradation of energy, 594–95
Degrees of freedom, 557–58
Dehumidifiers, 608 *pr*
Dehumidify a cellar, 537 *pr*
Del operator, 673 *fn*, A-16
Delayed neutrons, 1264
Delta particle, 1307
Deltoid muscle, 358 *pr*
Demagnetization, 828
Democritus, 493
Demodulator, 918
Dendrites, 736
Density (*defn*), 8, 366
charge, 623, 648, A-16
and floating, 377
of occupied states, 1205
probability, 1130, 1140, 1143, 1148, 1159, 1162–63, 1165, 1187
of states, 1204–7
of universe, 1350–51
of water, anomalous, 500–501
Depletion layer, 1212
Depth:
for atmosphere, 370
of field, 973
pressure and, 368
Depth finding, 481
Derivatives, 24–25, 30, A-6, inside back cover
partial, 200, 440
Descartes, René, 21
Descriptive laws (*defn*), 3
Design, ramps, 113 *pr*
Destructive interference, 444–46, 473–74, 999, 1009, 1187
and energy, 444
Detection:
of particles, 1248–49
of radiation, 1274
Detectors, of particles and radiation, 1248–49
Detergents, 387
Determinism, 159, 1135–36
Deuterium, 1227, 1257, 1263, 1267–71
Deuterium–tritium fusion (d–t), 1270
Deuteron, 1227, 1257
Dew point, 526
Diagrams:
Feynman, 1297, 1312
force, 97
free-body, 97–104
H–R, 1328, 1333
phase, 523–24
phasor, 885–86, 1002, 1029, 1041–42
potential, energy, 1189–92
PT, 523–24
PV, 522–23, 528, 551
ray, 933, 938, 964
for solving problems, 33, 34, 61, 67, 98, 104, 130, 175, 209, 244, 277, 334, 622, 963
Diamagnetism, 828–29
Diamond, 944
Diastole phase, 388
Diastolic pressure, 389
Dielectric constant, 703
Dielectric relaxation time, 776 *pr*
Dielectric strength, 703
Dielectrics, 703–5
molecular description of, 706–7
Diesel engine, 552, 575 *pr*, 607 *pr*
Differential cross section, 1261
Differential equation, second-order, 402
Diffraction, 450–51, 996, 1025–44, 1176
by circular opening, 1033–35
as distinguished from interference, 1033–34
in double-slit experiment, 1031–33
of electrons, 1107–8
Fraunhofer, 1026 *fn*
Fresnel, 1026 *fn*
of light, 996, 1025–44
as limit to resolution, 1033–37
by single slit, 1026–31
X-ray, 1043–44
in X-ray imaging, 1045–47
Diffraction factor, 1032
Diffraction grating, 1037–39
resolving power of, 1041–43

Diffraction limit of lens resolution, 1033–35
Diffraction patterns, 1026
of circular opening, 1033–34
of single slit, 1026–31
X-ray, 1043–44
Diffraction spot or disk, 1033–35
Diffuse reflection, 928
Diffusion, 530–32, 536 *pr*
applications of, 531–32
Fick's law of, 530
in living organisms, 531
Diffusion constant, 530
Diffusion equation, 530
Diffusion time, 531
Digital ammeter, 767, 770
Digital artifact, 971
Digital bits, 298 *pr*
Digital camera, 970–75
Digital circuits, 1218
Digital information, 856
Digital signal:
voltage, 676–79
with and without noise, 678
Digital-to-analog converter (DAC), 677, 780 *pr*
Digital transmission, 678–79
Digital TVs, 680–81
Digital video disc (DVD) players, 1177–78
Digital voltmeter, 767, 770
Digital zoom, 975, 987 *pr*
Dilation, time, 1064–69, 1075
Dimensional analysis, 14–15, 142 *pr*, 144 *pr*, 459 *pr*, 1127 *pr*, 1359 *pr*, A-12
Dimensions, 14–15
Diodes, 1151, 1213–14, 1249
forward-biased, 1213
junction, 1218
lasers
pn, 1216
semiconductor, 1177
light-emitting (LED), 680, 1215–16
photo-, 1101, 1215
reverse-biased, 1213
semiconductor, 1212–17
tunnel, 1151
zener, 1213–14
Diopter (D) (unit), 960
Dip, angle of, 784
Dipole antenna, 903–4
Dipole layer, 736–37
Dipole–dipole bonds, 1193
Dipole–induced dipole bonds, 1193
Dipoles and dipole moments:
of atoms, 1171–74
electric, 626, 629–31, 671–72, 706
magnetic, 793–94, 823
of nuclei, 1229
Dirac, P. A. M., 1128, 1161, 1204 *fn*, 1300
Dirac equation, 1300
Direct current (dc), 730 (*see also* Electric current)
Direction:
change in, for velocity, 88
of friction force in skid, 131
of net force, 127
and nonuniform circular motion, 133
and uniform circular motion, 123, 124
of vector, 22, 55, 61
Discharge:
capacitor, 762–63
corona, 666, 710 *pr*
Discharge tube, 1111
Discovery in science, 798
Disintegration, 1233
Disintegration energy (*defn*), 1234–35
Disks:
compact disc (or disk), 18 *pr*, 48 *pr*, 53 *pr*, 298 *pr*
of spine, 340–41
Disorder, order and, 593–94
Dispersion, 442, 942
Displacement, 22–23, 401, 411, 437
angular, 266, 413
resultant, 55–56
with unphysical solution, 43
vector, 22, 55–57, 62–63
in vibrational motion, 401
of wave, 437–41, 1130
Displacement current, 902
Dissipative forces, 207–10
energy conservation with, 208–10
Dissociation energy, 1188
Distance, 22, 24
astronomical, 1323, 1326, 1328, 1332–34
between atoms, 494
estimating, via SNIa and Redshift, 1333–34
force acting at a, 162
image, 929, 934, 946, 962–63
object, 929, 934, 946, 962–63
relativity of, 1070–76
table of, typical, 7
traveled, 22–23
wave amplitude decrease with, 436
Distance (tiny) measurement, 896 *pr*, 1010, 1021 *pr*
Distortion, by lenses, 985
Distribution, probability:
in atoms, 1130, 1140, 1143, 1148, 1159, 1162–63, 1165
in molecules, 1187
Distributive law, 308, 325 *pr*
Distributive property, 176, 188 *pr*
Diver, 304
Divergence, A-16
Divergence theorem, A-16
Diverging lens, 959 *ff*
Divine Comedy (Dante), 452 *pr*
Diving board, 331
Division, significant figures in, 4
DNA, 631–33, 1041, 1044, 1192–96, 1272, 1277–78
Dog whistle, 488 *pr*
Domains, magnetic, 825
Domes, 331, 351–52
Donor level, 1212
Doomsday asteroid, 114 *pr*
Door opener, automatic, 1101
Doorbell, 826
Doping of semiconductors, 1211 *ff*
Doppler, J. C., 476 *fn*
Doppler blood-flow meter, 479, 491 *pr*
Doppler echocardiography, 483
Doppler effect, 476–80
bats use, 479
for light, 480, 1085–86, 1338
in medical imaging, 483
in medicine, 479
for sound, 476–80
Doppler radar speed, 479
Doppler ultrasound image, 483
Dose, 1272–76
effective, 1273
Dosimetry, 1272–76
Dot (scalar) product, 176–77
Dots and dashes, 676 *fn*
Double helix, 632
Double-slit experiment:
electrons, 1129–31
light, 998–1002
intensity in pattern, 1002–4, 1031–33
Down quark, 1308
Drag coefficient, 398 *pr*
Drag force, 134–35, 184, 383, 398 *pr*
Drain, in MOSFET, 1218
Drain terminal, 857
DRAM, 716 *pr*, 857–58
Drift velocity, 732–35, 799
Drinking fountain, 398 *pr*
Driven oscillator, power transferred to, 425 *pr*
Driving, around curves, 130–31
Drum, 467
Dry cell, 719
Dry ice, 523
Dryness, in winter, 526
d–t (deuterium–tritium) fusion, 1270
Duality, wave–particle, 1106–8, 1119–20
Duck, bow wave, 480
Dulong and Petit value, 558
Dust, interstellar, 1325
DVD player, 1177–78
Dwarfs, white, 1326, 1328, 1330
Dye lasers, 1177
Dylan, Bob, 1352 *fn*
Dynamic lift, 383
Dynamic random access memory (DRAM), 716 *pr*, 857–58
Dynamic rope, 115 *pr*
Dynamics, 20, 86 *ff*
fluid, 378–88
hydro-, 378
of rotational motion, 274 *ff*
of uniform circular motion, 126–29
Dynamo, 846–48
Dyne (unit), 89, 173
Dynodes, 1248

Ear, 461
discomfort, 396 *pr*
response of, 467
sensitivity of, 464, 466–67
Earth:
artificial satellites of, 152–54
as concentric shells, 148–49, A-13–A-15
core of, 392 *pr*
electric field of, 642 *pr*, 658 *pr*
equatorial bulge, 327 *pr*
estimating radius of, 13, 18 *pr*
in geocentric vs. heliocentric systems, 155
gravitational field of, 642 *pr*
gravity due to, 38 *ff*, 125, 146–52
as inertial frame, 87, 151–52
interior, 481
magnetic field and magnetic poles of, 784
mass, radius, etc.: inside front cover
mass determination, 150
meteorite striking, 326 *pr*
model of, 392 *pr*
not inertial frame, 144 *pr*, 319–21
precession of axis, 327 *pr*
radius, 1, 7, 13, 18 *pr*
rocks and earliest life, 1248
seasons on, 566
soundings, 481
Sun's orbit in relation to, 159
Earthing, electrical, 612
Earthquake, 415, 418, 430, 435, 436, 437, 449–50, 453, 457
epicenter, 453 *pr*
intensity, 437
resonance, 418
Earthquake waves, 435, 437, 450, 453 *pr*
Eccentricity, 156
ECG, 663, 682–83
Echolocation, 434

Eclipse, 161
lunar, 161
solar, 161, 293 *pr*
Eddy currents:
electric, 850
in fluids, 379, 850, 861
Edison, Thomas, 679
Effective acceleration of gravity, 328 *pr*
Effective cross section, 1260
Effective dose, 1273
Effective length of musical instrument, 470 *fn*
Effective mass, 191 *pr*, 427 *pr*
Effective values, 731
Efficiency, 215, 579–82
car, 580
Carnot, 581–83
and Otto cycle, 584
steam engine, 582
in thermodynamic problem-solving, 600
Eiffel Tower, 511 *pr*
Einstein, Albert, 163, 494, 558, 1056, 1058, 1061–62, 1065, 1075, 1098, 1128, 1266, 1334–37, 1352
Einstein cross, 1336
Einstein ring, 1336
EKG, 663, 682–83
El Capitan, 83 *pr*, 392 *pr*
EL device, 1216
Elapsed time, 23
Elastic collisions, 235–39, 241–44
in one dimension, 236–41
Elastic cross section, 1260
Elastic limit, 342
Elastic modulus, 342–44, 461
and speed of sound waves, 433
Elastic potential energy, 199–200, 204–5, 210
Elastic region, 342
Elastic scattering, 1260
Elasticity, 342–45
Electric battery, 663, 718–21, 724, 748
Electric car, 598, 720, 742 *pr*, 744 *pr*, 854, 869
Electric cell, 718, 748
Electric charge, 610 *ff*
in atom, 611
bound and free, 706
conservation of, 610, 1240, 1301
continuous charge distributions, 622–25
and Coulomb's law, 613–18
of electron, 614
elementary, 614
free, 706
induced, 612–13, 706
motion of
in electric field, 628–29
in magnetic field, 788–93
point (*defn*), 615
quantization of, 614
test, 618–19
types of, 610
Electric circuits, 720–21, 729–32, 747–70, 875–89
ac, 730–32, 747 *fn*, 881–88
complete, 720
containing capacitors, 698–700, 759–64, 883 *ff*
dc, 747–70
digital, 1218
impedance matching of, 888
induced, 838–59, 869 *ff*
integrated, 1219
and Kirchhoff's rules, 754–57
LC, 877–81
LR, 875–77
LRC, 880–88
open, 720
parallel, 698, 729, 750
RC, 759–64
reactance, 881–84
rectifier, 1214
resonant, 887–88
series, 698, 749
time constants of, 760, 875–76
Electric conductivity, 725, 735
in nervous system, 736–37
Electric current, 717, 720–24, 729–36, 754 *ff*
alternating (ac), 730–32, 747 *fn*, 881–88
conduction (*defn*), 902
conventional, 721
density, 732–35
direct (dc) (*defn*), 730
displacement, 902
eddy, 850
hazards of, 764–66
induced, 839
leakage, 766
magnetic force on, 786–94
microscopic view of, 732–35
and Ohm's law, 722–24
peak, 730
produced by magnetic field, 839–40
produces magnetic field, 785–88, 825
protection from, 729, 764–6, 826, 859, 877
rms, 731–32
Electric dipole, 626, 629–31, 671–72, 706
Electric energy, 661–63, 674–75, 700–703, 726–28
resources, 599
stored in capacitor, 700–703
stored in electric field, 703
Electric field, 618–33, 643–52, 664–66, 672–73, 856
calculation of, 618–25, 647–52, 664, 672–73
and conductors, 627, 721 *fn*
continuous charge distributions, 622–25
in dielectric, 704–5
of and by dipole, 629–31
of Earth, 658 *pr*
in EM wave, 903–5
energy stored in, 703
and Gauss's law, 643–52
inside a wire, 735
motion of charged particle in, 628–29
produced by changing magnetic field, 839–40, 854–56
produces magnetic field, 899–902
relation to electric potential, 664–66, 672–73
Electric field lines, 626, 671
Electric flux, 644–45, 900
Electric force, 609, 613–18, 792
Coulomb's law for, 613–18
and ionization, 1271
in molecular biology, 631–33, 1192–96
Electric generator, 846–48
Electric guitar, 469
Electric hazards, 764–66
Electric motor, 795–96
counter emf in, 848–50
Electric plug, 765–66
Electric potential, 660–73
binary numbers, 676–79
digital, 676–79
of dipole, 671–72
due to point charges, 666–70
equipotential surfaces, 670–71
relation to electric field, 664–66, 672–73
signal voltage, 676–79
TV and computer monitors, 679–82
(*see also* Potential difference)
Electric potential energy, 660–63, 674–75, 700–703
Electric power, 726–30
in ac circuits, 731–32, 875–77, 882, 883, 886–89
generation, 846–48
in household circuits, 729–30
and impedance matching, 888
transmission of, 851–54
Electric quadrupole, 641 *pr*
Electric safety, 628, 729, 764–6, 826, 859, 877
Electric shielding, 628, 818
Electric shock, 764–66
Electric stove burner, 727
Electric toothbrush, 854
Electric vehicle, 742 *pr*, 744 *pr*
Electric wheelchair, 745 *pr*
Electrical grounding, 612, 721
Electricity, 609–925
hazards of, 764–66
static, 609 *fn*
three phase 889, 895 *pr*
Electrocardiogram (ECG, EKG), 660, 663, 682–83, 779 *pr*
Electrochemical series, 718
Electrode, 718
Electroluminescent (EL) device, 1216
Electrolyte, 718
Electromagnet, 826
Electromagnetic (EM) waves, 903–20
Doppler effect for, 475
(*see also* Light)
Electromagnetic energy, 1293
Electromagnetic force, 161–62, 792, 1241, 1296–98, 1304–5, 1313–16, 1334
Electromagnetic induction, 838 *ff*
Electromagnetic oscillations, 877–81, 888
Electromagnetic pumping, 802 *pr*
Electromagnetic spectrum, 910, 941–43
Electrometer, 613
Electromotive force (emf), 718, 748–49, 838–50 (*see also* Emf)
Electron:
as beta particle, 1234, 1237
as cathode rays, 679, 797
charge on, 614, 798–99
cloud, 1120, 1159, 1165, 1187–89
conduction, 611
defined, 1108
discovery of, 797–99
in double-slit experiment, 1130–31
electric force on, 615
as elementary particle, 1301–2
free, 611, 1141, 1202, 1210
mass of, 799, 1229
measurement of charge on, 798–99
measurement of e/m, 797–98
momentum of, 1077
motion of, in electric field, 628–29
in pair production, 1104–5
path in magnetic field, 790
photoelectron, 1100
sea of electrons, 1300
sharing, 1187
speed of, 734–35
spin, 825
in vacuum tube, 663
wave nature, 1131
wavelength of, 1107
Electron capture, 1239
Electron cloud, 1120, 1159, 1165, 1187–89

Electron configuration, 1167–68
Electron degeneracy, 1330
Electron diffraction, 1107–8
Electron lepton number, 1301, 1305, 1310
Electron microscope (EM), 1095, 1108, 1132, 1151, 1156 *pr*
Electron neutrino, 1304, 1305
Electron spin, 825, 1161, 1172–74, 1187
Electron transport layer (ETL), 1217
Electron volt (eV) (unit), 674–75, 1229
Electronic circuits, 1213–19
Electronic devices, 1210–19
Electronic pacemakers, 764, 871
Electroscope, 612–13, 718 *fn*
Electrostatic air cleaner, 710 *pr*
Electrostatic force, 613–18, 631–33, 1193
  defined, 615
Electrostatic potential energy, 661, 674–75
Electrostatic unit (esu), 614 *fn*
Electrostatics, 610–707
Electroweak era, 1346
Electroweak force, 162, 609 *fn*, 1313–16
Electroweak theory, 162, 1313–16
Elementary charge, 614
Elementary particle physics, 1289–1317
Elementary particles, 1289–1317
Elements, 493, 1167–68
  in compound lenses, 985
  origin of in universe, 1330–31
  Periodic Table of, 1167–68, inside back cover
  production of, 1330–31
  transmutation of, 1234, 1257–59
  transuranic, 1259
Elevation:
  air pressure variation, 370
  water boiling point change, 526–27
Elevator, 101, 108 *pr*
Ellipse, 156
EM, 1095, 1108, 1132, 1151, 1156 *pr*
EM waves, 475, 903–20 (*see also* Light)
Emergency blanket, 568 *pr*
Emf, 161–62, 718, 748–49, 838–50
  back, 848–50
  counter, 848–50
  of generator, 846–50
  Hall, 799–800
  induced, 838–50, 873
  motional, 845–46
  and photons, 1297
  RC circuit with, 761
  in series and in parallel, 757–59
  sources of, 748, 838–48
Emission spectra, 1096–97, 1111–12, 1115–18
  atomic, 1041, 1111
Emission tomography, 1279
Emissive layer (EML), 1217
Emissivity, 564
Emitter (transistor), 1218
EML, 1217
Empire State Building, 396 *pr*
Emulsion:
  on film, 970
  photographic, 1249
Endoergic reaction (*defn*), 1258
Endoscopes, 945
Endothermic reaction (*defn*), 1258
Energy, 172, 181–85, 195–217, 236–44, 281–85, 549–51, 660 *ff*
  as ability to do work, 181
  activation, 521, 1190, 1192
  and ATP, 1191–92
  binding, 1093 *pr*, 1116, 1188, 1190, 1192–93, 1230–32
  bond, 1187–88, 1192–93
  conservation of, 200–210, 543, 547, 549, 1137, 1235, 1238, 1240, 1302
  dark, 1290, 1300, 1348, 1351, 1352
  degradation of, 594–95
  disintegration, 1234–35
  dissociation, 1188
  electric, 661–63, 674–75, 700–703, 726–30
  in EM waves, 903–4, 913–15, 1293
  equipartition of, 557–58
  Fermi, 1204–6, 1210
  and first law of thermodynamics, 549–51
  from flywheel, 282
  geothermal, 599
  gravitational potential, 197–98, 202, 205–6, 210–12
  internal, 207, 540–41
  ionic cohesive, 1202
  ionization, 1116, 1118
  kinetic, 172, 181–85, 200–213, 215, 281–85, 1081–82
  in living cells, 216
  and mass, 1080–84
  mechanical, 200–206
  molecular kinetic, 518–19
  nuclear, 578 *fn*, 599, 1256–82
  nucleotide, 1194
  photon, 1098–1102
  potential, 197–212, 215–17, 660–63, 674–75, 700–703 (*see also* Electric potential; Potential energy)
  quantization of, 1097, 1113–19, 1143
  reaction (*defn*), 1258
  relation to work, 281–83, 1080
  relativistic, 1080–84
  rest, 1081–82, 1135
  rotational, 281–83 *ff*, 541, 1196–98, 1200–1201
  in simple harmonic oscillator, 408–10
  solar, 599
  storage in cells (ATP), 216
  thermal, 207, 540–41
  threshold, 1259, 1288 *pr*
  total binding, 1093 *pr*
  transformation of, 207, 213
  translational kinetic, 181–83
  transported by waves, 435–37
  unavailability of, 594–95
  unavailable to do work, 605 *pr*
  and uncertainty principle, 1133–34, 1148
  units of, 173, 182, 273 *fn*
  vacuum, 1352
  vibrational, 408–10, 541, 1198–1201
  work and, 181–85, 197–200, 210–14
  zero-point, 1143, 1149–50, 1156 *pr*, 1199
Energy bands, 1208–10
Energy conservation, law of, 1137, 1235, 1238, 1240, 1302
Energy conservation law, 207–12, 549–50
  general establishment, 207–8
  mechanical, 200–206
  work–energy principle vs., 210
Energy density:
  in electric field, 703–5
  in magnetic field, 875, 913
Energy gap, 1208–9
Energy levels:
  in atoms, 1113–19, 1160–62
  for fluorescence, 1174
  for lasers, 1175–78
  in molecules, 1196–1201
  nuclear, 1239–40
  in solids, 1208–9
  in square well, 1143
Energy states, in atoms, 1113–19
Energy transfer, heat as, 538–40
Engine:
  diesel, 552, 575 *pr*, 607 *pr*
  internal combustion, 512 *pr*, 579, 583–84
  power, 213–15
  steam, 578
Enriched uranium, 126, 1263
Entire universe, 1344
Entropy, 587–94
  and biological development, 594
  change in, for melting, 588
  as order to disorder, 593–94
  and second law of thermodynamics, 590–97
  as a state variable, 589
  statistical interpretation, 595–97
  and time's arrow, 593
Enzymes, 632, 633, 1192
Equally tempered chromatic scale, 468, 487 *pr*
Equation of continuity, 379–82
Equation of motion, 402–3
  for damped harmonic motion, 414
  for forced oscillation, 418–19
Equation of state, 502
  for an ideal gas, 504, 507
  Clausius, 527
  van der Waals, 527–28
Equations, in problem solving, 33, 34, 67, 68, 104
Equilibrium (*defn*), 216, 332 *ff*, 341–42
  conditions for, 332–34
  first condition for, 332–33
  force in, 332–34
  neutral, 216, 341
  and quasistatic processes, 551
  second condition for, 333–34
  stable, 216, 341
  static, 331–42
  thermal, 497, 543, 547
  unstable, 216, 341
Equilibrium distance, 1192, 1222 *pr*
Equilibrium equations, 334
Equilibrium position (vibrational motion), 400
Equilibrium state, 497, 502, 551
Equipartition of energy, 557–58
Equipotential lines, 670–71
Equipotential surface, 670–71
Equivalence, principle of, 163–64, 1334–35
Erg (unit), 173, 182
Error correction, 679
Error detection, 678
Escape lane for trucks, 79 *pr*, 110 *pr*
Escape speed from the Earth, 536 *pr*
Escape velocity, 212, 532 *pr*, 1351
  from Earth, 212
  from Moon, 212, 532 *pr*
  from solar system, 223 *pr*
Escher drawing, 218 *pr*
Estimated uncertainty, 4
Estimating, 11–13
  in problem solving, 33, 277
Eta (particle), 1305
Ether, 1058–61
ETL, 1217
Euclidean space, 1336–37

