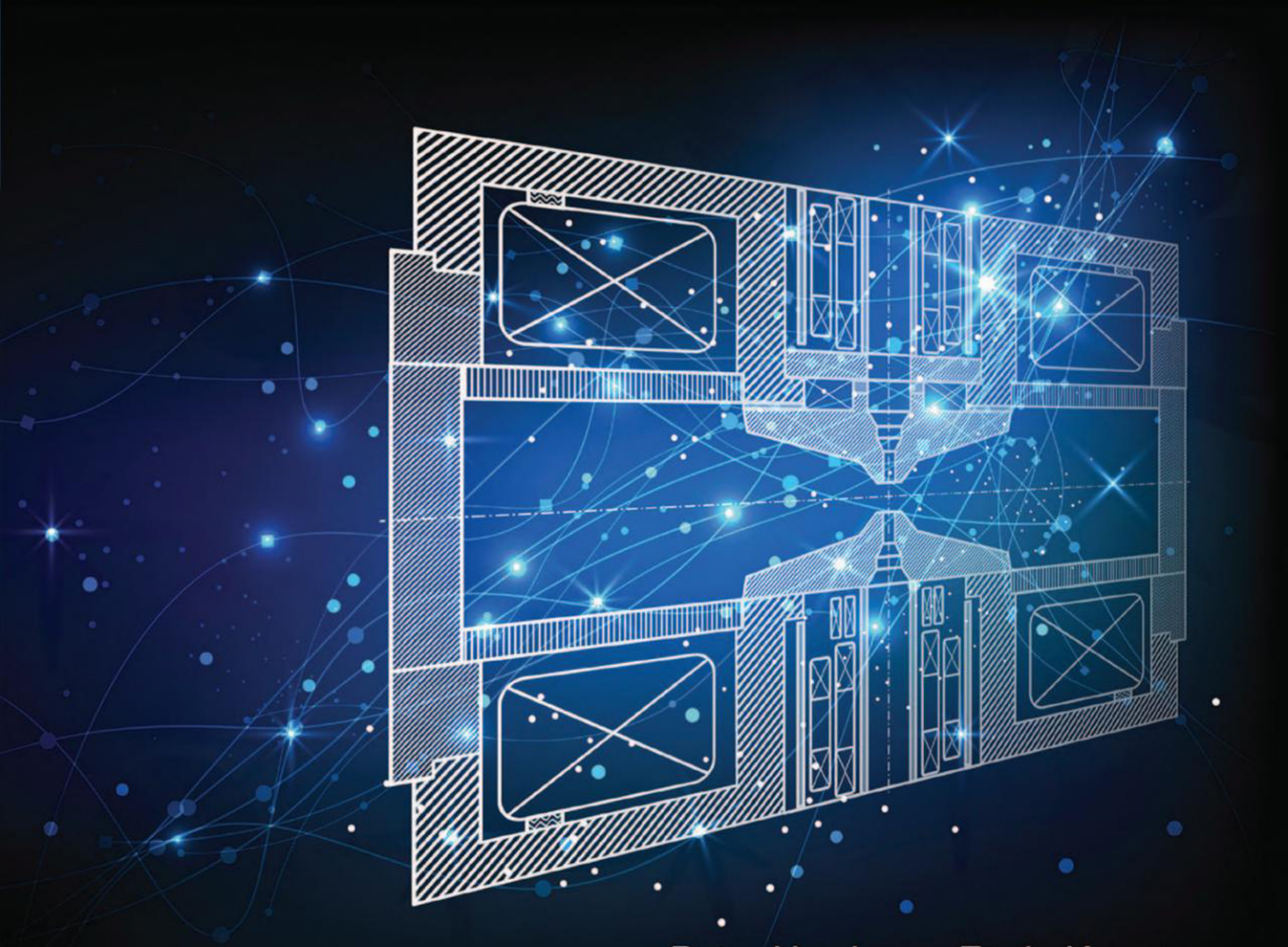


Peter Hawkes • Erwin Kasper

Principles of Electron Optics

Volume One: Basic Geometrical Optics

Second Edition

Principles of Electron Optics

Volume One: Basic Geometrical Optics

Principles of Electron Optics

Volume One: Basic Geometrical Optics

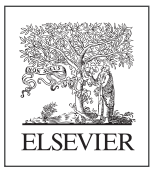
Second Edition

Peter Hawkes

CEMES—CNRS, Toulouse, France

Erwin Kasper

Formerly Institute of Applied Physics,
University of Tübingen, Germany



ACADEMIC PRESS

An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1800, San Diego, CA 92101-4495, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2018 Elsevier Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-08-102256-6

For Information on all Academic Press publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: John Fedor
Acquisition Editor: Anita Koch
Editorial Project Manager: Brianna Garcia
Production Project Manager: Vijayaraj Purushothaman
Cover Designer: Christian Bilbow

Typeset by MPS Limited, Chennai, India

Preface to the Second Edition

In the quarter century since the first edition of the *Principles* appeared, many branches of electron optics have been developed considerably, often driven by the success of aberration correction in the 1990s. At the end of the 1980s, we echoed the general opinion that quadrupole–octopole correctors had been given their chance, notably in Darmstadt and Chicago, but had failed to reach their goal. But the germs of the revolution in aberration studies that began in the next decade were already present: the degree of correction attained in Scherzer's project was impressive and Harald Rose had outlined a promising corrector based on sextupoles in 1981. What was missing were the indispensable diagnostic tools and fast feedback control circuitry that would turn these complicated systems into working correctors. These became available in the 1990s and the various correctors that have emerged are now examined in detail here. A different type of aberration corrector based on electron mirrors has also gained importance, especially in the low-energy-electron microscope (LEEM) and the photoemission electron microscope (PEEM). The chapters on the optics of electron mirrors have therefore been expanded to include the work of Dirk Preikszas and the parallel studies of the Russian school.

The third-order geometrical aberrations of round lenses were very fully covered in the first edition but aberration integrals for the fifth-order aberrations were not included. Such formulae had been published by one of us but a much improved set of aberration integrals has subsequently been derived by Zhixiong Liu. These are reproduced in Chapter 24, The Geometrical Aberrations of Round Lenses. The chapter on parasitic aberration too has needed much revision. We have preferred to retain the simple explanations in the first edition and then add new material covering recent work on the subject.

Another topic that has been extensively studied is electron emission, especially in the hands of Kevin Jensen, Christopher Edgcombe and Richard Forbes. We have incorporated some of their work and included many references covering aspects not discussed here. In particular, we describe the recommendations of Forbes and Deane for reformulating the Fowler–Nordheim theory.

In some cases, we have felt justified in removing material. Thus computer algebra, which was not very familiar in 1989, is now in routine use and several packages are available to perform it.

In the wake of aberration correction, monochromators have gained in importance. With the correction of spherical aberration, some way of reducing the effect of chromatic aberration was needed in order to benefit fully from the potential improvement in resolution. Certainly, correctors of chromatic aberration were known and have been implemented in practice but they introduce a further degree of complexity into microscope design. The alternative, much preferred by commercial microscope manufacturers, is to reduce the energy spread of the beam emerging from the source. Monochromator optics has therefore been perfected and we give some account of this in Part X. The optics of Wien filters was covered very superficially in the first edition. A new Chapter (38, The Wien Filter) now provides a much fuller treatment. In the case of topics still in rapid development, we have included only short accounts and many references. Ultrafast electron microscopy and multiple-beam systems for high-throughput electron lithography and scanning electron microscopy are the main examples of these.

Some material is admittedly of antiquarian interest only! This is particularly true of the many field models examined at length in Chapters 35 and 36, Electrostatic Lenses and Magnetic Lenses. We have nevertheless decided to retain them for they were a valuable feature of electron optics in the precomputer years when the mathematical skills that produced them were essential and they are thus part of the history of our subject.

There are innumerable minor changes and additions, not worth mentioning here individually. Many new references have been added and titles are now included in the lists of references.

In the Preface to the first edition, we claimed that *Principles* was the first attempt to cover the whole subject since Glaser's *Grundlagen der Elektronenoptik* appeared in 1952. A few substantial books on the subject have appeared since 1989, notably *Geometrical Charged-particle Optics* by Harald Rose and *Modern Map Methods in Particle Beam Physics* by Martin Berz as well the later volume by Berz, Kyoko Makino and Weishi Wan, *An Introduction to Beam Physics*, but none of these attempts the broad coverage of the present volumes. Nevertheless, they are essential complements to our text in that they deal with some subjects in greater detail or from a very different standpoint. Thus Rose, in a virtuoso performance, uses the eikonal theory systematically throughout and brings out clearly the importance of system symmetries, while Berz relies on the differential algebra that he has developed for charged-particle optics in several areas, notably accelerator optics as well as microscope optics. An introduction to this is included in Chapter 34, Numerical Calculation of Trajectories, Paraxial Properties and Aberrations.

Preface to the First Edition (Extracts)

The last attempt to cover systematically the whole of electron optics was made by the late Walter Glaser, whose *Grundlagen der Elektronenoptik* appeared in 1952; although a revised abridgement was published in the *Handbuch der Physik* 4 years later, we cannot but recognize that those volumes are closer to the birth of the subject, if we place this around 1930, than to the present day.

Furthermore, electron optics has been altered dramatically during these intervening decades by the proliferation of large fast computers. Analytic expressions for the aberration coefficients of superimposed deflection and round magnetic lens fields, for example, have been derived only recently, partly because the latest generation of microlithography devices required them but also because they could only be evaluated by numerical methods: the earlier practice of seeking models permitting hand calculation could never have served here. Again, computer calculations have shed considerable light on electron gun behaviour, as the length of Part IX testifies convincingly; in 1952, Glaser was able to condense his account of gun theory into four pages!

The growth of electron optics is not, however, solely due to the computer. Many systems that had not been thoroughly explored have now been analysed in detail and, in many cases, we have had to renounce the attempt to reproduce in detail new results, however interesting, to keep the number of pages within reasonable limits. This work should therefore be regarded as both a textbook and a source-book: the fundamentals of the subject are set out in detail, and there the student should find everything needed to master the basic ideas or to begin the analysis of some class of systems not yet explored; the principal electron optical components are likewise dealt with in great detail. Where optical elements that are not quite so common are concerned, however, we have felt at liberty to direct the reader to original articles and reviews, or specialist texts, to leave space for topics of wider interest.

The following chapters are, moreover, limited to geometric optics: wave optics is to be covered in a companion volume. With the Schrödinger equation as starting point, we shall examine the propagation of electron waves in electrostatic and magnetic fields and study image formation and resolution in the principal electron optical instruments. This demands some discussion of electron–specimen interactions. A chapter will be devoted to the four broad themes of image processing: discretization and coding; enhancement; restoration; and

analysis, description and pattern recognition. In another, we shall give an account of the steadily growing field of electron holography. Finally, we shall return to the optics of electron sources in order to understand the concept of coherence and we shall show how the notions of brightness, partial coherence and various associated spectral functions are interconnected.

Students of electron optics have been fortunate in that many excellent textbooks on the subject have appeared over the years, the first when the subject was still young (Brüche and Scherzer, 1934; Myers, 1939; Klemperer, 1939; Picht, 1939); these were followed in the 1940s by the encyclopaedic Zworykin et al. (1945), Cosslett (1946) and Gabor (1945). Many books on the subject appeared in the 1950s, of which the texts by Glaser already mentioned, Sturrock (1955) Grivet et al. (1955, 1958) and Kel'man and Yavor (1959) are the most important for our present purposes. Subsequently, however, the flow has shrunk to a trickle, new editions and short introductory texts dominating, with the exception of the multi-author volumes edited by Septier (1967, 1980, 1983); conversely, monographs on limited topics have become more common. Although certainly 'standing on the shoulders of giants', the present volumes do differ considerably from their many predecessors in that the developments of the past 20 years are accorded ample space.

For whom is this work intended? A knowledge of physics and mathematics to first degree level is assumed, though many reminders and brief recapitulations are included. It would be a suitable background text for a postgraduate or final year course in electron optics, and much of the material has indeed been taught for some years in the University of Tübingen; a course in the University of Cambridge likewise covered many of the principles. Its real purpose is, however, to provide a self-contained, detailed and above all modern account of electron optics for anyone involved with particle beams of modest current density in the energy range up to a few mega-electronvolts. Such a reader will find all the basic equations with their derivations, recent ideas concerning aberration studies, extensive discussion of the numerical methods needed to calculate the properties of specific systems and guidance to the literature of all the topics covered.

Composition of volumes such as these puts us in debt to a host of colleagues: many have permitted us to reproduce their results; the librarians of our institutes and the Librarian and Staff of the Cambridge Scientific Periodicals Library have been unflagging in their pursuit of recondite and elusive early papers; Mrs. Ströer has uncomplainingly word-processed hundreds of pages of mathematical and technical prose; Mrs. Maczkiewicz and Mr. Inial have taken great pains with the artwork as have Mrs. Bret and her colleagues with the references; Academic Press and Prof. Dr K.-H. Herrmann, Director of the Institut für Angewandte Physik der Universität Tübingen, have generously supported this work; the Zentrum für Datenverarbeitung has provided the text-editing facilities needed for TEX. To all of these we are extremely grateful.

Acknowledgments

In the first edition, we expressed our gratitude to several colleagues and librarians who helped in various ways and to those who typed the text and prepared the illustrations. To them, it is a pleasure to add Madame Isabelle Labau (CEMES-CNRS), Frau Susanne Hempel and Frau Gabriele Thomas (University of Tübingen), Frau Katrin Quetting (Fritz-Haber-Institut der MPG, Berlin), Madame Salomé Rohr (CERN Library, Geneva) and the ever-efficient personnel of the Technische Informationsbibliothek in Hannover.

Several colleagues pointed out errors or obscure passages in the first edition. We are most grateful to Erich Plies, the late Michal Lenc and Bohumila Lencová, the late Stella Yavor and her son Mikhail and Marijn Bronsgeest for their careful reading of the text. The advice of Shin Fujita, Dieter Kern and Tomáš Radlička has been much appreciated.

The expressions for aberration coefficients have been extended to include higher order aberrations of round lenses and those of electron mirrors. We are very grateful to Zhixiong Liu and Seitgerim Bimurzaev, who supplied lists of formulae that we could incorporate directly. We include a full account of the work of Dirk Preikszas on electron mirrors, who has been good enough to resuscitate his programs and provide aberration integrals for all the third-order geometrical aberrations (not hitherto published apart from the spherical aberration coefficient).

CHAPTER 1

Introduction

1.1 Organization of the Subject

The properties of beams of free electrons, released from a material source and propagating through a vacuum region in some device, are of interest in many diverse fields of instrumentation and technology. The study of such electron beams forms the subject of electron optics, which divides naturally into geometrical optics, when the wavelength is negligible, and wave optics, in which effects due to the finite wavelength are considered. This first volume is concerned with geometrical optics, a knowledge of which is needed to analyse an extremely wide range of instruments: cathode-ray tubes; the family of electron microscopes, which now includes the fixed-beam and scanning transmission instruments, the scanning electron microscope and the emission microscopes; electron spectrometers and mass spectrographs if we include charged particles other than electrons; image converters; electron interferometers and diffraction devices; electron welding machines; and electron-beam lithography devices. We could indeed include electron accelerators, such as betatrons and electron synchrotrons, but a rather different approach is often more useful in those machines. This list is by no means complete but it already demonstrates the great diversity of the possible applications of electron optics.

Over the years, a vast amount of knowledge about the many branches of electron optics has been accumulated and we have therefore had to be selective. The main emphasis is on the principles of electron optics, and technical details are only included to bridge the gap to the practical application of these principles. This seems justified, for the principles remain unaffected by the passage of time whereas instrumental development is so rapid that surveys and review articles are the best means of charting its progress.

The physical properties of electrons in a free beam may be classified as follows:

1. corpuscular properties;
2. wave properties;
3. macroscopic interactions;
4. microscopic or atomic interactions;
5. radiative properties.

A similar classification is given by Sturrock (1955).

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00001-8

© 2018 Elsevier Ltd. All rights reserved.

The *corpuscular properties* are described by classical (relativistic) mechanics, the electron being regarded as a charged particle acted on by electromagnetic forces. For almost all electron optical devices, extensive studies must be made on the assumption that these corpuscular properties alone are important and the present volume is almost wholly confined to the corresponding *geometrical* optics. Geometrical particle optics is very similar for all charged particles, ions in particular, and Parts I and X are written in such a way that many of the relations derived are valid for ions and electrons, or can be converted straightforwardly.

The rest-mass of the electron is extremely small, a characteristic that has important consequences for the technology associated with electron beams. Only quite modest voltages are needed to accelerate electrons to a very high velocity, and the time of flight between the departure of an electron from the cathode and its arrival at its destination in a typical device is so small that it can almost invariably be ignored completely. It is therefore quite sufficient to study the purely *geometrical* shape of the electron motion within the beam, although a time-like curve parameter may prove to be advantageous in numerical calculations.

A further consequence of the extremely small inertia of the beam electrons is that deflection by suitably placed magnetic or electrostatic fields occurs virtually instantaneously, in synchronism with the applied voltages or currents, unless the frequency involved is very high indeed. The performance of many devices relies upon this property. We shall almost always consider electron motion only in static, that is, *time-independent* fields. This is justified even when studying the deflector in a scanning device, the time of flight being so short that the applied field is quasistatic; the time dependence is then merely a common amplitude factor.

A knowledge of the *wave properties* of the electron is essential to understand the concept of resolution in electron microscopes, to analyse the interactions between electron beams and targets of all kinds, and to analyse the behaviour of electron interferometers and diffraction devices and of course to comprehend electron holography. These topics will occupy much of Volume 3.

The *macroscopic interactions* in an electron beam are a consequence of the fact that the latter may be regarded as a cloud of negative charges, which creates an electric field; this is superimposed on the external applied field and can thus alter the focusing properties of the device. In principle, of course, this occurs for every electron beam, but in reality such space charge effects are of importance only when the local beam intensity is very high. The space charge density can be treated as a macroscopic observable and the associated field calculation remains within the framework of classical electrostatics; we therefore call these interactions macroscopic. Such effects occur mainly in electron guns where the beam intensity can be high, and are therefore dealt with in Part IX, devoted to guns. Such effects are also extremely important in accelerators, but these are not within the scope of this book.

The *microscopic* or atomic interactions are the various scattering processes that occur within the beam on the atomic scale. Such processes arise when an electron-beam encounters a specimen or target and electron−electron collisions may also occur within the beam. The latter give rise to the Boersch effect, an anomalous broadening of the electron spectrum at beam waists, which are themselves enlarged. Classical collision theory is capable of providing an approximate explanation of this effect, which is examined briefly in Section 46.4 of Volume 2. Collisions between beam electrons and the atoms in a target can only be properly understood in terms of quantum mechanics; some space is devoted to this topic in Volume 3.

Finally, we come to the *radiative properties* of the electron, essentially the emission of bremsstrahlung when the acceleration is very high. This occurs mainly in high-voltage electron microscopes where particles with an energy of 1 MeV or more collide with the specimen placed in the path of the beam. Although the staff around the microscope must be protected from this radiation, the damage to the specimen is negligible in comparison with that inflicted by the mechanical bombardment. The bremsstrahlung caused by the acceleration of the electrons *in vacuo* becomes important only at the energies encountered in high-energy physics, which are beyond the scope of these volumes; we therefore ignore bremsstrahlung throughout.

Our theme is thus the study of the motion of electrons, regarded as classical charged particles of negligible extent, through static electric or magnetic fields. We begin with the derivation of the conservation laws for the electron motion and cast these into a form particularly well suited to electron optics. Various forms of the trajectory equations are established but these are not at all satisfactory for our purpose, which is the study not of single trajectories but of whole families of electron paths: not ballistics but optics. It is *Hamiltonian theory* that enables us to make the transition. As early as 1827, Hamilton drew his famous analogy between geometrical optics and classical mechanics; this tells us that, just as in optics, there must exist a mechanical characteristic function, or *eikonal*, with the property that the trajectories are always locally orthogonal to the surfaces of constant value of this function. This is true only in the absence of magnetic vector potentials; in the general case, when magnetic fields are present, the correct form of this orthogonality relation emerges from the Hamilton−Jacobi theory presented in Chapter 5. This theory is very important, for it forms the cornerstone of geometrical electron optics.

Most instrumental research is concerned with the design of new or improved electron optical systems, for which an accurate knowledge of the properties of families of rays traced through such systems is indispensable. This proceeds in two stages: first, the field distribution must be established, after which rays can be traced and quantities characteristic of the system calculated. A knowledge of the field distribution is usually needed only in the immediate vicinity of a curve in space, frequently a symmetry axis, known as the optic

axis; the beam is required to remain close to this axis to prevent the aberrations from degrading the performance of the device. Unfortunately, the required information about the field can rarely be obtained without solving a *boundary-value problem*, since the field will be generated by electrodes and magnetic materials, such as polepieces, at some distance from this axis.

In practice, field calculation is the most complicated part of numerical design and the principal methods are presented in considerable detail in Part II. *Series expansions* for electrostatic potentials and hence fields and for magnetic scalar and vector potentials are also listed since these are repeatedly needed in later chapters where the trajectory equations and aberration coefficients of various types of system are derived.

This thorough presentation of the physical and mathematical fundamentals leads naturally to the systematic investigation of electron optical components: how are these to be characterized, how can we code complex behaviour in terms of a few easily calculated parameters? Parts III and IV provide the traditional answers in terms of the paraxial approximation and the aberrations that measure departures from it. In the *paraxial approximation*, it is assumed that the electron trajectories remain so close to the optic axis that equations of motion linear in the off-axis coordinates describe them satisfactorily. Although this is an excellent first-order approximation, it is clearly an idealization, a consequence of which is that some electron optical systems appear to be free of any image defects and hence capable of producing a stigmatic, unblurred image or a sharp focus.

In reality, no system is free of aberrations. One of the major tasks of electron optics is to establish what types of aberrations can occur in any given system and then to reduce the most deleterious as far as possible. A long Part is therefore devoted to the *theory of aberrations*. Since all wave optical considerations are excluded from this volume, only *geometrical* and *chromatic aberrations* are investigated. The former are those that measure the discrepancy between the true point of arrival of an electron at its destination and the point predicted by the paraxial approximation, resulting from the inadequacies of the latter and any small imperfections in the system; the chromatic aberrations are those caused by the presence of electrons with different energies in the beam, arising from the small spread of the initial energies at the cathode surface or from the loss of various amounts of energy when traversing a thin specimen.

A separate Part is devoted to a similar analysis of deflection systems, of great practical importance for microlithography in the current quest for miniaturization. Such systems may be very complex, magnetic and electric deflection fields occupying the same region as a magnetic round lens field, and the number of degrees of freedom becomes very large. Both the theory and the experimental adjustment of such combinations reflect this complexity but it has proved necessary to resort to such intricate arrangements in modern electron-beam lithography machines, which are used to produce the semiconducting integrated circuits

required in computers. In Part VI the numerical techniques needed for tracing trajectories and computing aberrations in any system are presented, together with a short introduction to computer algebra, a tool that is very useful for establishing aberration integrals and evaluating these for the few models that permit a result to be obtained in closed form. In the next three Parts, VII–IX, the principles set out in the first volume are applied to the many different types of electron optical components—round lenses, quadrupoles, mirrors, cathode lenses, Wien filters—for each of which the general theory takes a special form. We have preferred to organize this material by component rather than by instrument, since the latter must inevitably be understood as an assembly of individual modules.

Part VII indicates what practical information is available in the literature concerning the optical properties of the various elements analysed in theory in earlier chapters, while in Part VIII we examine two special topics of sufficient importance to warrant separate treatment: aberration correction, essentially for systems of round lenses, and the theory of caustics.

Another topic of great complexity is the study of electron sources, usually known as electron guns. An entire Part (IX) is devoted to these, for although the degrees of freedom are not unduly large in number, the theoretical description requires concepts that are of little importance elsewhere and depend essentially on the purpose to which the gun is to be put. Thus the gun of an electron interferometer is very different from that of an electron welding machine. We have tried to impose a pattern on this complex and many-faceted topic by following the flight of the electron through the gun. First we examine the principal emission processes and the focusing effects in the neighbourhood of the cathode. This is followed by the theory of space charge, which may of course be important elsewhere than in the cathode region. We next introduce a number of quantities that are employed to characterize the beam farther from the cathode, and, in particular, brightness, emittance and the energy spectrum, which are very important when considering the suitability of the gun for specific tasks. The Part ends with a few remarks about the design of complete guns and such new subjects as multibeam systems and carbon nanotube emitters.

The final Part is devoted to systems in which the optic axis is curved, though in practice almost always a plane curve, a situation that arises in the electric or magnetic prisms of electron spectrometers. Monochromators are also examined here, though they do not necessarily have curved axes; they have become of great importance for microscopes the geometrical aberrations of which have been corrected. With this, most aspects of geometrical electron optics have been covered. The reader will notice that the emphasis throughout is on physical principles and on their theoretical formulation. Technical details are included only when they seem necessary to render the practical applications of these principles comprehensible. Inclusion of technological details would have made the book impossibly large and rapidly obsolete, for there are few branches of the subject that are not

in continual development. The lifetime of the underlying principles is, however, substantially longer.

The subject has acquired a very voluminous literature over the decades, so that a full bibliography would fill far too many pages. We have adopted a compromise towards these many publications: papers of especial relevance are mentioned in the body of the text but, in order not to interrupt the flow, the majority is grouped in annotated bibliographic appendices at the end of each book. In this way, the reader is directed to the literature of each topic but is not continually distracted by notes and references. Even so, we have made no systematic attempt to trace the history of the subject in these appendices and we therefore complete this introduction with a succinct account of the principal stages through which the subject has passed.

1.2 History

Electron optics was born in the 1920s. In 1925, Louis de Broglie argued convincingly that a wavelength should be associated with moving particles, electrons in particular; and in 1927, Hans Busch demonstrated that the action of an axially symmetric coil on electrons can be described in the language of geometrical optics, in terms of a focal length: "Eine kurze Spule hat also die Eigenschaft, die Kathodenstrahlen nach der Achse zu um einen Winkel γ abzulenken, der proportional der Achsenentfernung ... des Strahles ist. Genau die gleiche Eigenschaft besitzt aber für Lichtstrahlen eine Sammellinse";[1] this was an explicit statement of his conclusions adumbrated a year earlier (Busch, 1926). De Broglie's paper soon led to the experiments on electron diffraction of Davisson and Germer (1927a,b) and of Thomson and Reid (1927). Busch's idea of associating a lens-like character with a short magnetic field was tested by Max Knoll and his young student, Ernst Ruska (Ruska and Knoll, 1931), who went on to combine such lenses into the first electron microscope, built in the Electrotechnical Institute of the Berlin Technological University (Knoll and Ruska, 1932a,b).

An electron microscope has much in common with its light optical ancestor. It consists of a source of illumination, condenser lenses to direct the illuminating beam onto a suitably sized region of the specimen with an appropriate angular spread, an objective lens to provide a first magnification and projector lenses to magnify the intermediate images still further. In appearance and nature, however, each of these optical elements is very different from those of the familiar compound microscope. The source of illumination is now an electron gun, of which the simplest type is the thermionic triode. A filament is heated, thus releasing

[1] A short coil thus has the property of deflecting the cathode rays toward the axis through an angle γ, which is proportional to the distance of the ray from the axis. Exactly the same property obtains for light rays in a converging lens.

electrons which are accelerated to the desired energy by a suitably polarized anode. A third electrode, the wehnelt, placed between filament and anode, improves the performance of such sources considerably. These guns have now been superseded by cold field-emission guns or Schottky emission guns, in which electrons are extracted from the cathode by a high electric field. For the former, a much better vacuum is essential, typically 10^{-10} Torr ($\sim 10^{-8}$ Pa). Very recently, instruments for imaging rapid changes in specimen structure have been developed, in which the object is illuminated with a rapid sequence of very short electron pulses; here, the electrons are ejected from the cathode by bombardment.

The lenses are short stretches of rotationally symmetric magnetic field, created by a current-carrying coil enclosed in an iron yoke. The interior of the microscope must be evacuated to a pressure typically of the order of 10^{-6} Torr ($\sim 10^{-4}$ Pa) since electrons are scattered or halted by a very small amount of matter in their path. For the same reason, the specimen must be exceedingly thin (at most a few nanometres thick for a 100 kV instrument). In these conditions, the electrons are deflected or 'scattered' within the specimen but almost no electrons fail to emerge from the far side. The specimen is thus a 'phase object' and contrast is created at the image by various mechanisms analogous to those encountered in the phase-contrast microscope. This image is rendered visible by allowing the electrons to fall on a fluorescent screen or a recording medium, such as a photographic emulsion or CCD (charge-coupled device) detector. One aspect of electron lenses deserves special mention: their optical quality is astonishingly poor! They suffer from two lens defects that have been virtually eliminated from glass lenses: spherical aberration, a defect that severely limits the numerical aperture at which they can be operated and hence the resolution attainable; and chromatic aberration, by which we mean that their focusing power varies rapidly with the velocity of the incoming electrons. The high spherical aberration has the practical consequence of deteriorating the resolution of an electron microscope by some two orders of magnitude: with perfect lenses, the resolution limit might be expected to be of the order of picometres, whereas it is in reality of the order of hundreds of picometres (that is, of the order of ångströms). The harmful effects of chromatic aberration are avoided by using nearly monoenergetic electrons and stabilizing the lens currents to a very high degree, typically to one part in a million. Today, it is common to include a monochromator, a device that eliminates electrons with energies outside a chosen narrow range.

The first tentative studies of Ruska and Knoll, with which Bodo von Borries was soon associated, were sufficiently encouraging to initiate a decade of theoretical and empirical electron optics, during which the foundations of the theory were laid, largely by Walter Glaser and Otto Scherzer, and the magnetic electron microscope was perfected to such a point that a commercial model was put on the market by the German company Siemens in 1938. The British Metropolitan-Vickers company can, however, claim to have been the first commercial firm to supply a microscope, the custom-built EM1 instrument ordered by L.C. Martin for

Imperial College, London, where it was installed in 1936 (Martin et al., 1937); the resolution of the EM1 was not, however, superior to that of a light microscope (Mulvey, 1985).

Meanwhile, comparable work on an electrostatic instrument was being actively pursued in the research department of the Allgemeine Elektrizitätswerke-Gesellschaft (AEG). For full details of these activities, see Ruska's historical volume (1979, 1980) and the 'Selfportrait' of the AEG Research Institute prepared by Ramsauer (1941) with further editions in 1942 and 1943. The early development of the theory is fully chronicled in Glaser's *Grundlagen der Elektronenoptik* (1952).

Outside Germany, many electron microscope projects were launched in the 1930s, though it was not until the end of the Second World War that commercial production began on any scale. The prototypes built in England, the USA and Canada are described in various historical articles, especially Gabor (1957), Ruska (1957) and Mulvey (1962, 1967, 1973), and many references and reminiscences are to be found in Hawkes (1985) and Mulvey (1996). We must, however, make particular mention of the work of Ladislaus Marton, who constructed a series of simple instruments in Brussels, with which he obtained the earliest osmium-stained biological micrographs, the specimens being the long-leafed sundew and the root of the bird's-nest orchid (Marton, 1934a–c, 1935a,b, 1937). Soon after, first Driest and Müller (1935) and then Krause (1936) obtained biological electron micrographs with one of Ruska's microscopes that foreshadowed, albeit faintly, modern biological electron microscopy; Driest and Müller's images of the wing and leg of the common housefly were the first micrographs of unprepared biological specimens. A key role in the steps that led to the first commercial electron microscopes marketed by Siemens was played by Ernst Ruska's brother Helmut, a doctor convinced of the value of this new tool in medicine (Gelderblom and Krüger, 2014). He in turn persuaded his former clinical teacher Richard Siebeck, Director of the First Medical Clinic of the Berlin Charité, to assess its potential and his conclusions were highly influential in Siemens' decision. Siebeck's report is reproduced in Ruska's memoir (1979, 1980). Thirty-eight models of the first Siemens instrument were put into service, their fates are described by Wolpers (1991). The troubled history of the patenting of the first microscope is recapitulated in detail by Ruska (1984, 1986); Rüdenberg's account was published posthumously in 2010.

It was during the 1930s too that the field-emission microscope was developed by E.W. Müller, in one of the Siemens research laboratories in Berlin. In this instrument, a strong electric field is maintained at a tip and highly magnified details of the surface are visible in the image as a result of the differences in emission from point to point. Müller's first papers appeared in 1936 and 1937 and a historical account is to be found in Good and Müller (1956). More details of the development of this family of instruments are to be found in Müller and Tsong (1969, 1974), Müller (1960, 1975), Drechsler (1978) and Melmed (1996).

By the 1950s, electron microscopes were being produced in West Germany, England, France, Holland, Switzerland, Czechoslovakia and the Soviet Union, with more modest activity in other European countries, particularly Sweden. In the United States, RCA began manufacturing electron microscopes during the war years, and in Japan commercial production commenced in the late 1940s, though many prototypes were built during the first half of the decade; the Hitachi HU-4 was put on the market in 1947, for example, and the JEOL JEM-1 in 1949 (Sugata, 1968; Fujita, 1986).

Although our subject is not electron microscopy but electron optics, we must digress here to mention an important development in the years 1948−1952, which had a major effect on electron microscope design (Gettner and Ornstein, 1956). The accelerating voltages of the early microscopes were then of the order of 50−80 kV, which was too low to form a sharp image of a biological specimen if the latter was one or more micrometres in thickness. An increase in voltage therefore seemed imperative until in 1948, Pease and Baker succeeded in cutting sections only 0.3−0.5 μm thick with a modified Spencer 820 microtome, and by 1950 the figure had fallen to 0.2 μm. In 1949, Newman et al. introduced methacrylate as an embedding medium, the mechanical properties of which greatly facilitated section cutting. In 1952, Sjöstrand designed a new ultramicrotome (Sjöstrand, 1953) with which sections 20 nm in thickness could be cut reproducibly and "the problem of high resolution electron microscopy of sectioned material had been solved" (Sjöstrand, 1967); the immediate need for high-voltage microscopes in biology vanished and the first megavolt instruments were not built for about another decade. The first of these was a 1.5 MV instrument constructed in Toulouse (Dupouy et al., 1960; Dupouy, 1968, 1985) and this was rapidly followed by a 750 kV machine in Cambridge (Smith et al., 1966; Cosslett, 1981) and the commercial high-voltage microscopes of AEI, GESPA, Hitachi and JEOL. These were all giant versions of the familiar 100 kV instruments, however, and essentially represented only technological progress; their optics was distinctly conservative. Their great bulk and the need for special buildings to house them did, however, furnish one of the reasons for the interest in superconducting lenses that sprang up in the mid-1960s, another being the perfect stability of the magnetic field generated by a coil in the persistent-current mode (Laberrigue and Levinson, 1964; Fernández-Morán, 1965; Boersch et al., 1966; Siegel et al., 1966; Ozasa et al., 1966). Of the various designs, the shielding lens introduced by Dietrich et al. (1969) was clearly superior when it was important that both the specimen and its immediate environment be at very low temperature (see Weyl et al., 1972; Hardy, 1973; Dietrich, 1976; Hawkes and Valdrè, 1977; Riecke, 1982; Lefranc et al., 1982). Students of superconducting lenses were not, however, alone in enquiring whether the focusing power of the monster lenses in traditional high-voltage electron microscopes could not be obtained in some other way. We draw attention to the numerous 'unconventional' designs introduced by Mulvey and colleagues, compared and contrasted in Mulvey (1982, 1984), and to the laminated lenses of Murillo (Balladore and Murillo, 1977), in which the yoke is made of

highly inhomogeneous material in order to maintain the flux density constant over its cross-section.

The idea of forming an image not by irradiating a comparatively large specimen area and focusing this onto the image plane after suitable magnification but by scanning the specimen with a small probe, collecting a signal from the resulting interaction and using this signal to modulate the intensity of the spot of a cathode-ray tube scanned in synchronism goes back to the late 1930s. In 1938, von Ardenne described a primitive ancestor of the modern scanning (transmission) electron microscope, in which a probe size of some 10 nm was achieved but at the cost of a very small current indeed ($\sim$1 pA); the beam traversed the specimen and struck a photographic film attached to a drum which rotated and advanced appropriately (see von Ardenne, 1940, 1978, 1985). In 1942, an instrument in which secondary electrons from a thick target provided the image signal was developed by Zworykin et al. but it was not until 1953 that McMullan described the first of the series of scanning electron microscopes to be built under Charles Oatley's direction in Cambridge, which culminated in the commercial 'Stereoscan', marketed by the Cambridge Instrument Company in 1965 (Oatley et al., 1965, 1985; Oatley, 1982; Breton et al., 2004, Smith, 2013). A home-made instrument was constructed by André Léauté in 1944–6 (Léauté, 1946; Hawkes and McMullan, 2004). More recent versions of these instruments combine the properties of the X-ray microanalyser introduced by Castaing (1951) and perfected by Cosslett and Duncumb (1956) and Duncumb (1958), who added beam scanning, with the host of signals available in a scanning electron microscope with the result that analytical electron microscopy (AEM) has become a discipline in its own right (see for example, Botton and Prabhudev, 2018).

The next major instrumental development occurred in the early 1960s. A celebrated set of curves (Oatley et al., 1965) relating probe size, number of lines in the image and the time needed to record an image of acceptable quality had shown vividly that the resolution of the scanning microscope could never rival that of the transmission microscope owing to the inadequate performance of the thermionic gun. It was known that a field-emission gun would change this situation dramatically, making it possible to compress a useful current into a probe only a few ångströms in diameter, and it was in 1965 that Crewe first described a scanning transmission electron microscope (STEM) with a field-emission gun (Crewe et al., 1968; Crewe, 1970, 1973). Instrumental development began in three companies, AEI, Vacuum Generators (VG) and Siemens, of which only VG continued to market STEMs, finally ceasing production in 1996; their first commercial instrument was installed in 1974 (Wardell and Bovey, 2009; von Harrach, 2009). Those of Siemens and AEI were not pursued (Hawkes, 2009) and the present tendency is to offer a field-emission gun and a STEM mode as options with conventional transmission microscopes. A redesigned aberration-corrected STEM was developed by the Nion company; the first commercial model was delivered in 2007 (Krivanek et al., 2008).

The foregoing account evokes the main steps in electron optical instrumentation for image formation. Many other innovations might have been listed: the development of new types of gun (lanthanum and cerium hexaboride cathodes, carbon nanotube and other field emitters) and the introduction of various types of energy filter and analyser, in particular the focusing device of Castaing and Henry (1962), and monochromators. We now turn to the theory of the subject and single out the principal contributions. We have already mentioned the role of Busch in the founding of geometrical electron optics. The ideas of Hamiltonian mechanics were applied to electron motion by Walter Glaser, who derived expressions for the aberration coefficients with the aid of a characteristic function or eikonal, while Otto Scherzer obtained similar formulae by the 'trajectory method', in which equations of motion including aberration terms are derived and solved by the method of variation of parameters. In 1936, Scherzer published formulae for the coefficients of spherical and chromatic aberration that showed that these can never be made to change sign by skilful lens design; this result and the reactions to it are discussed in detail in Part IV. Eleven years later, it was again Scherzer (1947) who described several types of aberration correctors, capable in principle of cancelling these troublesome coefficients. The history of the 1950s is the history of early attempts to use such correctors. These continued with limited success and no commercial exploitation until the 1990s, which saw the major breakthroughs: first, a quadrupole–octopole corrector of spherical and chromatic aberration for the scanning electron microscope (Zach and Haider, 1995), then a sextupole corrector for the spherical aberration of the transmission electron microscope (Haider et al., 1998) and a quadrupole–octopole corrector for that of the scanning transmission electron microscope (Krivanek et al., 1997). Commercial models soon became available. Scherzer's were not the only suggestions for aberration correction, however; in 1948, Gabor described a method of correcting spherical aberration by an optical reconstruction technique, which he called holography. For technical reasons, this was unsuccessful at that time (the laser was yet to be invented) but with the development of bright electron sources and coherent light sources, both in-line and off-axis holography have subsequently been extensively developed, as we shall see in Volume 3.

We must return to the 1940s and 1950s to draw attention to some other landmarks in electron optics. In 1943, Grinberg published a very general form of the equations of motion of electrons in electric and magnetic fields and this was later extended to include aberrations by Vandakurov (1955, 1956a,b, 1957). Similar general equations were derived by Sturrock (1952), who developed and perfected Glaser's Hamiltonian approach to electron optics in many ways (Sturrock, 1955), and by Rose (1968), who analysed a more limited class of systems. The labour of transforming aberration integrals was reduced by Seman (1951, 1954, 1955, 1958a,b), who introduced a very ingenious method that replaces partial integration by differentiation. At about the same period, Lenz (1956, 1957) clarified the distinction between real and asymptotic aberration coefficients, already examined

briefly by Sturrock (1955). It was not until a decade later that the possibility of writing these asymptotic coefficients as polynomials of at worst fourth order in reciprocal magnification was noticed (Hawkes, 1968), an observation that renders computer-aided design of complex systems less arduous. The 1970s and 1980s saw the arrival of increasingly powerful computing techniques in electron optics, and these are now capable of solving almost all the problems that arise, including those branches of electron optics that had perforce been almost completely neglected by earlier theoreticians: electron guns are the most striking example. Parts II, VI and IX bear witness to the progress that has been made, much of it in the theoretical electron optics group of the University of Tübingen (Lenz, 1973; Kasper and Lenz, 1980; Kasper, 1982, 1984, 2001). Program suites for charged-particle optics have been developed by Eric Munro (MEBS, Munro's Electron Beam Systems) and Bohumila Lencová (SPOC, Software for Particle Optics Computations), notably EOD (Electron Optical Design). Other such programs in widespread use are SIMION, Frank Read's CPO and COSY INFINITY (Martin Berz, Kyoko Makino). Several special-purpose programs have also been written, such as MOPS, created by Dirk Preikszas to establish formulae for the aberration coefficients of electron mirrors, computer-aided design software developed by the Delft school and Anjam Khursheed's KEOS. References to all these are to be found in the appropriate chapters.

This short account of the history of electron optics and electron microscopes cannot but be invidious: we could have included all the other electron optical instruments and we could have charted the progress of the theory, equation by equation. This would, however, have left all too little space for the Principles of Electron Optics, to which we now turn. References to many other accounts of the history of the subject are to be found in *The Beginnings of Electron Microscopy* (Hawkes, 1985) and *The Growth of Electron Microscopy* (Mulvey, 1996). Among the historical articles not cited above are Reisner (1989), Rasmussen (1997), Müller (2009), Rose (2008) and Hawkes (2009, 2015). Biographies of several of the pioneers have been published: Lambert and Mulvey (1996, Ernst Ruska), Grümm and Schiske (1996, Walter Glaser), Mulvey and van de Laak-Tijssen (2001, Jan Le Poole), von Borries (1991, Bodo von Borries), Gelderblom and Krüger (2014, Helmut Ruska), Mulvey (1995, Dennis Gabor), Hawkes (1995, Tom Mulvey), Rose (1983) and Marko and Rose (2010, Otto Scherzer).

PART I

Classical Mechanics

CHAPTER 2

Relativistic Kinematics

In the following sections we shall examine the motion of a charged particle of rest mass m_0 and charge Q in an electromagnetic field characterized by the electric and magnetic field vectors $\boldsymbol{E}(\boldsymbol{r}, t)$ and $\boldsymbol{B}(\boldsymbol{r}, t)$, respectively. Whenever the specialization $Q = -e < 0$ for electrons is not made explicitly, the analysis is valid for the motion of any charged particle.

The derivation of useful trajectory equations for the motion of charged particles – and of all conservation laws satisfied by them – can be performed in a very general and elegant manner by means of variational calculus, as we shall show in Chapter 4, Variational Principles. First, however, we give a short introduction to relativistic kinematics, because this offers a better understanding of many of the optical aspects of the variational calculations.