European Center for Nuclear Research (CERN), 1293, 1294, 1314
Evaporation, 524, 548–49
  cooling with, 524, 548, 549
Event, 1062 *ff*
Event horizon, 1338
Everest, Mt., 150, 171 *pr*, 527
Evolution, 594
  stellar, 1329–32
Examination of motion, in problem solving, 67, 68
Exchange particles (carriers of force), 1296–98
Excited state:
  of atom, 1104, 1115 *ff*
  of nucleon, 1307
  of nucleus, 1239–40
Exciton, 1217
Exclusion principle, 1166–67, 1187, 1204, 1207, 1311, 1330, 1331
Exoergic reaction (*defn*), 1258
Exothermic reaction (*defn*), 1258
Expansion, 462
  free, 555, 591, 597
  holes, 499
  linear and volume, 342–45
  mathematical, A-1
  in structures, 498–500
  thermal, 497–501
  of universe, 1338–41, 1350–53
  in waves, 431
Expansion joints, 495, 501
Exponential curves, 759–62, 875, 1241–42
Exponential decay, 759–62, 876, 1241–42
  defined, 760
Exponents, A-1, inside back cover
Exposure time, 971–72
Extension cord, 730
Extensors, 339
External force, 231, 244, 249
Extragalactic (*defn*), 1325
Extrasolar planets, 158
  wobble of, 262 *pr*
Extraterrestrials, possible communication with, 923 *pr*
Eye:
  aberrations of, 985
  accommodation, 976
  bird's, 265
  defects of, 976–78, 985
  far and near points of, 976
  lens of, 976
  normal (*defn*), 976
  resolution of, 1035, 1037
  structure and function of, 975–78
Eyeglass lenses, 976–78
Eyepiece, 981

f-number, 972
f-stop (*defn*), 971–72
Factorial number, 605 *pr*, A-1 *fn*
Fahrenheit temperature scale, 496
Falkland Islands, battle of, 322 *pr*
Falling objects, 37–43
Fallout, radioactive, 1266
False-color images, 1046
Far field, 904
Far point of eye, 976
Farad (F) (unit of capacitance), 693
Faraday, Michael, 162, 618, 838–40
Faraday cage, 628
Faraday's law of induction, 840–41, 854–55, 903
Farsightedness, 976–77
Fat:
  body, 393 *pr*
  metabolizing, 575 *pr*
Femtometer (fm) (unit), 1228
Fermat, 1336
Fermat's principle, 956 *pr*, 1336
Fermi, Enrico, 13, 1106, 1129, 1167, 1204 *fn*, 1238, 1259, 1263, 1306–7
Fermi energy, 1204–7, 1210
Fermi factor, 1205
Fermi (fm) (unit), 1228
Fermi gas, 1204
Fermi level, 1204–7
Fermi speed, 1206
Fermi temperature, 1225 *pr*
Fermi–Dirac probability function, 1205, 1210
Fermi–Dirac statistics, 1204–7, 1209 *fn*
Fermilab, 1293, 1294
Fermions, 1167, 1204, 1311
Ferris wheel, 129, 170 *pr*, 264
Ferromagnetism and ferromagnetic materials, 783, 824–28
Fetus, image of, 482
Feynman, R., 1297
Feynman diagram, 1297, 1312
Fiber optics, 945
Fick's law, 530
Fictitious forces, 319–21
Field, 162
  conservative and nonconservative, 856
  electric, 618–33, 643–52, 664–66, 672–73, 856 (*see also* Electric field)
  football, 17 *pr*
  in elementary particles, 1296
  gravitational, 162–64, 627, 1334–41
  Higgs, 1314
  magnetic, 782–93, 811–29 (*see also* Magnetic field)
  soccer, 17 *pr*
  vector, 626
Fifth lumbar vertebra, 340
Film badge, 1249
Film projector, 989 *pr*
Film speed, 972 *fn*
Filter circuit, 884, 896 *pr*, 897 *pr*
Fine structure, 1128, 1159, 1161, 1174
Fine structure constant, 1174
Finger holds water in a straw, 371
Finite potential well, 1147–49
Fire hose, force on you, 395 *pr*
Fire truck, moving, 476
First harmonic, 446–47
First law of motion, 86–88
First law of thermodynamics, 538–67
  extended, 550–51
  and thermodynamic processes, 551–55
Fish finder, 489 *pr*
Fission, 238, 599
  nuclear, 1261–66
Fission bomb, 1266
Fission fragments, 1261–65
Fitzgerald, G. F., 1061
Fixed stars, 155
Flag pole, struck by lightning, 837 *pr*
Flash memory, 858
Flasher unit, 763
Flat screens, 680–81
Flatness, 1349
Flavor (of elementary particles), 1303, 1311
Flavor oscillation, 1303
Flexible cord, tension in, 99–102, 196
Flexors, 339
Flip coil, 867 *pr*
Floating gate, 858
Floating objects, 377
Floor vibrations, 406
Florence, Italy, 331, 352, 495
Flow:
  of fluids, 378–88
  laminar, 378, 379
  meter, Doppler, 479, 491 *pr*
  streamline, 378, 379
  in tubes, 379–82, 384–86
  turbulent, 379, 383
  volume rate of, 379
Flow rate, 379
Flow speed, measuring, 384
"Fluid displaced," in Archimedes' principle, 375
Fluid dynamics, 378–89
Fluid statics, 365–78
Fluids (*defn*), 365–89, 366
  in motion, 378–89
  (*see also* Flow, of fluids; Gases; Liquids; Pressure)
Fluorescence, 1174
Fluorescent lightbulb, 1175
  ballast, 853
Flute, 469–72
Flux:
  electric, 644–45, 900
  magnetic, 840 *ff*, 854–56, 902, 906
Flying buttresses, 350
Flywheel, 282, 301 *pr*
FM radio, 917–19
Focal length:
  of lens, 959–60, 967, 969–70, 975, 976
  of spherical mirror, 932, 937
Focal plane, 959
Focal point, 932, 937, 959–60, 976
Focus, 932
Focusing, of camera, 972–73
Fog formation, 526
Food, slower cooking at altitude, 526–27
Foot-candle (*defn*), 1016 *fn*
Foot (ft) (unit), 7
Foot-pounds (unit), 173, 182
Football field, 17 *pr*
Football kicks, 69, 71
Forbidden energy gap, 1208–9
Forbidden transitions, 1163, 1176 *fn*, 1199 *fn*, 1200
Force, 85–105, 161–62, 195–96, 228, 248–50, 1298, 1316
  addition of, 97, 149
  atomic, 207
  attractive, 1189–90, 1296
  on back, 340
  buoyant, 374–75
  but no work, 173
  centrifugal (pseudo), 127, 319
  centripetal, 126–29
  color, 1312–13, 1315
  conservative, 195–201
  contact, 86, 94, 97
  Coriolis, 319–21
  definition of, 89
  diagram, 97
  dissipative, 207–10
  drag, 134–35, 383, 398 *pr*
  electromagnetic, 161–62, 792, 1241, 1296–98, 1304–5, 1313–16, 1334
  electrostatic, 613–18, 631–33, 1193
  electroweak, 162, 609 *fn*, 1316
  in equilibrium, 332–34
  equilibrium equations for, 334
  exerted by inanimate object, 92
  external, 231, 244, 249
  fictitious, 319
  friction, 87, 88, 117–23, 196, 207–10
  of gravity, 86, 94–96, 145–64, 195, 197–98, 1298, 1316, 1317, 1322, 1331, 1334–38, 1350, 1352
  impulsive, 235
  inertial, 319
  internal, 231, 249

Force (*continued*)
long-range, 1233, 1334
magnetic, 782, 786–94
measurement of, 86
on Moon, 125
in muscles and joints, 295 *pr*, 339–41, 355 *pr*, 358 *pr*, 363 *pr*
net, 87–91, 97 *ff*
in Newton's laws, 85–105, 228, 231, 249–50
nonconservative, 196, 199, 207–9
normal, 94–96
nuclear, 161–62, 226 *pr*, 1232–33, 1238, 1296–1317, 1334
pseudoforce, 319
relation of acceleration to, 88–91
relation of momentum to, 228–29, 231, 234–35, 250, 251, 1077, 1079
repulsive, 1189–90, 1296
resistive, 134–35
restoring, 179, 400
short-range, 1233, 1334
strong nuclear, 161–62, 1232, 1259 *fn*, 1296–1317, 1334
types of, in nature, 161–62, 609 *fn*, 1298, 1316
units of, 89
van der Waals, 1192–96, 1202–3
velocity-dependent, 134–35
viscous, 385–86
weak nuclear, 161–62, 1233, 1238, 1298–1317, 1334
of wind, 263 *pr*
(*see also* Electric force)
Force constant, spring, 400
Force diagrams, 97
Force pumps, 374, 388
Forced convection, 563
Forced oscillations, 417–19
Forceps, 355 *pr*
Forearm, 295 *pr*, 339
Forward biased diode, 1213
Fossil-fuel steam plants, 599
Fossil fuels, 598
Foucault, J., 997
Four-dimensional space-time, 1072, 1336
Fourier analysis, 442, 472
Fourier integral, 442
Fourier's theorem, 442
14-nm technology, 1219
Fovea, 975
Foveon technology, 971
Fracture, 332
of bone, 346, 359 *pr*, 364 *pr*
Frame of reference, 21, 87–88, 318–21, 1056 *ff*
accelerating, 88, 90, 163–64, 318–21
inertial, 87–88, 90, 319, 1056 *ff*
noninertial, 88, 90, 319, 1056
rotating, 318–21
for Sun, 159
Sun as, 159
transformations between, 1073–76
Franklin, Benjamin, 610, 652
Franklin, Rosalind, 1044
Fraunhofer diffraction, 1026 *fn*
Free-body diagrams (*defn*), 97–104, 130, 277, 332, 334
Free charge, 706
Free-electron theory of metals, 1203–7
Free electrons, 611, 1141, 1202, 1210
Free expansion, 555, 591, 597
Free fall, 37–43, 154–55
Free particle, and Schrödinger equation, 1136–41
Freeway collapse, 418
Freezing, of a lake, 501 (*see also* Phase, changes of)
Freezing point, 496 *fn*
French Academy of Sciences, 7
Frequency, 125, 269, 401, 404, 430
angular, 404
of audible sound, 461–62, 467
beat, 475–76
of circular motion, 125
collision, 536 *pr*
cyclotron, 1292
fundamental, 447, 468–73
infrasonic, 462
of light, 909, 942, 943–44
natural, 404, 417, 446
resonant, 417, 418, 446–47
of rotation, 269
ultrasonic, 462, 482
of vibration, 401, 414, 446
of wave, 430, 436, 442
Frequency modulation (FM), 918, 919 *fn*
Fresnel, A., 1026
Fresnel diffraction, 1026 *fn*
Frets (guitar), 484 *pr*
Friction, 87, 116–23, 196, 207–10
coefficients of, 117–18
force of, 87, 88, 117–23
helping us to walk, 92
kinetic, 117 *ff*
rolling, 117, 144 *pr*, 289–90
static, 118, 286
Frictional damping, of waves, 437, 442
Fringe shift, 1060–61
Fringes, interference, 999–1001, 1060–61, 1179
Frisch, Otto, 1261
Frog muscle, 718
Fulcrum, 334, 340
Full moon, 161, 169 *pr*
Full-scale current sensitivity, 767
Full-wave rectifier, 1214, 1220 *pr*
Full width at half maximum (FWHM), 419
Fundamental constants: inside front cover
Fundamental frequency, 446–47, 468–72
Fundamental particle (*defn*), 1290, 1304–5, 1310, 1313
Furnaces, hot-air, 563
Fuse, 729–30
Fusion, nuclear, 599, 1266–71
in stars, 1267–69, 1329–31
Fusion bomb, 1269
Fusion reactor, 1269–71
FWHM, 419

g:
acceleration in, 41
measuring, 412
G, 147
g-factor, 1173
Gain, 490 *pr*
Galaxies, 1323–26, 1338–41, 1348, 1349, 1351–54
black hole at center of, 167 *pr*, 170 *pr*, 1326, 1337
clusters of, 1325, 1349, 1353
distant, 1333
mass of, 1324
Milky Way, 170 *pr*, 171 *pr*
origin of, 1349, 1353
redshift of, 1339–40
superclusters of, 1325–26
Galilean telescope, 980, 982
Galilean transformation, 1072–74
Galilean–Newtonian relativity, 1056–58, 1072–74
Galileo, 2, 20, 37–38, 54, 65–66, 156, 373, 374, 412, 495, 912, 928, 980, 1056, 1072–73, 1323
Galvani, Luigi, 718
Galvanometer, 767–78, 795, 867 *pr*
Gamma camera, 1278
Gamma decay, 1234, 1239–40
Gamma particle, 1234, 1239–40, 1271, 1296
Gamma ray, 1234, 1239–40, 1271, 1296
Gamow, George, 1343
Gas constant, 506
Gas-discharge tube, 1111
Gas lasers, 1177
Gas laws, 502–8
Gas tank:
overflow, 500
in Sun, 500
Gases, 365–89, 366, 502–31
adiabatic expansion of, 559–60
definition, 494
Fermi, 1204
ideal, 503–8, 516 *ff*
kinetic theory of, 516–31
molar specific heats for, 556–58
real, 522–24
vacuum in, 534 *pr*
vapors vs., 523
Gate, 681, 1218
Gate terminal, 857
Gauge bosons, 1290, 1304–5, 1309–12
Gauge pressure, 371
Gauge theory, 1313
Gauges, pressure, 372–73
Gauss, K. F., 643
Gauss (G) (unit), 787
Gauss's law, 643–52
for magnetism, 902, 903
Gauss's theorem, A-16
Gay-Lussac, Joseph, 503
Gay-Lussac's law, 503, 532 *pr*
Gazelle, 291 *pr*
Geiger counter, 690 *pr*, 1248
Gell-Mann, M., 1308
Gene, 631
General motion, 244, 283–90, 310–12
General theory of relativity, 163–64, 1322–23, 1334–36
General wave (non-sinusoidal), 439
Generator, 598
ac, 846–47
counter torque in, 849
dc, 847, 848
electric, 846–48
emf of, 846–50
Van de Graaff, 691 *pr*
Genetic code, 1194
Genetic information, 632
Geocentrism, 155–56, 159
Geodesic, 1336
Geological dating, 1247–48
Geology, 150, 378, 412, 415, 435, 452, 457, 599
Geometric optics, 927–85
Geometry, A-2
Geophysics, 435
Geostationary satellite, 153
Geothermal energy, 599
Germanium, 1210
Germer, L. H., 1107
GFCI, 766, 859
Gianotti, Fabiola, 1314
Giants, red, 1326, 1328, 1329–30
Giraffes, 396 *pr*
Glaser, D. A., 1249
Glashow, S., 1313
Glasses, eye, 976–78
Global positioning satellite (GPS), 18 *pr*, 169 *pr*, 924 *pr*, 1069
Global System for Mobile Communication (GSM), 920
Global warming, 598
and sea level rise, 511 *pr*

Glueballs, 1312 *fn*
Gluino, 1317
Gluons, 1290, 1298, 1304, 1305, 1310, 1311–14
Goblet, shattering, 417
Golden Gate bridge, 349, 363 *pr*
Golf putt, 52 *pr*
Gondola, for skiers, 187 *pr*
Gophers, 384
Gothic cathedrals, 350
GPS, 18 *pr*, 169 *pr*, 924 *pr*, 1069
Gradient, 386, 530, 561
  of electric potential, 672–73
Gradient operator (del), 673 *fn*
Gram (g) (unit), 8, 89
Grand unified era, 1345
Grand unified theories (GUT), 162, 1314–16
Grandfather clock, 421 *pr*
Graphical analysis, 30
  of linear motion, 30
Graphite, 1263
Grating, 1037–43
Gravimeters, 150
Gravitation, universal law of, 145–49, 210–12, 614, 1334
Gravitational assist, 216–17, 224 *pr*, 263 *pr*
Gravitational collapse, 1337
Gravitational constant (G), 147
Gravitational field, 162–64, 627, 1334–38
Gravitational force, 86, 94–96, 146–49 *ff*, 161–62, 1241, 1298, 1316, 1322, 1331, 1334–38, 1352
  due to spherical mass distribution, 148–49, A-13–A-15
Gravitational lensing, 1335–36
Gravitational mass, 163–64, 1335
Gravitational potential, 663, 671
Gravitational potential energy, 197–98, 201–14, 216–17
Gravitational redshift, 1339
Gravitational slingshot, 216–17, 263 *pr*
Gravitational waves, 1354
Gravitino, 1317
Graviton, 1298, 1317
Gravity, 37–43, 94, 145 *ff*, 195–98, 201–14, 216–17, 1298, 1316, 1322, 1331, 1352
  acceleration of, 37–43, 89 *fn*, 94, 149–51
  center of, 246, 248
  and curvature of space, 1334–38
  on Earth, 38 *ff*, 125, 146–52
  effect on light, 1335–36, 1337
  force of, 86, 94–96, 145–64, 1298, 1316, 1317, 1322, 1331, 1334–38, 1350, 1352
  in free-body diagrams, 97
  free fall under, 37–43, 154–55
  specific, 367
Gravity anomalies, 150
Gray (Gy) (unit), 1273
Great Britain, Roman houses in, 565
Greek alphabet: inside front cover
Greek temple, 344
Greenland's ice sheet, 19 *pr*
Grid electrode, 679
Grimaldi, F., 996, 1001
Ground fault, 859
Ground fault circuit interrupter (GFCI), 766, 859
Ground state, of atom, 1115
Ground wire, 764–65, 765–66
Grounding, electrical, 612, 721
Groves, Leslie, 1266
GSM, 920
Guanine (G), 631
Guest compounds, 1217
Guitar, 469, 484 *pr*, 487 *pr*
Guitar string, 446
GUT, 162, 1314–16
Guth, Alan, 1346, 1348
Gyration, radius of, 296 *pr*
Gyromagnetic ratio, 1173
Gyroscope, 317–18, 327 *pr*

h-bar ($\hbar$), 1133, 1162
Hadron era, 1346–47
Hadrons, 1304–5, 1308–13, 1346–47
Hahn, Otto, 1261
Hair dryer, 732
Hale-Bopp comet, 169 *pr*
Hale telescope, 982
Half-life, 1242–44
Half-wave rectification, 1214
Hall, E. H., 799
Hall effect, Hall emf, Hall field, Hall probe, 799–800, 1212
Hall voltage, 1212
Halley's comet, 168 *pr*
Halogens, 717, 1168
Hammer thrower, 296 *pr*
Hang glider, 112 *pr*
Hard drive, 18 *pr*, 293 *pr*
Hardening of arteries, 384, 386
Harmonic motion:
  damped, 414–17
  forced, 418
  simple, 402–10
Harmonic oscillator, 402–10, 1149, 1156 *pr*
Harmonic wave, 437–39
Harmonics, 447, 468–71
Hazards of electricity, 764–66
Headlights, 663, 727, 747
Headphones, use as detector, 484 *pr*
Headsets, 796
Hearing, 460, 464–67
  beat frequency, 475
  the ocean, 485 *pr*
  range of, 464
  threshold of, 464, 467
  (*see also* Sound)
Heart, 380, 386, 398 *pr*, 482
  defibrillator, 703, 712 *pr*, 764, 780 *pr*
  monitoring of, 682–83
  pacemaker, 764, 779 *pr*, 871
  as pump, 388–89
  valves, 388
Heart disease, 386
Heartbeats, number of, 12
Heat, 207, 539–67, 576–99
  calorimetry, 542–48
  compared to work, 549
  conduction, 561–62, 573 *pr*
  convection, 563
  defined, 540
  distinguished from internal energy and temperature, 540
  as energy transfer, 538–40
  in first law of thermodynamics, 549–51
  of fusion, 545
  heat loss through windows, 562, 568 *pr*
  latent, 545–49
  mechanical equivalent of, 540
  radiation, 563–66
  for thermodynamic problem-solving, 600
  of vaporization, 545
Heat capacity, 570 *pr* (*see also* Specific heat)
Heat death, 595
Heat engine, 577–80, 1264
  Carnot, 580–84
  efficiency of, 579–80
  internal combustion, 578–80
  operating temperatures, 578
  steam, 578
  and thermal pollution, 598
  why ΔT needed to drive, 579
Heat of fusion, 545
Heat of vaporization, 545
Heat pump, 584, 586–87, 603 *pr*
Heat reservoir (*defn*), 551, 580
Heat sinks, 568 *pr*
Heat transfer, 560–67
  conduction, 561–62
  convection, 563
  radiation, 563–67
Heating duct, 380
Heating element, 717, 726, 732
Heating systems, 563
  hot-water, 382
Heavily damped system, 415
Heavy elements, 1330–31
Heavy water, 1263
Hectare (*defn*), 10
Heisenberg, W., 1095, 1128, 1129
Heisenberg uncertainty principle, 1131–35, 1148, 1187
  and particle resonance, 1307
  and tunneling, 1236
Helical path, 329 *pr*
Helicopter, 54, 72, 83 *pr*
Heliocentrism, 155–56, 159
Helium, 1166, 1230–31, 1234, 1258, 1267
  balloons, 505
  I and II, 524
  primordial production of, 1347, 1348 *fn*
  and stellar evolution, 1329–31
Helium–neon laser, 1176–77
Helix, double, 632
Helmholtz coils, 836 *pr*
Henry, Joseph, 838
Henry (H) (unit), 870
Hertz, Heinrich, 909
Hertz (Hz) (unit), 269, 401
Hertzsprung–Russell diagram, 1328, 1333
Higgs, Peter, 1314
Higgs boson, 1290, 1304, 1314
Higgs field, 1314
High-altitude climbers, 571 *pr*
High definition (HD), 680
High-energy accelerators, 1290–96
High-energy physics, 1290–1317
High-frequency trading, 945 *fn*
High jump, 220 *pr*
High mass stars, 1330–31
High-pass filter, 884, 897 *pr*
High-voltage sparking, 893 *pr*
High wire, 109 *pr*
Highest occupied molecular orbitals (HOMO), 1217
Highway:
  banked and unbanked curves, 130–32
  buckling, 501
Hikers, 175, 218 *pr*, 220 *pr*, 225 *pr*
Hinge, 337
Hiroshima, 1266
H_2O molecule, 299 *pr*
Hockey, 98
Hole transport layer (HTL), 1217
Holes:
  expansion or contraction, 499
  in OLED functioning, 1217
  in semiconductors, 1209–12, 1218
Hologram and holography, 1178–79
HOMO, 1217
Homogeneous (universe), 1341
Hooke, Robert, 342, 1005 *fn*
Hooke's law, 179, 199, 342, 400
Horizon, 18 *pr*, 1344
  event, 1338
Horizontal (*defn*), 94 *fn*
Horizontal range, 70–72

Horsepower (hp) (unit), 213
Host material, 1217
Hot-air balloons, 492, 515 *pr*
Hot-air furnaces, 563
Hot-water heating system, 382
Hot-water radiator, 563
Hot wire, 764–65, 765–66
Hour (hr) (unit), 7
Household circuits, 729–30
H–R diagram, 1328, 1333
HST (*see* Hubble Space Telescope)
HTL, 1217
Hubble, Edwin, 1086, 1325, 1339
Hubble age, 1341
Hubble eXtreme Deep Field (XDF) photograph, 1322
Hubble parameter, 1339, 1341
Hubble Space Telescope (HST), 167 *pr*, 1034–35, 1336, 1340
Hubble Ultra Deep Field, 1340
Hubble's constant, 1339
Hubble's law, 1339, 1341, 1352
Human beings:
  balance for, 342
  circulatory system, 380
  ear, 460, 464–67
  response, 467
  hearing, 460, 464–67
  as highly ordered organisms, 594
  leg, as physical pendulum, 424 *pr*
Human centrifuge, 294 *pr*
Humidity, 524–26
  and comfort, 526
  relative, 524–25
Hurricane, 321, 324 *pr*, 394 *pr*
Huygens, C., 996
Huygens' principle, 996–98
Hydraulic brakes, 372
Hydraulic lift, 372
Hydraulic press, 393 *pr*
Hydrodynamics, 378–89
Hydroelectric power plant, 254 *pr*, 599
Hydrogen atom:
  Bohr theory of, 1113–19
  magnetic moment of, 794
  populations in, 1185 *pr*
  quantum mechanics of, 1159–65
  spectrum of, 1041, 1111–12
Hydrogen bomb, 1266, 1269
Hydrogen bond, 632, 1193, 1194
Hydrogen isotopes, 1227
Hydrogen-like atoms, 1113 *fn*, 1118, 1120
Hydrogen molecule, 1187–90, 1196, 1200
Hydrometer, 377
Hydrostatic paradox, 365, 390 *pr*, 391 *pr*
Hyperopia, 976
Hypothesis, 899
Hysteresis, 827–28
  loop, 827

Icarus, asteroid, 168 *pr*
Ice:
  life under, 500–501
  making, 586
Ice skater, 302, 304, 324 *pr*, 326 *pr*, 330 *pr*
Ideal gas, 1207
  defined, 504, 516
  internal energy of, 541
  kinetic theory of, 516–31, 1207
  pressure in, 518
Ideal gas law, 503–7, 522
  in terms of molecules, 506–7
Ideal gas temperature scale, 507–8, 582
Idealization, 551
Identical (electrons), 1167
Ignition:
  automobile, 663, 852
  fusion, 1271
ILC, 1295
Illuminance, 1016
Image:
  actual-sized, 952 *pr*
  CT, 1046
  formed by lens, 959 *ff*
  formed by plane mirror, 927–31
  formed by spherical mirror, 931–36, 982
  MRI, 1230, 1281–82
  NMR, 1230, 1279–82
  PET and SPECT, 1278–79
  real, 929, 933, 961
  seeing, 936, 937, 961
  as tiny diffraction pattern, 1033–35
  tomographic image reconstruction, 1046–47
  ultrasound, 482–83
  virtual, 929, 962
  X-ray, 1046
Image artifact, 971
Image distance, 929, 934, 946, 962–63
Imaging, medical, 482–83, 1278–82
Imbalance, rotational, 314–15
Impedance, 884, 886–88
Impedance matching, 888
Imperfect spring, 191 *pr*
Impulse, 234–35
Impulsive forces, 235
In-phase waves, 445, 999, 1005–9, 1038
Inanimate object, force exerted by, 92
Inch (in.) (unit), 7
Incidence, angle of, 444, 449, 927, 939
Incident ray, 444
Incident waves, 444, 449
Inclines, motion on, 103–4, 122
Incoherent source of light, 1002
Indefinite integrals, A-6–A-7
Indeterminacy principle, 1133 (*see also* Uncertainty principle)
Index of refraction, 939
  dependence on wavelength (dispersion), 942
  in Snell's law, 940
Indium–tin oxide (ITO), 1217
Induced current, 838–59, 869 *ff*
Induced electric charge, 612–13, 706
Induced emf, 838–46, 873
  counter, 848–50
  in electric generator, 846–48
  in transformer, 851–54
Inductance, 870–74
  in ac circuits, 875–89
  of coaxial cable, 874
  mutual, 870–71
  self-, 872–74
Induction:
  charging by, 612–13
  electromagnetic, 838 *ff*
  Faraday's law of, 840–41, 854–55, 903
  wireless transmission of electricity with, 854
Induction stove, 842
Inductive reactance, 882–83
Inductor, 872, 1219
  in circuits, 875–89
  energy stored in, 874–75
  reactance of, 882–83
Inelastic collisions, 236, 239–41
Inelastic scattering, 1260
Inertia, 87
  mass as measure of, 88
  moment of, 274–76
Inertial confinement, 1270, 1271
Inertial forces, 319–21
Inertial mass, 163, 1335
Inertial reference frame, 87–88, 90, 144 *pr*, 319, 1056 *ff*
  Earth as, 87, 151–52
  Earth is not, 144 *pr*
  equivalence of all, 1056–57, 1061–62
  transformations between, 1073–76
Infinitely deep square well potential, 1142–46
Inflationary scenario, 1346, 1348–50
Information, carried by signal voltage, 676
Information storage, 856–58
  magnetic storage, 856–57
  semiconductor memory, 857–58
Infrared (IR) radiation, 910, 941, 1041
Infrasonic waves, 462
Initial conditions, 403
Insertion, muscle, 339
Instantaneous acceleration, 29–30, 63–64
  angular acceleration, 267
  vector for, 63–64
Instantaneous axis, 284
Instantaneous speed, 24
Instantaneous velocity, 24–26, 63
  angular velocity, 266
  vector for, 63
Insulation:
  by clothes, 562
  thermal, 562
Insulators:
  electrical, 611, 724, 1208–10
  thermal, 561, 1208–10
Integral(s), 43–44, 178–79, A-6–A-7, A-16, A-17, inside back cover
  definite, A-7, A-9
  Fourier, 442
  indefinite, A-6–A-7
  line, 178
  as a sum, 178, 179, A-8–A-11
  surface, A-17
  volume, A-16
Integrated circuits, 18 *pr*, 1219
Integration by parts, 1146, 1164, A-6, A-7
Intensity, 464
  in interference and diffraction patterns, 1002–4, 1028–32
  of light, 941, 1016, 1130
  of Poynting vector, 914
  related to amplitude, 436, 466
  of sound, 464–67
  of wave, 436
Intensity levels, 464–67
Interference, 444, 473–76, 678, 995, 998–1009, 1031–33, 1187
  constructive, 444–45, 473–74, 999, 1009, 1187
  destructive, 444–46, 473–74, 999, 1009, 1187
  as distinguished from diffraction, 1033–34
  of electrons, 1130–31, 1187
  of light waves, 998–1009, 1032–33
  loudspeakers', 474
  of sound waves, 473–76
  in space, 473–74
  by thin films, 1004–9
  in time, 474–76
  of water waves, 445
  of waves on a string, 444

Interference factor, 1032
Interference fringes, 999–1001, 1060–61, 1179
Interference microscope, 1048
Interference pattern:
  double-slit, 998–1004, 1130–31
  including diffraction, 1031–33
  multiple slit, 1037–41
Interferometers, 1010, 1058–61
Interlaced sweeps, 680 *fn*
Intermodulation distortion, 441 *fn*
Internal combustion engine, 512 *pr*, 579, 580
Internal conversion, 1240
Internal energy, 207, 540–41, 549–59
  from atomic point of view, 207
  distinguished from heat and temperature, 540
  and free expansion, 555
  of ideal gas, 541
Internal force, 231, 249
Internal reflection, total, 456 *pr*, 943–45
Internal resistance, 748–49
International Linear Collider (ILC), 1295
International Thermonuclear Experimental Reactor (ITER), 1256, 1271
Interpolation, 4, A-3
Interstellar dust, 1325
Intravenous transfusions, 396 *pr*
Intrinsic luminosity, 1326, 1333
Intrinsic semiconductor, 1209, 1211
Invariant energy, 1084
Inverse square law, 147 *ff*, 211, 436, 465, 613–14, 652, 821, 904, 1016, 1181, 1280-81, 1293, 1338-39, A-15
Inverted population, 1176–77
Ion (*defn*), 611
Ionic bonds, 1188, 1190, 1202, 1203
Ionic cohesive energy, 1202
Ionization energy, 1116, 1118
Ionizing radiation (*defn*), 1271
IR radiation, 910, 941, 1041
Iris, 975
Irreversible process, 580, 582, 589, 591
ISO number, 972
Isobaric processes, 552–54
Isochoric processes, 552–54
Isolated system, 231, 542–43, 547
Isomer, 1240
Isotherm, 551
Isothermal processes, 551–53
Isotopes, 800, 1227–28, 1233–34
  mean life of, 1243, 1255 *pr*
  in medicine, 1277–78
  table of, A-18–A-21
Isotropic (universe), 1341
Isovolumetric process, 552–54
ITER, 1256, 1271
Iterative techniques, 1047, 1279
ITO, 1217

J (total angular momentum), 1173
J/ψ particle, 1135, 1310
Jars:
  into a campfire, 503
  and lids, 499
Jeans, J., 1097
JET, 1271
Jet engine, 607 *pr*
Jet plane:
  launching, 189 *pr*
  noise, 466
Jeweler's loupe, 980
Joint European Torus (JET), 1271
Joints, 339–41, 347–49, 358 *pr*
  method of, 348–49
Joule, James Prescott, 539
Joule (J) (unit), 173, 182, 273 *fn*, 674, 675, 727
  relation to calorie, 539–40
Joule's experiment, 539–40
Joyce, James, 1308 *fn*
jpeg, 678
Jump start, 758, 759
Jumper cables, 758–59
Junction diode, 1218
Junction rule, Kirchhoff's, 754
Junction transistor, 1218
Jupiter, 156, 158, 167 *pr*, *168 pr*, 169 *pr*, 216–17, 224 *pr*
  moons of, 158, 167 *pr*, 168 *pr*, 169 *pr*, 912, 980
  period and distance from Sun, 330 *pr*

K2, 9
K-capture, 1239
K lines, 1169–70
K particle (kaon), 1305, 1307
Kant, Immanuel, 1325
Kaon, 1305, 1307
Karate blow, 235
Keel, 383
Kelvin-Planck statement of the second law of thermodynamics, 583
Kelvin scale of temperature, 503, 597–98
Kelvin (unit), 503
Kepler, Johannes, 156, 159, 980 *fn*
Keplerian telescope, 980–81, 980 *fn*
Kepler's laws, 155–60
  second law derived, 316
Keyboard, computer, 695
Kilby, Jack, 1219
Kilo- (prefix), 8
Kilocalorie (unit), 539, 540
Kilogram (kg) (unit), 8, 88, 89
Kilometer (km) (unit), 8
Kilowatt-hour (kWh) (unit), 728
Kinematic equations, 31
  for constant acceleration, 31
Kinematics, 20–45, 54–76, 264–70
  definition, 20
  for rotational motion, 264–70
  for translational motion, 20–45, 54–76
  for uniform circular motion, 123–26
  vector kinematics, 62–75
Kinetic energy, 172, 181–85, 200, 281–85, 1081–82
  of CM, 284–86
  in collisions, 235–37, 239–41
  conservation of, 244
  and electric potential energy, 661
  of gas atoms and molecules, 518–19, 541, 557–58
  molecular, relation to temperature, 518–19, 541, 557–58
  near absolute zero, 520
  of photon, 1101
  relativistic, 1080–84
  rotational, 281–85
  translational, 172, 181–85, 200–213, 215
  of variables related to temperature, 518
Kinetic friction, 117 *ff*
  coefficient of, 117
Kinetic theory of gases, 493–94, 506–7, 516–32
  basic postulates of the, 517
  boiling, 525
  diffusion, 530–31
  evaporation, 524
  ideal gas, 516–22
  kinetic energy near absolute zero, 520
  mean free path, 528–29
  molecular speeds, distribution of, 520–22
  postulates of, 517
  of real gases, 522–25
  temperature of boiling water and altitude, 526–27
  van der Waals equation of state, 527–28
Kinetic theory of latent heats, 549
Kirchhoff, G. R., 754
Kirchhoff's rules, 754–57, 902 *fn*
  junction rule, 754
  loop rule, 754
Knot (unit), 7, 169 *pr*
Known quantities, 33, 34, 67, 68

L1 point, 160
Ladder, 338, 358 *pr*, 363 *pr*
Lag time, 876
Lagrange, Joseph-Louis, 160
Lagrange Point, 160
Lake:
  freezing of, 501
  level, rock overboard, 378
Lambda (particle), 1305, 1307
Laminar flow, 378–81
Land, Edwin, 1011
Lanthanides, 1168
Large Binocular Telescope, 982
Large Hadron Collider (LHC), 1289, 1293, 1294, 1314, 1317
Laser beam, 293 *pr*
Laser-Interferometer Gravitational-wave Observatory (LIGO), 1354
Lasers, 1175–78
  chemical, 1177
  gas, 1177
  helium–neon, 1176–77
  Neodymium-YAG, 923 *pr*
  pn diode lasers, 1216
  pulsed, 923 *pr*
  surgery, 1178
Lasers, pulsed, 923 *pr*, 1117
Latent heat, 545–49
  kinetic theory of, 549
Lateral magnification, 935, 936–38, 963, 983
Lattice structure, 494, 1202, 1211
Laue, Max von, 1044
Law (*defn*), 3
  causal, 159
  (*see specific laws by name*)
Law of cosines, 188 *pr*, A-4
Law of definite proportions, 493 *fn*
Law of inertia, 87
Law of reflection, 444
Law of refraction, 449
Law of sines, A-4
Lawn mower, 17 *pr*
Lawn sprinkler, 398 *pr*
Lawrence, E. O., 1291
Lawson, J. D., 1271
Lawson criterion, 1271
LC circuit, 877–81
LC oscillation, 877–81
LCD, 679, 680, 918–19, 970 *fn*, 1014–15
Leakage current, 766
"Leaning ladder problems," 363 *pr*
LED, 680, 1215–16
Legs:
  breaking, in jump, 364 *pr*
  force on, after jump, 257 *pr*
Length:
  focal, 932, 937, 959–60, 967, 969–70, 975, 976
  Planck, 15, 1345
  proper, 1070
  relativity of, 1070–76
  standard of, 1010
  standard unit of, 7, 8
Length contraction, 1070–71, 1075

Lennard-Jones potential, 224 *pr*, 226 *pr*
Lens, 958–85
  achromatic, 985
  axis of, 959
  coating of, 1008–9
  color-corrected, 985
  combination of, 966–67
  compound, 985
  contact, 978
  converging, 958 *ff*
  corrective, 976–78
  cylindrical, 977
  diverging, 959 *ff*
  of eye, 976
  eyeglass, 976–78
  eyepiece, 981
  focal length of, 959, 960, 967, 968
  magnetic, 1109
  magnification of, 964
  negative, 963
  normal, 975
  objective, 981–83
  ocular, 983
  positive, 963
  power of (diopters), 960
  resolution, 974
  resolution of, 974, 1033–37
  spherical, 948
  telephoto, 975
  thin (*defn*), 959
  wide-angle, 975, 985
  zoom, 975
Lens aberrations, 984–85, 1033, 1036
Lens elements, 985
Lensmaker's equation, 968–70
Lenz's law, 841–44
Leonardo da Vinci, 118
Lepton era, 1345, 1347
Lepton number, 1301–2, 1305–6, 1308, 1315
Leptons, 1290, 1296, 1301–2, 1304, 1305, 1308–10, 1312–14, 1317, 1346
Level:
  acceptor, 1212
  donor, 1212
  energy (*see* Energy levels)
  Fermi, 1204–7
  loudness, 467
  sound, 464–67
Level horizontal range, 70–72
Level range formula, 70–71
Lever, 183, 187 *pr*, 334
  length of, 272
Lever arm, 272
LHC, 1289, 1293, 1294, 1314, 1317
Lids and jars, 499, 503
Life, under ice, 500–501
Lifetime, 1305 (*see also* Mean life)
Lift, 383
Light, 909–10, 912–13, 926–1048
  coherent sources of, 1001
  color of, and wavelength, 941–43, 998, 1001, 1007
  dispersion of, 942
  Doppler shift for, 480, 1085–86, 1338
  as electromagnetic wave, 909–13
  frequencies of, 909, 942, 943–44
  gravitational deflection of, 1335–36, 1337
  incoherent sources of, 1002
  infrared (IR), 910, 941, 1041, 1052 *pr*
  intensity of, 941, 1016, 1130
  monochromatic (*defn*), 998
  as particles, 997, 1098–1106
  photon (particle) theory of, 1098–1106
  polarization of, 1010–14
  ray model of, 927 *ff*, 959 *ff*
  scattered by atmosphere, 1015–16
  spectrometer, 1040–41
  speed of, 7, 908–9, 912–13, 939, 997, 1057–58, 1062, 1081
  total internal reflection of, 1151
  ultraviolet (UV), 910, 941
  unpolarized, 1011
  velocity of, 908–9, 912–13, 939, 997, 1057–58, 1062, 1081
  visible, 910, 941–43
  wave, tunneling of, 1150–51
  wave-particle duality of, 1105–6
  wave theory of, 995–1048
  wavelengths of, 910, 941–43, 998, 1001, 1007
  white, 941–42
  (*see also* Diffraction; Intensity; Interference; Reflection; Refraction)
Light-emitting diode (LED), 680, 1215–16
Light-gathering power, 982
Light meter (photographic), 1101
Light pipe, 945
Light rays, 927 *ff*, 959 *ff*
Light sources, 1293
Light-year (ly) (unit), 17 *pr*, 1323
Lightbulb, 717, 719, 722, 726–27, 778 *pr*, 853, 1016–17, 1100
  burn out, 727
  fluorescent, 1175
Lightning, 728, 837 *pr*, 925 *pr*
  distance from strike, 461
Lightning rod, 666
Lightning storms, 716
LIGO, 1354
Limit of large *r*, 623
Linac, 1294
Linde, Andreí, 1346
Line integral, 178, 664, 694, 815, 903
Line spectrum, 1040–41, 1111 *ff*, 1128
Line voltage, 731
Linear acceleration, angular acceleration and, 268–69
Linear accelerator, 1294
Linear expansion (thermal), 498–99
  coefficient of, 498
Linear momentum, 227–53
Linear motion, 20–45
Linear polarization, 1010
Linear velocity, angular velocity and, 267–69
Linear waves, 435
Lines of force, 626, 783
Liquefaction, 502–4, 516, 523–26
Liquid crystal, 366, 524
Liquid crystal display (LCD), 679, 680, 970 *fn*, 1014–15
Liquid-drop model, 691 *pr*, 1261–62
Liquid-in-glass thermometer, 495, 496
Liquid nitrogen, 603 *pr*
Liquid scintillators, 1248
Liquids (*defn)*, 365–89, 494 (*see also* Phase, changes of)
Living organisms, diffusion in, 531–32
Lloyd's mirror, 1023 *pr*
Locomotive, 576
Loft, of down garments, 568 *pr*
Log table, A-3
Logarithms, 464, 467, A-2, A-3, inside back cover
Long jump, 80 *pr*, 81 *pr*, 225 *pr*
Long-range force, 1233, 1334
Longitudinal waves, 431 *ff*, 461–63
  and earthquakes, 435
  mathematical representation of, 462–63
  speed of, 433–35
  (*see also* Sound waves)
Lookback time, 1326, 1344
Loop rule, Kirchhoff's, 754
Lorentz, H. A., 1061, 1128
Lorentz equation, 792
Lorentz transformation, 1075–76
Los Alamos laboratory, 1266
Loudness, 461, 464, 467
  apparent, 465
  control, 467
  ear's response and, 467
  level, 467
  (*see* Intensity)
Loudspeakers, 406–7, 473, 474, 795–96, 884
  concert time delay, 490 *pr*
  output, 465
Loupe, jeweler's, 980
Low back pain, 341
Low mass stars, 1330
Low-pass filter, 884, 897 *pr*
Lowest unoccupied molecular orbitals (LUMO), 1217
LR circuit, 875–77
LRC circuit, 880–81, 885–87
LRO, 169 *pr*
Lucretius, 1352 *fn*
Lumbar vertebrae, 340
Lumen (lm) (unit), 1016
Luminosity (stars and galaxies), 1326, 1333
Luminous flux, 1016
Luminous intensity, 1016
LUMO, 1217
Lunar eclipse, 161
Lunar module, 222 *pr*
Lunar Reconnaissance Orbiter (LRO), 169 *pr*
Lungs, 388, 389
  alveoli in, 19 *pr*
Lyman series, 1112, 1116, 1117, 1169

m-RNA, 1194–95
M-theory, 1317
Mach, E., 480 *fn*
Mach number, 480
Macroscopic description of a system, 494
Macroscopic properties, 494
Macrostate of system, 595–96
Madelung constant, 1202
Magellanic clouds, 1325 *fn*
Magnet, 782–84, 824–26
  domains of, 825
  electro-, 826
  permanent, 825
  superconducting, 826
Magnetic bottle, 1270
Magnetic circuit breakers, 826
Magnetic confinement, 1270
Magnetic damping, 861 *pr*
Magnetic declination, 784
Magnetic deflection coils, 679
Magnetic dipoles and magnetic dipole moments, 793–94, 823, 1171–73
Magnetic domains, 825
Magnetic field, 782–93, 811–29
  of circular loop, 822–23
  definition of, 783, 787, 789
  determination of, 788, 816–23
  direction of, 783, 785, 791
  due to single moving charge, 824
  of Earth, 784
  energy stored in, 874–75
  hysteresis, 827–28

Magnetic field (*continued*)
induces emf, 839–54
motion of charged particle in, 788–93
produced by changing electric field, 899–902
produced by electric current, 785, 819–24 (*see also* Ampère's law)
produces electric field and current, 854–56
of solenoid, 819–21
sources of, 811–29
of straight wire, 787–88, 812–13
of toroid, 820
uniform, 784
Magnetic field lines, 783
Magnetic flux, 840 *ff*, 854–56, 902, 906
Magnetic force, 782, 786–94
on electric current, 786–89, 793–94
on moving electric charges, 788–93
Magnetic induction, 785 (*see also* Magnetic field)
Magnetic lens, 1109
Magnetic moment, 793–94, 823
Magnetic monopole, 783, 1350
Magnetic permeability, 812, 827
Magnetic poles, 782–84
of Earth, 784
Magnetic quantum number, 1160–61, 1171
Magnetic resonance imaging (MRI), 1230, 1281–82
Magnetic storage, of information, 856–57
Magnetic susceptibility (*defn*), 828
Magnetic tape and disks, 856
Magnetism, 782–875
Magnetization vector, 829
Magneto-optical atom trap, 837 *pr*
Magnetoresistive RAM (MRAM), 858
Magnification:
angular, 935, 936-38, 979, 981, 983
lateral, 935, 936-38, 963, 983
of electron microscope, 1109
of lens, 964
of lens combination, 966–67
of magnifying glass, 979–80
of microscope, 983–84, 1036, 1037, 1109
of mirror, 935
of telescope, 981, 1035
of thin lens, 963, 979, 980
sign conventions for, 935, 963
useful, 1037, 1109
Magnifier, simple, 958, 979–80
Magnifying glass, 958, 979–80
Magnifying mirror, 931, 937
Magnifying power, 979
total, 981
(*see also* Magnification)
Magnitude:
apparent, of star, 1360 *pr*
of cross product, 307
of electric force, 613
of vector, 22, 55, 61
and nonuniform circular motion, 133
and uniform circular motion, 123, 124
Main sequence (stars), 1328–30
Majorana, Ettore, 1303 *fn*
Majorana neutrinos, 1314
Majorana particles, 1303
Mammals, 291 *pr*
air circulation in burrows, 384
Manhattan Project, 1266
Manometer, 372–73, 497
Marching soldiers, 417
Marconi, Guglielmo, 917
Mars, 155, 157
atmosphere, 513 *pr*
Mass, 88–91, 163–64
and acceleration, 103, 121
of air in a room, 505
atomic, 493, 1136–38
center of, 244–48
critical, 1263–66
of electron, 799, 1229
of Galaxy, 1324
gravitational vs. inertial, 163, 1335
and luminosity, 1327
as measure of inertia, 88
molecular, 493, 504
of neutrinos, 1303–4
nuclear, 1228–30
of photon, 1101
precise definition of, 90–91
reduced, 1197
in relativity theory, 1079
rest, 1079
and speed, 37
of a spring, 191 *pr*
standard of, 8
units of, 8, 89
variable, systems of, 251–53
weight vs., 88, 94
Mass defect, 1231
Mass, effective, 191 *pr*, 427 *pr*
Mass excess (*defn*), 1231, 1254 *pr*
Mass number, 1227
Mass spectrometer (spectrograph), 800–801
Mass–energy, distribution in universe, 1350–53
Mass–energy transformation, 1080–84
Mathematical expansions, A-1
Mathematical signs and symbols: inside front cover
Mather, John, 1343
Matrix, 681
Matter:
anti-, 1299, 1316, 1318 *pr*
dark, 1290, 1317, 1348, 1351, 1352
passage of radiation through, 1271–72
states of, 366, 494
wave nature of, 1106–8, 1119–20
Matter-dominated universe, 1347, 1348
Matter waves, 1106–8, 1119–20, 1130 *ff*
Matter–antimatter problem, 1316
Maximum range, 70
Maxwell, James Clerk, 520–21, 899, 903, 905–6, 909, 1057–58
Maxwell distribution of molecular speeds, 520–22, 596, 1270
Maxwell's equations, 899, 903, 905–9, 1006 *fn*, 1055, 1057–58, 1062, 1075
differential form of, A-16–A-17
in free space, A-17
Maxwell's preferred reference frame, 1057–58
Mean free path, 528–29
Mean life, 1243, 1255 *pr*, 1305
of proton, 1315
Measurements, 3–6
of astronomical distances, 1323, 1328, 1332–34
of charge on electron, 798–99
of e/m, 797–98
electromagnetic, of blood flow, 491 *pr*, 845
flow speed, 384
force, 86
precision of, 6, 1131–33
pressure, 372–74
quantity affected by, 767–70
of radiation, 1272–76
of speed of light, 912–13
uncertainty in, 3–6, 1131–35
units of, 6–9
Mechanic, raising of engine by, 362 *pr*
Mechanical advantage, 102, 334, 372
Mechanical (analog) watch, 424 *pr*
Mechanical energy, 200–206
conservation of, 201–6
of simple harmonic oscillator, 408
Mechanical equivalent of heat, 540
Mechanical oscillations, 399
Mechanical waves, 428–51
Mechanics (*defn*), 20 (*see also* Motion)
Mediate, of forces, 1297
Medical applications:
of capacitors, 703
of Doppler technique, 479
of fiber optics, 945
of *RC* circuits, 763–64
Medical imaging, 482–83, 1278–82
Meitner, Lise, 1129, 1261
Melting point, 546, 548, 549 (*see also* Phase, changes of)
Memory, 715 *pr*, 716 *pr*
Mendeleev, Dmitri, 1167
Mercury barometer, 373, 535 *pr*
Mercury-in-glass thermometer, 495, 496
Mercury manometer, 497
Mercury (planet), 17 *pr*, 157, 158
Merry-go-round, 264, 267, 268, 293 *pr*, 318–20, 324 *pr*
Meson exchange, 1297–98
Meson lifetime, 1135
Mesons, 1297, 1298, 1301–2, 1304–7, 1309–12
Messenger RNA (m-RNA), 1194–95
Metabolism:
of fat, 575 *pr*
room comfort and, 565
Metal detector, 850
Metal-oxide semiconductor field-effect transistor (MOSFET), 1218–19
Metallic bond, 1202
Metals:
alkali, 1168
free-electron theory of, 1203–7
Metastable state, 1175, 1176, 1240
Meteorite, 223 *pr*, 326 *pr*
Meter (m) (unit), 7, 17 *pr*
Meters, electric, 767–70, 797
correction for resistance of, 768–69
Method of joints, 348–49
Method of mixtures, 544
Metric (SI) multipliers: inside front cover
Metric (SI) system, 8
Metronome, 424 *pr*
Mho (unit), 744 *pr*
Michelson, A. A., 912–13, 1010, 1058–61
Michelson interferometer, 1010, 1058–61
Michelson–Morley experiment, 1058–61
Microampere (μA) (unit), 720
Micrometer, 12
Micron (μm) (unit), 8
Microphones:
condenser, 695
induction used by, 858
magnetic, 856