2.1 The Lorentz Equation and General Considerations

The trajectory equation for the motion of charged particles is most simply represented in its Newtonian form

$$\frac{d\boldsymbol{g}}{dt} = \frac{d}{dt}(m\boldsymbol{v}) = Q\{\boldsymbol{E}(\boldsymbol{r}, t) + \boldsymbol{v} \times \boldsymbol{B}(\boldsymbol{r}, t)\} \tag{2.1}$$

$\boldsymbol{v} = d\boldsymbol{r}/dt$ being the velocity and $\boldsymbol{g} = m\boldsymbol{v}$ the *kinetic* momentum, which must be clearly distinguished from the *canonical* momentum, defined in Chapter 4, Variational Principles. The Lorentz force, given by the right-hand side of Eq. (2.1), is expressed in SI units, which will be used consistently throughout this volume. It is convenient to introduce the familiar abbreviations

$$\beta = \frac{v}{c} = \frac{|\boldsymbol{v}|}{c}, \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}} \tag{2.2}$$

whereupon Eq. (2.1) has the correct relativistic form if we represent the relativistic mass m by the well-known expression

$$m = \frac{m_0}{\sqrt{1 - \beta^2}} \equiv \gamma\, m_0 \tag{2.3}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00002-X
© 2018 Elsevier Ltd. All rights reserved.

The trajectory equation does not depend on the values of Q and m_0 separately but only on the ratio Q/m_0 as can be seen by rewriting Eq. (2.1) thus:

$$\frac{d}{dt}\left(\frac{\boldsymbol{v}}{\sqrt{1-\beta^2}}\right) = \frac{Q}{m_0}(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$$

In the special case of electron motion, we have

$$Q = -e = -1.602 \times 10^{-19}\ \text{C}, \quad \frac{e}{m_0} = 1.759 \times 10^{11}\ \text{C kg}^{-1} \tag{2.4}$$

The absolute value of the electron charge will always be denoted by e.

The practical evaluation of Eq. (2.1) requires the calculation of the field vectors $\boldsymbol{E}(\boldsymbol{r}, t)$ and $\boldsymbol{B}(\boldsymbol{r}, t)$ for arbitrary values of the position vector $\boldsymbol{r}$ and of the time t. The corresponding computer programs have to be written and executed before embarking on trajectory calculations. In practice, field calculation is the most complicated part of theoretical electron and ion optics. We shall deal with this subject in detail in Part II. In the present chapter we shall assume the fields to be known.

In the most general case of arbitrary electron optical devices, Eq. (2.1) can only be solved numerically, from given initial conditions. In order to derive more detailed laws, it is necessary to introduce simplifying symmetry conditions, which are assumed to be exactly valid. The inevitable departures from exact symmetry in practical devices will be discussed in later chapters.

2.2 Conservation of Energy

In most electron and ion optical devices, the applied fields are static, independent of time: $\boldsymbol{E} = \boldsymbol{E}(\boldsymbol{r})$, $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{r})$. It is then possible to derive a law for the conservation of particle energy. This is most easily done by scalar multiplication of both sides of Eq. (2.1) with $\boldsymbol{v}$. We shall denote derivatives with respect to time by dots. Using Eqs. (2.2) and (2.3) we obtain first

$$\boldsymbol{v} \cdot \frac{d}{dt}\left(\frac{m_0 \boldsymbol{v}}{\sqrt{1-\beta^2}}\right) = Q\boldsymbol{v} \cdot \boldsymbol{E}(\boldsymbol{r}), \quad (\boldsymbol{v} = \dot{\boldsymbol{r}})$$

Using the identity $\boldsymbol{v} \cdot \dot{\boldsymbol{v}} = v\dot{v}$, the left-hand side can be transformed to a total derivative:

$$\boldsymbol{v} \cdot \frac{d}{dt}(\gamma m_0 \boldsymbol{v}) = \gamma^3 m_0 v \dot{v} = \frac{d}{dt}\left(\gamma m_0 c^2\right)$$

In order to transform the right-hand side, we introduce the electrostatic potential $\Phi(\boldsymbol{r})$ and write

$$E(\boldsymbol{r}) = -\operatorname{grad} \Phi(\boldsymbol{r}) \tag{2.5}$$

The right-hand side then becomes a total derivative too:

$$Q\boldsymbol{v}\cdot\boldsymbol{E}(\boldsymbol{r}) = -Q\dot{\boldsymbol{r}}\cdot\operatorname{grad}\Phi(\boldsymbol{r}) = -\frac{d}{dt}(Q\Phi(\boldsymbol{r}))$$

Integration with respect to time and substitution for the factor γ results in:

$$E_0 := m_0c^2\left(\frac{1}{\sqrt{1-\beta^2}} - 1\right) + Q\Phi(\boldsymbol{r}) = \text{const} \tag{2.6}$$

The first term is the kinetic energy

$$T(v) := m_0c^2\left(\frac{1}{\sqrt{1-\beta^2}} - 1\right) \tag{2.7}$$

A power series expansion in v gives

$$T(v) = \frac{m_0}{2}v^2\left(1 + \frac{3}{4}\beta^2 + \frac{5}{8}\beta^4 + \cdots\right)$$

the first term being the familiar nonrelativistic approximation. The second term in Eq. (2.6) is the usual potential energy of classical mechanics,

$$V(\boldsymbol{r}) = Q\Phi(\boldsymbol{r}) \tag{2.8}$$

The functions $\Phi(\boldsymbol{r})$ and hence $V(\boldsymbol{r})$ are unique apart from the choice of an arbitrary additive constant. The total energy E_0 depends on the choice of this constant and on the initial conditions of the trajectory in question, a trivial point in theory but important in many practical situations.

2.3 The Acceleration Potential

For practical calculations, it is of great importance that virtually all *scalar* kinetic quantities can be represented as unique functions in space, the constant E_0 being a free parameter. The electrostatic potential $\Phi(\boldsymbol{r})$ is uniquely defined by its boundary values at the surfaces of the electrodes. In electron optics, the cathode surface in the electron gun is usually chosen to be the equipotential surface $\Phi(\boldsymbol{r}) = 0$. In this volume we shall adopt this most convenient gauge. The constant E_0 then has a very concrete physical meaning: it is the initial kinetic energy of the corresponding electron trajectory at the cathode surface. This is a small *positive* quantity of the order of an electronvolt. The simplification $E_0 = 0$, common in the literature on electron optics, is too strong a restriction, since this excludes the treatment of energy distributions in electron beams. Here, therefore, the convention $E_0 = 0$ will be

adopted in all practical calculations in which the electron energy distribution is not in question. In the remainder of the present chapter, however, the analysis will be completely general.

Once the function $\Phi(\boldsymbol{r})$ and the constant E_0 have been specified, all the other scalar functions are uniquely defined. The kinetic energy is given by

$$T(\boldsymbol{r}, E_0) = E_0 - Q\Phi(\boldsymbol{r}) \tag{2.9}$$

Since T can be rewritten as $T = m_0c^2(\gamma - 1)$, the dilatation factor γ is a function of position:

$$\gamma(\boldsymbol{r}, E_0) = \frac{1}{\sqrt{1-\beta^2}} = 1 + \frac{E_0 - Q\Phi(\boldsymbol{r})}{m_0c^2} \tag{2.10}$$

Eq. (2.9) also determines the absolute value of the kinetic momentum, $|\boldsymbol{g}| = g(\boldsymbol{r})$. In order to find the corresponding function, we first solve $g = m_0\upsilon(1-\beta^2)^{-1/2}$ for $\upsilon = \upsilon(g)$ and then substitute the resulting expression in Eq. (2.7), which yields the well-known formula

$$m_0c^2 + T = c\left\{(m_0c)^2 + g^2\right\}^{1/2} \equiv mc^2 \tag{2.11}$$

This expression will prove to be of great importance in the Hamiltonian theory. Here we solve Eq. (2.11) for g and obtain

$$\boldsymbol{g} \equiv |\boldsymbol{g}| = \sqrt{2m_0T\left(1 + \frac{T}{2m_0c^2}\right)} \tag{2.12}$$

Substituting for T from Eq. (2.9), we find

$$g(\boldsymbol{r}, E_0) = \sqrt{2m_0(E_0 - Q\Phi(\boldsymbol{r}))\left(1 + \frac{E_0 - Q\Phi(\boldsymbol{r})}{2m_0c^2}\right)} \tag{2.13}$$

By means of Eqs. (2.10) and (2.13), the absolute value of the velocity can also be determined as a function, $\upsilon(\boldsymbol{r}, E_0) = g/m_0\gamma$. In the nonrelativistic approximation, Eqs. (2.10), (2.13) and the expression for υ simplify to the well-known formulae

$$\gamma \approx 1, \quad g \approx \sqrt{2m_0(E_0 - Q\Phi)}, \quad \upsilon \approx \sqrt{\frac{2}{m_0}(E_0 - Q\Phi)} \tag{2.14}$$

Since the kinematic functions will be required very often in electron optical calculations, we introduce certain quantities to simplify the notation:

$$\Phi_0 := E_0/e \tag{2.15}$$

$$\varepsilon := \frac{e}{2m_0c^2} = 0.9785 \ \text{MV}^{-1} \tag{2.16}$$

$$\eta := \left(\frac{e}{2m_0}\right)^{1/2} = 2.965 \times 10^5 \ \ \mathrm{C}^{1/2} \ \ \mathrm{kg}^{-1/2} \tag{2.17}$$

Eq. (2.13) now becomes

$$g(\boldsymbol{r}, E_0) = \sqrt{2m_0 e(\Phi_0 + \Phi)\Big\{1 + \varepsilon(\Phi_0 + \Phi)\Big\}}$$

Since the radicand will be required very frequently, it is convenient to introduce a new function

$$\hat{\Phi}(\boldsymbol{r}, \Phi_0) = (\Phi_0 + \Phi(\boldsymbol{r}))\big\{1 + \varepsilon(\Phi_0 + \Phi(\boldsymbol{r}))\big\} \geq 0 \tag{2.18}$$

called the acceleration potential. The circumflex will be added to all functions and constants defined in terms of $\hat{\Phi}$ rather than Φ. It is now possible to express all the other kinematic functions in terms of $\hat{\Phi}$. Obviously,

$$g = \sqrt{2m_0 e\hat{\Phi}} \tag{2.19}$$

On substituting this in Eq. (2.11), we find

$$T = m_0 c^2\left(\sqrt{1 + 4\varepsilon\hat{\Phi}} - 1\right) \equiv \frac{2e\hat{\Phi}}{1 + \sqrt{1 + 4\varepsilon\hat{\Phi}}} \tag{2.20}$$

Since $T = m_0 c^2(\gamma - 1)$, the dilatation factor γ can be expressed as a function of $\hat{\Phi}$ or Φ:

$$\gamma = \sqrt{1 + 4\varepsilon\hat{\Phi}} = 1 + 2\varepsilon\Big\{\Phi(\mathbf{r}) + \Phi_0\Big\} \tag{2.21}$$

Using Eqs. (2.19), (2.21) and (2.17), the velocity becomes

$$\upsilon \equiv \frac{g}{m_0\gamma} = 2\eta\sqrt{\frac{\hat{\Phi}}{1 + 4\varepsilon\hat{\Phi}}} \tag{2.22}$$

Even the electric field strength $\boldsymbol{E}$ can be expressed in terms of $\hat{\Phi}$ and its gradient: differentiation of Eq. (2.18) results first in

$$\nabla\hat{\Phi} = 1 + 2\varepsilon(\Phi_0 + \Phi)\nabla\Phi$$

From Eq. (2.21), we see that the factor multiplying $\nabla\Phi$ is equal to γ, so that finally

$$\boldsymbol{E} = -\frac{1}{\gamma}\nabla\hat{\Phi} = -\left(1 + 4\varepsilon\hat{\Phi}\right)^{-1/2}\nabla\hat{\Phi} = \frac{\upsilon}{Q}\nabla g \tag{2.23}$$

the last form being *always* valid in static fields. Thus no independent kinematic function other than $\hat{\Phi}(\boldsymbol{r}, \Phi_0)$ is needed. The gain obtained by this simplification will be obvious later.

In order to lighten the notation, we shall omit the argument Φ_0 whenever we are not concerned with energy distributions and chromatic effects.

2.4 Definition of Coordinate Systems

Many of the subsequent calculations are most favourably carried out by explicit representation in some suitably chosen coordinate system. In order to avoid repetition, we introduce general definitions and standard notations here, which we shall use consistently provided there is no risk of confusion. Any necessary changes of the notation will be mentioned explicitly.

Cartesian coordinates are denoted by (x, y, z) or by subscripts (x_1, x_2, x_3); the Cartesian components of any vector and the unit vectors $\boldsymbol{i}$ in the three Cartesian directions will always be indicated by the corresponding subscripts. Even when using orthogonal curvilinear coordinates, the unit vector indicating the direction of the gradient of the coordinate in question will be denoted by $\boldsymbol{i}$ and the corresponding subscript. Examples are given below. In sums of Cartesian products, the familiar summation convention will be used wherever this does not cause confusion.

Any vector $\boldsymbol{a}$ is thus represented by the equivalent notations

$$\boldsymbol{a} = a_x\boldsymbol{i}_x + a_y\boldsymbol{i}_y + a_z\boldsymbol{i}_z = \sum_{j=1}^{3} a_j\boldsymbol{i}_j = a_j\boldsymbol{i}_j \tag{2.24}$$

and scalar products are written

$$\boldsymbol{a}\cdot\boldsymbol{b} = \sum_{j=1}^{3} a_j b_j = a_j b_j \tag{2.25}$$

As well as Cartesian coordinates, we shall need cylindrical coordinates (z, r, φ) and spherical coordinates (R, ϑ, φ). All these coordinate systems are related by the familiar transformations

$$\begin{aligned} &x = R\sin\vartheta\cos\varphi = r\cos\varphi \\ &y = R\sin\vartheta\sin\varphi = r\sin\varphi \\ &z = R\cos\vartheta, \quad r = R\sin\vartheta \\ &0 \le r \le R, \quad 0 \le \vartheta < \pi, \quad 0 \le \varphi < 2\pi \end{aligned} \tag{2.26}$$

The choice of notation for the spherical coordinate R is unusual but spherical coordinates are very rarely used in electron optics. The only important case is the treatment of cathode tips (Chapter 45 of Volume 2). Cylindrical coordinates are very frequently used and the corresponding notation is as simple as possible.

The element of length, ds, is given by

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ &= dz^2 + dr^2 + r^2 d\varphi^2 \\ &= dR^2 + R^2 d\vartheta^2 + R^2 \sin^2\vartheta \, d\varphi^2 \end{aligned}$$

The position vector in particular is given by

$$\boldsymbol{r} = x\boldsymbol{i}_x + y\boldsymbol{i}_y + z\boldsymbol{i}_z = z\boldsymbol{i}_z + r\boldsymbol{i}_r = R\boldsymbol{i}_R \tag{2.27a}$$

and the velocity by

$$\begin{aligned} \boldsymbol{v} = \dot{\boldsymbol{r}} &= \dot{x}\boldsymbol{i}_x + \dot{y}\boldsymbol{i}_y + \dot{z}\boldsymbol{i}_z = \dot{z}\boldsymbol{i}_z + \dot{r}\boldsymbol{i}_r + r\dot{\varphi}\boldsymbol{i}_\varphi \\ &= \dot{R}\boldsymbol{i}_R + R\dot{\vartheta}\boldsymbol{i}_\vartheta + R\sin\vartheta\dot{\varphi}\boldsymbol{i}_\varphi \end{aligned} \tag{2.27b}$$

The components of arbitrary vectors, characterized by the corresponding subscripts, have the value of the corresponding projections; for example, if $\boldsymbol{a} = a_R\boldsymbol{i}_R + a_\vartheta\boldsymbol{i}_\vartheta + a_\varphi\boldsymbol{i}_\varphi$, then $a_R = \boldsymbol{a} \cdot \boldsymbol{i}_R$. This is clearly different from the familiar, but less convenient, covariant formalism, which we do not need here. A simple consequence of our definitions is that

$$|\boldsymbol{a}|^2 = a_j a_j = a_z^2 + a_r^2 + a_\varphi^2 = a_R^2 + a_\vartheta^2 + a_\varphi^2 \tag{2.28}$$

No metric tensor is needed in this context. To facilitate the evaluation of vector products, it is useful to remember that the basic vectors, $\boldsymbol{i}_z$, $\boldsymbol{i}_r$, $\boldsymbol{i}_\varphi$ and $\boldsymbol{i}_R$, $\boldsymbol{i}_\vartheta$, $\boldsymbol{i}_\varphi$ respectively, form positively orientated orthonormalized bases.

In order to avoid giving explicit Cartesian representations of vector differentiations in different vector spaces, we introduce the familiar abbreviation

$$\frac{\partial}{\partial \boldsymbol{u}} = \boldsymbol{i}_x \frac{\partial}{\partial u_x} + \boldsymbol{i}_y \frac{\partial}{\partial u_y} + \boldsymbol{i}_z \frac{\partial}{\partial u_z} = \boldsymbol{i}_j \frac{\partial}{\partial u_j} \tag{2.29}$$

for differentiation in the space in question. Ordinary vector differentiation is denoted by $\nabla = \partial/\partial \boldsymbol{r}$. Some common representations in cylindrical coordinates are as follows:

$$\nabla = \boldsymbol{i}_z \frac{\partial}{\partial z} + \boldsymbol{i}_r \frac{\partial}{\partial r} + \boldsymbol{i}_\varphi \frac{1}{r}\frac{\partial}{\partial \varphi} \tag{2.29a}$$

$$\nabla^2 = \frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2} \tag{2.29b}$$

These are extensively used in Part II. Partial derivatives are denoted in various different ways, as shown in the following examples:

$$\frac{\partial^2 V}{\partial x^2} \equiv \partial^2_{xx} V \equiv V_{|xx}$$

$$\frac{\partial^2 V}{\partial x \partial y} \equiv \partial^2_{xy} V \equiv V_{|xy}$$

Expressions of this kind will appear intermittently in later chapters.

2.5 Conservation of Axial Angular Momentum

This conservation law may be obtained very elegantly by use of the Lagrange formalism but the following elementary presentation brings out its physical meaning clearly.

We now assume that the electron optical device is rotationally symmetric about an optic axis. This assumption is valid in all perfect round lenses and mirrors. The fields in these devices are usually also static, but this additional assumption is not necessary. The following considerations remain valid in time-dependent systems and in ion optics. As everywhere in physics, the conservation of axial angular momentum is a direct consequence of the assumption of rotational symmetry.

On forming the vector product of $\boldsymbol{r}$ with Eq. (2.1), we see that the expression on the left-hand side is the derivative of the familiar mechanical angular momentum vector:

$$\frac{d}{dt}(m\boldsymbol{r} \times \boldsymbol{v}) = Q\boldsymbol{r} \times (\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{2.30}$$

On the right-hand side, only the component parallel to the axis can be represented as a total derivative, and we therefore confine our attention to the evaluation of this component.

In view of the assumed symmetry, we may anticipate that cylindrical coordinates (z, r, φ) will be most appropriate, the z-axis coinciding with the optic axis. The axial component of Eq. (2.30) then becomes

$$\frac{d}{dt}\left(mr^2\dot{\varphi}\right) = Qr\left\{E_\varphi + (\boldsymbol{v} \times \boldsymbol{B})_\varphi\right\} = Qr(E_\varphi + \dot{z}B_r - \dot{r}B_z) \tag{2.31}$$

In order to transform the expression on the right-hand side into a total derivative with respect to time, it is necessary to introduce the *magnetic flux function* $\Psi(z, r, t)$. This is defined as the magnetic flux through a coaxial circular disc C, located in a plane $z = \text{const}$ and of radius r. The coordinates (z, r, φ) specify the instantaneous position of the charged particle on its trajectory. This is illustrated in Fig. 2.1.

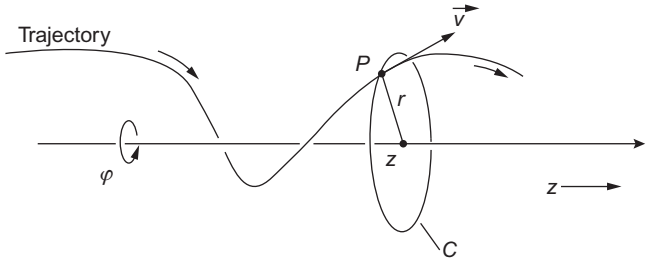


Figure 2.1
Trajectory of a charged particle; a coaxial circular disc C is associated with an arbitrary point P with coordinates $(z, r, \varphi; t)$. This circle travels along the axis and its radius varies as the point P moves along the trajectory.

This function $\Psi(z, r, t)$ is easily evaluated by expressing the surface element $d\boldsymbol{a} = da\,\boldsymbol{i}_z = 2\pi r' dr' \boldsymbol{i}_z$ in cylindrical coordinates, giving

$$\Psi(z, r, t) = \int_C \boldsymbol{B} \cdot d\boldsymbol{a} = 2\pi \int_0^r r' B_z(z, r', t) dr' \tag{2.32}$$

Differentiation with respect to r gives immediately

$$B_z(z, r, t) = \frac{1}{2\pi r} \frac{\partial \Psi}{\partial r} \tag{2.33}$$

Using the condition div $\boldsymbol{B} = 0$, the second representation

$$B_r(z, r, t) = -\frac{1}{2\pi r} \frac{\partial \Psi}{\partial z} \tag{2.34}$$

can be derived (see Section 6.4). Finally the integral induction law $\oint \boldsymbol{E} \cdot d\boldsymbol{r} = -\partial \Psi / \partial t$, applied to the circumference of the disc C, results in

$$E_\varphi(z, r, t) = -\frac{1}{2\pi r} \frac{\partial \Psi}{\partial t} \tag{2.35}$$

Introducing Eqs. (2.33), (2.34) and (2.35) into (2.31) we obtain:

$$\frac{d}{dt}\left(m r^2 \dot{\varphi}\right) = -\frac{Q}{2\pi}\left(\frac{\partial \Psi}{\partial t} + \dot{z}\frac{\partial \Psi}{\partial z} + \dot{r}\frac{\partial \Psi}{\partial r}\right) = -\frac{Q}{2\pi}\frac{d\Psi}{dt} \tag{2.36}$$

Integration with respect to time leads to the conservation law for the axial angular momentum:

$$N := \frac{m_0 r^2 \dot{\varphi}}{\sqrt{1 - \beta^2}} + \frac{Q}{2\pi}\Psi(z, r, t) = \text{const} \tag{2.37}$$

This expression differs from that familiar in classical mechanics by the presence of the important second term, which contains the electromagnetic interaction in integral form. Its physical origin and meaning appear very clearly on comparing Eq. (2.31) with (2.36). The *electric* torque, QrE_φ, is an induction effect in a *fixed* loop (specified by z and r). The *magnetic* part of the torque, $Qr(\boldsymbol{v} \times \boldsymbol{B})_\varphi$, is the corresponding induction effect in a loop *moving* with the particle and is hence a consequence of the implicit variation of Ψ with time: $-Q\boldsymbol{v} \cdot \nabla\Psi/2\pi$. This latter part remains nonzero even in static magnetic round lenses where it forces the particle beam to rotate about the optic axis, a fact of great importance in the physics of magnetic lenses.