Microscope, 983–84, 1035–37
atomic force, 1151
compound, 983–84
and contrast, 1048
electron, 1095, 1108, 1132, 1151, 1156 *pr*
interference, 1048
magnification of, 983–84, 1037, 1109
phase-contrast, 1048
resolving power of, 1036
scanning tunneling electron (STM), 1151, 1156 *pr*
useful magnification, 1037, 1109
Microscopic description of a system, 494, 516 *ff*
Microscopic properties, 494, 516 *ff*
Microstate of a system, 595–96, 605 *pr*
Microwave background radiation, 1126
cosmic, 1342–44, 1348, 1349, 1353
Microwave oven, 925
Microwaves, 910, 1342–43
Middle C, 468, 485 *pr*
Mile (unit), 7
nautical (unit), 7, 19 *pr*, 169 *pr*
Milky Way galaxy, 170 *pr*, 171 *pr*, 1323–24
Milliampere (mA) (unit), 720
Milliarcsec (unit), 262 *pr*
Millikan, R. A., 798, 1099
Millikan oil-drop experiment, 798
Millimeter (mm) (unit), 8
Millimeters of mercury, 373, 389
Mine shaft, 374
Minerals, exploration for, 150, 481
Minimum speed, for uniform circular motion, 127–28
Minute (min) (unit), 7
Mirage, 998
Mirror equation, 934–38
Mirrors, 927–38
aberrations of, 982 *fn*, 984–85
concave, 931–37, 982
convex, 931, 938
focal length of, 931–32, 936–38
Lloyd's, 1023 *pr*
magnifying, 931, 936–37
plane, 927–31
used in telescope, 982
Missing orders, 1052 *pr*
Mixed layer, 511 *pr*
Mizar, 1054 *pr*
MKS system (*defn*), 8
mm-Hg (unit), 373, 389
mm-H_2O (unit), 373
mm of water (unit), 373
Mobiles, 361 *pr*
Models, 3, 21
of solar system
geocentric, 155–56
heliocentric, 155–56
Moderator, 238, 1263–64
Modern physics (*defn*), 2, 1056
Modulation:
amplitude, 918
frequency, 918, 919 *fn*
Module landing on the Moon, 222 *pr*
Moduli of elasticity, 343, 433
Molar specific heats
for gases, 556–58
of solids, 558–59
Mole (22.4 L), 505
Mole (unit), 504
volume of, for ideal gas, 505
Molecular biology, electric force in, 631–33, 1192–96
Molecular kinetic energy, 518–19, 541, 557–58
Molecular mass, 493, 504
Molecular rotation, 1196–97, 1199–1201
Molecular spectra, 1196–1201
Molecular speeds, 516–22
distribution of, 520–22
temperature vs., 521
Molecular temperature, 516–19
Molecular vibration, 1198–1201
Molecular weight, 493
Molecules, 493, 516–22, 527–31, 1186–1201
bonding in, 1186–89
in a breath, 507, 514 *pr*
in human cells, 519
polar, 611, 629–30, 1189
potential energy diagrams for, 1189–92
spectra of, 1196–1201
weak bonds between, 1192–96
Moment arm, 272
Moment of a force about an axis, 272
Moment of inertia, 274–76, 296 *pr*, 297 *pr*, 413
determining, 279–81, 413
parallel-axis theorem, 280–81
perpendicular-axis theorem, 281
Momentum, 227–53
angular, 303–7, 309–18, 1113
center of mass (CM), 244–48
in collisions, 230–44
conservation of angular, 303–5, 315–17
conservation of linear, 230–33, 236–44, 250, 1301–2
invariant energy, 1084
linear, 227–54
of photon, 1102
relation of force to, 228–29, 231, 234–35, 249–53, 1077, 1079
relativistic, 1077–79, 1080, 1084
uncertainty in measurement of, 1132
Monitors, TV and computer, 679–82
Monochromatic aberration, 985
Monochromatic light (*defn*), 998
Moon (of Earth), 19 *pr*, 329 *pr*, 898, 920, 928, 1323
acceleration due to gravity on, 51 *pr*, 166 *pr*
acceleration toward the Earth, 125
centripetal acceleration of, 125, 146
face of, toward Earth, 329 *pr*
force on, 146, 148
full moon, 161, 169 *pr*
in geocentric system, 155
module landing on, 222 *pr*
new moon, 161
and ocean tides, 159
orbit of, 161
phases of, 161
revolution and phases of, 161
rising of, 161, 169 *pr*
sidereal period, 161, 169 *pr*
synodic period, 161, 169 *pr*
work on, 176
Moonrise, 161
Moons of Jupiter, 156, 158, 167 *pr*, 168 *pr*
Morley, E. W., 1058–61
Morse code, 676 *fn*
Morse potential, 1224 *pr*
Moseley, H. G. J., 1169
Moseley plot, 1169
MOSFET, 1218–19
Motion, 1055–87
of charged particle in electric field, 628–29
circular, 123–34
at constant acceleration, 30–43, 65–73
damped, 414–17
description of (kinematics), 20–45, 54–76
and frames of reference, 21–22
in free fall, 37–43, 154–55
graphical analysis of linear, 30
harmonic, 402–10, 414–17
on inclines, 103–4
Kepler's laws of planetary, 155–60, 316
linear, 20–45
Newton's laws of, 86–94, 97–104, 116 *ff*, 228, 231, 249, 250, 275–79, 310–11, 1077, 1129, 1136, 1137
nonuniform circular, 133–34
oscillatory, 399 *ff*
overdamped, 415
parabolic, 65, 73
periodic (*defn*), 400
planetary, 155–60
projectile, 54, 65–73
rectilinear, 20–45
and reference frames, 21
relative, 73–75, 1055–87
rolling, 283–90
rotational, 264–321
satellite, 145, 152–54
simple harmonic (SHM), 402–10
thermal, 519
translational, 21, 98
underdamped, 415
uniform circular, 123–26
uniformly accelerated, 30–43
at variable acceleration, 43–44
vibrational, 399 *ff*
of waves, 428–51
Motion sensor, 489 *pr*
Motional emf, 845–46
Motor:
ac, 796
back emf in, 848–50
electric, 795–96
Mountain climbers, 110 *pr*, 114 *pr*, 115 *pr*, 193 *pr*, 370
Mountaineering, 143 *pr*, 193 *pr*
Moving picture, 680
MRAM, 858
MRI, 1230, 1281–82
Mt. Everest, 150, 171 *pr*, 527
Mt. Whitney, 527
Mu meson, 1301–2, 1304, 1305
Mud, car in, 115 *pr*
Multielectron atoms, 1166–71
Multigrade oils, 385
Multimeter, 769
Multiplication:
significant figures in, 4
of vectors, 58, 176–77, 307–9
Multiplication factor, 1264
Muon, 1301–2, 1304, 1305
Muon lepton number, 1301–5, 1310
Muon neutrino, 1304, 1305
Muscle(s), 273, 295 *pr*
forces in joints and, 295 *pr*, 339–41, 355 *pr*, 358 *pr*, 363 *pr*
frog, 718
insertions of, 339
Musical instruments, 447, 457 *pr*, 458 *pr*, 460, 467–73
tuning, 468
warming up, 472
Musical notes, 468
Mutation, 633
Mutual inductance, 870–71
Myopia, 976

n-type semiconductor, 1211–14
Nagasaki, 1266
Nanometer, 1212 *fn*
Napoleon I, 159 *fn*, 718
Natural abundances, 1227
Natural convection, 563

Natural frequency, 404, 417, 446 (*see also* Resonant frequency)
Natural logarithms, A-2
Natural radioactive background, 1237, 1274
Natural radioactivity, 1233
Natural resonant frequencies, 466–71
Natural stride, 424 *pr*
Nautical mile (unit), 7, 19 *pr*, 169 *pr*
Navstar Global Positioning System (GPS), 169 *pr*
Nd:YAG laser, 1177
Near Earth Asteroid Rendezvous (NEAR) spacecraft, 169 *pr*
Near-Earth orbit, 168 *pr*
Near field, 904
Near point, of eye, 976
Nearsightedness, 976–78
Nebulae, 1325
Negative, of a vector (*defn*), 57
Negative, photographic, 970 *fn*
Negative curvature, 1337, 1350
Negative electric charge (*defn*), 610, 721
Negative lens, 963
Negative sea of electrons, 1300
Negative work, 174
Neon tubes, 1158
Neptune, 158
  gravitational slingshot from, 217
  period and distance from Sun, 330 *pr*
Neptunium, 1259
Nerve pulse, 736–37, 790
Nervi, Pier Luigi, 352
Nervous system, electrical conduction in, 736–37
Net displacement, 55–56
Net force, 87–91, 97 *ff*
  with uniform circular motion, 126–27
Net resistance, 750
Net work determination, 174
Neuron, 736
Neutral atom, 1229
Neutral equilibrium, 216, 341
Neutral wire, 765
Neutrino flavor oscillation, 1303
Neutrinoless double beta decay, 1321 *pr*
Neutrinos, 1237–39, 1290, 1301–5, 1347
  mass of, 1303–4, 1305
  types of, 1301–4
Neutron, 238, 611, 1227, 1290, 1305
  delayed, 1264
  in nuclear reactions, 1261–66
  role in fission, 1261 *ff*
  thermal, 1261
Neutron activation analysis, 1288 *pr*
Neutron cross section, 1261
Neutron degeneracy, 1331
Neutron number, 1227
Neutron physics, 1259
Neutron star, 305, 330 *pr*, 1223 *pr*, 1326, 1331
New moon, 161
Newton, Isaac, 20, 87, 88, 91, 145–46, 161, 618, 982 *fn*, 997, 1005 *fn*, 1056, 1334, 1337 *fn*
Newton (N) (unit), 89
Newtonian focus, 982
Newtonian mechanics, 85–164
Newton's first law of motion, 86–88
Newton's law of universal gravitation, 145–49, 210–12, 614, 1334
Newton's laws of motion, 85–104, 116 *ff*, 208, 228, 231, 249–50, 275–79, 303, 310–11, 1077, 1129, 1136, 1137
Newton's rings, 1005–6
Newton's second law, 88–91, 93, 96–104, 130, 228, 231, 249–53, 1057, 1077
  for rotation, 275–79, 286–89, 303, 310–11
  for a system, 249–50
  for a system of objects, 249–50
  for a system of particles, 310–11
  for translation, 277
Newton's synthesis, 159
Newton's third law of motion, 91–94
NMR, 1230, 1279–82
Noah's ark, 19 *pr*
Noble gases, 1168, 1203
Nodes, 446–47, 468–71
Noise, 472–73
  and digital transmission, 678–79
  jet plane, 466
Noise-canceling headphones, 484 *pr*
Non-Euclidean space, 1336–37
Non-metal floor, 715
Non-sinusoidal wave, mathematical representation of, 439
Nonconductors, 611, 703–7, 724
Nonconservative field, 856
Nonconservative forces, 195–96, 199, 207–9
Noninductive winding, 872
Noninertial reference frames, 88, 90, 319–21
Nonlinear device, 1213
Nonlinear spring, 190 *pr*
Nonohmic device, 722
Nonreflecting glass, 1008–9
Nonrelativistic quantum mechanics, 1138, 1140
Nonuniform circular motion, 133–34
Nonvolatile memory (NVM), 858
"Normal" body temperature, 496
Normal eye (*defn*), 976
Normal force, 94–96
  and friction force, 117
Normal lens, 975
Normal X-ray images, 1045
Normalization condition, 1138, 1141 *fn*, 1143–46
Normalization constant, 1144
North pole, Earth, 784
North pole, of magnet, 783
Notes, musical, 468
Notre Dame, in Paris, 350
Nova, 1326, 1331
Noyce, Robert, 1219
npn transistors, 1218
Nuclear angular momentum, 1229
Nuclear binding energy, 1230–33
Nuclear collision, 238, 239, 243
Nuclear decay, 1081
Nuclear energy, 578 *fn*, 599, 1256–82
Nuclear fission, 238, 1261–66
Nuclear forces, 161–62, 226 *pr*, 1232–33, 1238, 1296–1317, 1334
Nuclear fusion, 1266–71, 1329–31
Nuclear magnetic moments, 1230
Nuclear magnetic resonance (NMR), 1230, 1279–82
Nuclear magneton, 1230
Nuclear masses, 1227 *fn*
Nuclear medicine, 1276–78
Nuclear physics, 1226–82
Nuclear power, 1264–66
Nuclear power plants, 847, 1264–65
Nuclear radioactive decay, 255 *pr*
Nuclear radius, 1228
Nuclear reactions, 1257–64
Nuclear reactors, 238, 1263–66, 1269–71
Nuclear spin, 1229
Nuclear structure, 1227–30
Nuclear weapons testing, 1266
Nucleon, 1227, 1290, 1314, 1346–47
Nucleon number, conservation of, 1240, 1301–2
Nucleosynthesis, 1329–31, 1347
Nucleotide bases, 631, 1194
Nucleus, 1227 *fn*
  compound, 1261–62
  daughter and parent (*defn*), 1234
  half-lives of, 1240–44
  liquid-drop model of, 691 *pr*
  masses of, 1227–30
  radioactive decay of unstable, 1233–48
  size of, 1228
  structure and properties of, 1227–30
Nuclide (*defn*), 1227
Null result, 1058, 1061, 1075
Numbers:
  binary, 676–79
  decimal, 676
Numerical integration, as a sum, 178, 179, A-8–A-11
NVM, 858

Object distance, 929, 934, 946, 962–63
Objective lens, 981–83, 1036
Object(s):
  force exerted by, 91
  in Newton's third law, 94
  in problem solving, 33, 34, 67, 68
  velocity subscript for, 73
Observable universe, 1344
Observations, 2, 1056
  and uncertainty, 1132
Occhialini, G., 1298
Occupied states, density of, 1205
Ocean currents, 563
Ocean tides, 159, 170 *pr*
Ocean waves, 429, 435, 443, 459 *pr*
  speed of, 459 *pr*
  velocity of, 459 *pr*
Oersted, H. C., 785–86
Off-axis astigmatism, 984
Ohm, G. S., 722
Ohm (Ω) (unit), 722
Ohmmeter, 769, 795
Ohm's law, 722–24, 735, 750, 755
Oil-drop experiment, 798
Oil(s):
  exploration for, 150, 481
  multigrade, 385
  viscosity of, 385
OLED, 1216–17
Olympic Games of 1936, 225 *pr*
Omega (particle), 1305
One-dimensional Schrödinger equation, 1137–50
One-dimensional wave equation, 440–41
One-dimensional waves, 435, 437
One over r^2 law, 436
Onnes, H. K., 735
Open circuit, 720, 730
Open system, 542
Open tube, 470–72

Open-tube manometer, 372–73
Opening a tight lid, 499
Operating temperatures, heat engines, 578
Operational definitions, 815
Oppenheimer, J. Robert, 1266
Optical coating, 1008–9
Optical illusion, 940, 998
Optical instruments, 970–85, 1009, 1033–43
Optical pumping, 1176
Optical sound track, 1101
Optical tweezers, 108 *pr*, 916
Optical zoom, 975, 987 *pr*
Optics:
  fiber, 945
  geometric, 927–85
  law of reflection, 444
  physical, 995–1048
Orbit:
  of the Moon, 161, 169 *pr*
  near-Earth, 168 *pr*
  satellite, 226 *pr*
Orbital angular momentum, in atoms, 1160–61, 1173–74
Orbital period, 152
Orbital quantum number, 1160
Orbitals, 1166
  highest occupied molecular, 1217
  lowest unoccupied molecular, 1217
Orchestra, tuning, 475
Order and disorder, 593–94
Order of interference or diffraction pattern, 999–1001, 1038–39, 1041, 1044, 1052 *pr*
Order of magnitude (*defn*), 7
Order-of-magnitude estimate, 11–13, 104
Ordinary conversation, 464, 465
Organ pipe, 469–72
Organic light-emitting diode (OLED), 1216–17
Origin, 21, 67
Orion, 1325
Oscillations, 399–419
  of air columns, 470–72
  damped harmonic motion, 414–17
  displacement, 401
  forced, 417–19
  mechanical, 399
  of molecules, 557–58
  of physical pendulum, 412–13
  simple harmonic motion (SHM), 402–10
  as source of waves, 430
  of a spring, 400–401
  on strings, 446–48, 467–69
  of torsion pendulum, 413
  as wave source, 430
Oscillator, simple harmonic, 402–10, 1149, 1156 *pr*
Oscillators, 879, 880, 887, 888
  electromagnetic, 877–881
  sawtooth, 763, 780 *pr*
Oscilloscope, 682
Osteoporosis, diagnosis of, 1103
Otto cycle, 583–84, 603 *pr*
Out of phase, waves, 445
Out-of-phase waves, 999, 1009, 1038
Overdamped system, 415
Overexposure, 972
Overflow, gas tank, 500
Overloading, of fuses, 729
Overtones, 447, 468–71
Owens, Jesse, 225 *pr*
Oximeter, pulse, 1216
Oxygen cylinder, 514 *pr*

p-type semiconductor, 1211–14
P waves, 435, 437, 450
Pacemaker, heart, 764, 871
Packet, wave, 1141
Packing of atoms, 1202
Paige, Satchel, 172
Paintball, 191 *pr*
Pair production, 1104–5
Pantheon, dome of, 351
Paper, thickness of, 12
Paper roll, 299 *pr*
Parabola, 54, 73, 349
Parabolic mirror, 932
Parachutist, 362 *pr*
Paradiso (Dante), 452 *pr*
Parallax, 1332–34
Parallel-axis theorem, 280–81
Parallel circuits, 698, 729, 750
Parallel emf, 757–59
Parallel LRC circuit, 896 *pr*
Parallelogram method of adding vectors, 57
Paramagnetism, 828–29
Paraxial rays (*defn*), 932
Parcel of air, 573 *pr*
Parent nucleus (*defn*), 1234
Parity bits, 678
Parsec (pc) (unit), 1333
Partial derivatives, 200, 440
Partial ionic character, 1188–89
Partial pressure, 525–26
Partially destructive interference, 445
Particle accelerators, 1290–96
Particle classification, 1304–6
Particle detectors, 1214, 1248–49, 1295
Particle exchange, 1296–98, 1312
Particle interactions, 1301 *ff*
Particle physics, 1289–1317
Particle resonance, 1306–7
Particle–antiparticle pair, 1300
Particle(s):
  definition of, 21, 98
  detection of, 1248–49
  elementary, 1289–1317
  fundamental (*defn*), 1290
  velocity of wave vs., 429
Particulate pollution, 18 *pr*
Pascal, Blaise, 367, 372, 397 *pr*
Pascal (Pa) (unit), 367, 373
Pascal's principle, 371–72
Pashen series, 1112, 1116
Passive solar heating, 599
Path independence, of entropy change, 588, 589
Pauli, Wolfgang, 1128, 1129, 1166, 1238
Pauli exclusion principle, 1166–67, 1187, 1204, 1207, 1311, 1330, 1331
Peak current, 730
Peak voltage, 730
Peak widths, of diffraction grating, 1041–43
Peaks, tallest, 9–10
Pendulum, 14, 129, 206, 240–41, 411–13
  ballistic, 240–41
  conical, 129
  damped, 416
  grandfather clock, 421 *pr*
  physical, 412–13
  simple, 14, 206, 411–12
  torsion, 413
Pendulum clock, 412
Penetration, barrier, 1149–51, 1236
Penzias, Arno, 1342
Percent body fat, 393 *pr*
Percent uncertainty, 4
  and significant figures, 6
Perfect cosmological principle, 1341
Performance, coefficient of (COP), 585–87
Perfume atomizer, 383
Period, 125, 269, 401, 404, 430
  of circular motion, 125
  of Moon, 161
  orbital, 152
  of pendulums, 14, 412, A-12
  of planets, 156–57
  of rotation, 269
  sidereal, 161, 169 *pr*
  of Sun, 324 *pr*
  synodic, 161, 169 *pr*
  of vibration, 401, 404
  of wave, 430
Periodic motion, 400 *ff*
Periodic Table, 1167–68, 1227 *fn*, inside back cover
Periodic wave, 430
Perlmutter, Saul, 1352
Permeability, magnetic, 812, 827
Permittivity, 614, 704
Perpendicular-axis theorem, 281
Perpendicular forces, 94
Perpendicular vectors, adding, 56
Perturbations, 158
PET, 1278–79
Phase, 445
  in ac circuit, 881–88
  of matter, 366, 494
  of waves, 438, 445, 999, 1005–9, 1038
Phase angle, 403, 438, 885–86
Phase changes, 522–24, 545, 547
Phase constant, 403, 438, 1140 *fn*, 1142
Phase-contrast microscope, 1048
Phase diagram, 523–24
Phase shift, 1006, 1009
Phase transitions, 522–27, 545–49
Phase velocity, 438, 439
Phases of the Moon, 161
Phasor diagram:
  ac circuits, 885–86
  interference and diffraction of light, 1002, 1029, 1041–42
Phon (unit), 467
Phonons, 1215 *fn*
Phosphor, 1248
Phosphorescence, 1175
Photino, 1317
Photocathode, 1248
Photocell, 689 *pr*, 1098
Photocell circuit, 1098, 1101
Photocopier, 619
Photodiode, 1101, 1215
Photoelectric effect, 1098–1101, 1104, 1271
Photographic emulsion, 1249
Photographic film, 971
Photomultiplier (PM) tube, 1248–49
Photon, 1098–1106, 1130, 1167, 1290, 1296–97, 1299, 1304–5, 1310, 1346–47
  absorption of, 1174–75
  decoupled (early universe), 1343, 1348
  and emf, 1297
  energy of, 1101
  mass of, 1101
  mediation of (force), 1297
  momentum of, 1102
  virtual, 1297
Photon exchange, 1296–98
Photon interactions, 1104–5
Photon theory of light, 1098–1106
Photosynthesis, 531, 1101
Photovoltaic (solar) cells, 599, 605 *pr*, 1214–15
Physical pendulum, 412–13
  human leg as, 424 *pr*
Physics:
  classical (*defn*), 2, 1056
  modern (*defn*), 2, 1056
  relation to other fields, 2

Pi meson, 1298, 1305, 1306, 1309–12
Piano, 13, 469, 473, 475
Piano string, 447, 460, 468
Piano tuner, 13, 475
Piano wire, 343
Pick-up nuclear reaction, 1285 *pr*
Piezoelectric materials, 894 *pr*
Pinball machine, 423 *pr*
Pincushion distortion, 985
Ping-Pong ball, in jet of air, 383
Pins, 347–49
  structural, 346
Pion, 1298, 1305, 1306, 1309–12
Pipe:
  light, 945
  vibrating air columns in, 467 *ff*
Pipe organs, 469–72
Pisa:
  cathedral, 412
  tower, 38
Pitch, 461, 468, 469, 471–73
Pitched baseball, 384
Pitcher, 172
Pixels, 680–81, 970, 974, 1014, 1046
Planck, Max, 1097, 1128
Planck length, 15, 1127 *pr*, 1345
Planck time, 18 *pr*, 1127 *pr*, 1316, 1345
Planck's constant, 1097–98, 1134
Planck's quantum hypothesis, 1097
Plane:
  focal, 959
  mirror, 927–31
Plane geometry, A-2
Plane polarization, 1010
Plane waves, 443, 905, 1140–41
Planetary motion, 155–60, 316
Planets (*defn*), 155–60, 167 *pr*, 250, 262 *pr*, 330 *pr*
  angular momentum, 316
  extrasolar, 158
  period of, 156–57
Plants, watering, in the evening, 568 *pr*
Plasma (*defn*), 366, 1256, 1270
Plasma globe, 897 *pr*
Plastic region, 342
Plate tectonics, 378
Platform, 324 *pr*
  rotating, 318–20, 329 *pr*
Plum-pudding model of atom, 1110
Plumb bob, 171 *pr*
Pluto, 158, 1323
Plutonium, 1259, 1265, 1266
PM tube, 1248–49
pn diode lasers, 1216
pn junction, 1212–17
pn junction diode, 1212–17, 1249
pn junction laser, 1177
pnp transistor, 1218
Point:
  boiling, 496, 525, 545
  breaking, 342
  critical, 523
  dew, 526
  far, 976
  focal, 932, 937, 959–60, 976
  freezing, 496 *fn*
  Lagrange, 160
  mathematical, 21
  melting, 546, 548, 549
  near, 976
  sublimation, 523
  triple, 508, 523
  turning, 215
Point charge (*defn*), 615
  potential, 666–70
Point particle, 21, 98
Point rule, Kirchhoff's, 902 *ff*
Point source, 927
Pointed arch, 350–51
Poise (P) (unit), 385
Poiseuille, J. L., 386
Poiseuille's equation, 386
Poisson, Siméon, 1026
Polar molecules, 611, 629–30, 632, 706, 1188–89
Polar satellites, 153
Polarization, 995, 1010–14
  circular, 925 *pr*
  linear, 1010
  plane, 1010
  by reflection, 1013–14
Polarizer, Polaroid as, 1011
Polarizing angle, 1013, 1014
Polaroid sheet, 1011
Polaroids, 1011–13
Pole vault, 189 *pr*, 194, 203–4
Poles, magnetic, 782–84
  of Earth, 784
Police siren, 478
Pollution, 18 *pr*, 576, 598–600
Poloidal field, 1271
Polonium, 1233
Polymers, 1216
Pontecorvo, Bruno, 1303 *fn*
Pool depth, apparent, 941
Pope, Alexander, 1337 *fn*
Popping a wheelie, 300 *pr*
Population, inverted, 1176–77
Position, 21, 28
  angular, 265, 1134
  average, 1147
  uncertainty in, 1132–35
  vector for, 62–65
Positive curvature, 1337, 1350
Positive electric charge (*defn*), 610
Positive holes, 1211
Positive lens, 963
Positron, 1104–5, 1239, 1279, 1290, 1300
Positron emission tomography (PET), 1278–79
Post-and-beam construction, 344
Posture adjustment, 342
Potato chip bag, 513 *pr*
Potential (*see* Electric potential)
Potential difference, electric, 661 *ff* (*see also* Electric potential; Voltage)
Potential energy, 197–212, 215–17
  definition, 197, 199
  diagrams, 1189–92
  elastic, 199–200, 204–5, 210, 408–9
  electric, 660–63, 668–69, 674–75, 700–703
  general definition, 199
  gravitational, 197–98, 201–14, 216–17
  in metal crystal, 1207
  for molecules, 1189–92, 1198, 1202
  for nucleus, 1151, 1236
  only changes matter, 201
  related to force, 199–200
  in Schrödinger equation, 1138–40, 1142–49
  for square well and barriers, 1142–49
  in 3-D, 200
Potential energy diagrams, 215–16
  simple harmonic motion, 409
  turning points, 215–16
Potential well, 215, 1142–49
Potentiometer, 779 *pr*
Potter's wheel, 295 *pr*
Pound (lb) (unit), 89
Powell, C. F., 1298
Power factor (ac circuit), 886
Power generation, 598–600, 846–47
  geothermal, 599
  hydroelectric, 599
  nuclear, 599
  solar, 599
  wind, 599
  (*see also* Electric power)
Power (mechanical), 213–15, 726–32, 886
  climbing stairs, 213
  energy vs., 213
  as rate energy is transformed or transferred, 213
  rating of an engine, 214–15
  transferred to driven oscillator, 425 *pr*
  units of, 213
Power (optical):
  of a lens, 960
  magnifying power, 979, 981
Power outages, 780 *pr*
Power plants:
  fossil-fuel, 599
  nuclear, 847, 1264–65
Power reactor, 1264
Power transmission, 851–54
Powers of ten, 5, 7
Poynting, J. H., 913 *fn*
Poynting vector, 913–15
Prairie dogs, 384
Precession, 317–18
  of Earth's axis, 327 *pr*
  of the equinoxes, 327 *pr*
Precipitator, 710 *pr*
Precision, 6
  accuracy vs., 4, 6
Prefixes, metric, 8
Presbyopia, 977
Prescriptive laws (*defn*), 3
Pressure, 367–71
  absolute, 371
  in an ideal gas, 518
  atmospheric, 370–74
  conversion factors, 373
  at a faucet, 369
  for fluids, 365, 367–89
  in a gas, 371, 502–3, 518, 522–28
  gauge, 371
  head, 369
  hydraulic, 372
  on living cell, 371
  measurement of, 372–74
  partial, 525–26
  and Pascal's principle, 371–72
  radiation, 915–16
  units of, 367, 373
  vapor, 524–26
Pressure amplitude, 466–67
Pressure cooker, 535 *pr*
Pressure gauges, 372–73
Pressure gradient, 386, 561
Pressure head, 369
Pressure variation amplitude, 463
Pressure waves, 435, 462–63, 466–67, 470–71
Prestressed concrete, 346
Primary coil, 851
Principal axis, 932
Principal quantum number, 1113 *ff*, 1160–62
Principia (Newton), 87, 145
Principle, 3, 30, 34 (*see also specific principles by name*)
Principle of complementarity, 1105–6
Principle of correspondence, 1086, 1119, 1129
Principle of equipartition of energy, 557–58
Principle of equivalence, 163–64, 1334–35
Principle of superposition, 441–42 *ff*, 472–73, 616, 620, 625
Printers, computer, 619
Prism, 941–42