In the vicinity of the optic axis, this rotation is of most interest and can be easily calculated. In a sufficiently small paraxial domain, the magnetic field can be considered as radially homogeneous; the magnetic flux is then $\Psi = \pi r^2 B(z, t)$, the function $B(z, t)$ being the axial flux density. Furthermore, we have $\dot{r}^2 + r^2\dot{\varphi}^2 \ll \dot{z}^2$. From Eq. (2.37), with $m(z, t) = m_0(1 - \dot{z}^2/c^2)^{-1/2}$, we then obtain

$$N = r^2\left\{m(z,\ t)\dot{\varphi} + \frac{1}{2}QB(z,\ t)\right\}$$

We now confine our considerations to trajectories that intersect the optic axis at some point. For such trajectories, N vanishes and we obtain

$$\dot{\varphi} = -\frac{QB(z,t)}{2m(z,t)} \tag{2.38}$$

This is the local and instantaneous value of the familiar *Larmor* frequency.

In the important case of electron trajectories in static round lenses, it is convenient to represent the azimuth as a function of z. On the optic axis the relation

$$m\dot{z} = \sqrt{2m_0 e\hat{\phi}_0(z)}$$

is satisfied, $\phi_0(z)$ being the axial acceleration potential. Using $d\varphi/dz = \dot{\varphi}/\dot{z}$ and integrating with respect to z we find

$$\varphi(z) = \eta\int_{z_0}^{z} \hat{\phi}_0^{-1/2}(z')B(z')dz' + \varphi(z_0) \tag{2.39}$$

In Part III we shall show that the trajectory equations simplify considerably in a coordinate system that is twisted round the optic axis by this local angle of rotation.

CHAPTER 3

Different Forms of Trajectory Equations

The Lorentz equation (2.1) is generally valid but not always convenient. In very many practical applications, expressions for the trajectories of the form $\boldsymbol{r} = \boldsymbol{r}(t)$ are of no interest; it is their geometrical shape that is required. In devices in which the fields $\boldsymbol{E}$ and $\boldsymbol{B}$ depend explicitly on the time, it is rarely possible to eliminate the latter from the trajectory equation. We therefore confine the following discussion to systems with static fields. In order to avoid repetition, we shall give here only those forms of the trajectory equations that cannot (or at least, only with greater difficulty) be derived from variational principles but are yet of practical importance.

3.1 Parametric Representation in Terms of the Arc-Length

The differential arc-length is given by $ds := |d\boldsymbol{r}| = \upsilon dt$. The transformation of the differential operator is hence given by $d/dt = \upsilon d/ds$. Introducing this into Eq. (2.1) and noting that $\boldsymbol{\upsilon}/\upsilon = d\boldsymbol{r}/ds$, we obtain

$$\frac{d\boldsymbol{g}}{ds} = \frac{Q\boldsymbol{E}}{\upsilon} + Q\frac{d\boldsymbol{r}}{ds} \times \boldsymbol{B}$$

It is now of great importance that υ and $g = m\upsilon$ are unique functions of $\boldsymbol{r}$, as explained in Chapter 2, Relativistic Kinematics. Thus the time is already eliminated. By means of Eq. (2.23), the electric field vector can also be eliminated, giving

$$\frac{d}{ds}\left\{g(\boldsymbol{r})\frac{d\boldsymbol{r}}{ds}\right\} = \text{grad}\ \ g(\boldsymbol{r}) + Q\frac{d\boldsymbol{r}}{ds} \times \boldsymbol{B}(\boldsymbol{r}) \tag{3.1}$$

In the absence of the magnetic term, this trajectory equation is even valid for the propagation of light in matter provided that the geometric approximation is adequate (see Born and Wolf, 1959 (3.2.2); Kasper, 1972). In this case the kinetic momentum is given by $g = \hbar k_0 n(\boldsymbol{r}, k_0)$, k_0 being the mean wave number (2π/wavelength) in the vacuum and $n(\boldsymbol{r}, k_0)$ the corresponding index of refraction in the material; $\hbar = h/2\pi$ where h is Planck's constant. Hence Eq. (3.1) is a very general trajectory equation.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00003-1
© 2018 Elsevier Ltd. All rights reserved.

Some useful relations can be derived from Eq. (3.1) by calculating its components with respect to the orthonormal basis vectors

$$\boldsymbol{t} \coloneqq \frac{d\boldsymbol{r}}{ds}, \quad \boldsymbol{n} \coloneqq R\frac{d\boldsymbol{t}}{ds}, \quad \boldsymbol{b} \coloneqq \boldsymbol{t} \times \boldsymbol{n} \tag{3.2}$$

the local tangent, principal normal and binormal respectively, R being the (positive) radius of curvature. Scalar multiplication of Eq. (3.1) with each of these unit vectors in turn, using $dg/ds = \boldsymbol{t} \cdot \text{grad}\ g$ and some elementary vector operations, yields:

$$\frac{g}{R} = \boldsymbol{n} \cdot \text{grad}\ \ g - Q\boldsymbol{b} \cdot \boldsymbol{B} \equiv Q(\boldsymbol{n} \cdot \boldsymbol{E}/\upsilon - \boldsymbol{b} \cdot \boldsymbol{B}) \tag{3.3}$$

$$0 = \boldsymbol{b} \cdot \text{grad}\ \ g + Q\boldsymbol{n} \cdot \boldsymbol{B} \equiv Q(\boldsymbol{b} \cdot \boldsymbol{E}/\upsilon + \boldsymbol{n} \cdot \boldsymbol{B}) \tag{3.4}$$

In a purely *magnetic* field ($E \equiv 0$), Eq. (3.4) shows that the vector $\boldsymbol{n}$ is always orthogonal to $\boldsymbol{B}$. From Eq. (3.3) the (absolute) curvature is found to be

$$\frac{1}{R} = -\frac{Q}{g}\boldsymbol{b} \cdot \boldsymbol{B} \equiv \frac{1}{g}|Q\boldsymbol{t} \times \boldsymbol{B}| \tag{3.5}$$

For electrons the curvature can be rewritten with the aid of the formulae of Section 2.3 as

$$\frac{1}{R} = \frac{\eta}{\sqrt{\hat{\phi}}}|\boldsymbol{t} \times \boldsymbol{B}| \tag{3.6}$$

where $\hat{U} \coloneqq \hat{\Phi}$ is now a *constant* acceleration potential. Eq. (3.1) then simplifies to:

$$\frac{d^2\boldsymbol{r}}{ds^2} = \frac{\eta}{\sqrt{\hat{U}}}\boldsymbol{B}(\boldsymbol{r}) \times \frac{d\boldsymbol{r}}{ds} \tag{3.7}$$

For the circular motion of charged particles in a homogeneous magnetic field, the familiar relation

$$BR = g|Q|^{-1} \tag{3.7a}$$

is a simple consequence of Eq. (3.5). This 'BR-product' is of importance in the design of spectrometers and analysers (see Part X).

In a purely *electrostatic* field the binormal $\boldsymbol{b}$ is always orthogonal to $\boldsymbol{E}$, and Eq. (3.3) can then be rewritten in the familiar self-evident form

$$\frac{m\upsilon^2}{R} = |QE_n| = Q|\boldsymbol{t} \times \boldsymbol{E}| \tag{3.8}$$

The dependence of the left-hand side on the acceleration potential can be obtained by use of the kinematic functions, given in Section 2.3.

3.2 Relativistic Proper-Time Representation

The arc-length is a highly unsuitable trajectory parameter for numerical trajectory computations in electron guns and mirrors, since the radius of curvature varies over an extremely wide range of values. This implies that the integration step length Δs must also vary considerably. The time would be a much better parameter but has the disadvantage that the range of values is now too small. Other parameterizations such as that of Kel'man et al. (1972, 1973a,b) are of specific interest for mirrors. These will be dealt with in Chapter 18, Electron Mirrors.

A favourable parameter, which exhibits none of the disadvantages mentioned above and is very advantageous, particularly in the numerical investigation of electron guns, is defined by (Kasper, 1985):

$$d\tau = u\sqrt{1-\beta^2}dt = udt/\gamma = udt' \quad (u = \text{const}) \tag{3.9}$$

dt' being the relativistic proper-time element, observable in the frame of reference of the electron. The observable nature of the proper-time element is, of course, purely abstract, but the trajectory equation can be slightly simplified by introducing Eq. (3.9). A particularly suitable choice of the constant factor u is

$$u = \sqrt{\frac{2e}{m_0}\hat{U}} = 2\eta\sqrt{\hat{U}} \tag{3.10}$$

the second constant $\hat{U}$ being the relativistic acceleration potential at some suitably chosen fixed point of reference. From Eqs. (3.9) and (3.10), it is obvious that u has the dimensions of a speed and thus $d\tau$ that of a length, though $d\tau$ is proportional in magnitude to dt'.

Eq. (3.1) can now be straightforwardly transformed by means of the operator relation

$$\frac{1}{\sqrt{1-\beta^2}}\frac{d}{dt} = u\frac{d}{d\tau}$$

The kinematic momentum is given by

$$\boldsymbol{g} = m_0 u\frac{d\boldsymbol{r}}{d\tau} = (2m_0 e\hat{U})^{1/2}\frac{d\boldsymbol{r}}{d\tau} \tag{3.11}$$

which is already a simplification, since the factor before the derivative is now a constant, rather than a complicated function. Multiplying by the factor $\gamma = (1-\beta^2)^{-1/2}$ and using the relations given above, the Lorentz equation (2.1) for electrons ($Q = -e$) now transforms into

$$m_0 u^2 \frac{d^2\boldsymbol{r}}{d\tau^2} = -e\gamma\boldsymbol{E} + e\boldsymbol{B} \times \frac{d\boldsymbol{r}}{d\tau}$$

The electric term is eliminated by means of Eq. (2.23). Using Eq. (3.10) we obtain finally

$$\frac{d^2\boldsymbol{r}}{d\tau^2} = \frac{1}{2}\,\mathrm{grad}\left(\frac{\hat{\Phi}(\boldsymbol{r})}{\hat{U}}\right) + \frac{\eta}{\sqrt{\hat{U}}}\boldsymbol{B}(\boldsymbol{r}) \times \frac{d\boldsymbol{r}}{d\tau} \tag{3.12}$$

The normalization constant $\hat{U}$ can be set arbitrarily to any positive value. In devices having an asymptotically field-free domain, we may choose any reference point $\boldsymbol{R}_0$ located in this domain: $\hat{U} = \hat{\Phi}(\boldsymbol{R}_0)$. This implies that $d\tau = ds$ in that domain. In an electron microscope a good choice for the reference point $\boldsymbol{R}_0$ is the centre of the recording screen. Then $\hat{\Phi}(\boldsymbol{R}_0)$ is usually the maximum of $\hat{\Phi}$. Without loss of generality, the starting point of each electron trajectory at the cathode surface may be the point with $\tau = 0$. The final value of τ is then slightly longer than the length L of the device, roughly $\tau_{\max} \sim 1.5L$. There is thus no need for any special precautions to avoid inconvenient scales.

The conservation laws for energy and axial angular momentum can also be represented in a very convenient form. By scalar multiplication of Eq. (3.12) with $d\boldsymbol{r}/d\tau$ and integration with respect to τ we find

$$\left|\frac{d\boldsymbol{r}}{d\tau}\right|^2 = \frac{\hat{\Phi}(\boldsymbol{r})}{\hat{U}} \tag{3.13}$$

With the conventions adopted in Section 2.3, the constant of integration must be zero.

For motion in static fields, Eq. (2.37) can be simplified with the aid of kinematic functions, the result being

$$r^2 \frac{d\varphi}{d\tau} \equiv x\frac{dy}{d\tau} - y\frac{dx}{d\tau} = \frac{\eta}{\sqrt{\hat{U}}}\left(\frac{N}{e} + \frac{\Psi(z,r)}{2\pi}\right) \tag{3.14}$$

These conservation laws are useful as additional checks of the accuracy in numerical computations. In practice, the evaluation of Eq. (3.12) has proved to be the most successful method of calculating Lorentz trajectories.

3.3 The Cartesian Representation

The various representations of the trajectory equation derived hitherto are suitable for calculating individual trajectories from given initial conditions but they are not at all suitable for developing a systematic theory of focusing and aberrations. For this, a Cartesian representation $x = x(z)$, $y = y(z)$ is preferable. Such a formalism is possible if the electron optical device in question has a straight optic axis, if $\hat{\Phi} > 20$ eV and if the slopes $x'(z)$, $y'(z)$

of *all* trajectories remain finite. The last two conditions are not satisfied in electron guns and mirrors and any entirely satisfactory theory of the aberrations in these devices must take this into account.

In what follows, we shall use the explicit Cartesian representation Eq. (2.27) with $x = x(z)$, $y = y(z)$. Differentiation with respect to z will be denoted by a prime, thus

$$\rho \boldsymbol{t}(z) \equiv \boldsymbol{r}'(z) = x'(z)\boldsymbol{i}_x + y'(z)\boldsymbol{i}_y + \boldsymbol{i}_z \tag{3.15}$$

Since $|\boldsymbol{t}| = 1$, the absolute value of $\boldsymbol{r}'$ is:

$$\rho \coloneqq |\boldsymbol{r}'| = \sqrt{1 + x'^2 + y'^2} \tag{3.16}$$

This function has a very simple geometric meaning: $\rho^{-1} = \boldsymbol{t} \cdot \boldsymbol{i}_z = \cos\,\alpha$, α being the angle between the local tangent and the optic axis. This holds even for skew trajectories.

The required Cartesian representation of the trajectory equation is most easily obtained by substituting

$$\frac{d}{ds} = \frac{dz}{ds}\frac{d}{dz} = \frac{1}{\rho}\frac{d}{dz}$$

in Eq. (3.1), giving

$$\frac{1}{\rho}\frac{d}{dz}\left(\frac{g}{\rho}\frac{d\boldsymbol{r}}{dz}\right) = \text{grad}\, g + \frac{Q}{\rho}\boldsymbol{r}' \times \boldsymbol{B}$$

Expanding the derivative on the left-hand side yields

$$\frac{g}{\rho^2}\boldsymbol{r}'' + \frac{\boldsymbol{r}'}{\rho}\frac{d}{dz}\left(\frac{g}{\rho}\right) = \text{grad}\, g + \frac{Q}{\rho}\boldsymbol{r}' \times \boldsymbol{B} \tag{3.17}$$

These are three scalar differential equations for the two functions $x(z)$, $y(z)$; the third equation

$$\frac{1}{\rho}\frac{d}{dz}\left(\frac{g}{\rho}\right) = \frac{\partial g}{\partial z} + \frac{Q}{\rho}\boldsymbol{i}_z \cdot (\boldsymbol{r}' \times \boldsymbol{B}) \tag{3.18}$$

is therefore dependent on the first two and may be omitted. In fact, it is possible to derive the x- and y-components of Eq. (3.17) directly from a two-dimensional variational principle but not Eq. (3.18). Here we shall use Eq. (3.18) to simplify Eq. (3.17) by eliminating the second term. Multiplying Eq. (3.18) by $\boldsymbol{r}'$ and subtracting the result from Eq. (3.17), we find

$$\frac{g}{\rho^2}\boldsymbol{r}'' = \text{grad}\, g - \boldsymbol{r}'\frac{\partial g}{\partial z} + \frac{Q}{\rho}\left\{\boldsymbol{r}' \times \boldsymbol{B} - \boldsymbol{i}_z \cdot (\boldsymbol{r}' \times \boldsymbol{B})\boldsymbol{r}'\right\}$$

The third component of this is a trivial identity. The two components of interest are given explicitly by

$$x'' = \frac{\rho^2}{g}\left(\frac{\partial g}{\partial x} - x'\frac{\partial g}{\partial z}\right) + \frac{Q\rho}{g}\left\{y'(B_z + x'B_x) - B_y(1 + x'^2)\right\}$$
$$y'' = \frac{\rho^2}{g}\left(\frac{\partial g}{\partial y} - y'\frac{\partial g}{\partial z}\right) + \frac{Q\rho}{g}\left\{-x'(B_z + y'B_y) + B_x(1 + y'^2)\right\}$$

The magnetic terms can be rewritten in a more compact form by introducing the tangential component of B:

$$B_t \coloneqq \boldsymbol{t} \cdot \boldsymbol{B} = \frac{1}{\rho}(B_z + x'B_x + y'B_y) \tag{3.19}$$

and we finally obtain

$$\begin{aligned} x'' &= \frac{\rho^2}{g}\left(\frac{\partial g}{\partial x} - x'\frac{\partial g}{\partial z}\right) + \frac{Q\rho^2}{g}(y'B_t - \rho B_y) \\ y'' &= \frac{\rho^2}{g}\left(\frac{\partial g}{\partial y} - y'\frac{\partial g}{\partial z}\right) - \frac{Q\rho^2}{g}(x'B_t - \rho B_x) \end{aligned} \tag{3.20}$$

These trajectory equations are valid for all charged particles, provided that the conditions mentioned above are satisfied. With $Q = 0$ and $g = \hbar k_0 n(\boldsymbol{r})$, they are even valid for light rays, $n(\boldsymbol{r})$ being the optical index of refraction. The vacuum momentum $\hbar k_0$ cancels out and we obtain the ray equations

$$x'' = \frac{\rho^2}{n}\left(\frac{\partial n}{\partial x} - x'\frac{\partial n}{\partial z}\right), \quad y'' = \frac{\rho^2}{n}\left(\frac{\partial n}{\partial y} - y'\frac{\partial n}{\partial z}\right) \tag{3.21}$$

For electron trajectories, Eqs. (2.19) and (2.17) may be used, whereupon Eq. (3.20) become

$$\begin{aligned} x'' &= \frac{\rho^2}{2\hat{\Phi}}\left(\frac{\partial \hat{\Phi}}{\partial x} - x'\frac{\partial \hat{\Phi}}{\partial z}\right) + \frac{\eta\rho^2}{\sqrt{\hat{\Phi}}}(\rho B_y - y'B_t) \\ y'' &= \frac{\rho^2}{2\hat{\Phi}}\left(\frac{\partial \hat{\Phi}}{\partial y} - y'\frac{\partial \hat{\Phi}}{\partial z}\right) - \frac{\eta\rho^2}{\sqrt{\hat{\Phi}}}(\rho B_x - x'B_t) \end{aligned} \tag{3.22}$$

These trajectory equations are equally well suited for the numerical computation of individual trajectories and for the development of a systematic theory of focusing and

aberrations. It must be emphasized that only the existence of a straight optic axis is required, not rotational symmetry about this axis. Thus not only can round lenses be considered here, but also stigmators, systems of multipole lenses and deflection units such as saddle coils.

3.4 Scaling Rules

A number of simple scaling rules can be derived for trajectories in static fields. Since the time-dependent form of the trajectories is now of no interest, we confine the discussion to the time-independent form. Scale changes are most easily performed on Eq. (3.1). We shall consider only two very important special cases.

For purely *electrostatic* fields, Eq. (3.1) becomes linear in $g(\boldsymbol{r})$, which means that Eq. (3.1) is unaffected by a scale transform $g(\boldsymbol{r}) = g_0 g^*(\boldsymbol{r})$ of the kinematic momentum. Any experimental change of scales, however, is made by alteration of the electrode potentials and this affects the kinematic momentum only indirectly. From Eq. (2.13), it is obvious that a simple rule can only be expected in the nonrelativistic case $|E_0 - Q\Phi(\boldsymbol{r})| \ll m_0 c^2$. Since the initial energy E_0 is not constant in a particle beam, a unique change of scale common to all the particles is only possible if $E_0 \ll |Q\Phi|$. For electrons and negatively charged ions, Eq. (2.13) then simplifies to:

$$g(\boldsymbol{r}) = \sqrt{2m_0|Q|\Phi(\boldsymbol{r})}, \quad \Phi \geq 0 \tag{3.23}$$

which is only valid sufficiently far from the cathode. In these circumstances, a linear scale transform $\Phi(\boldsymbol{r}) = U_0 \Phi^*(\boldsymbol{r})$ of the potential is equivalent to such a transform of the kinematic momentum, the relation between the scaling factors being given by $g_0 = {U_0}^{1/2}$. Introducing Eq. (3.23) into (3.1), the factor $(2m_0|Q|)^{1/2}$ is seen to cancel; the trajectory equation then simplifies to

$$\frac{d}{ds}\left(\sqrt{\Phi(\boldsymbol{r})}\frac{d\boldsymbol{r}}{ds}\right) = \operatorname{grad}\sqrt{\Phi(\boldsymbol{r})} \tag{3.24}$$

which can be rewritten as

$$2\Phi(\boldsymbol{r})\frac{d^2\boldsymbol{r}}{ds^2} = \left\{\frac{d\boldsymbol{r}}{ds} \times \nabla\Phi(\boldsymbol{r})\right\} \times \frac{d\boldsymbol{r}}{ds} \tag{3.25}$$

Here the relation $d\Phi/ds = (d\boldsymbol{r}/ds) \cdot \nabla\Phi$ has been used. This trajectory equation is *linear* in $\Phi(\boldsymbol{r})$. A linear scale change $\Phi = U_0\Phi^*$ can be performed most easily by alteration of all the electrode potentials by the same factor, the cathode being $\Phi^* = 0$. From Eq. (3.25) it is obvious that the geometric shape of the trajectories depends neither on Q and m_0 nor on the scale factor U_0. These constants affect only the time of propagation, which is of little interest.