Prism binoculars, 944, 982
Probability:
  and entropy, 595–97
  in kinetic theory, 516–22
  in nuclear decay, 1240
  in quantum mechanics, 1130, 1131, 1135–36, 1145, 1159, 1163–65, 1187–89
Probability density (probability distribution)
  in atoms, 1130, 1140, 1143, 1148, 1159, 1162–63, 1165
  in molecules, 1187–89
Probability function, Fermi–Dirac, 1205, 1210
Problem-solving strategies, 33–36, 61, 67, 98, 104, 130, 175, 209, 244, 277, 334, 547, 600, 622, 755, 791, 818, 843, 964, 1009
Processes:
  isobaric, 552–54
  isochoric, 552–54
  isothermal, 551–53
  isovolumetric, 552–54
  reversible and irreversible (*defn*), 580–83, 588–91
  spontaneous, 577 *fn*
Projectile, horizontal range of, 70–72
Projectile motion, 54, 65–73
  kinematic equations for, 67
  parabolic, 73
Propagate (*defn*), 432
Proper length, 1070
Proper time, 1066, 1321 *pr*
Proportional limit, 342
Protein molecule, 631
Proteins:
  shape of, 1195–96
  synthesis of, 1194–95
Proton, 1227 *fn*, 1257, 1266–69, 1277, 1290, 1305
  decay of, 1305, 1315–16
  electric force exerted by, 615
  mean life of, 1315
Proton centers, 1277
Proton decay, 1305, 1315–16
Proton therapy, 1277
Proton–proton chain, 1267–68, 1329
Proton–proton collision, 243
Protostar, 1329
Protractor (for angles), 5
Proxima Centauri, 1323
Pseudoforce, 127, 319–21
Pseudovector, 270, 307 *fn*
Psi, in Schrödinger equation, *defn*, 1136–40
psi (unit), 367, 373
PT diagram, 523–24
Ptolemy, 155, 156
Pulley, 102, 188 *pr*, 218 *pr*, 277, 278
Pulse, wave, 429
Pulse-echo technique, 482, 1281
Pulse oximeter, 1216
Pulsed lasers, 923 *pr*, 1117
Pumps, 374, 388–89
  centrifugal, 388–89
  heat, 586–87
Punting a football, 69, 71
Pupil, 975
Purely rotational motion, 265
Pushing a desk, 172
Putting, in golf, 52 *pr*
PV diagrams, 522, 528, 551–54, 560
Pyrex glass, 509 *pr*
Pythagorean theorem, A-2, A-4

Q-value (disintegration energy), 1234–35
Q-value (quality factor) of a resonant system, 419, 425 *pr*, 896 *pr*
Q-value (reaction energy), 1258
QCD, 1298, 1311–14
QED, 1297
QF, 1273
QSOs, 1326, 1336
Quadratic equation, 40
Quadratic formula, 42, A-1, inside back cover
Quadrupole, electric, 641 *pr*
Quality factor (Q-value) of a resonant system, 419, 425 *pr*, 896 *pr*
Quality factor (QF) of radiation, 1273
Quality of sound, 472–73
Quantities, base, 8, 9
Quantization:
  of angular momentum, 1113, 1160–61
  of electric charge, 614
  of energy, 1097, 1113–19, 1143
Quantization error, 677
Quantization loss, 677
Quantum chromodynamics (QCD), 1298, 1311–14
Quantum condition, Bohr's, 1113, 1120
Quantum electrodynamics (QED), 1297
Quantum fluctuations, 1349
Quantum hypothesis, Planck's, 1097
Quantum mechanics, 1120, 1128–1219
  of atoms, 1158–79
  Copenhagen interpretation of, 1136
  of molecules and solids, 1186–1219
Quantum numbers, 1097, 1113–15, 1143, 1160–63, 1166–67, 1196–1201
  principal, 1113 *ff*
Quantum (quanta) of energy, 1097
Quantum theory, 1056, 1095–1120, 1128–1219
  of atoms, 1113–20, 1158–79
  of blackbody radiation, 1096–97
  of light, 1095–1106
  of specific heat, 558
Quarks, 614 *fn*, 1290, 1296–98, 1305, 1308–13, 1345, 1347
  confinement, 1312, 1345–47
Quartz oscillator clock, 487 *pr*
Quasars (quasi-stellar objects, QSOs), 1326, 1336
Quasistatic process (*defn*), 551
Quintessence, 1352

r, limit of large, 623
R-values, of thermal insulation, 562
Rabbits, 384
Rad equivalent man (rem), 1273
Rad (unit), 1273
Radar, 480, 481, 910
  in weather forecasts, 480
Radar guns, 475
Radial acceleration, 124 *ff*, 133
Radial forces, 130
Radial probability distribution, 1163–65
Radian (rad), measure for angles, 265–66
Radiant flux, 1016
Radiation, electromagnetic, 563–66
  blackbody, 1096–97, 1327, 1342
  body cooling by, 564
  cosmic microwave background, 1342–44, 1348, 1349, 1353
  emissivity of, 564
  gamma, 1234, 1239–40, 1271
  infrared (IR), 910, 941, 1041
  microwave, 910
  seasons and, 566
  solar constant and, 566
  from the Sun, 563, 566
  synchrotron, 1293
  thermal, 563–67
  ultraviolet (UV), 910, 941
  X-ray, 910, 1043–44, 1054 *pr*, 1170 (*see also* X-rays)
Radiation, nuclear:
  activity of, 1241, 1243–44, 1272
  alpha, 1234–37, 1240
  beta, 1234, 1237–40, 1331
  detection of, 1274
  dosimetry for, 1272–76
  gamma, 1234, 1239–40, 1271
  human exposure to, 1274–76
  ionizing (*defn*), 1271
  measurement of, 1272–76
  medical uses of, 1276–78
  types of, 1234, 1240
Radiation biology, 1276–78
Radiation-dominated universe, 1347–48
Radiation dosimetry, 1272–76
Radiation era, 1347–48
Radiation field, 904
Radiation film badge, 1274
Radiation pressure, 915–16
Radiation sickness, 1274
Radiation therapy, 1276–77
Radiator, hot-water, 563
Radio, 917–20
Radio signals, 452 *pr*, 453 *pr*
Radio waves, 910, 1036
Radioactive background, natural, 1237, 1274
Radioactive dating, 784, 1246–48
Radioactive decay, 255 *pr*, 257 *pr*, 1233–49
Radioactive decay constant, 1240–41
Radioactive decay law, 1241, 1242
Radioactive decay series, 1245–46
Radioactive fallout, 1266
Radioactive tracers, 1277–78
Radioactive waste, 1264–66
Radioactivity, 1226–49
  artificial (*defn*), 1233
  natural (*defn*), 1233
Radiofrequency (RF) signal, 917, 1280–81
Radioisotope (*defn*), 1233
Radionuclide (*defn*), 1233, 1272
Radiotelescope, 1036
Radium, 1233, 1234, 1245
Radius:
  of curvature (*defn*), 124, 134
  of Earth estimate, 13, 18 *pr*
  of gyration, 296 *pr*
  of nuclei, 1228
Radon, 1234, 1274, 1276
Rae Lakes, 11
Rainbow, 942, 957 *pr*
RAM (random access memory), 692–93, 716 *pr*, 857
Ramps:
  design, 113 *pr*
  ramp vs. stair analogy, 1097
Random access memory (RAM), 692–93, 716 *pr*, 857
Random motion, 631, 633
Range:
  of projectile, 70–71
  of validity, 33, 71, 104
Rapid estimating, 11–13
Rapid transit system, 53 *pr*
Rare-earth solid-state lasers, 1177
Rarefactions, in waves, 431
Rate of nuclear decay, 1240–44

Ray, 443–44, 927 *ff*, 959 *ff*
paraxial (*defn*), 932
Ray diagram, 933, 938, 964
Ray model of light, 927 *ff*, 959 *ff*
Ray tracing, 927 *ff*, 959 *ff*
Rayleigh, Lord, 1034, 1097
Rayleigh criterion, 1034–35
Rayleigh–Jeans theory, 1097
RBE, 1273
RC circuit, 759–64
Reactance, 872, 881–84
capacitive, 884
inductive, 882–83
(*see also* Impedance)
Reaction energy, 1258
Reaction time, 876
Reactions:
chain, 1262–64, 1266
endoergic, 1258
endothermic, 1258
exoergic, 1258
exothermic, 1258
nuclear, 1257–64
rate of, 521
slow-neutron, 1258
subcritical, 1264, 1266
supercritical, 1264, 1266
Reactors, nuclear, 1263–66, 1269–71
Read/Write head, 856–57
Reading, in problem solving, 67, 68, 104
Reading cells, 857
Real gases, 522–24
Real image, 929, 933, 961
Real processes, irreversibility of, 580
Rearview mirror, curved, 938
Reasonableness check, 33, 34, 104
Rebounder, 50 *pr*
Receivers, radio and television, 917–19
Reciprocating compressor, 574 *pr*
Reciprocating pump, 388
Recoil, 233
Recombination epoch, 1348
Rectangular coordinates, 21
Rectifiers, 1214, 1220 *pr*
Rectilinear motion, 20–45
Recurrent novae, 1331
Recycling, of metals, 861 *pr*
Red giants, 1326, 1328, 1329–30
Redshift, 1086, 1334, 1339–40, 1344
in cosmology, 480
Redshift parameter, 1339
Reduced mass, 1197
Reference frames, 21, 87–88, 209, 318–21, 1056 *ff*
accelerating, 88, 90, 163–64, 318–21
inertial, 87–88, 90, 319, 1056 *ff*
noninertial, 88, 90, 319–21, 1056
rotating, 318–21
for Sun, 159
Sun as, 159
transformations between, 1073–76
velocity subscript for, 73
Reference level, 198
Reflected ray, 444
Reflected wave, 444
Reflecting telescope, 982
Reflection:
angle of, 444, 927
diffuse, 928
law of, 444, 927
and lens coating, 1009
of light, 926, 927–31
phase changes during, 1004–9
polarization by, 1013–14
specular, 928
from thin films, 1004–9
total internal, 456 *pr*, 943–45
of vectors, 270
of waves on a cord, 443
Reflection coefficient, 1150, 1157 *pr*
Reflection grating, 1037
Reflectors, 951 *pr*
Refracting telescope, 980–81
Refraction, 449–50, 939–41, 997–98
angle of, 449, 939
of earthquake waves, 450
index of, 939
law of, 449, 940, 997–98
of light, 939–41, 997–98
and Snell's law, 939–41
at spherical surface, 946–48
by thin lenses, 959–62
of water waves, 449
Refresh rate, 682
Refrigerators, 584–86
Carnot, 606 *pr*
coefficient of performance (COP), 585, 603 *pr*
Regelation, 533 *pr*
Reheating, 1346
Reines, Frederick, 1303 *fn*
Reinforced concrete, 346
Relationships, in problem solving, 104
Relative biological effectiveness (RBE), 1273
Relative humidity, 525–26
Relative motion, 73–75, 1055–87
Relative permeability, 828
Relative velocity, 73–75, 1063 *ff*, 1073 *ff*
Relativistic addition of velocities, 1076
Relativistic energy, 1080–84
Relativistic mass, 1079
Relativistic momentum, 1077–79, 1084
derivation of, 1077–79
Relativity:
Einstein's theory of, 2
Galilean–Newtonian, 1056–58, 1072–74
general theory of, 163–64, 1322–23, 1334–36
special theory of, 1055–87, 1334
constancy of speed of light, 1062
four-dimensional space-time, 1072
impact of, 1086–87
and length, 1070–71
and Lorentz transformation, 1073–76
and mass, 1079
mass–energy relation in, 1080–84
postulates of, 1061–62
simultaneity in, 1062–63
and time, 1063–69, 1072
Relativity principle, 1056–57, 1061 *ff*
Relaxation time, 415
Relay, 830 *pr*
Rem (unit), 1273
Repeating cycles, 578
Repulsive forces, 1189–90, 1296
Rereading, in problem solving, 33, 34
Research reactor, 1264
Resistance and resistors, 722–24, 727, 881
in ac circuit, 881 *ff*
with capacitor, 759–64, 880–88
color code, 723
definition of resistance, 722
and electric currents, 718 *ff*
with inductor, 875–77, 880–88
internal, in battery, 748–49
in *LRC* circuit, 880–88
of meter, 768–69
net, 750
in series and parallel, 749–53
shunt, 767
shunt resistors, 767
and superconductivity, 735–36
wire-wound resistors, 723
Resistance thermometer, 726
Resistive force, 134–35
Resistivity, 724–26
temperature coefficient of, 726
Resolution:
in analog-to-digital conversions, 677
of diffraction grating, 1041–44
of electron microscope, 1109
of eye, 1035, 1037
of high-energy accelerators, 1290–91
of lens, 974, 1033–37
of light microscope, 1037
limits of, 1033–37
and pixels, 974
of telescope, 1036
of vectors, 58–61, 98, 99
Resolving power, 1036, 1043
Resonance, 417–19, 446
in ac circuit, 887–88
elementary particle, 1306–7
nuclear magnetic, 1230, 1279–82
Resonant frequency, 417, 446, 468–71, 887
Resonant modes of vibration, 446
Resonant oscillation, 417–19
Resonant peak, width of, 419
Rest energy, 1081–82, 1135
Rest mass, 1079
Resting potential, 736–37
Restitution, coefficient of, 259 *pr*
Restoring force, 179, 400, 404
Resultant displacement, 55–56
Resultant vector, 55–57, 60–61
Retentivity (magnetic), 827
Retina, 975
Reverse-biased diode, 1213
Reversible cycle, 580–83, 589
Reversible process, 580–83, 588, 590
Revolutions per second (rev/s), 269
Reynold's number, 395 *pr*
RF signal, 917, 1280–81
Rho (particle), 1305
Ribosome, 1195
Richards, P., 1343
Riess, Adam, 1352
Rifle recoil, 233
Right-hand rule, 270, 307, 785, 786, 789, 791, 813, 843
Rigid box, particle in, 1142–46
Rigid object (*defn*), 265
rotational motion of, 264–90, 312–15
translational motion of, 248–50, 284–86
Ripple voltage, 1214, 1224 *pr*
Rivets, 511 *pr*
rms (root-mean-square):
current, 731–32
speed, 519, 522
voltage, 731–32
RNA, 1194–95
Robot arm, work done by, 180
Rock climbing, 115 *pr*, 143 *pr*, 193 *pr*
Rockets, 85, 92, 102, 233, 250–52, 395 *pr*
propulsion of, 233, 251–52
Rock(s):
dating oldest Earth, 1248
throwing overboard, 378
Roemer, Ole, 912
Roentgen, W. C., 1043
Roentgen (R) (unit), 1273

Roller coaster, 198, 202–3, 208, 209, 221 *pr*, 225 *pr*
Rolling down an incline, 284–87
Rolling friction, 117, 144 *pr*, 289–90
Rolling motion, 283–90
  instantaneous axis of, 284
  total kinetic energy, 284
  why slows down, 289–90
  without slipping, 283–88
Rollover, 328 *pr*
Rolls of paper, 299 *pr*
Roman bridge, 2
Roman Forum, 350
Romans, 350, 351
Rome, Italy, 2, 351, 352
Roof truss, 347
Room comfort, metabolism and, 565
Room temperature, 511 *pr*
Root-mean-square (rms):
  current, 731–32
  speed, 519, 522
  voltage, 731–32
Rope:
  climbing, 115 *pr*
  stretchable, 193 *pr*
Rotating frames of reference, 318–21
Rotating platform, 318–20, 329 *pr*
Rotating turntable, 307
Rotation, 264–321
  axis of (*defn*), 265
  frequency of (*defn*), 269
  of rigid body, 264–90, 312–15
Rotational angular momentum quantum number, 1196–97, 1200–1201
Rotational imbalance, 314–15
Rotational inertia, 274–76 (*see also* Moment of inertia)
Rotational kinetic energy, 281–83
  molecular, 541, 557–58
Rotational motion, 264–321
  plus translational motion, 283–84
Rotational transitions, 1196–97
Rotational work, 282
Rotor, 795, 848
Rough calculations, 11–13, 104
Round arch, 350–51
Rubidium–strontium dating, 1253 *pr*
Ruby laser, 1176
Rulers, 4
Runways, 32
Russell, Bertrand, 1108
Rutherford, Ernest, 1110, 1228, 1234, 1257, 1288 *pr*
Rutherford's model of the atom, 1110–11
Rydberg constant, 1112, 1117
Rydberg states, 1184 *pr*

S waves, 435
SAE, viscosity numbers, 385 *fn*
Safety, electric hazards and, 764–66
Safety factor, 345–46
Sag of beam, 346
Sailboats, against the wind, 383
Salam, A., 1313
Samples, in analog-to-digital conversions, 677
Sampling rate, 677, 678
San Francisco, 13
Satellite dish, 919
Satellites, 145, 152–55, 186 *pr*, 223 *pr*, 296 *pr*
  geostationary, 153
  global positioning, 18 *pr*, 169 *pr*, 1069
  orbit of, 226 *pr*
  polar, 153
Saturated vapor pressure, 525–26
Saturation, 524, 827
Saturn, 157, 158, 169 *pr*, 263 *pr*
  gravitational slingshot from, 217
  period and distance from Sun, 330 *pr*
Sauna, 536 *pr*
Savart, Felix, 821
Sawtooth oscillator, 763, 780 *pr*
Sawtooth voltage, 763
Scalar components, 58
Scalar (*defn*), 55
Scalar (dot) product, 176–77
Scalar quantities, 55
Scale, musical, 468
Scale, spring, 86
Scale factor of universe, 1340
Scanning electron microscope (SEM), 1095, 1108
Scanning tunneling electron microscope (STM), 1109, 1151, 1156 *pr*
Scattering:
  elastic, 1260
  of light, by atmosphere, 1015–16
  of X-rays, Bragg, 1179
Schmidt, Brian, 1352
Schrödinger, Erwin, 1095, 1128, 1129
Schrödinger equation, 1136–49, 1159–60, 1198, 1207
Schwarzschild radius, 1337, 1360 *pr*
Science, dynamic nature of, 2–3
Scientific notation, 5–6
Scintigram, 1278
Scintillation counter, 1248
Scintillator, 1248, 1278
Scratch shot, 263 *pr*
Scuba diving, 514 *pr*, 536 *pr*, 575 *pr*
Scuba tank, 394 *pr*, 512 *pr*, 513 *pr*, 515 *pr*, 535 *pr*, 536 *pr*
SDSS, 1353
Sea level rise, 511 *pr*
Sea of electrons, 1300
Search coil, 867 *pr*
Seashells, 485 *pr*, 487 *pr*
Seashore, 435, 443
Seasons, 165 *pr*, 566
Second derivative, 30
Second harmonic, 446–47, 470
Second law of motion, 88–91, 93, 96–104, 228–29, 231, 249–53, 1057, 1077
  for rotation, 275–79, 310–11
  for a system, 249–50
  for a system of objects, 249–50
  for a system of particles, 249, 310–11
  for translation, 277
Second law of thermodynamics, 576–99
  and Carnot efficiency, 581–83
  Clausius statement of, 577, 585
  and efficiency, 579–80
  and entropy, 587–94, 600
  general statement of, 592–93, 597
  heat engine, 577–80
  and irreversible processes, 580–83
  Kelvin-Planck statement of, 583
  refrigerators, air conditioners, and heat pumps, 584–87
  reversible processes, 580–83
  and statistical interpretation of entropy, 595–97
  and time's arrow, 593
Second-order differential equation (*defn*), 402
Second (s) (unit), 7
Secondary coil, 851
SEER rating, 587
Seesaw, balancing a, 335
Segrè, Emilio, 1299
Seismic reflection, 457 *pr*
Seismograph, 859
Selection rules, 1162–63, 1196, 1199, 1200
Self-inductance, 872–74
Self-sustaining chain reaction, 1263–66
SEM, 1095, 1108
Semicircular arch, 350–51
Semiconductor detector, 1249
Semiconductor diode lasers, 1177
Semiconductor diodes, 1212–17
Semiconductor doping, 1210–12
Semiconductor memory, 857–58
Semiconductor particle detectors, 1215
Semiconductors, 611, 724, 1209–19
  compound, 1215
  intrinsic, 1209, 1211
  n and *p* types, 1211–14
  resistivity of, 724
  silicon wafer, 1249
Sensitivity:
  of the ear, 464, 466–67
  full-scale current, 767
  of human ear, 466, 467
  of meters, 768
  of voltmeter, 769
Separation of variables, 1139
Series electric circuit, 698, 749
Series emf, 757–59
Shattering goblet, 417
Shear modulus, 343, 345
Shear strength, 346
Shear stress, 344–45
Shells, atomic, 1167
Shielded cable, 818, 874, 911
Shielding, electrical, 628, 818
SHM (*see* Simple harmonic motion)
SHO (*see* Simple harmonic oscillator)
Shock absorbers, 399, 415
Shock waves, 480–81
Shockley, W., 1218
Short circuit, 729, 730
Short-range forces, 1233, 1334
Short (short circuit), 729, 730
Shortstop, 322 *pr*
Shunt resistor, 767
Shutter speed, 971–73
SI derived units: inside front cover
SI units, 8–9
  new, 8–9
Sidereal period, 161, 169 *pr*
Siemens (S) (unit), 744 *pr*
Sievert (Sv) (unit), 1273
Sigma (particle), 1305
Sign conventions, 549, 578, 588–89, 947
  geometric optics, 935, 963
Signal-to-noise ratio, 486 *pr*, 490 *pr*, 679
Signal voltage, 676–79
Significant figures, 4–5
  percent uncertainty vs., 6
Significant figures rule, 4
Signs, in problem solving, 61
Silicon, 1209 *ff*
Silicon wafer semiconductor, 1249
Silicon wafers, 18 *pr*
Similar triangles, A-2–A-3
Simple harmonic motion (SHM), 402–10
  applied to pendulums, 411–12
  damped, 414–17
  energy, 408–10
  including springs' mass, 402
  related to uniform circular motion, 410
  sinusoidal nature of, 402

Simple harmonic oscillator (SHO), 402–10, 1149, 1156 *pr*
acceleration of, 405
energy in, 408–10, 1156 *pr*
molecular vibration as, 1199–1200
velocity and acceleration of, 405
"Simple machines"
lever, 187 *pr*, 334
pulley, 101–2
Simple magnifier, 979–80
Simple pendulum, 14, 206, 399, 411–12
with damping, 416
period of, 411
period of, and Galileo, 412
Simultaneity, 1062–64
Sine, 59
"Singing rod" demonstration, 491 *pr*
Single-lens reflex (SLR) camera, 975
Single moving charge, magnetic field due to, 824
Single photon emission computed tomography (SPECT), 1278–79
Single photon emission tomography (SPET), 1278
Single-slit diffraction, 1026–31
Singularity, 1337
Sinusoidal curve, 402 *ff*
Sinusoidal motion, 402, 403, 405, 406
Sinusoidal traveling wave, 437–39
Siphon, 390 *pr*
Sit-up, 291 *pr*
Skater, 302, 304, 324 *pr*, 330 *pr*
Skeet, 262 *pr*
Sketches, for problem solving, 98, 99
Ski jump, 225 *pr*
Ski runs, 194
Skidding car, 131–32, 190 *pr*
Skier, 116, 121, 155, 194, 220 *pr*, 224 *pr*
Skin, heat conduction, 573 *pr*
skis, cross-country, 136 *pr*
Sky diver, 108 *pr*
SLAC, 1294
Sled, falling, 233
Slepton, 1317
Sliding, rolling vs., 284, 288
Slingshot, gravitational, 216–17, 263 *pr*
Sloan Digital Sky Survey (SDSS), 1353
Slope, of a curve, 25
Slow-neutron reaction, 1258
SLR camera, 975
Slug (unit), 89
Small Sports Palace in Rome, 352
Smoke detector, 1237
Smoke up a chimney, 384
Smoot, George, 1343
SN1987a, 1303, 1331
Snell, W., 940
Snell's law, 940–41, 946, 969, 997
SNIa (type Ia) supernovae, 1332, 1333, 1352
Snow, high-altitude climbers and, 571 *pr*
Snowboarder, 54, 139 *pr*
Soap bubble, 396 *pr*, 995, 1004, 1007–8
Soaps, 387
Soccer ball curve, 384
Soccer field, 17 *pr*
Sodium chloride, bonding in, 1188, 1190–91, 1202
Sodium pump, 736 *fn*
Solar absorption spectrum, 1040–41, 1111
Solar and Heliospheric Observatory (SOHO) satellite, 160
Solar cells (photovoltaic), 599, 605 *pr*, 1214–15
Solar constant, 566
Solar eclipse, 161, 293 *pr*
Solar energy, 599, 924 *pr*
Solar heating, 599
Solar neutrino problem, 1302–3
Solar power, 599, 924 *pr*
Solar pressure, 916
Solar sail, 916
Solar system, 216
escape velocity from, 223 *pr*
formation of, 330 *pr*
models, 155–56
Soldiers, marching, 417
Solenoid, 811, 819–21, 826–28, 873
Solid angle, 8 *fn*, 1016 *fn*
Solid state devise (SSD), 858
Solid-state lighting, 1216
Solid-state physics, 1202–19
Solids, 342 *ff*, 366, 1202–10
amorphous, 1202
band theory of, 1208–10
bonding in, 1202–3
definition, 494
energy levels in, 1208–10
molar specific heats for, 558–59
(*see also* Phase, changes of)
Solutions:
in problem solving, 33, 98, 99, 104, 277, 334, 547
unphysical, 34, 43
Solvay Conference, 1128
Sonar, 481–82, 489 *pr*
Sonic boom, 481
Sonograms, 482
Sound, 460–83
audible range of, 461–62
and beats, 474–76
dBs of, 464–67
detection of, 460
Doppler effect of, 476–80
ear's response to, 467
infrasonic, 462
intensity of, 464–67
on the street, 464
interference of, 473–76
level of, 464–67
loudness of, 461, 464, 465, 467
mathematical representation of wave, 462–63
of the ocean, 487 *pr*
pitch of, 461
pressure amplitude of, 463, 466–67
quality, 472–73
shock waves of, 480–81
and sonic boom, 481
sound level of, 464–67
sources of, 460, 467–72
speed of, 461, 462, 910
supersonic, 462, 480–81
timbre of, 472–73
tone color of, 472–73
ultrasonic, 462, 482–83
Sound barrier, 480–81
breaking, 481
Sound frequency, highest, 536 *pr*
Sound level, 8 *fn*, 464–67
Sound navigation ranging, 481 *fn*
Sound spectrum, 472–73
Sound track, optical, 1101
Sound waves, 460–83
displacement, 466
pressure vs. displacement, 463
(*see also* Sound)
Sounding board, 469
Sounding box, 469
Soundings, 481
Source activity, 1272
Source terminal, 857
Source(s):
of emf, 748, 838–48
in MOSFET, 1218
of sound, 460, 467–72
South pole, Earth, 784
South pole, of magnet, 783
Space:
absolute, 1057, 1061–62
curvature of, 163–64, 1336–38, 1349–51
Euclidean and non-Euclidean, 1336–37
relativity of, 1070–76
Space perception, by animals, 434
Space quantization, 1161
Space station, 136 *pr*, 141 *pr*, 145, 155, 168
Space travel, 1068
Spacecraft, 85, 216–17, 224 *pr*, 263 *pr*, 293 *pr*
Spacers, 495, 501
Space–time (4-D), 1072
curvature of, 1336–38, 1349–50
Space–time interval, 1072
Spark plug, 869
Sparking, high voltage 666, 728, 893
Speaker wires, 725
Special theory of relativity, 1055–87, 1334 (*see also* Relativity)
Specific gravity, 367, 377
Specific heat, 541–44, 546–48, 556–59
of air, 574 *pr*
defined, 541
of gases, 556–58
high, for water, 542
molar, 556–59
for solids, 558–59
SPECT, 1278
Spectrometer:
light, 1040–41
mass, 800–801
Spectroscope and spectroscopy, 1040–41, 1052 *pr*
Spectroscopic notation, 1174
Spectrum, 1038
absorption, 1040–41, 1111, 1200
atomic emission, 1040, 1111–12, 1116–18
band, 1196, 1200–1201
continuous, 1040, 1096
electromagnetic, 910, 941–43
emitted by hot object, 1096–97
line, 1040–41, 1111 *ff*, 1128
molecular, 1196–1201
visible light, 941–43
X-ray, 1169–71
Specular reflection, 928
Speed, 10, 14, 22–23
of air molecules, 519
average, 22–23, 520, 522
of EM waves, 907–9, 911
Fermi, 1206
instantaneous, 24
of light (*see separate entry below*)
of longitudinal waves, 433–35
mass and, 37
minimum, to maintain circular motion, 127–28
of molecule, 516–22
most probable, 522
of ocean waves, 459 *pr*
rms (root-mean-square), 519, 522
of top surface of water, 394 *pr*
of transverse waves, 432–33
of water waves, 459 *pr*
of waves, 430, 432–35

Speed of light, 7, 8, 15, 908–9, 912–13, 939, 997, 1057–58, 1062, 1081
constancy of, 1062
measurement of, 912–13
as ultimate speed, 1079
Speed of light principle, 1062
Speed of sound, 461, 462, 480–81
infrasonic, 462
supersonic, 462, 480–81
(*see also* Velocity, of sound)
Speed traps, 475
SPET, 1278
Sphere:
celestial, 155
gravitational force exerted on particle outside, 147
Spherical aberration, 932, 984, 985, 1033, 1037
Spherical lens, 948
Spherical mirrors, image formed by, 931–36, 982, 985
Spherical shells, Earth, 148–49, A-13–A-15
Spherical wave, 436, 443
Spider, 405
Spider web, 405
Spiderman, 191 *pr*
Spin:
boson, 1311
down, 1161, 1280
electron, 825, 1161, 1172–74, 1187
fermion, 1311
nuclear, 1229
up, 1161, 1280
Spin angular momentum, 1161
Spin-echo technique, 1281
Spin-orbit interaction, 1161, 1174
Spin quantum number, 1161
Spinal column, 340
Spinal disks, 340–41
Spine, 340–41
Spinning:
bicycle wheel, 324 *pr*
top, 317–18
Spiral galaxy, 1325
Splitting of atomic energy levels, 1208, 1280
Spontaneous processes, 577 *fn*
Spool, unrolling, 299 *pr*
Sports, 204
Spring, 179–80, 184
car, 404
effect of its own mass, 191 *pr*
hung vertically, 401
nonlinear or imperfect, 191 *pr*
potential energy of, 199, 204–5, 408–9
push start, 408
vibration of, 400 *ff*
work on, 179–80, 184
Spring constant, 179, 400
Spring equation, 179, 400
Spring scale, 86
Spring stiffness constant, 179, 400
Sprinkler, lawn, 398 *pr*
Sprockets, 192 *pr*, 299 *pr*
SPS, 1294
Spyglass, 982
Square feet (ft^2), 10
Square meters (m^2), 10
Square wave, 442
Square well potential, infinitely deep, 1142–46
Squark, 1317
Squash game, 574 *pr*
SSD, 858
Stability, 341–42
of particles, 1306–7
Stable equilibrium, 216, 341
Stable nucleus, 1233
Standard candle, 1333
Standard conditions (STP), 504–5
Standard length, 1010
Standard Model:
cosmological, 1345–48
elementary particles, 1290, 1311–14
Standard temperature and pressure (STP), 504
Standard temperature scale, 497, 507–8
Standards, 6–9
for length, 7
for mass, 8
for time, 7
Standing waves, 446–49, 467–71
fundamental frequency of, 447
mathematical representation of, 448–49
natural frequencies of, 446
resonant frequencies of, 446–47
and sources of sound, 467–72
tube diameter correction, 470
Stanford Linear Accelerator Center (SLAC), 1294
Star clusters, 1325
Stars, 1267–68, 1323–34 *ff*
Betelgeuse, 566
black holes, 164, 167 *pr*, 170 *pr*, 1326, 1331, 1337–38, 1350, 1359 *pr*
clusters of, 1325
color of, 1096, 1328
distance to, 1332–34
distant, 1333
evolution of, 1329–32
fixed, 155
H–R diagram, 1328, 1330, 1333
magnitude of, 1360 *pr*
neutron, 305, 1223 *pr*, 1326, 1331
planets around other, 158
quasars, 1326, 1336
radius, 566
red giants, 1326, 1328–30
size of, 566
source of energy of, 1267–68, 1329–31
Sun (*see* Sun)
supernovae, 1303–4, 1326, 1331–34
temperature of, 1327
types of, 1326 *ff*
variable, 1333
white dwarfs, 1326, 1328, 1330, 1360 *pr*
State (*defn*), 502
for an ideal gas, 504
bound, 1147
changes of, 522–27, 545–49
energy, in atoms, 1113–19
equation of, 502
of matter, 366, 494
metastable, 1175, 1176, 1240
as physical condition of system, 502
van der Waals, 527–28
State variables (*defn*), 502, 550, 588–89
Static electricity, 609–707
Static equilibrium, 331–42
Static friction, 118, 286–89
coefficient of, 117–18
Static rope, 115 *pr*
Static stability factor, 328 *pr*
Statics, 331–52
Stationary states in atom, 1113–20
Statistics:
Bose–Einstein, 1204 *fn*
entropy and, 595–97
Fermi–Dirac, 1204–7
Stator, 848
Steady-state model of universe, 1341
Stealth aircraft, 1022 *pr*
Steam engine, 576, 578, 606 *pr*
Steam power plants, 1265
Stefan-Boltzmann constant, 564
Stefan-Boltzmann law (or equation), 564, 1327
Stellar evolution, 1329–32
Stellar fusion, 1267–69
Step:
motion over, 299 *pr*
and wheel, 299 *pr*
Step-down transformer, 851
Step-up transformer, 851
Stepladder, 358 *pr*
Stereo, 919 *fn*
Sterilization, 1277
Stern-Gerlach experiment, 1172–73
Stimulated emission, 1175–78
Stirling cycle, 607 *pr*
STM, 1109, 1151, 1156 *pr*
Stock market illegal trading at speed of light, 945 *fn*
Stokes's theorem, A-16–A-17
Stomata, 531
Stone arch, 350–51
Stopping a car, 28, 35, 183, 288–89
Stopping potential, 1098
Stopping voltage, 1098
Storage, of information, 857
Storage rings, 1294
Stove, induction, 842
STP, 504
Strain, 344
Strain gauge, 746 *pr*
Strange particles, 1307, 1308
Strange quark, 1308
Strangeness, 1305 *fn*, 1307–8
conservation of, 1307
Strassman, Fritz, 1261
Straw, 371, 374
Streamline (*defn*), 378, 380
Streamline flow, 378–81
Strength of materials, 342, 346
Stresses, 344
compressive, 344–45
shear, 344–45
tensile, 344–45
thermal, 501
Stretchable ropes, 193 *pr*
"String telephone," 484 *pr*
String theories, 1317
Stringed instruments, 447, 467–69
Strings, vibrating, 446–49, 467–69
Stripping nuclear reaction, 1285 *pr*
Stromboli, 73
Strong bonds, 1187–89, 1192–93, 1202–3
Strong nuclear force, 161–62, 1232, 1259 *fn*, 1296–1317, 1334
and elementary particles, 1296–1317
Strongly interacting particles (*defn*), 1305
Structure:
fine, 1128, 1159, 1161, 1174
of universe, 1348–49
Struts, 347–48
Stunt driver, 80 *pr*
Subclavian artery, 384
Subcritical reactions, 1264, 1266
Sublimation, 523
Sublimation point, 523
Subpixels, 680
Subscripts, for velocity, 73
Subshells, atomic, 1167, 1168
Subtraction:
significant figures in, 5
of vectors, 57–58

Suction, 374
Sum, of vectors, 55–61, 89, 97
Sun, 1267–68, 1324, 1326–30
  energy source of, 1267–68, 1329
  in geocentric vs. heliocentric systems, 155
  mass determination, 157–58
  period, 324 *pr*
  radiation from, 563, 566
  as reference frame, 159
  reference frame of, 159
  surface temperature of, 1096
Super Proton Synchrotron (SPS), 1294
Supercluster, 1325–26
Superconducting magnets, 826
Superconductivity, 735–36
Supercritical reactions, 1264, 1266
Superdome (New Orleans, LA), 352
Superfluidity, 524
Supernovae, 1303–4, 1326, 1331–34
  as source of elements on Earth, 1331
  type Ia, 1332, 1333, 1352
Superposition principle, 441–42 *ff*, 472–73, 616, 620, 1266 *pr*
Supersaturated vapor, 526
Supersonic speed, 462, 480
Superstring theory, 1317
Supersymmetry, 1314, 1317
Supply voltage (*defn*), 676
Surface area formulas, A-1, inside back cover
Surface charge density, 706
Surface of last scattering, 1344
Surface tension, 386–88
Surface waves, 435, 443
Surfactants, 387
Surge protector, 877; (*see also* 729, 766, 826, 859)
Surgery, laser, 1178
Suspension bridge, 349, 363 *pr*
SUSYs, 1317
SUV rollover, 328 *pr*
Sway, tall buildings, 426 *pr*
Sweet spot, baseball bat, 330 *pr*
Swing, pushing child on a, 417
Symmetry, 12, 41, 97, 146, 167 *pr*, 216, 242, 247, 314, 334, 348, 351, 613 *fn*, 615, 621, 623, 630, 631, 645, 647, 649–51, 699, 702, 788, 816–22, 855, 899, 901, 906, 935, 968, 1003, 1078, 1106, 1308, 1315, 1317, 1345
Symmetry breaking, 1315, 1345
Synapse, 736
Synchrocyclotron, 1293
Synchrotron, 1293
Synchrotron radiation, 1293
Synodic period, 161, 169 *pr*
System (*defn*), 100, 198, 231, 244, 303, 542
  closed, 542
  isolated, 231, 542–43
  open, 542
  other, Kepler's laws for, 158
  for potential energy, 198
  as set of objects, 100
  for thermodynamic problem-solving, 600
  of units, 8–9
  of variable mass, 251–53
  and work-energy principle vs. energy conservation, 210
Système International (SI), 8–9, inside front cover
Systole phase, 388
Systolic pressure, 389

t-RNA, 1194–95
Tacoma Narrows Bridge, 418
Tail-to-tip method of adding vectors, 56–57
Tall buildings, sway, 426 *pr*
Tangent, 25, 59
Tangential acceleration, 133–34, 267–68
Tape measure, 510 *pr*, 513 *pr*
Tape recorder, 828, 856–57
Tarzan, 220 *pr*
Tau lepton, 1301, 1304, 1305, 1310
Tau lepton number, 1301–3, 1305, 1310
Tau neutrino, 1304, 1305
Teapot, 543, 565
Technetium-99, 1278
Technology generation, 1219
Teeth, sprocket, 299 *pr*
Telephone, cell, 852, 898, 911, 920, 1015
Telephoto lens, 975
Telescope(s), 980–82, 1035–36
  Arecibo, 1036
  astronomical, 980–82
  Galilean, 980, 982
  Hale, 982
  Hubble Space (HST), 1034–35, 1336, 1340
  Keplerian, 980–81
  Large Binocular, 982
  magnification of, 981
  reflecting, 982
  refracting, 980–81
  resolution of, 1035–36
  space, 1034–35, 1336, 1340
  terrestrial, 982
Television, 679–82, 918–20, 1015
TEM, 1109
Temperature, 495–97, 502–8, 597–98
  absolute, 502–3, 507–8, 597–98
  of boiling water, and altitude, 526–27
  in calorimetry problem-solving, 547
  Celsius (or centigrade), 496–97
  chemical reactions depending on, 521
  converting, 496
  critical, 523
  Curie, 825, 829
  difference in, as driver of heat engine, 579
  distinguished from heat and internal energy, 540
  Fahrenheit, 496
  Fermi, 1225 *pr*
  human body, 496, 549
  ideal gas scale, 507–8, 582
  Kelvin, 503, 507–8, 597–98
  molecular interpretation of, 516–19
  molecular speeds vs., 521
  "normal" body, 496
  operating (of heat engine), 578
  and rate of chemical reaction, 521
  relation to average kinetic energy of molecules, 518
  relation to molecular kinetic energy, 518–19, 541, 557–58
  relation to molecular velocities, 516–22
  scales of, 496, 502–3, 507–8, 582
  of star, 1327
  in thermodynamic problem-solving, 600
  transition, 735
Temperature coefficient of resistivity, 724, 726
Temperature gradient, 561
Tendon, 339, 358 *pr*
Tennis, 229, 234
  serve, 81 *pr*, 229, 234
Tennis ball curve, 384
Tensile strength, 346
Tensile stress, 344–45
Tension:
  electric, 718
  in flexible cord, 99–102, 196
  as stress, 344–45
Terminal, of battery, 719, 758, 759, 780 *pr*
Terminal velocity, 38 *fn*, 134–35
Terminal voltage, 748–49
Terrestrial telescope, 982
Tesla (T) (unit), 787
Test charge, 618–19
Testing, of ideas/theories, 2
Tetherball, 136 *pr*
Tevatron, 1293, 1294
TFT, 681
TFTR, 1271
Themal insulation, *R*-values, 562
Theories (general), 2, 3
Theories of everything, 1317
Therm (unit), 539
Thermal conductivity, 561
Thermal conductors, 561
Thermal contact, 497
Thermal energy (*defn*), 207, 540
  transformation of electric to, 726
  (*see also* Internal energy)
Thermal equilibrium, 497, 502, 538, 543, 547
Thermal expansion, 497–501
  anomalous behavior of water below 4° C, 500–501
  coefficients of, 498, 499
  linear expansion, 498–99
  volume expansion, 499–500
Thermal insulation, 562
Thermal insulators, 561
Thermal motion, 519, 631
Thermal neutron, 1261
Thermal noise, 678
Thermal pollution, 576, 598, 600
Thermal radiation, 563–66
Thermal resistance, 562
Thermal stresses, 501
Thermal windows, 562
Thermionic emission, 679
Thermistor, 726
Thermodynamic probability, 596–97
Thermodynamic temperature scale, 597–98
Thermodynamics, 538–99
  first law of, 538–67
  second law of (*see* Second law of thermodynamics)
  third law of, 588 *fn*, 598
  zeroth law of, 497
Thermograms, 567
Thermography, 567
Thermoluminescent dosimeter (TLD) badge, 1274
Thermometers, 495, 497, 507–8
  bimetallic-strip, 495
  constant-volume gas, 497
  liquid-in-glass, 495, 496
  mercury-in-glass thermometer, 495, 496
  resistance, 726
  in shade, 568 *pr*
Thermonuclear devices, 1269
Thermonuclear runaway, 1332
Thermos bottle, 568 *pr*
Thermostat, 509 *pr*
Thin-film interference, 1004–9
Thin-film transistor (TFT), 681
Thin lens equation, 962–65
Thin lenses, 959 *ff*
Third law of motion, 91–94
Third law of thermodynamics, 588 *fn*, 598
Thomson, G. P., 1107
Thomson, J. J., 797–98, 1107, 1108
Thought experiment, 1063 *ff*
  definition, 1063

Three-dimensional waves, 435, 436
Three Mile Island, 1265
Three-phase ac, 895 *pr*, 889
Three-way lightbulb, 778 *pr*
Threshold energy, 1259, 1288 *pr*
Threshold of hearing, 464, 467
Threshold of pain, 464, 467
Thrust, 252, 395 *pr*
Thunder, 461
Thymine (T), 631
TIA, 384
Tidal wave, 430
Tides, ocean, 159, 170 *pr*
Tight lid, opening, 499
Tightrope walker, 291 *pr*, 362 *pr*
Timbre, 472
Time:
  absolute, 1057
  characteristic expansion, 1341
  interference in, 474–76
  lookback, 1344
  Planck, 18 *pr*, 1127 *pr*, 1316, 1345
  proper, 1066, 1321 *pr*
  relativity of, 1062–69, 1072–76
  standard unit of, 7
Time constant, 760, 875–76, 1243
Time-dependent Schrödinger equation, 1138–40
Time dilation, 1064–69, 1075
Time-independent Schrödinger equation, 1136–40
Time intervals, 23, 33, 34, 67, 68
Time's arrow, 593, 601 *pr*
Tire gauge, 372
Tire iron, 272
Tire pressure:
  cold, 506
  low, 514 *pr*
Tire swing, 420 *pr*
Tokamak, 1270–71
Tokamak Fusion Test Reactor (TFTR), 1271
Tomographic image reconstruction, 1046–47
Tomography, 1045, 1278–79
Ton, 8, 17 *pr*, inside front cover
Tone color, 472
Toner, 619
Toothbrush, electric, 854
Top, spinning, 317–18
Top quark, 1308
Topness, 1309
Topographic map, 671
Toroid, 820, 827
Toroidal field, 1271
Torque, 271–76 *ff*, 308 *ff*
  counter, 849
  on current loop, 793–94
  on dipole, 630
  general definition of, 308
  relation of angular momentum and, 310–15
  in statics problems, 334
  vector, 308–9
Torr (unit), 373
Torricelli, Evangelista, 373, 374, 382
Torricelli's theorem, 382
Torsion balance, 613
Torsion pendulum, 413
Torso, 340
Total angular momentum, 1173
Total binding energy, 1093 *pr*, 1231
Total cross section, 1260
Total internal reflection, 456 *pr*, 943–45, 1151
Total kinetic energy, 284–86
Total linear momentum, of system, 250
Total magnifying power, 981
Total reaction cross reaction, 1260
Total solar eclipse, 293 *pr*
Tower crane, 355 *pr*
Townsend, J. S., 798
Tracers, 1277–78
Tracheae, 536 *pr*
Track width, 328 *pr*
Trade winds, 321
Tragic substitution, 2, 332, 346–47
Trampoline, 221 *pr*
Transducers, 894 *pr*
Transfer-RNA (t-RNA), 1194–95
Transformation of energy, 207, 213
Transformations:
  Galilean, 1072–74
  Lorentz, 1075–76
Transformer, 851–54, 871
Transformer equation, 851
Transfusions, intravenous, 396 *pr*
Transient ischemic attack (TIA), 384
Transistors, 1212, 1218–19
Transition elements, 1168
Transition temperature, 735
Transitions, in atoms, 1113–19, 1161, 1169–71, 1175–77
  forbidden and allowed, 1162–63, 1176 *fn*, 1196–97, 1199, 1200
  for lasers, 1175–77
  in molecules, 1196–1201
  in nuclei, 1239–40
  in semiconductors, 1209–12
Translational kinetic energy, 181–85
Translational motion, 21–253
  and center of mass (CM), 248–50, 284–86
  kinematics for, 20–45, 54–76
  object as point particle in problems of, 98
Transmission axis, 1011
Transmission coefficient, 1150
Transmission electron microscope (TEM), 1108
Transmission grating, 1038 *ff*
Transmission lines, 853, 912
Transmission of electricity, 853
  wireless, 854
Transmission rate, 676, 678
Transmutation of elements, 1234, 1257–59
Transuranic elements, 1259
Transverse waves, 431 *ff*
  and earthquakes, 435
  EM waves, 905
  speed of, 432–33
Traveling sinusoidal wave, 437–39
Traveling wave, 437–39
Traverse, climber's, 110 *pr*
Trees, offsetting carbon dioxide, 608 *pr*
Triangle, on a curved surface, 1336
Triangles:
  congruent, A-2
  similar, A-2
Triangulation, 12, 1332 *fn*
Triceps muscle, 295 *pr*
Trigonometric functions and identities, 59, 60, A-4–A-5, inside back cover
Trigonometric table, A-5
Triple point, 523
  of water, 508
Tritium, 1227, 1254 *pr*, 1270
Tritium dating, 1254 *pr*
Triton, 1227
Trough, 430
Trumpets, 465
Truss bridge, 347–49
Trusses, 347–49, 360 *pr*
TS diagram, 605 *pr*, 607 *pr*
Tsunami, 430, 459 *pr*
Tubes:
  closed, 470–72
  closed at both ends, 487 *pr*
  flow in, 379–82, 384–86
  open, 470–72
  vibrating column of air in, 467 *ff*
Tug-of-war, 255 *pr*
Tumors, 482
Tuning:
  an instrument, 468
  orchestra, 475
  a piano, 475
Tuning forks, 474, 476
Tunnel diode, 1151
Tunneling, 858
  of light wave, 1150–51
  through a barrier, 1149–51, 1236
Turbine, 254 *pr*, 598, 847
Turbulence, 379, 383, 386
Turbulent flow, 379, 383, 386
Turn signal, automobile, 763
Turning points, in potential energy diagrams, 215–16
Turntable, 294 *pr*, 318–20, 324 *pr*
  rotating, 307
Tuscany, Grand Duke of, 37
Twin paradox, 1068
Two-dimensional waves, 428, 435, 443–45, 449–50
Type Ia supernovae (SNIa), 1332, 1333, 1352
Typhoons, 321

UA1 detector, 1298
UHD (Ultra high definition), 680
Ultimate speed, 1079
Ultimate strength, 342, 345–46
Ultracapacitors, 714 *pr*
Ultracentrifuge, 126
Ultra high definition (UHD), 680
Ultrasonic frequencies, 462, 482–83
Ultrasonic waves, 462, 479, 482–83
Ultrasound, 482
Ultrasound medical imaging, 482–83
Ultraviolet (UV) light, 910, 941
Unavailability of energy, 594–95
Unbanked curves, 130–32
Uncertainty (in measurements), 3–6, 1131–35
  estimated, 4
  percent, 4, 6
Uncertainty principle, 1131–35, 1148, 1187
  and particle resonance, 1307
  and tunneling, 1236
Underdamped system, 415
Underexposure, 972
Underground:
  air circulation, 384
  animals, 384
Underwater vision, 978
Unification distance, 1321 *pr*
Unification scale, 1315
Unified atomic mass units (u), 8, 493, 1229
Unified (basis of forces), 1313
Unified theories, grand (GUT), 162, 1314–16
Uniform circular motion, 123–29
  dynamics of, 126–29
  kinematics of, 123–26
  simple harmonic motion related to, 410
Uniform sphere, gravitational force exerted on particle outside, 147
Uniformly accelerated motion, 30 *ff*, 65 *ff*