These conclusions are together known as the *electrostatic principle*, which also holds for positively charged particles, provided that $\Phi(\boldsymbol{r})$ is replaced by $|\Phi(\boldsymbol{r})|$ and $\Phi(\boldsymbol{r}) = 0$ is now the emitting anode surface. In practice, a consequence of this principle is that an electrostatic microscope can be operated with different kinds of particles and fixed *ratios* of the voltage differences between the acceleration electrodes and the source.

The consequences of a geometric scale change can also be derived from Eq. (3.25). A transform

$$\boldsymbol{r} = a\boldsymbol{r}^*, \quad s = as^*, \quad \Phi^*(\boldsymbol{r}^*) = \Phi(\boldsymbol{r})/U_0 \tag{3.26}$$

does not affect the geometric shape of the trajectories, since Eq. (3.25) is invariant with respect to Eq. (3.26). The magnitude of the trajectories is proportional to the distances between the electrodes if the *shape* of the field is unaltered.

In purely *magnetostatic* devices the solutions of the trajectory equations depend on Q/m_0. We shall confine our considerations to electron motion; we must then investigate the effect of scale changes on Eq. (3.7). The constant $\hat{U}$ will now be the relativistic acceleration potential. Introducing the scale transforms

$$\boldsymbol{r} = a\boldsymbol{r}^*, \quad s = as^*, \quad \hat{U} = \hat{U}_0\hat{U}^*, \quad B(\boldsymbol{r}) = B_0B^*(\boldsymbol{r}^*) \tag{3.27}$$

into Eq. (3.7) we find that this trajectory equation remains invariant if the condition

$$aB_0 = \sqrt{\hat{U}_0} \tag{3.28}$$

is satisfied. This can be put into a more practical form by introducing scales for the field-producing currents. The magnetic field strength B is related to the electric current I by an expression of the form $B = \mu I/l$, l being some typical length and μ a permeability. An appropriate scale transform is now given by

$$B = B_0B^*, \quad I = I_0I^*, \quad l = al^*, \quad aB_0 = I_0 \tag{3.29}$$

which must be compatible with Eq. (3.27). Eq. (3.28) now simplifies to

$$I_0^2 = \hat{U}_0 \tag{3.30}$$

This simple scaling rule is of great help in the practical design of magnetic lenses.

CHAPTER 4

Variational Principles

All our analysis has so far been based on the Lorentz equation (2.1). This equation is, however, identical with the Euler–Lagrange equations of Hamilton's variational principle which may hence be regarded as more fundamental. This principle may be stated thus:

$$W := \int_{t_0}^{t_1} L(\boldsymbol{r}, \boldsymbol{v}, t)dt = \text{extr.} \tag{4.1}$$

where 'extr.' denotes an extremum or at least a stationary value. The necessary constraints are that t_0, t_1, $\boldsymbol{r}(t_0)$ and $\boldsymbol{r}(t_1)$ remain fixed and that the variable of integration must not be varied: $\delta t = 0$.

4.1 The Lagrange Formalism

In charged particle dynamics the integrand L, the Lagrangian, takes the form

$$L = m_0 c^2 \left(1 - \sqrt{1 - \beta^2}\right) + Q(\boldsymbol{v}\cdot\boldsymbol{A} - \Phi) \tag{4.2}$$

$\Phi(\boldsymbol{r}, t)$ and $\boldsymbol{A}(\boldsymbol{r}, t)$ being the electromagnetic potentials (e.g., Goldstein, 1959; Goldstein et al., 2001). These are related to the field vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ by

$$\boldsymbol{B}(\boldsymbol{r}, t) = \text{curl } \boldsymbol{A}(\boldsymbol{r}, t) \tag{4.3}$$

$$\boldsymbol{E}(\boldsymbol{r}, t) = -\text{ grad } \Phi(\boldsymbol{r}, t) - \frac{\partial}{\partial t}\boldsymbol{A}(\boldsymbol{r}, t) \tag{4.4}$$

These relations do *not* provide a unique definition of the potentials; in other words, the same field vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ may be obtained from *different* sets of potentials. The consequences of adopting different gauges for Φ and $\boldsymbol{A}$ are discussed in Section 5.5. Since only $\boldsymbol{E}$ and $\boldsymbol{B}$ have physical significance, but not Φ and $\boldsymbol{A}$, the results of all calculations should be presented in a gauge-invariant form.

It is convenient to rewrite the Lagrangian, L (Eq. 4.2), as $L = T^* - V^*$ with

$$T^*(v) := m_0 c^2 (1 - \sqrt{1 - \beta^2}) \tag{4.5}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00004-3
© 2018 Elsevier Ltd. All rights reserved.

$$V^*(\boldsymbol{r}, \boldsymbol{v}, t) \coloneqq Q\Phi(\boldsymbol{r}, t) - Q\boldsymbol{v}\cdot\boldsymbol{A}(\boldsymbol{r}, t) \tag{4.6}$$

since this closely resembles the familiar classical form $L = T - V$. The function $T^*(v)$, known as the kinetic potential, is closely related to the kinetic energy T, by $T^* = T(1 - \beta^2)^{1/2}$ (see Eq. 2.7). The function V^* is a generalization of the familiar potential energy V Eq. (2.8), since it is now velocity-dependent.

Since there are no geometric constraints in charged particle dynamics, we adopt Cartesian coordinates. The Euler–Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_j}\right) - \frac{\partial L}{\partial x_j} = 0 \quad (j = 1, 2, 3) \tag{4.7a}$$

are obviously equivalent to

$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial v_j}\right) = \frac{d}{dt}\left(\frac{\partial V^*}{\partial v_j}\right) - \frac{\partial V^*}{\partial x_j} \tag{4.7b}$$

Performing the differentiations on the left-hand side, we first obtain

$$\frac{dT^*}{dv} = \frac{m_0 v}{\sqrt{1-\beta^2}} \equiv g(v) \tag{4.8}$$

But $\partial v/\partial v_j = v_j/v$ and so

$$\frac{\partial T^*}{\partial v_i} = \frac{dT^*}{dv}\frac{\partial v}{\partial v_j} = \frac{m_0 v_j}{\sqrt{1-\beta^2}} = g_j, \quad j = 1, 2, 3$$

This result is obviously the Cartesian representation of the well-known expression for the kinematic momentum. By using Eq. (2.29) with $\boldsymbol{u} = \boldsymbol{v}$, the kinematic momentum can be rewritten as

$$\boldsymbol{g}\,(\boldsymbol{v}) = \frac{\partial T^*}{\partial \boldsymbol{v}} = \frac{m_0 \boldsymbol{v}}{\sqrt{1-\beta^2}} \tag{4.9}$$

and (4.7b) as

$$\dot{\boldsymbol{g}} = \frac{d}{dt}\left(\frac{\partial V^*}{\partial \boldsymbol{v}}\right) - \frac{\partial V^*}{\partial \boldsymbol{r}} =: \boldsymbol{F} \tag{4.10}$$

We now evaluate the expression $\boldsymbol{F}$ on the right-hand side:

$$\frac{\partial V^*}{\partial \boldsymbol{v}} = -Q\boldsymbol{A}, \quad \frac{\partial V^*}{d\boldsymbol{r}} = Q\nabla\Phi - Q\frac{\partial}{\partial \boldsymbol{r}}(\boldsymbol{v}\cdot\boldsymbol{A})$$

and thus

$$\boldsymbol{F} = Q\left(-\frac{d\boldsymbol{A}}{dt} - \nabla\Phi + \frac{\partial}{\partial \boldsymbol{r}}(\boldsymbol{v}\cdot\boldsymbol{A})\right)$$

The total derivative $d\boldsymbol{A}/dt$ is given by

$$\frac{d}{dt}\boldsymbol{A}(\boldsymbol{r}(t),t) = \frac{\partial \boldsymbol{A}}{\partial t} + \dot{x}_j \frac{\partial}{\partial x_j}\boldsymbol{A} \equiv \frac{\partial \boldsymbol{A}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{A}$$

We now make use of the vector identity

$$\boldsymbol{v} \times \text{curl } \boldsymbol{A}(\boldsymbol{r}) = \text{grad } \boldsymbol{v}\cdot\boldsymbol{A}(\boldsymbol{r}) - (\boldsymbol{v}\cdot\nabla)\boldsymbol{A}(\boldsymbol{r})$$

valid for any constant $\boldsymbol{v}$ and any vector function $\boldsymbol{A}(\boldsymbol{r})$. (This may easily be verified by writing the terms out in Cartesian coordinates or, more simply still, by introducing tensor notation.) Recalling that $\boldsymbol{v}$ is to be treated as a constant in differentiations with respect to $\boldsymbol{r}$, we find

$$\boldsymbol{F} = Q\left(-\nabla\Phi - \frac{\partial \boldsymbol{A}}{\partial t} + \boldsymbol{v} \times \text{curl } \boldsymbol{A}\right)$$

Substituting for $\boldsymbol{E}$ and $\boldsymbol{B}$ from Eqs. (4.4) and (4.3), we obtain finally

$$\boldsymbol{F} \coloneqq \frac{d}{dt}\left(\frac{\partial V^*}{\partial \boldsymbol{v}}\right) - \frac{\partial V^*}{\partial \boldsymbol{r}} = Q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{4.11}$$

which means that the Lorentz force is the *functional* derivative of the generalized potential V^* given by Eq. (4.6). Eq. (4.10) in combination with Eq. (4.11) is identical with Eq. (2.1).

This calculation is only slightly longer than the usual derivation of the Lorentz equation by direct evaluation of Eq. (4.7a) without making the separation $L = T^* - V^*$, but it brings out clearly the physical meaning of the various expressions since it uses as few gauge-dependent quantities as possible. The direct evaluation of Eq. (4.7a) leads, as a first step, to the definition of the *canonical* momentum

$$p_j = \frac{\partial L}{\partial \dot{x}_j}, \quad j = 1, 2, 3$$

or, in vector notation and using Eq. (4.2)

$$\boldsymbol{p} = \frac{\partial L}{\partial \boldsymbol{v}} = \boldsymbol{g} + Q\boldsymbol{A}(\boldsymbol{r}, t) \tag{4.12}$$

Though familiar and frequently used in theoretical physics, this quantity has *no* physical meaning as an observable, since it is gauge-dependent.

A great advantage of the variational calculus is the fact that the value of the integral appearing in Eq. (4.1) is invariant with respect to transformations of the coordinates and of

the variable of integration. Since it is usually easier to perform these transformations on the Lagrangian than on the corresponding Lorentz equation, new forms of trajectory equations can be derived straightforwardly. Not every useful form can be obtained in this way, however. For instance, it is impossible to derive Eq. (3.1) as Euler−Lagrange equations, since the arc-length does not satisfy the necessary constraints and is thus not a permissible variable of integration in the action integral W. An excellent example of the beneficial use of coordinate transformations is given below.

4.2 General Rotationally Symmetric Systems

We assume that the system in question is rotationally symmetric about an optic axis and we introduce the corresponding cylindrical coordinates. The components of $\boldsymbol{E}$ and $\boldsymbol{B}$ must not depend on the azimuth φ, and it may be assumed that the same is true of A_z, A_r, A_φ and Φ. In cylindrical coordinates, the Lagrangian Eq. (4.2) is given explicitly by

$$L = m_0 c^2 \left[1 - \left\{ 1 - (\dot{z}^2 + \dot{r}^2 + r^2 \dot{\varphi}^2) c^{-2} \right\}^{1/2} \right] + Q(\dot{z} A_z + \dot{r} A_r + r \dot{\varphi} A_\varphi - \Phi) \tag{4.13}$$

A_z, A_r, A_φ and Φ being functions of z, r and t only. Hence L does not depend explicitly on φ, which means that φ is a cyclic variable. The corresponding canonical momentum is therefore a constant of motion

$$p_\varphi = \partial L / \partial \dot{\varphi} = \text{const}$$

or

$$p_\varphi = \frac{m_0 r^2 \dot{\varphi}}{\sqrt{1 - \beta^2}} + Q r A_\varphi(z, r, t) = \text{const}$$

Integrating Eq. (4.3) over the circular disk C introduced in Section 2.5 and using Stokes's theorem we find

$$\oint \boldsymbol{A} \cdot d\boldsymbol{r} = 2\pi r A_\phi = \int_C \boldsymbol{B} \cdot d\boldsymbol{a} = \Psi(z, r, t)$$

Thus the constant p_φ is identical with the axial angular momentum N of Eq. (2.37). More suggestively, we can use Eq. (4.12) to represent the axial angular momentum as

$$N = (\boldsymbol{r} \times \boldsymbol{p})_z \tag{4.14}$$

but this does *not* mean that $\boldsymbol{p}$ has become an observable quantity, even though we can measure N. We note that although the rotationally symmetric gauge adopted for the potentials is most convenient, it is not absolutely necessary. If some other (unsymmetric) gauge were used, Eq. (4.14) would not hold, whereas Eq. (2.37) always remains valid. Moreover, the derivation of the conservation law given in Section 2.5 gives more physical

insight than does the formalism presented here. The evaluation of the remaining two Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}}\right) - \frac{\partial L}{\partial r} = 0, \quad \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{z}}\right) = \frac{\partial L}{\partial z} \tag{4.15}$$

is straightforward. Expressing Eqs. (4.3) and (4.4) in cylindrical coordinates, we find

$$\begin{aligned} \frac{d}{dt}(m\dot{r}) - m\dot{r}\dot{\varphi}^2 &= Q(E_r + r\dot{\varphi}B_z - \dot{z}B_\phi) \\ \frac{d}{dt}(m\dot{z}) &= Q(E_z - r\dot{\varphi}B_r + \dot{r}B_\varphi) \end{aligned} \tag{4.16}$$

$$m = \frac{m_0}{\sqrt{1-\beta^2}} = m_0\left\{1-(\dot{z}^2+\dot{r}^2+r^2\dot{\varphi}^2)c^{-2}\right\}^{-1/2} \tag{4.17}$$

The remaining calculations are left to the reader. The same results can, of course, be obtained by transforming the Lorentz equation in the appropriate fashion.

The conservation law $p_\varphi = N = \text{const}$ may be used to eliminate $\dot{\varphi}$ from Eq. (4.16). Better still, $\dot{\varphi}$ may be eliminated from Eq. (4.13) before evaluating Eq. (4.15). Owing to the dependence of the mass on $\dot{\varphi}$ in Eq. (4.17), the resulting formulae are highly complicated. It is thus advantageous to eliminate $\dot{\varphi}$ from Eq. (4.13) only in the nonrelativistic approximation $m = m_0$. Moreover, we make the simplifying assumption that all the terms in B_φ can be neglected, since this field component is only produced by the particle beam itself and is always very weak in comparison with the external magnetic field. $B_\varphi \equiv 0$ is most easily satisfied by the gauge $A_z \equiv A_r \equiv 0$, which will henceforward be adopted. From Eq. (4.4), we see that E_z and E_r are now represented by a quasistationary approximation:

$$E_z = -\frac{\partial \Phi}{\partial z}, \quad E_r = -\frac{\partial \Phi}{\partial r} \quad \text{with} \quad \Phi = \Phi(z, r, t)$$

a simplification that is justified even in technical applications involving high-frequency devices. The essential induction effect is incorporated in the components E_φ and A_φ and thus in the dependence of Φ on time.

With all these simplifications the Lagrangian now becomes

$$L = \frac{1}{2}m_0(\dot{z}^2 + \dot{r}^2 + r^2\dot{\varphi}^2) + Q(r\dot{\varphi}A_\varphi - \Phi) \tag{4.18}$$

Solving Eq. (2.37) for $\dot{\varphi}$ in the nonrelativistic limit, we obtain

$$\dot{\varphi} = \frac{N - Q\Psi(z, r, t)/2\pi}{m_0 r^2} \tag{4.19}$$

On inserting this in Eq. (4.18) we obtain a Lagrangian that is a function of z, r, t, $\dot{z}$ and $\dot{r}$ only but is a very cumbersome expression. We can get a more compact form by making the Legendre transform

$$L^* = L - \dot{\varphi} p_\varphi \equiv L - \dot{\varphi} N \tag{4.20}$$

before substituting for $\dot{\varphi}$. This transform does not change the final equations of motion since the corresponding action integrals differ from each other only by a fixed constant and have thus the same extremal trajectories:

$$W^* = \int_{t_0}^{t_1} L^* dt = \int_{t_0}^{t_1} L dt - N(\varphi_1 - \varphi_0)$$

Substitution of Eq. (4.19) in L^* now results in

$$L^* = \frac{m_0}{2}(\dot{z}^2 + \dot{r}^2) - X(z, r, t) \tag{4.21}$$

with the *effective* potential energy

$$X(z, r, t) = Q\Phi(z, r, t) + \frac{\{N - Q\Psi(z, r, t)/2\pi\}^2}{2m_0 r^2} \tag{4.22}$$

The latter contains two contributions, the familiar electric term $Q\Phi$ and a *centrifugal* potential. The latter contains the terms involving $\dot{\varphi}$ and has the value $m_0 r^2 \dot{\varphi}^2/2$. It differs from the familiar classical form in possessing a contribution from the magnetic flux Ψ. The final form of the trajectory equations is now obtained by writing down the Euler–Lagrange equation of Eq. (4.21) (Störmer, 1904, 1906a–d, 1933):

$$m_0 \ddot{r} = -\frac{\partial X}{\partial r}, \quad m_0 \ddot{z} = -\frac{\partial X}{\partial z} \tag{4.23}$$

The third equation is Eq. (4.19), which can be rewritten in compact form if the quantity N, though constant with respect to time, is regarded as a free parameter. It is readily seen that

$$\dot{\varphi} = -\frac{\partial L^*}{\partial N} = \frac{\partial X}{\partial N} \tag{4.24}$$

In this sense L^* is a Routhian function.

In static fields, for which X is a function of z and r only, the law of conservation of energy can be simplified to

$$E_0 = \frac{m_0}{2}(\dot{z}^2 + \dot{r}^2) + X(z, r) = \text{const} \tag{4.25}$$

as can be easily verified.

This example demonstrates that the use of the Lagrange formalism in electron optics can be quite advantageous, since the elementary derivation of Eq. (4.23) turns out to be more complicated if the electromagnetic fields are time-dependent. These equations are very useful in studies of particle motion in high-frequency devices, for which alternative simple forms of the trajectory equations are not available.

4.3 The Canonical Formalism

As well as the Lagrangian, the Hamiltonian function is of great importance. The latter is needed in Hamilton–Jacobi theory, which is outlined in Chapter 5, Hamiltonian Optics; the Hamiltonian itself and the associated canonical equations of motion are needed in the theory of electron emission from cathodes (see Chapter 44 of Volume 2).

In vector notation, the Legendre transform between the Lagrange function $L(\boldsymbol{r}, \boldsymbol{v}, t)$ and the Hamilton function $H(\boldsymbol{r}, \boldsymbol{p}, t)$ has the form

$$H(\boldsymbol{r},\boldsymbol{p},t) \coloneqq \boldsymbol{p}\cdot\boldsymbol{v} - L(\boldsymbol{r},\boldsymbol{v},t) \tag{4.26}$$

in which the velocity $\boldsymbol{v}$ has to be expressed in terms of the canonical momentum $\boldsymbol{p}$. This expression is obtained by solving Eq. (4.12) for $\boldsymbol{v}$, the first step being

$$\boldsymbol{v} = \frac{1}{m}\boldsymbol{g} = \frac{1}{m}(\boldsymbol{p} - Q\boldsymbol{A})$$

The relativistic mass can be expressed as a function of $\boldsymbol{p}$ with the aid of Eq. (2.11) and we finally obtain

$$\boldsymbol{v} = \frac{c(\boldsymbol{p} - Q\boldsymbol{A})}{\sqrt{(m_0c)^2 + (\boldsymbol{p} - Q\boldsymbol{A})^2}} \tag{4.27}$$

Introducing this into Eq. (4.26), we obtain the Hamilton function

$$H = c\sqrt{(m_0c)^2 + (\boldsymbol{p} - Q\boldsymbol{A})^2} - m_0c^2 + Q\Phi \tag{4.28}$$

This expression is always valid, even in systems with time-dependent electromagnetic potentials. The canonical equations of motion are given by

$$\boldsymbol{v} \equiv \dot{\boldsymbol{r}} = \frac{\partial H}{\partial \boldsymbol{p}} \tag{4.29}$$

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{r}} \tag{4.30}$$

Eqs. (4.29) and (4.30) can easily be verified by performing the necessary differentiations. Eq. (4.29) proves to be identical with (4.27), while (4.30) is equivalent to the Lagrange equation $\dot{\boldsymbol{p}} = \partial L/\partial \boldsymbol{r}$. More generally, the canonical equations can be derived from a variational principle of least action in *phase space* (defined as the union of the vector spaces of $\boldsymbol{r}$ and $\boldsymbol{p}$). This variational principle will not be investigated here.