Uniformly accelerated rotational motion, 270–71
Unit conversion, 9–10, inside front cover
Unit vectors, 62, 177
Units of measurement, 6
converting, 9–10, inside front cover
prefixes, 8
for pressure, 367, 373
in problem solving, 33, 34, 104
and standards, 6–9
systems of, 8–9
in thermodynamic problem-solving, 600
Universal gas constant, 504
Universal law of gravitation, 145–49, 210–12, 614, 1334
Universe, 19 *pr*
age of, 1315 *fn*, 1342
Big Bang theory of, 1316, 1341 *ff*
CDM model of, 1353–54
creation of, 18 *pr*
critical density of, 1350–51
curvature of, 1336–37, 1349–50
entire, 1344
expanding, 1338–41, 1350–53
finite or infinite, 1323, 1337–38, 1342, 1350
future of, 1350–53
homogeneous, 1341
inflationary scenario of, 1346, 1348–50
isotropic, 1341
matter-dominated, 1348–50
observable, 1344
origin of elements in, 1330–31
radiation-dominated, 1347–48
Standard Model of, 1345–48
steady-state model of, 1341
Unknown quantities, 67, 68, 244
Unphysical solutions, 34, 43
Unpolarized light, 1011
Unstable equilibrium, 216, 341
Unstable nucleus, 1233 *ff*
Unwanted floor vibrations, 405
Up quark, 1308
Updated subpixels, 682
Uranium:
in dating, 1245–48
enriched, 126, 1263
fission of, 1261–66
in reactors, 1261–66
Uranus, 158
gravitational slingshot from, 217
period and distance from Sun, 330 *pr*
Useful magnification, 1037
UV light, 910, 941
Vacuum, 374, 380
in gas, 534 *pr*
Vacuum energy, 1352
Vacuum pump, 374, 388
Vacuum state, 1300, 1349
Vacuum tube, electron in, 663
Valence, 1168
Valence band, 1208–10
Validity, range of, 33
Valves:
heart, 388
pump, 388
Van de Graaff generator, 628, 691 *pr*
van der Waals, J. D., 527
van der Waals bonds and forces, 1192–96, 1202–3
van der Waals equation of state, 527–28
van der Waals gas, 527–28
Vapor:
gas vs., 523 (*see also* Gases)
supersaturated, 526
Vapor pressure, 524–26
Vaporization, latent heat of, 545–49
Variable acceleration, 43–44
Variable capacitance, 715 *pr*
Variable mass systems, 251–53
Variable stars, 1333
Variables, keeping constant, 503
Vector angular velocity, 306–7, 312, 314, 325 *pr*
Vector cross product, 307–9
Vector displacement, 22, 55–57, 62–63
Vector equations, 56
Vector field, 626
Vector form of Coulomb's law, 618
Vector model (atoms), 1184 *pr*
Vector product, 307–9
Vectors, 22, 55–65, 176–77, 307–8
addition of, 55–61
angular momentum, 306, 309
average acceleration, 63
average velocity, 63
components of, 58–62
cross product, 307–9
direction of, 22, 55
displacement, 22, 55–57
force, addition of, 89
instantaneous acceleration, 63
instantaneous velocity, 63
kinematics, 62–75
magnetization, 829
magnitude of, 22, 55
multiplication, by a scalar, 58
multiplication of, 176–77, 307–9
negative of (*defn*), 57
parallelogram method of adding, 57
position, 62–65
Poynting, 913–15
pseudo-, 270
under reflection, 270
resolution of, 58–61, 98, 99
resultant, 55–57, 60–61
scalar (dot) product, 176–77
subtraction of, 57–58
sum, 55–61, 89, 97, 149, 230
tail-to-tip method of adding, 56–57
torque, 308–9
unit, 62
Vehicle static stability factor, 328 *pr*
Veins, 389
Velocity, 14, 22–26, 63
and acceleration, 28, 40
addition of, 73–75, 1076
angular, 266–71
average, 22–24, 63
drift, 732–35, 799
of EM waves, 905–9
escape, 212, 223 *pr*, 1351
of galaxies, 480
instantaneous, 24–26, 63
of light, 908–9, 912–13, 939, 997, 1057–58, 1062, 1081
of molecules, 519
of ocean waves, 459 *pr*
phase, 438, 439
relative, 73–75
relativistic addition of, 1076
rms, 519–22
and slope, 25–26
of sound, 461, 462
supersonic, 462, 480
terminal, 38 *fn*, 134–35
of water waves, 459 *pr*
of waves, 430
Velocity-dependent forces, 134–35
Velocity gradient, 385
Velocity selector, 792
Veneziano, Gabriele, 1317
Ventricle, heart, 388–89
Ventricular fibrillation, 703, 764
Venturi meter, 384, 395 *pr*
Venturi tube, 384
Venus, 155, 156, 167 *pr*, 980
Vertebra, 340
Vertebral artery, 384
Vertical circle, revolving ball in, 128–29
Vertical (*defn*), 94 *fn*
Vertical spring, 401
Vibrating strings, 446–49, 468–69
Vibration, 399–419
of air columns, 469–72
of floor, 406
forced, 417–19
molecular, 541, 557–58, 1198–1201
as source of waves, 430
of spring, 400 *ff*
on strings, 446–48, 467–69
(*see also* Oscillations)
Vibrational energy, 408–10
molecular, 541, 558, 1198–1201
Vibrational quantum number, 1199
Vibrational transition, 1198–1201
Vibrato, 490 *pr*
Vinyl records, 679 *fn*
Violin, 468, 469, 473, 490 *pr*
Violin string, 446, 473
Virtual image, 929, 962
Virtual particles, 1297
Virtual photon, 1297
Virus, 7
Viscosity, 379, 385–86
coefficient of, 385
of oils, 385
Viscous force, 385–86
Visible light, wavelengths of, 910, 941–43
Visible spectrum, 941–43
Volatile memory, 858
Volleyball net, 357 *pr*
Volt-Ohm-Meter/Volt-Ohm-Milliammeter (VOM), 769
Volt (V) (unit), 662
Volta, Alessandro, 662, 693, 718–19
Voltage, 660, 662 *ff*, 719 *ff*, 748
base bias, 1218
bias, 1213
breakdown, 666
electric field related to, 664–65, 672–73
Hall, 1212
hazards of, 764–66
measuring, 767–70
peak, 730
ripple, 1214
rms, 731
signal, 676–79
supply, 676
terminal, 748–49
(*see also* Electric potential)
Voltage divider, 774 *pr*, 778 *pr*
Voltage drop, 755 (*see also* Voltage)
Voltage gain (*defn*), 1218
Voltaic battery, 718–19
Voltmeter, 767–70, 795
digital, 767, 770
Volume, 11
change under pressure, 345
work done in changes of, 552–54
Volume expansion (thermal), 499–501
coefficient of, 499
Volume formulas, A-1, inside back cover

Volume holograms, 1179
Volume rate of flow, 379
VOM, 769
von Laue, Max, 1044
Voyager 2 probe, 217
$W^{\pm}$ particles, 1298, 1304–6, 1310, 1313
Walking, 92
  the plank, 357 *pr*
Walls, radiant heating of, 565
Washing a car, 229
Watches, 1015
Water:
  anomalous behavior below 4°C, 500–501
  below 4°C, 500–501
  boiling, 203, 526–27
  cohesion of, 387–88
  density, 366–67, 377, 500–501
  dipole moment of, 672
  and electric shock, 765
  expansion of, 500–501
  from a faucet, 390 *pr*, 397 *pr*
  heavy, 1263
  holding, in a straw, 371
  latent heats of, 547
  molecule, 1189, 1190
  polar nature of, 611, 630, 672, 1189
  properties of: inside front cover
  saturated vapor pressure, 525
  specific gravity of, 367, 377
  specific heat of, 542
  thermal expansion of, 500–501
  triple point of, 508, 523
Water molecule, 299 *pr*
Water supply, 369
Water waves, 428, 429, 435, 443, 445, 449–50
  speed of, 459 *pr*
  velocity of, 459 *pr*
Waterwheel, 329 *pr*
Watson, J., 1044
Watt, James, 213 *fn*
Watt (W) (unit), 213, 727
Wave displacement, 437 *ff*, 1130
Wave equation, 440–41, 908–9
  Schrödinger, 1136–49, 1159–60, 1198, 1207
Wave forms, 448–49
Wave front, 443–44, 996
Wave function, 1129–31, 1137–51
  for H atom, 1159, 1160, 1163–66, 1187
  for square well, 1142–49
Wave intensity, 436–37, 464–67, 913–15, 1002–4, 1028–33
Wave interference, 444–45, 473–76, 995, 998–1009, 1031–33, 1187
Wave motion (*see* Light; Sound; Waves)
Wave nature of electron, 1131
Wave nature of matter, 1106–8, 1119–20
Wave number (*defn*), 438
Wave packet, 459 *pr*, 1141
Wave-particle duality:
  of light, 1105–6
  of matter, 1106–8, 1119–20, 1129–33
Wave speed, 430, 432–35, 442
Wave theory of light, 995–1048
Wave velocity, 430, 905–9 (*see also* Light; Sound)
Waveform, 472
Wavelength (*defn*), 430
  absorption, 1118
  Compton, 1103
  cutoff, 1170–71
  de Broglie, 1106–7, 1119–20, 1130, 1137, 1291
  depending on index of refraction, 942, 997
  as limit to resolution, 1036, 1291
  of material particles, 1106–8, 1119–20
Wave(s), 428–51, 903 *ff*, 909 *ff*, 995–1048
  amplitude decrease with distance, 436
  amplitude of, 401, 430, 437 *ff*, 1130
  bow, 480–81
  breaking at seashore, 435
  complex, 442, 472
  composite, 442, 472
  compression, 431, 435
  compressions, 431
  continuous (*defn*), 430
  on a cord, 429–30, 437–44, 446–49
  crest, 430
  diffraction of, 450–51, 996, 1025–44
  dispersion, 442, 942
  displacement of, 437–41
  earthquake, 435, 437, 450, 453 *pr*
  electromagnetic, 903–20 (*see also* Light)
  energy transported by, 435–37
  expansions, 431
  expansions in, 431
  frequency, 430
  front, 443, 996
  function, 1129–31, 1137–50, 1159, 1163–66
  general (non-sinusoidal), 439
  gravitational, 1354
  harmonic (*defn*), 438
  in-phase, 445
  incident, 444, 449
  infrasonic, 462
  intensity, 436–37, 464–67, 913–15
  interference of, 444–45, 473–76, 995, 998–1009, 1031–33, 1187
  light, 907–13, 995–1048, 1150–51 (*see also* Light)
  linear, 435
  longitudinal (*defn*), 431–32
  mathematical representation of, 437–39, 462–63
  of matter, 1106–8, 1119–20, 1130 *ff*
  mechanical, 428–51
  motion of, 428–51
  number, 438
  one-dimensional, 435, 437
  out-of-phase, 445
  P, 435, 437, 450
  packet, 1141
  particle velocity vs., 429
  period of, 430
  periodic (*defn*), 430
  phase of, 438, 445
  plane, 443–44, 905, 1140–41
  polarization of, 995, 1010–14
  power, 436
  pressure, 435, 462 *ff*
  pulse, 429–30, 443, 444
  radio, 910, 1036
  rarefactions, 431
  reflection of, 443–44, 455 *pr*
  refraction of, 449–50
  S, 435
  shear, 435
  shock, 480–81
  sinusoidal traveling, 437–39
  sound, 460–83, 910
  pressure vs. displacement, 463
  source of, oscillations as, 430
  speed of, 430, 432–35, 442 (*see also* Speed of light)
  spherical, 436, 443
  square, 442
  standing, 446–49, 467–72
  on a string, 446–49, 467–69
  superposition, 441–47
  surface, 435, 443
  three-dimensional, 435, 436
  tidal, 430
  transmission, 443, 455 *pr*
  transverse, 431–33, 435, 905
  traveling sinusoidal, 437–39
  trough, 430
  and tunneling, 1150–51
  two-dimensional, 435
  types of, 431–35 (*see also* Light)
  ultrasonic, 462, 479, 482–83
  velocity of, 429, 430, 905–9
  water, 428, 429, 435, 443, 445, 449–50, 459 *pr*
  (*see also* Light)
Weak bonds, 632, 633, 1192–96, 1203
Weak charge, 1313
Weak nuclear force, 161–62, 1233, 1238, 1298–1317, 1334
Weakly interacting massive particles (WIMPs), 1352
Weather, 321, 526, 573 *pr*
  and Coriolis force, 319–21, 328 *pr*
  forecasting of, 475, 480
  parcel of air, 573 *pr*
Weber (Wb) (unit), 840
Weight, 86, 88, 94–96, 149–51
  apparent, 154–55
  as a force, 88, 94
  force of gravity, 86, 94–96, 149–51
  mass compared to, 88, 94
Weightlessness, 154–55
Weinberg, S., 1313
Well:
  finite potential, 1147–49
  infinite potential, 1142–49
Wess, J., 1317
Wet surface of glass, 387
Wheatstone bridge, 778 *pr*
Wheel balancing, 314–15
Wheelchair, electric, 745 *pr*
Wheelie, 300 *pr*
Whirlpool galaxy, 1325
White dwarf star, 167 *pr*, 324 *pr*, 1326, 1328, 1330, 1360 *pr*
White light, 941–42
White-light holograms, 1179
White-light LED, 1215
Whitney, Mt., 527
Whole-body dose, 1275
Wide-angle lens, 975, 985
Width, of resonance, 1307
Wien, W., 1096–97
Wien's displacement law, 1096, 1327
Wien's radiation theory, 1096–97
Wilkinson, D., 1343
Wilkinson Microwave Anisotropy Probe (WMAP), 1343
Wilson, Robert, 1293 *fn*, 1342
WIMPs, 1352
Wind:
  as convection, 563
  force of, 263 *pr*
  pattern, 321
  topple building, 360 *pr*
Wind instruments, 469–72
Wind power, 599
Windings, 796
Windows:
  heat loss through, 562, 568 *pr*
  thermal, 562
Windshield wipers, 763
Wine bottles, 512 *pr*
Wing, airplane, 383
Wire, ground, 764–65, 766
Wire drift chamber, 1249

Wire-wound resistors, 723
Wireless communication, 898, 917–20
Wireless transmission of electricity, 854
Witten, Edward, 1317
WMAP, 1343
Wobble, of extrasolar planets, 262 *pr*
Work, 172–85, 195–200, 207–14, 282, 539–40, 549–60
  ability to do, 181
  on a backpack, 175
  to bring positive charges together, 667
  on a car, 183
  compared to heat, 549
  defined, 173, 176, 178, 207, 549 *ff*
  done by a constant force, 173–75
  done by a gas, 552 *ff*
  done by a varying force, 177–80
  done by gravity, 174–76, 195, 197–98
  done by spring force, 179–80, 184
  done depends on the process, 553
  done in rotation, 282
  done on a baseball, 172, 182
  done when energy transferred or transformed, 207, 213
  and first law of thermodynamics, 549–60
  force but no, 173
  from heat engines, 578
  on the Moon, 176
  negative, 174
  and power, 213–15
  relation to energy, 181–85, 197–200, 207–11, 213–14, 282
  rotational, 282
  to stop a car, 183
  for thermodynamic problem-solving, 600
  units of, 173
  in volume changes, 552–54
Work-energy principle, 172, 181–85, 200, 282, 1080
  derivation of, 185
  energy conservation vs., 210
  general, 208–10
  as reformulation of Newton's laws, 182
Work function, 1099, 1207
Working substance (*defn*), 578
Wrench, 294 *pr*
Wright, Thomas, 1323
Writing, data, 857

X-ray crystallography, 1044
X-ray diffraction, 1043–44
X-ray scattering, 1103
X-rays, 910, 1043–44, 1169–71, 1240, 1278–79
  and atomic number, 1169–71
  characteristic, 1169–70
  in electromagnetic spectrum, 910
  spectra, 1169–71
  X-ray imaging, 1045–47
Xerox (*see* Photocopier)
Xi (particle), 1305

YBCO superconductor, 735
Yerkes Observatory, 981
Yo-yo, 287, 301 *pr*
Yosemite Falls, 207
Yosemite National Park, 83 *pr*, 207, 392 *pr*, 516
Young, Thomas, 998, 1001
Young's double-slit experiment, 998–1004, 1031–33, 1130–31
Young's modulus, 342–44
Yttrium, barium, copper, oxygen superconductor (YBCO), 735
Yukawa, Hideki, 1296–98
Yukawa particle, 1296–98
Yukawa potential, 226 *pr*

Z (atomic number), 1166, 1169–71, 1227
Z^0 particle, 1156 *pr*, 1298, 1304–6, 1310, 1313
Z-particle decay, 1298
Zeeman effect, 809 *pr*, 1161, 1172, 1173
Zener diode, 1213–14
Zero, absolute, temperature of, 502–3, 598
Zero-point energy, 1143, 1149–50, 1156 *pr*, 1199
Zeroth law of thermodynamics, 497
Zoom:
  digital, 975, 987 *pr*
  optical, 975, 987 *pr*
Zoom lens, 975
Zumino, B., 1317
Zweig, G., 1308

Photo Credits

Cover photos top left clockwise Lucas Kane Photography, LLC; Douglas Giancoli; Richard Price/Getty Images; Giuseppe Molesini; Richard Megna/Fundamental Photographs

p. iii top left Douglas Giancoli **p. iii bottom right** DIMITAR DILKOFF/Getty Images **p. iv top left** Earth Imaging/ GettyImages **p. iv bottom right** NOAA **p. v** Marevision/AGE Fotostock **p. vi top right** Mike Timo/Getty Images **p. vi bottom right** Kevin Burke/Corbis **p. vii top left** George Resch/Fundamental Photographs **p. vii bottom right** Douglas Giancoli **p. viii** Fundamental Photographs **p. ix top left** Douglas Giancoli **p. ix bottom right** Giuseppe Molesini **p. xi** Douglas Giancoli **p. xix** Douglas Giancoli

CO–1 Justin Frable/Reuters **1–1a** Antranig M. Ouzoonian, P.E./Weidlinger Associates, Inc. **1–1b** fotoVoyager/Getty Images **1–2** Mary Teresa Giancoli **1–3a/b** Paul Silverman/ Fundamental Photographs **1–4** Pearson Education/Travis Amos **1–5a** Douglas Giancoli **1–5b** MECKES/OTTAWA/ Photo Researchers, Inc. **1–6** Art Wolfe/Getty Images **1–7** Douglas Giancoli **1–8** Larry Voight/Photo Researchers Inc. **1–11** University of Chicago/AIP/Science Source **1–13** Evan Sklar/Food Pix/Getty Images **1–15** David Parker/Science Source

CO–2 NASA/Bill Ingalls **2–8** John E. Gilmore, III **2–20** fStop Images GmbH/Alamy Stock Photo **2–21** Staten Island Academy **2–26** Scala/Art Resource **2–27** Harold E. Edgerton/© 2010 MIT. Courtesy of MIT Museum. **2–30** Fred de Noyelle/Getty Images **2–32** jsaunders84/Getty Images

CO–3 Lucas Kane Photography, LLC **3–19a** Berenice Abbott/ Getty Images **3–19b** Reproduced with permission from the American Association of Physics Teachers **3–21** Richard Megna/Fundamental Photographs, NYC **3–30a** Reproduced with permission from the American Association of Physics Teachers **3–30b** Westend61 GmbH/Alamy Stock Photo

CO–4 John F. Kennedy Space Center/NASA **4–1** Daly and Newton/Getty Images **4–4** Bettmann/Getty Images **4–5** Gerard Vandystadt/Agence Vandystadt/Science Source **4–7** David Jones/Science Source **4–10** Gene Blevins/Reuters **4–29** Lars Ternblad/Amana japan **4–31** Reinhard Rittmeyer **4–32** Prof. Nicholas Murgo **4–35** Prof. Martin Hackworth **4–38** Brian Bahr/Getty Images **4–72** Tyler Stableford/Getty Images

CO–5 left DIMITAR DILKOFF/Getty Images **CO–5 right** Grant Faint/Getty Images **5–8a** Agence Zoom/ Getty Images **5–16c** Peterfactors/Fotolia **5–22** Guido Alberto Rossi **5–45** C. Grzimek/Tierbild Okapia/Science Source **5–48** AF archive/Alamy Stock Photo **5–52** Phil Bender/Cedar Point Park LLC

CO–6 Earth Imaging/GettyImages **6–8** Douglas Giancoli **6–10** NASA **6–14** NASA **6–15a** Christopher Phufl/AP Images **6–15b** Brad Barket/BET/Getty Images **6–15c** Sean M. Haffey/ Getty Images **6–16** David Nunuk/Photo Researchers, Inc. **6–20** Mary Teresa Giancoli **6–23** NASA

CO–7 New York Times Co./Getty Images **7–26a** U.S. Department of Defense photo by Airman Kristopher Wilson,U.S. Navy **7–30** AF archive/Alamy Stock Photo

CO–8 MIT Museum / Harold Egerton **8–10** MIT Museum **8–11** John Marshall Mantel/AP Images **8–15a** Inga Spence/Getty Images **8–15b** Arch White/Alamy Stock Photo **8–21** Nick Rowe/PhotoDisc/Getty Images **8–26** The M.C. Escher Company B.V. **8–48** Bettmann/Getty Images

CO–9 Richard Megna/Fundamental Photographs NYC **9–1** ADRIAN DENNIS/Getty Images **9–8a** hkeita/ Shutterstock **9–8b** Kathy Willens/Associated Press **9–11** Ted Kinsman/Science Source **9–14** Richard Megna/ Fundamental Photographs **9–17** Lawrence Berkeley Laboratory/Science Photo Library/Science Source **9–20** Brookhaven National Laboratory **9–22** Berenice Abbott/ Photo Researchers, Inc.

CO–10 David R. Frazier/The image Works **10–8** Mary Teresa Giancoli **10–12a** Richard Megna/Fundamental Photographs NYC **10–12b** Kari Marttila/Alamy Stock Photo **10–32b** Richard Megna/Fundamental Photographs NYC **10–43** incamerastock/Alamy Stock Photo **10–44** The Canadian Press/Clement Allard/AP Images **10–51** Karl Weatherly/Getty Images **10–59** Tom Stewart/ Corbis/Getty Images **10–64** REUTERS/Kai Pfaffenbach **10–83** NurPhoto/Contributor/Getty Images

CO–11 Yuko Kawaguti **11–27c** NOAA **11–28** Stepehen Dunn/Getty Images **11–30** AMELIE-BENOIST/BSIP/AGE Fotostock

CO–12 Douglas Giancoli **12–1** Pete Leabo/AP Images **12–20** Douglas Giancoli **12–22a/b** Mary Teresa Giancoli **12–25** Michael Doolittle/Alamy Stock Photo **12–32a** Prochasson Frederic/Alamy Stock Photo **12–33a** Douglas Giancoli **12–33b** Citizen of the Planet/Alamy Stock Photo **12–35** Douglas Giancoli **12–37a** Everett Art/Shutterstock **12–37b** Douglas Giancoli **12–38** Susana Guzman/Alamy Stock Photo **12–39a** Italian Government Tourist Board; Courtesy of Yosemite Hospitality **12–51a** James Lemass/Photo Library/ Getty Images

CO–13 Marevision/AGE Fotostock **13–12** Betmann/Corbis/ Getty Images **13–21a** GARY SETTLES/Science Source **13–21b** Gary S. Settles/Science Source **13–29a** Mike Brake/ Shutterstock **13–35a** Lester V. Bergman/Getty images **13–35b** Biophoto Associates/Science Source **13–36** Larry West/Getty Images **13–38a** Alan Blank/Photoshot Holdings Ltd./Bruce Coleman **13–38b** Shipa Hampole/Reproduced with permission from the American Association of Physics Teachers **13–48** Douglas Giancoli **13–51** Adam Jones/ Science Source **13–52** Anna Russell/Reproduced with permission from the American Association of Physics Teachers **13–56** NASA

CO–14 Ford Motor Company **14–7** Paul Springett A/Alamy Stock Photo **14–9** Douglas Giancoli **14–13** Richard Megna/Fundamental Photographs, NYC **14–15** Douglas Giancoli **14–22** Taylor Devices, Inc. **14–25** Richard Megna/Fundamental Photographs, NYC **14–26a** AP Images **14–26b** Paul X. Scott **14–28** Andre Gallant/Getty Images

CO–15 Douglas Giancoli **15–23a** Douglas Giancoli **15–29** Jacynthroode/Getty Images **15–31** Richard Megna/Fundamental Photographs, NYC **15–41** Richard Megna/Fundamental Photographs, NYC

CO–16 Scala/Art Resource, NY **16–5** Willie Maldonado/Getty Images **16–10** Radius Images/Alamy Stock Photo **16–9a** Ben Clark/GettyImages **16–9b** Tony Cenicola/The New York Times **16–22** David Frazier/Sciencesource **16–24a** SuperStock **16–24b** Frohne, Estate of Mary V. **16–25b** Gary S. Settles/Photo Researchers, Inc. **16–26** Science Source **16–33** Science Photo Library/Superstock **16–43** Douglas Giancoli

CO–17 left Carlos Caetano/Shutterstock **CO–17 right** Richard Price/Getty Images **17–3** Robert & Jean Pollock/Science Source **17–4** Istituto e Museo di Storia della Scienza **17–6** Image Courtesy of Tec-Science **17–7b** JunAh/Shutterstock **17–12** Mary Teresa Giancoli **17–16** D-BASE/Getty Images **17–17** Stockbyte/Getty Images **17–20** Reed Kaestner/Getty Images **17–24** Jan Wlodarczyk/AGE Fotostock

CO–18 Dave Houser **18–8** Richard Megna/Fundamental Photographs **18–9a** NORBERT ROSING/National Geographic **18–9b** OJO Images Ltd/Alamy Stock Photo **18–9c** Tortoon/Fotolia **18–14a/b/c** Mary Teresa Giancoli, Photographer **18–16** Ashley Cooper/Alamy Stock Photo

CO–19 Mike Timo/Getty Images **19–27a/b** Science Source **19–29** Frohne, Estate of Mary V. **19–31** Stacey Bates/Shutterstock **19–38** Tobias Titz/Getty Images

CO–20 Frank Herholdt/Getty Images **20–1a/b/c** Douglas Giancoli **20–14a** Carsten Koall/Getty Images **20–14b** Kevin Burke/Corbis **20–14c** Woodfall, David **Table 20–3 top left** Larry Lee Photography/GettyImages **Table 20–3 mid-top** Michael Collier **Table 20–3 mid-bottom** micheldenijs/Getty Images **Table 20–3 bottom left** schmidt-z/GettyImages **20–18** Photo Researchers, Inc. **20–21** schmidt-z/Getty Images **20–22** Michael Collier

CO–21 George Resch/Fundamental Photographs **21–40** Peter Menzel/Science Source **21–47** Dr. Gopal Murti/Science Source **21–50** Matthew Claspill/Reproduced with permission from the American Association of Physics Teachers

CO–23 Emily Michot/MCT/Newscom **23–19** Douglas Giancoli **23–31 left** Douglas Giancoli **23–31 right** Pearson Education **23–35** beerkoff/Shutterstock **23–48** Douglas Giancoli

CO–24 Eric Schrader **24–14** Eric Schrader **24–15** Mikael Vaisanen/GettyImages

CO–25 Douglas Giancoli **25–1** Science Source **25–2** Photo 12/Alamy Stock Photo **25–6a** Richard Megna/Fundamental Photographs **25–11** Songsak Paname/iStock/Getty Images Plus **25–15** Richard Megna/Fundamental Photographs **25–16a** Paul Silverman/Fundamental Photographs **25–16b** Matthew Ashmore/EyeEm/Getty Images **25–18** clintspencer/Getty images **25–33** Jerry Marshall **25–35** Scott Smith/Getty Images

CO–26 Patrik Stollarz/Staff **26–14** David R. Frazier Photolibrary, Inc./Alamy Stock Photo **26–21** Apogee/Science Source **26–24a** Photodisc/Getty Images **26–24b** Pearson **26–25** Pearson **26–27a** Pearson **26–27b** Olaf Döring/imageBROKER **26–64** AGE Fotostock **26–77** Pearson

CO–27 Richard Megna/Fundamental Photographs **27–1** Tim Ridley/Dorling Kindersley **27–4a** Stephen Oliver/Dorling Kindersley **27–6** Mary Teresa Giancoli **27–8a/b** Richard Megna/Fundamental Photographs **27–19** Richard Megna/Fundamental Photographs **27–21b** Science Source

CO–28 Richard Megna/Fundamental Photographs **28–25** Richard Megna/Fundamental Photographs **28–33** Clive Streeter/Science Museum, London/Dorling Kindersley

CO–29 Richard Megna/Fundamental Photographs **29–8** Myrleen Pearson/Alamy Stock Photo **29–13** Jeff Hunter/Getty Images **29–17** Rick Bowmer/AP Images **29–22** Robert F. Bukaty/AP Images **29–23** Tom Carter/Alamy Stock Photo **29–29a** Terence Kearey **29–36a** Phil Degginger/Alamy Stock Photo **29–36b** Eric Schrader/Fundamental Photographs

CO–30 STAN HONDA/Staff/Getty Images

CO–31 NASA **31–1** New York Public Library/Science Source **31–13** The Image Works **31–16** Time Life Pictures/Getty Images **31–23** David J. Green/Alamy Stock Photo **31–24** Donald M. Baida **31–27** NASA

CO–32 Douglas Giancoli **32–6** Douglas Giancoli **32–11a** Mary Teresa Giancoli **32–11b** Paul Silverman/Fundamental Photographs **32–20** Federico Rostagno/Alamy Stock Photo **32–22a** Paul Burns/Getty Images **32–23** Douglas Giancoli **32–27** Science Source **32–30b** Douglas Giancoli **32–33** Cultura RM/Alamy Stock Photo **32–36** Science Source **32–42** Douglas Giancoli **32–44** Mary Teresa Giancoli **32–72** Sarah Lampen/Reproduced with permission from the American Association of Physics Teachers

CO–33 Douglas Giancoli **33–1** Douglas Giancoli **33–2c/d** Douglas Giancoli **33–4** Kari Erik Marttila Photography **33–7a/b** Douglas Giancoli **33–21** Mary Teresa Giancoli **33–22a/b** Mary Teresa Giancoli **33–24** PhotoPlus Magazine/Contributor/Getty Images **33–35a/b** Istituto e Museo di Storia della Scienza **33–37** Yerkes Observatory, Photo Services **33–38c** Sandy Huffaker/Getty Images **33–38d** Joe McNally/Contributor/Getty Images **33–40b** Leica Microsystems **33–45** Ron Chapple Photography **33–49** Reproduced with permission from the American Association of Physics Teachers **33–50** Matt Buck/Reproduced with permission from the American Association of Physics Teachers **33–51** National Weather Service **33–53a/b** Scott Dudley

CO–34 Giuseppe Molesini **34–4** KENT WOOD/Science Source **34–9** Richard Megna/Fundamental Photographs **34–16a** Giuseppe Molesini **34–16b** Richard Megna/Fundamental Photographs **34–16c** Paul Silverman/Fundamental Photographs **34–18b** Ken Kay/Fundamental Photographs **34–22** Kristen Brochmann/Fundamental Photographs **34–31** Diane Hirsch/Fundamental Photograph **34–34a/b** Jiaren-Lau Photography **34–39** Suunto Oy **34–40** Pearson Education, Inc

CO–35 Richard Megna/Fundamental Photographs **35–2a** P.M. Rinard/American Journal of Physics **35–2b** Richard Megna/Fundamental Photographs **35–2c** Ken Kay/Fundamental Photographs **35–3** Durk Gardenier/Alamy Stock Photo **35–12a/b** Richard Megna/Fundamental Photographs **35–13a/b** SpringerNature **35–16a** Space Telescope Science Institute **35–16b** NASA/ESO/NOAJ/G.Paglioli **35–17** David Parker/Science Source **35–22** Wabash Instrument Corp/Fundamental Photographs **35–27** John Wiley & Sons, Inc. **35–30** OMIKRON/Science Source **35–34a** Martin M. Rotker **35–34b** Living Art Enterprises, LLC/Science Source.