An important law can be established concerning the variation of the Hamiltonian with time. Taking the total derivative with respect to time of both sides of Eq. (4.26) we have first

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} + \dot{\boldsymbol{r}} \cdot \frac{\partial H}{\partial \boldsymbol{r}} + \dot{\boldsymbol{p}} \cdot \frac{\partial H}{\partial \boldsymbol{p}} = \dot{\boldsymbol{p}} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \dot{\boldsymbol{p}} - \frac{\partial L}{\partial t} - \dot{\boldsymbol{r}} \cdot \frac{\partial L}{\partial \boldsymbol{r}} - \dot{\boldsymbol{v}} \cdot \frac{\partial L}{\partial \boldsymbol{v}}$$

Using the canonical equations (4.29) and (4.30), the first part of this equation reduces to $dH/dt = \partial H/\partial t$. Definition (4.12) together with the Lagrange equation shows that the expression on the far right-hand side reduces to $-\partial L/\partial t$, thus

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \tag{4.31}$$

The most important consequence of this relation is that the value of H is *conserved* in all systems with *static* electromagnetic fields, since the only explicit dependence of H and L on time occurs in the potentials $\Phi(\boldsymbol{r}, t)$ and $A(\boldsymbol{r}, t)$. By comparison of Eq. (4.28) with Eqs. (2.6), (2.7), (2.8) and (2.11), it can be seen that the first two terms on the right-hand side represent the kinetic energy T as a function of $\boldsymbol{p}$ and $\boldsymbol{r}$, while the last term $Q\Phi$ is the potential energy $V(\boldsymbol{r})$, so that H is the same as the total energy E_0 of the motion. According to the conventions adopted in Section 2.3, this quantity $H = E_0$ is the kinetic starting energy of the electron at the cathode surface; it thus has a very concrete and important meaning, since this quantity will be used in the statistical analysis of the emission process.

4.4 The Time-Independent Form of the Variational Principle

Since the value of H is conserved in all static systems, we can cast the variational principle Eq. (4.1) into a very attractive form by eliminating L. Solving Eq. (4.26) for L and introducing the resulting expression into Eq. (4.1), we find

$$W = \int_{t_0}^{t_1} (\boldsymbol{p} \cdot \boldsymbol{v} - H) dt = \text{extr.}$$

In all *static* systems, the contribution

$$\int_{t_0}^{t_1} H \, dt = E_0(t_1 - t_0)$$

is a fixed constant, which does not affect the equation of motion and may hence be omitted. We thus obtain the (reduced) principle of least action:

$$\overline{S} \coloneqq \int_{t_0}^{t_1} \boldsymbol{p}\cdot\boldsymbol{v}\, dt = W + E_0(t_1 - t_0) = \text{extr.} \tag{4.32}$$

Elimination of the time and the introduction of any other legitimate variable of integration u is now straightforward:

$$\overline{S} = \int_{u_0}^{u_1} \boldsymbol{p}\cdot\frac{d\boldsymbol{r}}{du}du = \int_{r_0}^{r_1} \boldsymbol{p}\cdot d\boldsymbol{r} = \text{extr.} \tag{4.33}$$

The second representation shows that $\overline{S}$ is *invariant* with respect to the choice of the parameter.

In systems with a straight optic axis, apart from mirrors, the axial coordinate z is always the most convenient choice of variable of integration. Using Eqs. (4.12) and (3.15) we obtain

$$\overline{S} = \int_{z_0}^{z_1} (\boldsymbol{g} + Q\boldsymbol{A})\cdot\boldsymbol{r}'(z)dz \tag{4.33a}$$

Since the direction of the kinematic momentum vector $\boldsymbol{g}$ must be the same as that of the local tangent $\boldsymbol{t}$, we may use

$$\boldsymbol{g}\cdot\boldsymbol{r}' = g(\boldsymbol{r})\boldsymbol{t}\cdot\boldsymbol{r}' = \rho g(\boldsymbol{r})$$

where ρ is given by Eq. (3.16). Bringing all this together and writing out the expression for $\overline{S}$ explicitly, we obtain finally

$$\overline{S} = \int_{z_0}^{z_1} \overline{M}(x, y, z, x', y')\, dz \tag{4.34}$$

with

$$\overline{M}(x, y, z, x', y') = \sqrt{1 + x'^2 + y'^2}\, g(\boldsymbol{r}) + Q(x'A_x + y'A_y + A_z) \tag{4.35}$$

For electron motion, Eq. (2.19) may be used. The corresponding Euler equations

$$\frac{d}{dz}\left(\frac{\partial\overline{M}}{\partial x'}\right) = \frac{\partial\overline{M}}{\partial x} \quad , \quad \frac{d}{dz}\left(\frac{\partial\overline{M}}{\partial y'}\right) = \frac{\partial\overline{M}}{\partial y} \tag{4.36}$$

are the x- and y-components of Eq. (3.17) and will not be discussed here. An important way of developing a theory of aberrations consists in expanding Eq. (4.35) as a Taylor series before evaluating Eq. (4.36) and then developing an appropriate perturbation calculus.

4.5 Static Rotationally Symmetric Systems

A good example of the simplification achieved by eliminating the time as a trajectory parameter is the static rotationally symmetric system. Moreover, the Störmer equation of motion thus obtained is of some importance. It is advantageous to transform Eq. (4.35) into cylindrical coordinates. In static systems, the only surviving component of $\boldsymbol{A}$ is $A_\varphi = \Psi(z, r)/2\pi r$, Ψ being the magnetic flux. Thus we immediately obtain

$$\overline{M}(z, r, r', \varphi') = g(z, r)\sqrt{1 + r'^2 + r^2\varphi'^2} + Q\varphi'\Psi(z, r)/2\pi$$

Since φ is cyclic, the conservation of axial angular momentum is now expressed by

$$N = p_{\varphi'} = \frac{\partial \overline{M}}{\partial \varphi'} = \frac{g(z, r)r^2\varphi'}{(1 + r'^2 + r^2\varphi'^2)^{1/2}} + \frac{Q\Psi(z, r)}{2\pi} = \text{const} \tag{4.37}$$

On comparison with Eq. (2.37), it is clear that the value of N is *invariant* with respect to the parametric transform. Every canonical momentum can be shown to be invariant with respect to every parametric transform in variational theory.

Even for relativistic motion, it is easy to eliminate φ' by solving Eq. (4.37) for φ', the result being

$$\varphi' = \frac{\sqrt{1 + r'^2}}{r^2\mu(z, r)}\{N - Q\Psi(z, r)/2\pi\} \tag{4.38}$$

with

$$\mu(z, r) := \left\{g^2(z, r) - r^{-2}(N - Q\Psi/2\pi)^2\right\}^{1/2} \tag{4.39}$$

Elementary calculations show that

$$\rho = \sqrt{1 + r'^2 + r^2\varphi'^2} = g(z, r)\mu^{-1}\sqrt{1 + r'^2}$$

The integrand of Eq. (4.34) can be simplified by means of the Legendre transform $M^* = \overline{M} - \varphi' N$ (cf. 4.20), the result being

$$M^*(z, r, r') = \sqrt{1 + r'^2}\mu(z, r)$$

Evaluation of the Euler equation

$$\frac{d}{dz}\left(\frac{\partial M^*}{\partial r'}\right) = \frac{\partial M^*}{\partial r}$$

yields

$$\frac{d}{dz}\left(\frac{r'\mu}{\sqrt{1 + r'^2}}\right) = \sqrt{1 + r'^2}\frac{\partial \mu}{\partial r}$$

which can be further simplified to the Störmer equation

$$r''(z) = \frac{1 + r'^2}{\mu(z, r)} \left(\frac{\partial \mu}{\partial r} - r' \frac{\partial \mu}{\partial z} \right) \tag{4.40}$$

This trajectory equation is similar in structure to Eq. (3.21) and to the electric terms of Eq. (3.20). The function $\mu(z, r)$ is the length of the meridional projection of $\boldsymbol{g}$, since Eq. (4.39) can be rewritten as $\mu = |\boldsymbol{i}_z g_z + \boldsymbol{i}_r g_r|$. When considering the motion of electrons, it is helpful to introduce the acceleration potential by means of Eq. (2.19). We find

$$r'' = \frac{1 + r'^2}{2\tilde{\Phi}(z, r)} \left(\frac{\partial \tilde{\Phi}}{\partial r} - r' \frac{\partial \tilde{\Phi}}{\partial z} \right) \tag{4.41}$$

with the effective potential

$$\tilde{\Phi}(z, r) = \hat{\Phi}(z, r) - \frac{\left\{ N + e\Psi(z, r)/2\pi \right\}^2}{2m_0 e r^2} \tag{4.42}$$

the second term being again a centrifugal contribution.

On comparing these equations with the corresponding formulae in Section 4.2, the reader will notice that no approximations have been necessary here and the exact calculation has not become at all complicated. The essential simplification is a consequence of the fact that, in the time-independent representation, the relativistic mass is a simple function $m(z, r)$; the need to use Eq. (4.17) in the time-dependent situation highly complicates the calculations.

CHAPTER 5

Hamiltonian Optics

In Chapter 4, Variational Principles, we investigated the various types of variational principles and showed how they can be made to yield suitable forms of trajectory equations. These trajectory equations are to be solved by methods that are explained in later chapters and the solutions will provide us with a certain measure of physical understanding. Though this is a possible way of investigating electron optical devices, it is not entirely satisfactory. In geometric light optics, Hamilton's theory of characteristic functions, in which the rays of light are treated as trajectories orthogonal to eikonal or characteristic functions, was a major advance, making it possible to investigate whole bundles of trajectories instead of individual ones. This is the main difference between optical and purely mechanical or ballistic treatments.

In charged particle optics, the analogue of the eikonal theory is well-known under the name of Hamilton–Jacobi theory, which we shall now consider in detail. In the standard textbooks on classical mechanics, this theory is derived by means of canonical transformations, which is a very general but undeniably elaborate method (see e.g., Goldstein, 1959, Chapter 8; Goldstein et al. 2011). We shall not follow their example. The presentation that follows is considerably simpler. The scholarly text of Radlička (2008) compares and contrasts the Hamiltonian, Lie and eikonal theories together with discussion of differential algebra.

5.1 Introduction of the Characteristic Function

We again set out from Eq. (4.1). For given boundary values, t_0 and t_1, of t and hence of the vectors $\boldsymbol{r}_0 \coloneqq \boldsymbol{r}(t_0)$ and $\boldsymbol{r}_1 \coloneqq \boldsymbol{r}(t_1)$, a function $W(\boldsymbol{r}_0, t_0; \boldsymbol{r}_1, t_1)$ can be defined to be the stationary value of Eq. (4.1), obtained by integration over a *physical* trajectory, a solution of the corresponding Euler equations. This definition may be complicated by the presence of singularities and ambiguities, which will be investigated later. For the moment, we shall assume that the value of the integral expression is a unique and differentiable function of its arguments. We shall see that it is appropriate to regard this function W as a characteristic of the system and we shall indeed refer to it as a *characteristic function.*

We now consider differential variations of the endpoint, the starting point being unaltered. In order to distinguish these variations from the notation used for integration,

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00005-5
© 2018 Elsevier Ltd. All rights reserved.

we denote them by Δt_1 and $\Delta \boldsymbol{r}_1$ rather than dt and $d\boldsymbol{r}$. Since the curves connecting the fixed starting point and shifted endpoint must correspond to physical trajectories, the condition

$$\Delta \boldsymbol{r}_1 = \boldsymbol{v}(t_1)\Delta t_1 = \boldsymbol{v}_1 \Delta t_1$$

must be satisfied. The corresponding variation of W is then given by $\Delta W = L_1 \Delta t_1 = L(\boldsymbol{r}_1, \boldsymbol{v}_1, t_1)\Delta t_1$. The Lagrangian L_1 can be expressed in terms of the Hamiltonian H_1 at the terminal point 1, the result being

$$\Delta W = (\boldsymbol{p}_1 \cdot \boldsymbol{v}_1 - H_1)\Delta t_1 = \boldsymbol{p}_1 \cdot \Delta \boldsymbol{r}_1 - H_1 \Delta t_1 \tag{5.1}$$

On the other hand we can expand the difference

$$\Delta W \coloneqq W\,(\boldsymbol{r}_0, t_0;\ \ \boldsymbol{r}_1 + \Delta \boldsymbol{r}_1, t_1 + \Delta t_1) - W\,(\boldsymbol{r}_0, t_0;\ \ \boldsymbol{r}_1, t_1)$$

as a Taylor series. Retaining only the first order terms, we obtain

$$\Delta W = \Delta \boldsymbol{r}_1 \cdot \frac{\partial W}{\partial \boldsymbol{r}_1} + \Delta t_1 \frac{\partial W}{dt_1} \tag{5.2}$$

The two expressions for ΔW must be identical for all increments Δt_1 and $\Delta \boldsymbol{r}_1$ that represent a physical motion. Certainly we can choose arbitrary values of Δt_1 provided that a continuous range of values leads from t_1 to $t_1 + \Delta t_1$. It is now of importance that Eq. (5.1) must be identical with Eq. (5.2) in every respect, that is, for *all* acceptable configurations of $\boldsymbol{r}_0$, t_0, $\boldsymbol{r}_1$ and t_1. This implies that $\Delta \boldsymbol{r}_1$ must be regarded as *independent* of Δt_1, even though $\boldsymbol{r}_1 = \boldsymbol{v}_1 \Delta t_1$; the velocity $\boldsymbol{v}_1$ may be an arbitrary vector only subject to $|\boldsymbol{v}_1| < c$. From Eqs (5.1) and (5.2), we then obtain the necessary and sufficient conditions

$$\boldsymbol{p}_1 = \frac{\partial W}{\partial \boldsymbol{r}_1}, \quad H_1 = -\frac{\partial W}{\partial t_1} \tag{5.3}$$

By considering variations of the starting point, we likewise find

$$\boldsymbol{p}_0 = -\frac{\partial W}{\partial \boldsymbol{r}_0}, \quad H_0 = \frac{\partial W}{\partial t_0} \tag{5.4}$$

Thus the variation ΔW due to alterations of both sets of arguments is given by Hamilton's central equation:

$$\Delta W = \Big[\boldsymbol{p} \cdot \Delta \boldsymbol{r} - H \Delta t\Big]_0^1 \equiv \boldsymbol{p}_1 \cdot \Delta \boldsymbol{r}_1 - \boldsymbol{p}_0 \cdot \Delta \boldsymbol{r}_0 - H_1 \Delta t_1 + H_0 \Delta t_0 \tag{5.5}$$

Since W is a continuously differentiable function of its arguments, Eq. (5.5) must hold even for completely arbitrary differential increments.

5.2 The Hamilton–Jacobi Equation

In the following discussion, we shall assume that the starting coordinates $\boldsymbol{r}_0$, t_0 are uniquely specified and remain unaltered, while the terminal coordinates $\boldsymbol{r}_1$, t_1 may vary within physically allowed domains. In order to simplify the notation, we shall omit the arguments $\boldsymbol{r}_0$ and t_0 whenever they are not explicitly needed. Furthermore, we shall omit the subscript 1 of the terminal coordinates, and the ordinary vector notation for the gradient will refer to differentiation with respect to $\boldsymbol{r}_1$.

Introducing Eq. (5.3) into the Hamiltonian $H(\boldsymbol{r}, \boldsymbol{p}, t)$, we immediately obtain the Hamilton–Jacobi equation

$$H(\boldsymbol{r},\ \text{grad}\ W,\ t) = -\frac{\partial W}{\partial t} \tag{5.6}$$

W now being a function of $\boldsymbol{r}$ and t. Recalling Eq. (4.28), (5.6) is given explicitly by

$$c\left\{(m_0 c)^2 + (\text{grad}\ W - Q\boldsymbol{A})^2\right\}^{1/2} - m_0 c^2 + Q\Phi + \frac{\partial W}{\partial t} = 0 \tag{5.7}$$

Φ and $\boldsymbol{A}$ being functions of $\boldsymbol{r}$ and t, like W. This is a partial differential equation of first order for the function $W(\boldsymbol{r}, t)$. Since only the derivatives of W appear in Eq. (5.7) and not the function itself, the solution of Eq. (5.7) may contain an arbitrary additive constant. A reasonable and simple normalization is $W(\boldsymbol{r}_0, t_0; \boldsymbol{r}_0, t_0) = 0$, since the solution then fits the original definition of W, namely as the integral representation of the action. From among the many solutions of (5.7), we shall select all those that satisfy this condition. They correspond to the paths of all trajectories that start at the point $\boldsymbol{r}_0$ at the time t_0.

Even after imposing this restriction on the set of solutions, the practical solution of Eq. (5.7) will be extremely complicated in the general case; further simplification is therefore necessary. In all systems with static potentials $\Phi(\boldsymbol{r})$, $\boldsymbol{A}(\boldsymbol{r})$, the separation (in full notation)

$$W(\boldsymbol{r}_0, t_0, \boldsymbol{r}, t) = \overline{S}(\boldsymbol{r}_0, \boldsymbol{r}) - E_0(t - t_0) \tag{5.8}$$

is possible and advantageous, Eq. (5.7) then simplifying to

$$c\left\{(m_0 c)^2 + (\text{grad}\ \overline{S} - Q\boldsymbol{A}(r))^2\right\}^{1/2} - m_0 c^2 + Q\Phi(\boldsymbol{r}) = E_0 \tag{5.9}$$

The *point characteristic function* $\overline{S}(\boldsymbol{r}_0, \boldsymbol{r})$ introduced by this separation is exactly the same as the function $\overline{S}$ appearing in the integral representations Eqs (4.32) and (4.33) if the normalization condition for W is satisfied. Thus, instead of solving Eq. (5.9) directly with $\overline{S}(\boldsymbol{r}_0, \boldsymbol{r}_0) = 0$, $\overline{S}$ may in practice be obtained by evaluating Eq. (4.33).

Familiar though it is in theoretical physics, Eq. (5.9) is an inconvenient form of the reduced Hamilton–Jacobi equation. By means of Eq. (2.13) we can simplify it to

$$\left\{\text{grad}\,\overline{S}(\boldsymbol{r}) - Q\boldsymbol{A}(\boldsymbol{r})\right\}^2 = g^2(\boldsymbol{r}) \tag{5.10}$$

which is still exact. In the case of electron propagation, we can rewrite it as

$$\left\{\text{grad}\,\overline{S}(\boldsymbol{r}) + e\boldsymbol{A}(\boldsymbol{r})\right\}^2 = 2m_0 e\hat{\Phi}(\boldsymbol{r}) \tag{5.11}$$

In all practical calculations we shall use this form. Despite this simplification, the practical solution may still be very arduous.

5.3 The Analogy With Light Optics

Having formally introduced the characteristic function, we now discuss its physical meaning. For simplicity, we first assume that the vector potential $\boldsymbol{A}(\boldsymbol{r})$ vanishes. In the static case, the first equation in Eq. (5.3) then reduces to

$$\boldsymbol{p}(\boldsymbol{r}) \equiv \boldsymbol{g}(\boldsymbol{r}) = \text{grad}\,\overline{S}(\boldsymbol{r}) \tag{5.12}$$

and Eq. (5.11) to

$$\left\{\text{grad}\,\overline{S}(\boldsymbol{r})\right\}^2 = \left\{\boldsymbol{g}(\boldsymbol{r})\right\}^2 = \boldsymbol{g}^2(\boldsymbol{r}) = 2m_0 e\hat{\Phi}(\boldsymbol{r}) \tag{5.13}$$

the last term of this equation being valid for electrons. The physical meaning of Eqs (5.12) and (5.13) is quite clear: Eq. (5.12) expresses the vector $\boldsymbol{g} = m\boldsymbol{v}$ as a function of $\boldsymbol{r}$, while Eq. (5.13) is the condition that the length of this vector is in agreement with relativistic kinematics. The truly new result, going beyond relativistic kinematics, is that the local direction of the vector $\boldsymbol{g}$ is always orthogonal to the corresponding surface $\overline{S}(\boldsymbol{r}) = \text{const}$ (5.12). Since the vector $\boldsymbol{g} = g(\boldsymbol{r})\boldsymbol{t}(\boldsymbol{r})$ always points in the same direction as the local tangent $\boldsymbol{t}(\boldsymbol{r})$ of the trajectory in question, we can draw the following conclusion: the particle trajectories are orthogonal to the set of surfaces $\overline{S}(\boldsymbol{r}) = \text{const}$.

This statement is illustrated in Fig. 5.1. The analogy with geometric light optics is now complete. We may use the relation $\boldsymbol{g} = \hbar k_0 n(\boldsymbol{r})\boldsymbol{t}(\boldsymbol{r})$, $\boldsymbol{t}(\boldsymbol{r})$ being the local tangent vector of a ray passing through the point $\boldsymbol{r}$. The vacuum momentum $\hbar k_0$ cancels out if we introduce an *eikonal function* $L(\boldsymbol{r})$ by writing $\overline{S}(\boldsymbol{r}) =: \hbar k_0 L(\boldsymbol{r})$, and we obtain

$$n(\boldsymbol{r})t(\boldsymbol{r}) = \text{grad}\,L(\boldsymbol{r}) \tag{5.14}$$

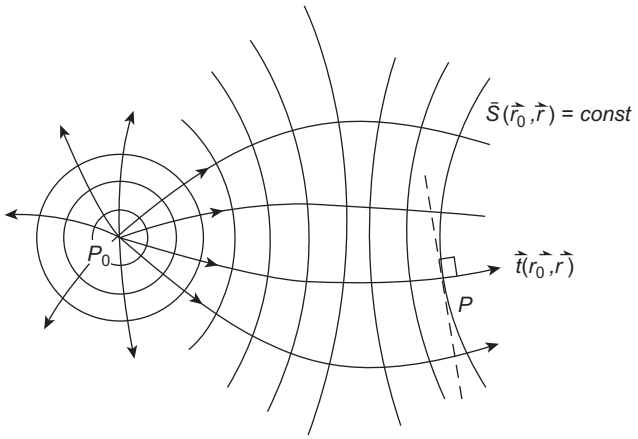


Figure 5.1
Simplified representation of the point characteristic function $\bar{S}(\boldsymbol{r}_0, \boldsymbol{r})$ in an isotropic medium. The point P_0, with position vector $\boldsymbol{r}_0$, may be regarded as a point source, from which the various rays emanate. A number of surfaces $\bar{S} = \text{const}$ are shown in a two-dimensional section. The value of $\bar{S}$ at any arbitrary point P is equal to the variational integral along the trajectory from P_0 to P. The local tangent $\boldsymbol{t}$ at P is orthogonal to the corresponding surface $\bar{S} = \text{const}$.

$$\left\{\text{grad}\, L(\boldsymbol{r})\right\}^2 = n^2(\boldsymbol{r}) \tag{5.15}$$

Eq. (5.14) expresses Hamilton's statement that light rays are the normals to a family of surfaces $L(\boldsymbol{r}) = \text{const}$ known as wavefronts or eikonal surfaces (see e.g., Born and Wolf (1959) Eq. (3.1.15b), where Eq. (5.15) is referred to as the eikonal equation, recalling the work of Bruns (1895) on 'Das Eikonal'). The eikonal itself is an optical length. Explicitly, the point eikonal

$$L(r_0, r_1) = \int_{P_0}^{P_1} n(\boldsymbol{r})ds \tag{5.16}$$

is the optical length between the points P_0 and P_1 with position vectors $\boldsymbol{r}_0$ and $\boldsymbol{r}_1$, respectively. The integration is to be performed along a physical trajectory. According to Fermat's principle, this function takes the same value along all trajectories connecting P_0 and P_1 that are continuously deformable into one another, and is stationary when evaluated along a trajectory. For rotationally symmetric dioptric systems, this stationary value is a true minimum (Sturrock, 1955, p. 60). This result is exactly analogous to Eq. (4.32).