CO–36 Douglas Giancoli **36–1** Bettman/Getty Images **36–8** Douglas Giancoli **36–16** Douglas Giancoli

CO–37 left Paul Gunning/Science Source **CO–37 right** Professors Pietro M. Motta & Silvia Correr/Science Source **37–10** Photograph by Samuel Goudsmit, courtesy AIP Emilio Segre Visual Archives, Goudsmit Collection **37–11** Education Development Center **37–14a** Lee D. Simon/Science Source **37–14b** Oliver Meckes/Max–Planck-Institut-Tubingen/Photo Researchers, Inc. **37–14c** Science Photo Library/Alamy Stock Photo **37–19b** Richard Megna/Fundamental Photographs **37–20a/b/c** Wabash Instrument Corp./Fundamental Photographs

This page intentionally left blank

Fundamental Constants

Quantity	Symbol	Approximate Value	Current Best Value[†]
Speed of light in vacuum	c	3.00×10^8 m/s	2.99792458×10^8 m/s
Gravitational constant	G	6.67×10^{-11} N · m^2/kg^2	$6.67430(15) \times 10^{-11}$ N · m^2/kg^2
Avogadro's number	N_A	6.02×10^{23} mol^{-1}	$6.02214076 \times 10^{23}$ mol^{-1}
Gas constant	R	8.314 J/mol · K = 1.99 cal/mol · K = 0.0821 L · atm/mol · K	8.314462618 J/mol · K
Boltzmann's constant	k	1.38×10^{-23} J/K	1.380649×10^{-23} J/K
Charge on electron	e	1.60×10^{-19} C	$1.602176634 \times 10^{-19}$ C
Stefan-Boltzmann constant	σ	5.67×10^{-8} W/m^2 · K^4	$5.670374419 \times 10^{-8}$ W/m^2 · K^4
Permittivity of free space	ϵ_0	8.85×10^{-12} C^2/N · m^2	$8.8541878128(13) \times 10^{-12}$ C^2/N · m^2
Permeability of free space	μ_0	1.26×10^{-6} T · m/A	$1.25663706212(19) \times 10^{-6}$ T · m/A
Planck's constant	h	6.63×10^{-34} J · s	$6.62607015 \times 10^{-34}$ J · s
Electron rest mass	m_e	9.11×10^{-31} kg = 0.000549 u = 0.511 MeV/c^2	$9.1093837015(28) \times 10^{-31}$ kg = $5.48579909065(16) \times 10^{-4}$ u
Proton rest mass	m_p	1.6726×10^{-27} kg = 1.00728 u = 938.27 MeV/c^2	$1.67262192369(51) \times 10^{-27}$ kg = 1.007276466621(53) u
Neutron rest mass	m_n	1.6749×10^{-27} kg = 1.008665 u = 939.57 MeV/c^2	$1.67492749804(95) \times 10^{-27}$ kg = 1.00866491595(49) u
Atomic mass unit (1 u)		1.6605×10^{-27} kg = 931.49 MeV/c^2	$1.66053906660(50) \times 10^{-27}$ kg = 931.49410242(28) MeV/c^2

[†]Numbers in parentheses indicate one-standard-deviation experimental uncertainties in final digits (2019, new SI). Values without parentheses are exact (i.e., defined quantities).

Other Useful Data

Joule equivalent (1 cal)	4.186 J
Absolute zero (0 K)	−273.15°C
Acceleration due to gravity at Earth's surface (avg.)	9.80 m/s^2 (= g)
Speed of sound in air (20°C)	343 m/s
Density of air (dry)	1.29 kg/m^3
Earth: Mass	5.98×10^{24} kg
Radius (mean)	6.38×10^3 km
Moon: Mass	7.35×10^{22} kg
Radius (mean)	1.74×10^3 km
Sun: Mass	1.99×10^{30} kg
Radius (mean)	6.96×10^5 km
Earth–Sun distance (mean)	149.60×10^6 km
Earth–Moon distance (mean)	384×10^3 km

The Greek Alphabet

Alpha	A	α	Nu	N	ν
Beta	B	β	Xi	Ξ	ξ
Gamma	Γ	γ	Omicron	O	o
Delta	Δ	δ	Pi	Π	π
Epsilon	E	ϵ, ε	Rho	P	ρ
Zeta	Z	ζ	Sigma	Σ	σ
Eta	H	η	Tau	T	τ
Theta	Θ	θ	Upsilon	Y	υ
Iota	I	ι	Phi	Φ	ϕ, φ
Kappa	K	κ	Chi	X	χ
Lambda	Λ	λ	Psi	Ψ	ψ
Mu	M	μ	Omega	Ω	ω

Values of Some Numbers

$\pi = 3.1415927$	$\sqrt{2} = 1.4142136$	$\ln 2 = 0.6931472$	$\log_{10} e = 0.4342945$
$e = 2.7182818$	$\sqrt{3} = 1.7320508$	$\ln 10 = 2.3025851$	1 rad = 57.2957795°

Mathematical Signs and Symbols

$\propto$	is proportional to	$\leq$	is less than or equal to
$=$	is equal to	$\geq$	is greater than or equal to
$\approx$	is approximately equal to	$\sum$	sum of
$\neq$	is not equal to	$\bar{x}$	average value of x
$>$	is greater than	Δx	change in x
$\gg$	is much greater than	$\Delta x \to 0$	Δx approaches zero
$<$	is less than	$n!$	$n(n-1)(n-2)\ldots(1)$
$\ll$	is much less than		

Properties of Water

Density (4°C)	1.000×10^3 kg/m^3
Heat of fusion (0°C)	334 kJ/kg (79.8 kcal/kg)
Heat of vaporization (100°C)	2260 kJ/kg (539.9 kcal/kg)
Specific heat (15°C)	4186 J/kg · C° (1.00 kcal/kg · C°)
Index of refraction	1.33

Unit Conversions (Equivalents)

Length

1 in. = 2.54 cm (defined)
1 cm = 0.3937 in.
1 ft = 30.48 cm
1 m = 39.37 in. = 3.281 ft
1 mi = 5280 ft = 1.609 km
1 km = 0.6214 mi
1 nautical mile = 1.151 mi = 6076 ft = 1.852 km
1 fermi = 1 femtometer (fm) = 10^{-15} m
1 angstrom (Å) = 10^{-10} m = 0.1 nm
1 light-year (ly) = 9.461×10^{15} m
1 parsec = 3.26 ly = 3.09×10^{16} m

Volume

1 liter (L) = 1000 mL = 1000 cm^3 = 1.0×10^{-3} m^3 = 1.057 qt (U.S.) = 61.02 $in.^3$
1 gal (U.S.) = 4 qt (U.S.) = 231 $in.^3$ = 3.785 L = 0.8327 gal (British)
1 quart (U.S.) = 2 pints (U.S.) = 946 mL
1 pint (British) = 1.20 pints (U.S.) = 568 mL
1 m^3 = 35.31 ft^3

Speed

1 mi/h = 1.4667 ft/s = 1.6093 km/h = 0.4470 m/s
1 km/h = 0.2778 m/s = 0.6214 mi/h
1 ft/s = 0.3048 m/s = 0.6818 mi/h = 1.0973 km/h
1 m/s = 3.281 ft/s = 3.600 km/h = 2.237 mi/h
1 knot = 1 nautical mile/h = 1.151 mi/h = 1.852 km/h = 0.5144 m/s

Angle

1 radian (rad) = 57.30° = 57°18′
1° = 0.01745 rad
1 rev/min (rpm) = 0.1047 rad/s

Time

1 day = 8.640×10^4 s
1 year = 365.242 days = 3.156×10^7 s

Mass

1 atomic mass unit (u) = 1.6605×10^{-27} kg
1 kg = 0.06852 slug
1 ton (metric) = 1000 kg
[1 kg has a weight of 2.20 lb where $g = 9.80\ m/s^2$.]

Force

1 lb = 4.44822 N
1 N = 10^5 dyne = 0.2248 lb
1 ton (U.S.) = 2000 lbs

Energy and Work

1 J = 10^7 ergs = 0.7376 ft · lb
1 ft · lb = 1.356 J = 1.29×10^{-3} Btu = 3.24×10^{-4} kcal
1 kcal = 4.19×10^3 J = 3.97 Btu
1 eV = 1.6022×10^{-19} J
1 kWh = 3.600×10^6 J = 860 kcal
1 Btu = 1.056×10^3 J

Power

1 W = 1 J/s = 0.7376 ft · lb/s = 3.41 Btu/h
1 hp = 550 ft · lb/s = 746 W

Pressure

1 atm = 1.01325 bar = 1.01325×10^5 N/m^2 = 14.7 $lb/in.^2$ = 760 torr
1 $lb/in.^2$ = 6.895×10^3 N/m^2
1 Pa = 1 N/m^2 = 1.450×10^{-4} $lb/in.^2$

Some SI Units in Terms of Base Units

Quantity	Unit	Abbreviation	In Terms of Base Units†
Force	newton	N	$kg \cdot m/s^2$
Energy and work	joule	J	$kg \cdot m^2/s^2$
Power	watt	W	$kg \cdot m^2/s^3$
Pressure	pascal	Pa	$kg/(m \cdot s^2)$
Frequency	hertz	Hz	s^{-1}
Electric charge	coulomb	C	$A \cdot s$
Electric potential	volt	V	$kg \cdot m^2/(A \cdot s^3)$
Electric resistance	ohm	Ω	$kg \cdot m^2/(A^2 \cdot s^3)$
Capacitance	farad	F	$A^2 \cdot s^4/(kg \cdot m^2)$
Magnetic field	tesla	T	$kg/(A \cdot s^2)$
Magnetic flux	weber	Wb	$kg \cdot m^2/(A \cdot s^2)$
Inductance	henry	H	$kg \cdot m^2/(A^2 \cdot s^2)$

† kg = kilogram (mass), m = meter (length), s = second (time), A = ampere (electric current).

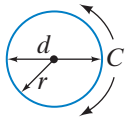
Metric (SI) Multipliers

Prefix	Abbreviation	Value
yotta	Y	10^{24}
zeta	Z	10^{21}
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^9
mega	M	10^6
kilo	k	10^3
hecto	h	10^2
deka	da	10^1
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}
zepto	z	10^{-21}
yocto	y	10^{-24}

Useful Geometry Formulas — Areas, Volumes

Circumference of circle $C = \pi d = 2\pi r$

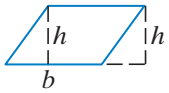
Area of circle $A = \pi r^2 = \dfrac{\pi d^2}{4}$



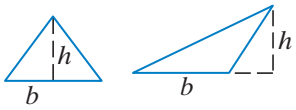
Area of rectangle $A = \ell w$



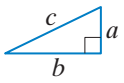
Area of parallelogram $A = bh$



Area of triangle $A = \frac{1}{2}hb$



Right triangle (Pythagoras) $c^2 = a^2 + b^2$

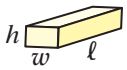


Sphere: surface area $A = 4\pi r^2$

volume $V = \frac{4}{3}\pi r^3$

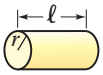


Rectangular solid: volume $V = \ell wh$



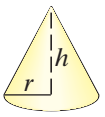
Cylinder (right): surface area $A = 2\pi r\ell + 2\pi r^2$

volume $V = \pi r^2\ell$



Right circular cone: surface area $A = \pi r^2 + \pi r\sqrt{r^2 + h^2}$

volume $V = \frac{1}{3}\pi r^2 h$



Exponents

$(a^n)(a^m) = a^{n+m}$ [Example: $(a^3)(a^2) = a^5$]

$(a^n)(b^n) = (ab)^n$ [Example: $(a^3)(b^3) = (ab)^3$]

$(a^n)^m = a^{nm}$ [Example: $(a^3)^2 = a^6$; Example: $(a^{\frac{1}{4}})^4 = a$]

$a^{-1} = \dfrac{1}{a}$ $\quad a^{-n} = \dfrac{1}{a^n}$ $\quad a^0 = 1$

$a^{\frac{1}{2}} = \sqrt{a}$ $\quad a^{\frac{1}{4}} = \sqrt{\sqrt{a}}$

$(a^n)(a^{-m}) = \dfrac{a^n}{a^m} = a^{n-m}$ [Ex.: $(a^5)(a^{-2}) = a^3$]

$\dfrac{a^n}{b^n} = \left(\dfrac{a}{b}\right)^n$

Logarithms [Appendix A–7; Table A–1]

If $y = 10^x$, then $x = \log_{10} y = \log y$.

If $y = e^x$, then $x = \log_e y = \ln y$.

$\log(ab) = \log a + \log b$

$\log\left(\dfrac{a}{b}\right) = \log a - \log b$

$\log a^n = n \log a$

Some Derivatives and Integrals†

$\dfrac{d}{dx}x^n = nx^{n-1}$ $\qquad \displaystyle\int \sin ax\,dx = -\frac{1}{a}\cos ax$

$\dfrac{d}{dx}\sin ax = a\cos ax$ $\qquad \displaystyle\int \cos ax\,dx = \frac{1}{a}\sin ax$

$\dfrac{d}{dx}\cos ax = -a\sin ax$ $\qquad \displaystyle\int \frac{1}{x}\,dx = \ln x$

$\displaystyle\int x^m\,dx = \frac{1}{m+1}x^{m+1}$ $\qquad \displaystyle\int e^{ax}\,dx = \frac{1}{a}e^{ax}$

† See Appendix B for more.

Quadratic Formula

Equation with unknown x, in the form

$$ax^2 + bx + c = 0,$$

has solutions

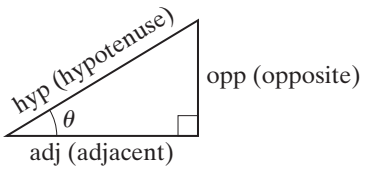
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Binomial Expansion

$$(1 \pm x)^n = 1 \pm nx + \frac{n(n-1)}{2\cdot 1}x^2 \pm \frac{n(n-1)(n-2)}{3\cdot 2\cdot 1}x^3 + \cdots \qquad [\text{for } x^2 < 1]$$

$$\approx 1 + nx \qquad \text{if } x \ll 1$$

Trigonometric Formulas [Appendix A–9]



$\sin\theta = \dfrac{\text{opp}}{\text{hyp}}$ $\quad \cos\theta = \dfrac{\text{adj}}{\text{hyp}}$ $\quad \tan\theta = \dfrac{\text{opp}}{\text{adj}}$

$\text{adj}^2 + \text{opp}^2 = \text{hyp}^2$ (Pythagorean theorem)

$\tan\theta = \dfrac{\sin\theta}{\cos\theta}$

$\sin^2\theta + \cos^2\theta = 1$

$\sin 2\theta = 2\sin\theta\cos\theta$

$\cos 2\theta = (\cos^2\theta - \sin^2\theta) = (1 - 2\sin^2\theta) = (2\cos^2\theta - 1)$

$\sin(180° - \theta) = \sin\theta$; radians: $\sin(\pi - \theta) = \sin\theta$

$\cos(180° - \theta) = -\cos\theta$

$\sin(90° - \theta) = \cos\theta$

$\cos(90° - \theta) = \sin\theta$

$\sin\frac{1}{2}\theta = \sqrt{(1 - \cos\theta)/2}$ $\qquad \cos\frac{1}{2}\theta = \sqrt{(1 + \cos\theta)/2}$

$\sin\theta \approx \theta$ [for small $\theta \lesssim 0.2$ rad]

$\cos\theta \approx 1 - \dfrac{\theta^2}{2}$ [for small $\theta \lesssim 0.2$ rad]

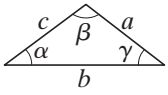
$\sin(A \pm B) = \sin A\cos B \pm \cos A\sin B$

$\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B$

For any triangle:

$c^2 = a^2 + b^2 - 2ab\cos\gamma$ (law of cosines)

$\dfrac{\sin\alpha}{a} = \dfrac{\sin\beta}{b} = \dfrac{\sin\gamma}{c}$ (law of sines)



Periodic Table of the Elements[§]

Masses averaged over natural abundances

Group I	Group II	Transition Elements										Group III	Group IV	Group V	Group VI	Group VII	Group VIII
H 1 1.00796 ← averaged over natural abundances of $^{1}_{1}H$, $^{2}_{1}H$ $1s^1$																	**He** 2 4.002602 $1s^2$
Li 3 6.941 $2s^1$	**Be** 4 9.0121831 $2s^2$											**B** 5 10.814 $2p^1$	**C** 6 12.0109 $2p^2$	**N** 7 14.00675 $2p^3$	**O** 8 15.99936 $2p^4$	**F** 9 18.99840316 $2p^5$	**Ne** 10 20.1797 $2p^6$
Na 11 22.98976928 $3s^1$	**Mg** 12 24.305 $3s^2$											**Al** 13 26.9815384 $3p^1$	**Si** 14 28.085 $3p^2$	**P** 15 30.9737620 $3p^3$	**S** 16 32.064 $3p^4$	**Cl** 17 35.453 $3p^5$	**Ar** 18 39.948 $3p^6$
K 19 39.0983 $4s^1$	**Ca** 20 40.078 $4s^2$	**Sc** 21 44.955908 $3d^14s^2$	**Ti** 22 47.867 $3d^24s^2$	**V** 23 50.9415 $3d^34s^2$	**Cr** 24 51.9961 $3d^54s^1$	**Mn** 25 54.938043 $3d^54s^2$	**Fe** 26 55.845 $3d^64s^2$	**Co** 27 58.933194 $3d^74s^2$	**Ni** 28 58.6934 $3d^84s^2$	**Cu** 29 63.546 $3d^{10}4s^1$	**Zn** 30 65.38 $3d^{10}4s^2$	**Ga** 31 69.723 $4p^1$	**Ge** 32 72.630 $4p^2$	**As** 33 74.921595 $4p^3$	**Se** 34 78.971 $4p^4$	**Br** 35 79.904 $4p^5$	**Kr** 36 83.798 $4p^6$
Rb 37 85.4678 $5s^1$	**Sr** 38 87.62 $5s^2$	**Y** 39 88.90584 $4d^15s^2$	**Zr** 40 91.224 $4d^25s^2$	**Nb** 41 92.90637 $4d^45s^1$	**Mo** 42 95.95 $4d^55s^1$	**Tc** 43 (98) $4d^55s^2$	**Ru** 44 101.07 $4d^75s^1$	**Rh** 45 102.90549 $4d^85s^1$	**Pd** 46 106.42 $4d^{10}5s^0$	**Ag** 47 107.8682 $4d^{10}5s^1$	**Cd** 48 112.414 $4d^{10}5s^2$	**In** 49 114.818 $5p^1$	**Sn** 50 118.710 $5p^2$	**Sb** 51 121.760 $5p^3$	**Te** 52 127.60 $5p^4$	**I** 53 126.90447 $5p^5$	**Xe** 54 131.293 $5p^6$
Cs 55 132.9054519 $6s^1$	**Ba** 56 137.327 $6s^2$	57–71[†]	**Hf** 72 178.49 $5d^26s^2$	**Ta** 73 180.94788 $5d^36s^2$	**W** 74 183.84 $5d^46s^2$	**Re** 75 186.207 $5d^56s^2$	**Os** 76 190.23 $5d^66s^2$	**Ir** 77 192.217 $5d^76s^2$	**Pt** 78 195.084 $5d^96s^1$	**Au** 79 196.966570 $5d^{10}6s^1$	**Hg** 80 200.592 $5d^{10}6s^2$	**Tl** 81 204.383 $6p^1$	**Pb** 82 207.2 $6p^2$	**Bi** 83 208.98040 $6p^3$	**Po** 84 (209) $6p^4$	**At** 85 (210) $6p^5$	**Rn** 86 (222) $6p^6$
Fr 87 (223) $7s^1$	**Ra** 88 (226) $7s^2$	89–103[‡]	**Rf** 104 (267) $6d^27s^2$	**Db** 105 (268) $6d^37s^2$	**Sg** 106 (269) $6d^47s^2$	**Bh** 107 (271) $6d^57s^2$	**Hs** 108 (277) $6d^67s^2$	**Mt** 109 (276) $6d^77s^2$	**Ds** 110 (281) $6d^97s^1$	**Rg** 111 (282) $6d^{10}7s^1$	**Cn** 112 (285) $6d^{10}7s^2$	**Nh** 113 (286) $6d^{10}7s^27p^1$	**Fl** 114 (289) $6d^{10}7s^27p^2$	**Mc** 115 (290) $6d^{10}7s^27p^3$	**Lv** 116 (293) $6d^{10}7s^27p^4$	**Ts** 117 (294) $6d^{10}7s^27p^5$	**Og** 118 (294) $6d^{10}7s^27p^6$

[†]**Lanthanide Series**	**La** 57 138.90547 $5d^16s^2$	**Ce** 58 140.116 $4f^15d^16s^2$	**Pr** 59 140.90766 $4f^35d^06s^2$	**Nd** 60 144.242 $4f^45d^06s^2$	**Pm** 61 (145) $4f^55d^06s^2$	**Sm** 62 150.36 $4f^65d^06s^2$	**Eu** 63 151.964 $4f^75d^06s^2$	**Gd** 64 157.25 $4f^75d^16s^2$	**Tb** 65 158.925354 $4f^95d^06s^2$	**Dy** 66 162.500 $4f^{10}5d^06s^2$	**Ho** 67 164.930328 $4f^{11}5d^06s^2$	**Er** 68 167.259 $4f^{12}5d^06s^2$	**Tm** 69 168.934218 $4f^{13}5d^06s^2$	**Yb** 70 173.045 $4f^{14}5d^06s^2$	**Lu** 71 174.9668 $4f^{14}5d^16s^2$
[‡]**Actinide Series**	**Ac** 89 (227) $6d^17s^2$	**Th** 90 232.0377 $6d^27s^2$	**Pa** 91 231.03588 $5f^26d^17s^2$	**U** 92 238.02891 $5f^36d^17s^2$	**Np** 93 (237) $5f^46d^17s^2$	**Pu** 94 (244) $5f^66d^07s^2$	**Am** 95 (243) $5f^76d^07s^2$	**Cm** 96 (247) $5f^76d^17s^2$	**Bk** 97 (247) $5f^96d^07s^2$	**Cf** 98 (251) $5f^{10}6d^07s^2$	**Es** 99 (252) $5f^{11}6d^07s^2$	**Fm** 100 (257) $5f^{12}6d^07s^2$	**Md** 101 (258) $5f^{13}6d^07s^2$	**No** 102 (259) $5f^{14}6d^07s^2$	**Lr** 103 (262) $5f^{14}6d^17s^2$

[§] Atomic mass values averaged over isotopes in percentages they occur at Earth's surface (2019). Individual isotopes are found in Appendix G. For many unstable elements, mass of the longest-lived known isotope is given in parentheses.