It is possible and of potential interest to introduce an electron optical index of refraction in such a way that Eqs (5.12) and (5.13) are in formal agreement with (5.14) and (5.15), respectively. Since the index of refraction may be defined in light optics as the ratio of the

momentum in the medium in question and *in vacuo*, we may similarly define the electron optical index of refraction to be

$$n_E(\boldsymbol{r}) = \frac{g(\boldsymbol{r})}{G} = \sqrt{\frac{\hat{\Phi}(\boldsymbol{r})}{\hat{U}}} \tag{5.17}$$

$G = (2m_0 e\hat{U})^{1/2}$ being some suitable constant momentum. We now have complete formal agreement if we write $\overline{S} = GL$ instead of $\overline{S} = \hbar k_0 L$. There is, however, an important difference between light optics and electron optics. With respect to the propagation of light, the vacuum is the privileged medium of reference as it is the only medium free of dispersion and absorption. In electron optics there is no privileged medium of reference, the choice of the positive constant G being arbitrary; different choices are to be found in the literature. A definition of an electron optical index of refraction such as Eq. (5.17) thus offers no particular advantage and it seems more sensible to use Eqs (5.12) and (5.13) in their original form, since the kinematic momentum has a direct experimental significance.

A more serious difference between light optics and electron optics is that, in the nonrelativistic approximation, the condition $\nabla^2(n^2_E) = 0$ is satisfied in all source-free domains, whereas no such condition obtains in light optics. Although the index $n(\boldsymbol{r})$ for the refraction of light is a piecewise constant function in all the principal optical devices, the free choice of the lens surfaces makes aberration correction possible. The fact that Laplace's equation must be satisfied is, however, such a strong restriction that some of the aberrations in electron optical devices are rigorously incapable of correction (see Part IV). The relativistic terms in Eq. (5.17) do not alter this situation.

5.4 The Influence of Vector Potentials

The analogy between light optics and electron optics breaks down completely when vector potentials have to be considered. We then have

$$\boldsymbol{p}(\boldsymbol{r}) = \boldsymbol{g}(\boldsymbol{r}) + Q\boldsymbol{A}(\boldsymbol{r}) = \operatorname{grad} \overline{S}(\boldsymbol{r}) \tag{5.18}$$

Since the vectors $\boldsymbol{g}$ and $\boldsymbol{A}$ are in general *not* parallel to each other and $\boldsymbol{g}$ has the direction of the local tangent, $\boldsymbol{p}$ does not always point in this tangential direction. The rays of particles are hence *no longer* orthogonal trajectories of surfaces $\overline{S}(\boldsymbol{r}) = \text{const}$. This is shown in Fig. 5.2.

The principle of least action Eq. (4.33) can be rewritten as $\delta\overline{S} = 0$ or

$$\overline{S}(\boldsymbol{r}_0, \boldsymbol{r}_1) = \int_{r_0}^{r_1} \boldsymbol{p} \cdot d\boldsymbol{r} = \int_{P_0}^{P_1} \mu(\boldsymbol{r}, \boldsymbol{t}) ds = \text{extr.} \tag{5.19}$$

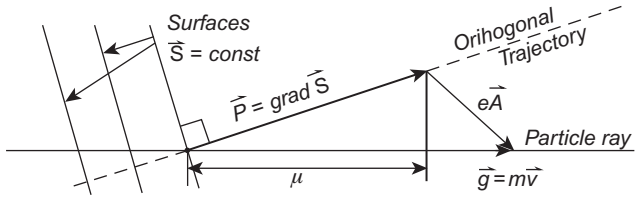


Figure 5.2
Influence of the vector potential $\boldsymbol{A}$ on the characteristic function $\overline{S}$.

with

$$\mu(\boldsymbol{r},\boldsymbol{t}) \coloneqq \boldsymbol{t}\cdot\boldsymbol{p} = g(\boldsymbol{r}) + Q\boldsymbol{t}\cdot\boldsymbol{A}(\boldsymbol{r}) \tag{5.20}$$

and for electrons:

$$\mu(\boldsymbol{r},\boldsymbol{t}) = \sqrt{2m_0 e\hat{\Phi}(\boldsymbol{r})} - e\boldsymbol{t}\cdot\boldsymbol{A}(\boldsymbol{r}) \tag{5.21}$$

The meaning of this quantity is shown in Fig. 5.2. The generalization of Eq. (5.17) is now

$$n_E(\boldsymbol{r},\boldsymbol{t}) \coloneqq \sqrt{\frac{\hat{\Phi}(\boldsymbol{r})}{\hat{U}}} - \frac{\eta}{\hat{U}^{1/2}}\boldsymbol{t}\cdot\boldsymbol{A}(\boldsymbol{r}) \tag{5.22}$$

Expressions of this form or proportional to it are commonly encountered in the literature. The square-root term is regarded as an isotropic contribution to the index of refraction, while the vector potential term—owing to its dependence on $\boldsymbol{t}$—is an anisotropic contribution. It must be emphasized that all these considerations are of a purely formal character and have no experimental significance, as will soon be obvious. Nevertheless, the point characteristic function $\overline{S}(\boldsymbol{r}_0,\boldsymbol{r}_1)$, defined by Eq. (5.19), does retain—disregarding any constant normalization factor—the character of the optical distance between the points $\boldsymbol{r}_0$ and $\boldsymbol{r}_1$. Various choices of the normalization factor (essentially $\hat{U}$ in Eq. 5.22) are to be found in the literature (see Picht, 1939, 1957, Eqs (3.10−11) or Picht, 1963, Section 3.8; Glaser, 1952 Sections 9−10; Grivet, 1965, Chapter 6; Kel'man and Yavor, 1959, 1968, Section 4 of Chapter 1). A system of units—not adopted here (but see Section 24.8)—has been devised by Sturrock (1955, Section 1.2) to eliminate e, m_0 and c from the equations. We emphasize that the choice of any particular normalization is only a question of convenience and has no physical meaning.

It is possible and obviously sensible to choose a gauge for $\boldsymbol{A}(\boldsymbol{r})$ such that $\boldsymbol{A}(\boldsymbol{r})$ vanishes in all domains in which $\boldsymbol{B}$ is zero. Thus at least in the field-free domains in front of and behind magnetic devices, we can make use of the orthogonality between trajectories and eikonal surfaces. The Aharonov−Bohm effect, for which this gauge breaks down, is described in Sections 59.6 and especially 62.4 of Volume 3.

5.5 Gauge Transformations

In the previous chapters we have frequently used electromagnetic potentials but, apart from the purely electrostatic potential, never specified their gauge. We now discuss the influence of different gauges on various physical quantities.

Eqs (4.3) and (4.4) can be satisfied by different pairs of potentials Φ, $\boldsymbol{A}$ and Φ', $\boldsymbol{A}'$, say, provided that these pairs are related by

$$\Phi = \Phi' - \frac{\partial}{\partial t} F(\boldsymbol{r}, t), \quad \boldsymbol{A} = \boldsymbol{A}' + \text{grad } F(\boldsymbol{r}, t) \tag{5.23}$$

This is called a gauge transformation. The function $F(\boldsymbol{r}, t)$ is arbitrary so long as it is sufficiently differentiable. The other Maxwell equations impose further restrictions but do not completely eliminate the freedom of choice of F. Since only the field vectors have experimental significance and not the potentials, all quantities that depend in any way on $F(\boldsymbol{r}, t)$ have *no* experimental significance.

The kinematic functions, introduced in Chapter 2, Relativistic Kinematics, are gauge-invariant since they are essentially related to the kinetic energy and not to the potential. In the variational formalism, the gauge-dependent quantities are the generalized potential V^*, the canonical momentum $\boldsymbol{p}$, the Lagrangian L and the Hamiltonian H, the corresponding transforms being given by

$$V^* = V^{*\prime} - Q\left(\frac{\partial F}{\partial t} + \boldsymbol{v} \cdot \text{grad } F\right) \equiv V^{*\prime} - Q\frac{dF}{dt} \tag{5.24}$$

$$L = L' + Q\frac{dF}{dt} \tag{5.25}$$

$$\boldsymbol{p} = \boldsymbol{p}' + Q \text{ grad } F \tag{5.26}$$

$$H = H' - Q\frac{\partial F}{\partial t} \tag{5.27}$$

The transform of the variational integral Eq. (4.1) involves integration of a total derivative, the result being

$$W = W' + \Big[QF(\boldsymbol{r}, t)\Big]_0^1 \tag{5.28}$$

If the endpoints are kept fixed, W and W' differ only by a constant and thus have the same extremals as solutions of the corresponding Euler–Lagrange equations. If, however, the endpoints are regarded as variables, Eq. (5.28) shows that the action W is essentially gauge-dependent. The characteristic functions are therefore gauge-dependent and are not observable quantities. Only for the propagation of particles in

purely electrostatic fields can Eqs (5.12) and (5.13) be regarded as having a physical meaning, as they have a gauge-invariant form. In Eq. (5.20), however, the gauge-dependence is now obvious and so definition Eq. (5.22) has no particular advantage.

5.6 Poincaré's Integral Invariant

We now return to Eq. (5.5). In time-independent systems, variations of the time are of no interest and we thus choose $\Delta t_1 = \Delta t_0 = 0$. Since spatial variations involve only the static point characteristic function $\overline{S}(\boldsymbol{r}_0, \boldsymbol{r}_1)$, Eq. (5.5) then simplifies to

$$\Delta \overline{S} = \boldsymbol{p}_1 \cdot \Delta \boldsymbol{r}_1 - \boldsymbol{p}_0 \cdot \Delta \boldsymbol{r}_0 \tag{5.29}$$

We now consider a one-parameter family of nonintersecting rays, each ray being uniquely characterized by a well-defined value of some parameter u, as shown in Fig. 5.3. This means that the points with vectors $\boldsymbol{r}_0(u)$ and $\boldsymbol{r}_1(u)$ are located on the same ray. It is useful to introduce derivatives with respect to u, for instance $\Delta \boldsymbol{r}_0 = \Delta u \cdot d\boldsymbol{r}_0/du$, with similar expressions for $\Delta \boldsymbol{r}_1$ and $\Delta \overline{S}$. The quantity Δu then cancels out from Eq. (5.29), so that

$$\frac{d\overline{S}}{du} = \boldsymbol{p}_1(u) \cdot \frac{d\boldsymbol{r}_1}{du} - \boldsymbol{p}_0(u) \cdot \frac{d\boldsymbol{r}_0}{du} \tag{5.30}$$

is exactly valid.

In the next step we consider a tube of nonintersecting rays, its mantle surface now being a one-parameter family, as shown in Fig. 5.4. On this surface we choose two closed loops Γ_0 and Γ_1 with parametric representations $\boldsymbol{r}_0(u)$ and $\boldsymbol{r}_1(u)$ for $u_\alpha \leq u \leq u_\beta$, respectively. Integration of (5.30) over the whole interval of u results in

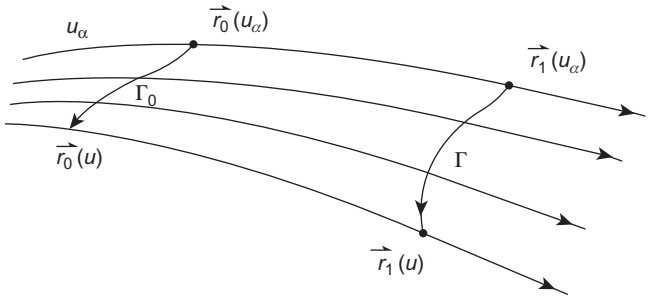


Figure 5.3
A one-parameter family of rays, each ray being uniquely characterized by a specific value of the parameter u. Two paths of integration, Γ_0 and Γ, are also shown.

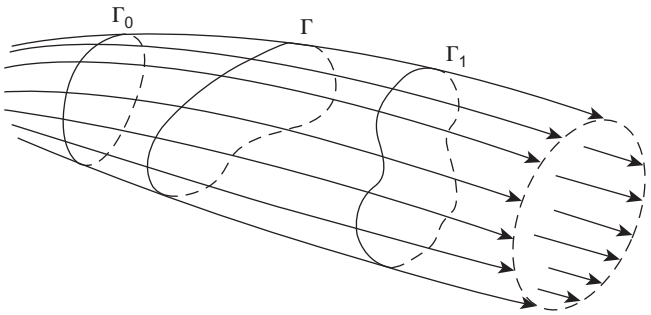


Figure 5.4
Tube of nonintersecting rays and three closed paths of integration, Γ_0, Γ_1 and Γ round its surface.

$$\int_{u_\alpha}^{u_\beta} \frac{d\overline{S}}{du} du = \int_{u_\alpha}^{u_\beta} \boldsymbol{p}_1(u) \cdot \frac{d\boldsymbol{r}_1}{du} du - \int_{u_\alpha}^{u_\beta} \boldsymbol{p}_0(u) \cdot \frac{d\boldsymbol{r}_0}{du} du$$

Since $\boldsymbol{r}_j(u_\beta) = \boldsymbol{r}_j(u_\alpha)$, $(j = 0, 1)$, and $\overline{S}$ is a unique function of its arguments, the expression on the left-hand side vanishes. Thus

$$I \coloneqq \int_{u_\alpha}^{u_\beta} \boldsymbol{p}_1(u) \cdot \frac{d\boldsymbol{r}_1}{du} du = \int_{u_\alpha}^{u_\beta} \boldsymbol{p}_0(u) \cdot \frac{d\boldsymbol{r}_0}{du} du$$

is invariant. The parametric representation facilitates the evaluation of these integrals but is not absolutely necessary; the value of I is invariant with respect to parametric transforms. Since Γ_0 and Γ_1 are arbitrary loops, the expression

$$I = \oint_\Gamma \boldsymbol{p} \cdot d\boldsymbol{r} \tag{5.31}$$

has the same value for any closed loop Γ on the surface of the tube. This is *Poincaré's integral invariance theorem.* This quantity is even invariant with respect to gauge transformations, since it can be rewritten as

$$I = \oint_\Gamma \boldsymbol{g} \cdot d\boldsymbol{r} + Q\Psi_\Gamma \tag{5.32}$$

where Ψ_Γ is the magnetic flux through Γ:

$$\Psi_\Gamma = \oint_\Gamma \boldsymbol{A} \cdot dr = \int_{(\Gamma)} \boldsymbol{B} \cdot d\boldsymbol{a}$$

The flux term in Eq. (5.32) gives rise to a phase shift of magnitude $Q\Psi_\Gamma/\hbar$ in the wave-optical interference patterns produced by electron optical biprism devices. This phase shift,

known as the Aharonov–Bohm effect (Ehrenberg and Siday, 1949; Aharonov and Bohm, 1959), is discussed in detail in Section 62.4 of Volume 3.

Another interesting consequence of Eq. (5.31), is the existence of the point characteristic function. The invariance of the integral expression given by Eq. (5.31) can be derived by following an alternative route, using canonical transforms (Goldstein, 1959), and may thus be regarded as fundamental. Let us now assume that the whole bundle of rays shown in Fig. 5.4 and also all rays propagating laminarly in the interior of the tube intersect at some point $\boldsymbol{r}_0$. Then any closed loop Γ may be contracted to this point $\boldsymbol{r}_0$, and hence the integral I vanishes. From $\oint \boldsymbol{p}\cdot d\boldsymbol{r} \equiv 0$ (even in the interior), we can deduce that curl $\boldsymbol{p} \equiv 0$. Hence there must be a function $U(\boldsymbol{r}) \neq 0$ such that $\boldsymbol{p} = \text{grad}\ U$. With the reasonable assumption $U(\boldsymbol{r}_0) = 0$ we recover the point characteristic function

$$U(\boldsymbol{r}_1) =: \overline{S}(\boldsymbol{r}_0, \boldsymbol{r}_1) = \int_{r_0}^{r_1} \boldsymbol{p}(\boldsymbol{r})\cdot d\boldsymbol{r}$$

Since the value of this integral is independent of the path of integration between $\boldsymbol{r}_0$ and $\boldsymbol{r}_1$, it must be identical with that of Eq. (5.19).

It is often preferable to use the invariance theorem in its differential form. This can be obtained easily by considering *congruences* of rays. These are two-parameter manifolds or families of rays, represented by functions $\boldsymbol{r}(u, v; s)$, u and v being the parameters in question and s the arc-length. For instance, all monoenergetic rays emerging from a 'point source' at $\boldsymbol{r}_0$ form a congruence, the parameters u and v then being angles characterizing the starting direction. The definition is, however, more general. It is easily seen that the generalization of Eq. (5.30) for a congruence is given by

$$\begin{aligned} \frac{\partial \overline{S}}{\partial u} &= \boldsymbol{p}_1 \cdot \frac{\partial \boldsymbol{r}_1}{\partial u} - \boldsymbol{p}_0 \cdot \frac{\partial \boldsymbol{r}_0}{\partial u} \\ \frac{\partial \overline{S}}{\partial v} &= \boldsymbol{p}_1 \cdot \frac{\partial \boldsymbol{r}_1}{\partial v} - \boldsymbol{p}_0 \cdot \frac{\partial \boldsymbol{r}_0}{\partial v} \end{aligned} \tag{5.33}$$

since all vector quantities are now functions of u and v (neglecting the irrelevant dependence on s). From the condition that $\partial^2 \overline{S}/\partial u \partial v$, calculated in different ways, must be the same continuous function, we obtain

$$\frac{\partial \boldsymbol{p}_1}{\partial v} \cdot \frac{\partial \boldsymbol{r}_1}{du} - \frac{\partial \boldsymbol{p}_0}{\partial v} \cdot \frac{\partial \boldsymbol{r}_0}{\partial u} = \frac{\partial \boldsymbol{p}_1}{\partial u} \cdot \frac{\partial \boldsymbol{r}_1}{\partial v} - \frac{\partial \boldsymbol{p}_0}{\partial u} \cdot \frac{\partial \boldsymbol{r}_0}{\partial v}$$

or rearranging

$$\frac{\partial \boldsymbol{r}_1}{\partial u} \cdot \frac{\partial \boldsymbol{p}_1}{\partial v} - \frac{\partial \boldsymbol{p}_1}{\partial u} \cdot \frac{\partial \boldsymbol{r}_1}{\partial v} = \frac{\partial \boldsymbol{r}_0}{\partial u} \cdot \frac{\partial \boldsymbol{p}_0}{\partial v} - \frac{\partial \boldsymbol{p}_0}{\partial u} \cdot \frac{\partial \boldsymbol{r}_0}{\partial v}$$

Since the points $\boldsymbol{r}_0$ and $\boldsymbol{r}_1$ may be chosen arbitrarily so long as both are located on the same trajectory specified by the values of u and v, the expressions on each side do not depend on position but only on u and v, and are thus constant along each ray. This constant is the familiar Lagrange bracket

$$\{u, v\}: = \frac{\partial \boldsymbol{r}}{\partial u} \cdot \frac{\partial \boldsymbol{p}}{\partial v} - \frac{\partial \boldsymbol{p}}{\partial u} \cdot \frac{\partial \boldsymbol{r}}{\partial v} = \text{const} \tag{5.34}$$

A congruence is said to be *normal* if $\{u, v\} \equiv 0$. This is equivalent to $I = 0$, since the Poincaré invariant can be obtained from Eq. (5.34) by integration (Sturrock, 1955, Section 2.3). In every normal congruence, there is thus a family of surfaces $U(\boldsymbol{r}) = \text{const}$ such that $\boldsymbol{p} = \text{grad } U$, U usually being the point characteristic function $\overline{S}$. As is illustrated in Fig. 5.2, this does not automatically imply orthogonality with respect to the trajectories themselves.

5.7 The Problem of Uniqueness

In the preceding considerations we have tacitly assumed that the solutions of the Hamilton–Jacobi equation are unique and regular. This is very often not the case. A simple example is shown in Fig. 5.5. In a beam, there may exist an envelope surface, formed by a one-parameter family of rays. This surface, which is known as a caustic, usually has sharp edges, and its extension depends on the positions and the shapes of any apertures confining the beam. The caustic represents a singularity of the point characteristic function, since it separates the domain of no solution from that with two solutions, where the rays intersect. The example illustrated is highly simplified; in realistic electron optical devices, caustics may have a very complicated structure. We shall therefore not investigate them here in a general manner; instead we shall treat some concrete examples later, see Chapter 42 of

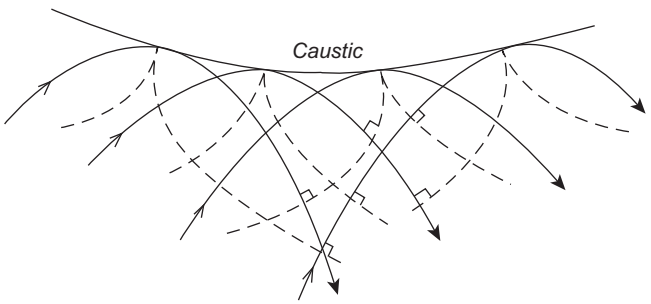


Figure 5.5

Particle trajectories (full lines) forming a caustic and wavefronts $\overline{S}(\boldsymbol{r}) = \text{const}$ (broken lines) in a two-dimensional section through a beam in an isotropic medium. The domain beyond the caustic is inaccessible to the particles; within the caustic, the trajectories may intersect. At the caustic, the lines $\overline{S} = \text{const}$ form cusps.

Volume 2. The foregoing theory remains valid in domains accessible to the beam, the vicinity of caustics being excluded.

In domains in which different functional branches of the point characteristic function overlap, each branch is to be treated separately.

5.8 Lie Algebra

A method of analysing the behaviour of charged particles in electric and magnetic fields based on Lie algebra has been found valuable in accelerator optics, where it was introduced by Dragt and further developed by him and Forest (Dragt, 1982, 1987, 1990; Dragt et al., 1986; Forest, 1998). They subsequently extended it to the domain of electron optics (Dragt and Forest, 1986; Dragt et al., 1986; Dragt, 1987), where Ximen (1995), Hu and Tang (1998, 1999), Hu et al. (1999) and Matsuya et al. (1995) have employed it. Here, we cannot give more than an introduction; a very readable account is to be found in the article of Dragt and Forest (1986). Moreover, another full presentation has been published by Radlička (2008), in which it is compared closely with other approaches, notably the trajectory and eikonal methods.

The basic element of the Lie algebraic approach is the Poisson bracket $[f, g]$ of two functions f and g (Poisson, 1809):

$$[f, g] = \frac{\partial f}{\partial \boldsymbol{p}} \frac{\partial g}{\partial \boldsymbol{q}} - \frac{\partial f}{\partial \boldsymbol{q}} \frac{\partial g}{\partial \boldsymbol{p}} \tag{5.35}$$

in which $\boldsymbol{q}$ here denotes the position vector $(x\ y)^{\mathrm{T}}$ and $\boldsymbol{p}$ is the canonical momentum Eq. (4.12). (We follow Dragt and Radlička in denoting the position vector by $\boldsymbol{q}$ for the benefit of readers who wish to go more deeply into their publications.) A linear operator, the *Lie operator*, generates such a Poisson bracket:

$$f \rightarrow :f: \quad \text{where} \quad :f:g = [f, g] \tag{5.36}$$

Powers of $:f:$ are defined in terms of imbricated Poisson brackets,

$$(:f:)^2 g = [f,\ [f, g]]$$

and $(:f:)^0$ is the identity operator, $(:f:)^0 g = g$. The exponential function $\exp(:f:)$ plays a central role and $\exp(:f:)g$ is known as the Lie transformation of g. We have

$$\exp(:f:) = \sum_0^\infty \frac{(:f:)^n}{n!} \tag{5.37}$$

and so

$$\exp(:f:)g = g + [f,g] + \frac{[f,[f,g]]}{2} + \cdots$$

The relation between $\boldsymbol{w} \coloneqq (\boldsymbol{q}, \boldsymbol{p})$ in two planes z_1 and z_2 is represented by a transfer map $\boldsymbol{M}$,

$$\boldsymbol{w}_2 = \boldsymbol{M}\boldsymbol{w}_1 \tag{5.38a}$$

and provided that $\boldsymbol{w} = 0$ corresponds to a trajectory (the optic axis), the map has the form

$$\boldsymbol{M} = \prod_{j=2}^{\infty} \exp(:f_j:) \tag{5.38b}$$

in which each function f_j is a homogeneous polynomial of degree j in $\boldsymbol{w}_1$. We shall find that $:f_2:$ corresponds to paraxial properties and the next terms to aberrations. We now specialize to monochromatic beams in a system with a straight optic axis. It is convenient to write

$$P_x = p_x/p_0, \quad P_y = p_y/p_0, \quad Q_x = x, \quad Q_y = y$$

in which p_0 is the momentum on the axis. For the Poisson brackets, we define

$$[Q_x, Q_y] = [P_x, P_y] = 0, \quad [Q_i, P_j] = \delta_{ij}$$

Dragt and Forest consider three simple forms of $\boldsymbol{M} = \exp(:f_2:)$ to illustrate the use of this algebra: $f_2 = -\frac{L}{2}(\boldsymbol{P}_1)^2$, $f_2 = -\frac{k}{2}(\boldsymbol{Q}_1)^2$ and $f_2 = \varphi(\boldsymbol{Q}_1 \times \boldsymbol{P}_1)\boldsymbol{i}_z = \varphi(Q_{x,1}P_{y,1} - Q_{y,1}P_{x,1})$. The first represents a translation and the second, the action of a thin lens. The third represents rotation through an angle φ as we now show. For the four Lie operations, we find

$$\begin{aligned} :f_2:Q_{x,1} &= \varphi Q_{y,1} \quad :f_2:Q_{y,1} = -\varphi Q_{x,1} \\ :f_2:P_{x,1} &= \varphi P_{y,1} \quad :f_2:P_{y,1} = -\varphi P_{x,1} \end{aligned} \tag{5.39}$$

We now evaluate the corresponding quantities in the plane z_2:

$$\begin{aligned} Q_{x,2} = \boldsymbol{M}Q_{x,1} &= \exp(:f_2:)Q_{x,1} = \{1 + :f_2: + (:f_2:)^2/2 + (:f_2:)^3/6 + \cdots\}Q_{x,1} \\ &= (1 - \varphi^2/2 + \cdots)Q_{x,1} + (\varphi - \varphi^3/6 + \cdots)Q_{y,1} \\ &= Q_{x,1}\cos\varphi + Q_{y,1}\sin\varphi \end{aligned}$$

and similarly

$$\begin{aligned} Q_{y,2} &= Q_{y,1}\cos\varphi - Q_{x,1}\sin\varphi \\ P_{x,2} &= P_{x,1}\cos\varphi + P_{y,1}\sin\varphi \\ P_{y,2} &= P_{y,1}\cos\varphi - P_{x,1}\sin\varphi \end{aligned}$$

This map therefore does indeed represent a rotation φ about the optic axis and the map $\boldsymbol{M} = \exp(:f_2:)$ contains enough information to characterize the paraxial optics of a system with a straight axis.

It is convenient to discuss the effect of aberrations here but some familiarity with Chapter 24, The Geometrical Aberrations of Round Lenses, is assumed. We limit the discussion to systems with rotational symmetry, which implies that $\boldsymbol{P}$ and $\boldsymbol{Q}$ occur only as the combinations $\boldsymbol{P}^2$, $\boldsymbol{Q}^2$, $\boldsymbol{P}\cdot\boldsymbol{Q}$ and $\boldsymbol{Q}\times\boldsymbol{P}$. Since this excludes homogeneous polynomials of odd order, $\boldsymbol{M}$ must take the form

$$\boldsymbol{M} = \exp(:f_2:)\exp(:f_4:)\exp(:f_6:)\cdots \tag{5.40}$$

For the third-order aberrations $\boldsymbol{w}^{(3)}$, we have

$$\boldsymbol{w}^{(3)} = \exp(:f_4:)\,\boldsymbol{w}^{(1)} = \boldsymbol{w}^{(1)} + \left[f_4, \boldsymbol{w}^{(1)}\right] + \frac{1}{2}\left[f_4,\ \left[f_4, \boldsymbol{w}^{(1)}\right]\right] + \cdots \tag{5.41}$$

and only the second term need be retained. The function f_4 will have the following form, in which we have anticipated the standard notation for the aberration coefficients introduced in Chapter 24:

$$\begin{aligned} &C_s(\boldsymbol{P}^2)^2 + K\boldsymbol{P}^2(\boldsymbol{P}\cdot\boldsymbol{Q}) + k\boldsymbol{P}^2\{(\boldsymbol{Q}\times\boldsymbol{P})\cdot\boldsymbol{i}_z\} + A(\boldsymbol{P}\cdot\boldsymbol{Q})^2 + a(\boldsymbol{P}\cdot\boldsymbol{Q})\{(\boldsymbol{Q}\times\boldsymbol{P})\cdot\boldsymbol{i}_z\} \\ &+ F\boldsymbol{P}^2\boldsymbol{Q}^2 + D\boldsymbol{Q}^2(\boldsymbol{P}\cdot\boldsymbol{Q}) + d\boldsymbol{Q}^2\{(\boldsymbol{Q}\times\boldsymbol{P})\cdot\boldsymbol{i}_z\} + \lambda(\boldsymbol{Q}^2)^2 \end{aligned} \tag{5.42}$$

Consider, for example, the term $C_s(\boldsymbol{P}^2)^2$. In the image plane, we have

$$Q_x^{(3)} = Q_x^{(1)} + [C_s(\boldsymbol{P}^2)^2,\ \ Q_x^{(1)}] = -C_s\frac{\partial(\boldsymbol{P}^2)^2}{\partial P_x} = -4C_s\boldsymbol{P}^2P_x \tag{5.43}$$

which clearly has the same nature as spherical aberration. The other terms can be identified with coma (*K, k*), astigmatism (*A, a*), field curvature (*F*) and distortion (*D, d*).

We shall not go further into this approach, which is thoroughly explored by Dragt and Forest; the design of aberration correctors is particularly interesting. We shall however, say a few words about the presentation of Radlička, which sheds a slightly different light on the steps leading to the aberration coefficients. The equation of motion for a Hamiltonian H is written

$$\boldsymbol{w}' = [\boldsymbol{w}, H]$$

with $\boldsymbol{w} = (\boldsymbol{q}\ \boldsymbol{p})^{\mathrm{T}}$. In the case of round magnetic lenses, the Hamiltonian can be written as

$$H = H_2 + H_4 + \ldots$$

where H_2, for example, has the easily recognizable form

$$H_2 = \frac{\boldsymbol{P}^2}{2\hat{\phi}^{1/2}} + \frac{\eta^2 B^2}{8\hat{\phi}}\boldsymbol{Q}^2 \tag{5.44}$$

in the rotating coordinate system, where $\boldsymbol{Q}$, $\boldsymbol{P}$ replace $\boldsymbol{q}$, $\boldsymbol{p}$. H_4 likewise has a familiar appearance:

$$\begin{aligned} H_4 &= \frac{1}{4}L_1(\boldsymbol{Q}^2)^2 + \frac{1}{2\hat{\phi}}L_2\boldsymbol{Q}^2\boldsymbol{P}^2 + \frac{1}{4\hat{\phi}^2}L_3(\boldsymbol{P}^2)^2 + \left(-\frac{1}{\hat{\phi}}RL_z + C_P\boldsymbol{Q}^2 + \frac{C_Q}{\hat{\phi}}\boldsymbol{P}^2\right)L_z \\ L_z^2 &= \boldsymbol{P}^2\boldsymbol{Q}^2 - (\boldsymbol{Q}\cdot\boldsymbol{P})^2 \end{aligned} \tag{5.45}$$

in which L_1, L_2, L_3 and R are defined in Eq. (24.3); the functions P and Q of Eq. (24.3) are here written C_P and C_Q to prevent confusion.

The aberrations are obtained with the aid of the *interaction Hamiltonian*, which here takes the form

$$H_4^{\text{int}} = H_4(\boldsymbol{Q}(\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{P}}), \boldsymbol{P}(\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{P}})) \tag{5.46}$$

in which $\tilde{\boldsymbol{Q}}$, $\tilde{\boldsymbol{P}}$ are obtained from

$$\begin{pmatrix} \boldsymbol{Q} \\ \\ \boldsymbol{P} \end{pmatrix} = \begin{pmatrix} g(z)\hat{1} & \hat{\phi}_o^{-1/2}h(z)\hat{1} \\ \\ \hat{\phi}^{1/2}g'(z)\hat{1} & \left(\dfrac{\hat{\phi}}{\hat{\phi}_o}\right)^{1/2}h'(z)\hat{1} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{Q}} \\ \\ \tilde{\boldsymbol{P}} \end{pmatrix}$$

where $\hat{1}$ denotes the identity matrix. From the interaction Hamiltonian we can calculate

$$g_4 = -\int_{z_o}^{z} H_4^{\text{int}}(\boldsymbol{Q}^{[2]}, \boldsymbol{P}^{[2]}, z)dz$$

The exponent [2] indicates the second perturbation beyond the paraxial approximation; in the case of axially symmetric systems, this brings us to the third-order aberrations, as there are no second-order aberrations, which would correspond to exponent [1]). In the image plane

$$\boldsymbol{Q}_{(3)} = M(\boldsymbol{Q}_0) - \frac{\partial g_4(\boldsymbol{Q}_0\boldsymbol{P}_0)}{\partial \boldsymbol{P}_0} \tag{5.47}$$

which leads to the expressions for the aberration coefficients given by the eikonal method. Other special cases are examined in great detail by Radlička, a valuable feature of his full account of a relatively little-known branch of aberration theory.

5.9 Summary

In this part we have dealt with the general theoretical fundamentals of electron optics. We have derived various forms of trajectory equations, kinematic functions and conservation laws. We have investigated various kinds of variational principles and these have permitted us to derive the general theory of Hamiltonian optics. All the relations obtained are essential tools in the investigation of particular aspects of electron optical devices. Since a knowledge of the applied electromagnetic fields is needed before the trajectory equations can be solved, we must now interrupt the purely electron optical discussions and deal with field calculations. In Part III, we shall return to the theory of electron propagation.

PART II

Calculation of Static Fields

CHAPTER 6

Basic Concepts and Equations

6.1 General Considerations

In this Part we shall deal with the calculation of electrostatic or magnetostatic fields, that is to say, time-independent fields. This major simplification is justified, since in the vast majority of practical electron optical devices the applied fields are static. Even in electric and magnetic deflection units, the frequencies of the time-dependent fields are so low that a quasistatic approximation is entirely justified. By this, we mean that all field functions can be separated in the form $F(\boldsymbol{r}, t) = f_1(\boldsymbol{r})f_2(t)$, where $f_1(\boldsymbol{r})$ is practically independent of the frequency and may be calculated as a static field. Special high-frequency devices for which these assumptions do not hold are not treated in this book.

Charged particles usually propagate *in vacuo* in a very narrow domain far distant from any material walls. The only exceptions are the immediate vicinity of emitting surfaces in electron guns, the surfaces of mirrors, specimens, apertures and recording devices. Specimens, apertures and recording media are of little interest in the present context, because they are usually located in field-free domains or are assumed to have no effect on the field distribution. The vicinities of cathodes and reflecting surfaces will be excluded from the following discussion. With these exceptions, the space through which particles travel will be referred to as the *extended paraxial domain.* Usually this is a narrow but long tube around the optic axis of the system in question, see Fig. 6.1. One of the aims of this Part is to derive suitable series expansions for the field in this paraxial domain, since these are fundamental for the investigation of trajectories, focusing properties and aberration effects in practical devices.

Although a concrete knowledge of the field in the appropriate extended paraxial domain would be quite sufficient for all further electron optical considerations (always excluding cathodes and mirrors), this knowledge cannot be obtained without making a complete calculation of the field within the whole device. The reason for this is that a static field within a given domain can only be calculated as the solution of a boundary-value problem, as is further outlined in Chapter 8, Boundary-Value Problems. In electron optics, the boundaries are the surfaces of electrodes or polepieces, which are of great importance though they are usually far from the paraxial domain. Thus the second major aim of this Part is to present techniques for solving boundary-value problems.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00006-7
© 2018 Elsevier Ltd. All rights reserved.

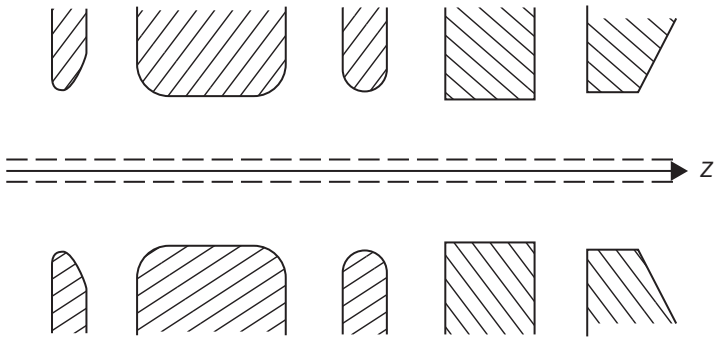


Figure 6.1
Axial section through part of some electron optical device; the paraxial region is indicated by dashed lines.

Throughout this Part we shall assume that electrodes and magnetic polepieces have isotropic material properties. The corresponding material coefficients may be functions of position. This is mainly the case when saturation effects arise in ferromagnetic polepieces. We shall adopt the following standard electrodynamic notation:

$\boldsymbol{E}$: electric field strength;	$\boldsymbol{D}$: displacement vector;
$\boldsymbol{H}$: magnetic field strength;	$\boldsymbol{B}$: magnetic flux density;
ε, ε_0: permittivity;	μ, μ_0: permeability;
$\nu = 1/\mu$: magnetic reluctance;	$\boldsymbol{A}$: vector potential;
ρ: space charge density;	$\boldsymbol{j}$: electric current density;
σ: surface charge density;	$\boldsymbol{\omega}$: surface current density.

Scalar potential functions are denoted in different ways, as they will appear frequently in different contexts; very often they have only a formal mathematical meaning.

6.2 Field Equations

In the case of stationary fields, Maxwell's equations reduce to

$$\begin{aligned} &\operatorname{curl} \boldsymbol{E} = 0, \quad \operatorname{curl} \boldsymbol{H} = \boldsymbol{j} \\ &\operatorname{div} \boldsymbol{D} = \rho, \quad \operatorname{div} \boldsymbol{B} = 0, \quad (\operatorname{div} \boldsymbol{j} = 0) \end{aligned} \tag{6.1}$$

These are to be complemented by the material equations

$$\boldsymbol{D} = \varepsilon \boldsymbol{E}, \quad \boldsymbol{B} = \mu \boldsymbol{H}, \quad (\text{or } \boldsymbol{H} = \nu \boldsymbol{B}) \tag{6.2}$$

In ferromagnetic materials, the reluctance ν is a function of $B = |\boldsymbol{B}|$, hence

$$\boldsymbol{H} = \nu(B)\boldsymbol{B} \tag{6.3}$$

In Eq. (6.1) the space charge density $\rho(\boldsymbol{r})$ and the current density $\boldsymbol{j}(\boldsymbol{r})$ are regarded as given functions of position. The determination of space charge distributions will be treated in Chapter 46 of Volume 2. Electric fields in the interior of conducting materials are not considered here. The source-free Maxwell equations permit us to introduce electromagnetic potentials,

$$\boldsymbol{E} = -\operatorname{grad} \Phi(\boldsymbol{r}), \quad \boldsymbol{B} = \operatorname{curl} \boldsymbol{A}(\boldsymbol{r}) \tag{6.4}$$

these equations being special cases of Eqs. (4.3) and (4.4). Combining Eq. (6.4) with (6.1) and (6.2), we obtain partial differential equations of Poisson's type. For the electrostatic potential we have

$$\operatorname{div} \left\{ \varepsilon(\boldsymbol{r}) \operatorname{grad} \Phi(\boldsymbol{r}) \right\} = -\rho(\boldsymbol{r}) \tag{6.5}$$

In homogeneous dielectric media, where ε is constant, Eq. (6.5) reduces to

$$\nabla^2 \Phi(\boldsymbol{r}) = -\rho(\boldsymbol{r})/\varepsilon \tag{6.6}$$

and in domains free of space charge, to Laplace's equation

$$\nabla^2 \Phi(\boldsymbol{r}) = 0 \tag{6.7}$$

The derivation of Poisson's equation for the vector potential proceeds as follows. In the general case, combination of $\nabla \times \boldsymbol{H} = \boldsymbol{j}$, $\boldsymbol{H} = \nu \boldsymbol{B}$ with $\nu = \nu(B)$ and $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ results in

$$\operatorname{curl} \left\{ \nu \left(\left| \operatorname{curl} \boldsymbol{A} \right| \right) \cdot \operatorname{curl} \boldsymbol{A}(\boldsymbol{r}) \right\} = \boldsymbol{j}(\boldsymbol{r}) \tag{6.8}$$

This complicated nonlinear vector Poisson equation has to be solved in ferromagnetic domains with saturation effects. When the permeability $\mu = \nu^{-1}$ is constant, this differential equation can be simplified considerably. Using the vector differential identity

$$\operatorname{curl} \operatorname{curl} \boldsymbol{A} = \operatorname{grad} \operatorname{div} \boldsymbol{A} - \nabla^2 \boldsymbol{A}$$

valid only in Cartesian representations, we obtain first

$$-\operatorname{grad} \operatorname{div} \boldsymbol{A} + \nabla^2 \boldsymbol{A} = -\mu \boldsymbol{j}(\boldsymbol{r})$$

A further important simplification can be achieved by choosing a gauge for $\boldsymbol{A}$ that satisfies

$$\operatorname{div} \boldsymbol{A}(\boldsymbol{r}) \equiv 0 \tag{6.9}$$

We then finally obtain the vector Poisson equation

$$\nabla^2 \boldsymbol{A}(\boldsymbol{r}) = -\mu \boldsymbol{j}(\boldsymbol{r}) \tag{6.10}$$

Yet another simplification is possible in current-free domains. We then have $\nabla \times \boldsymbol{H} = 0$ and it is always permissible to write

$$\boldsymbol{H}(\boldsymbol{r}) = -\operatorname{grad} \chi(\boldsymbol{r}) \tag{6.11}$$

$\chi(\boldsymbol{r})$ being the scalar magnetic potential. Since μ is constant, we have div $\boldsymbol{H} = 0$ and χ satisfies the Laplace equation

$$\nabla^2 \chi(\boldsymbol{r}) = 0 \tag{6.12}$$

The simplification thus achieved lies in the fact that only one scalar differential equation is to be solved instead of three coupled by Eq. (6.9). The representation (6.11), (6.12) is of course less general than that in terms of the vector potential, since the condition $\boldsymbol{j}(\boldsymbol{r}) = 0$ cannot be true throughout the whole space. The domain of solution of Eq. (6.12) is thus to be confined in such a way that $\boldsymbol{j} = 0$ and that the solution for χ remains unique. Important examples are given in Chapter 8, Boundary-Value Problems and Chapter 9, Integral Equations.

Although Eq. (6.11) is the correct form of the gradient representation of $\boldsymbol{H}$, it is rather inconvenient, since it is the flux density $\boldsymbol{B}$, and not $\boldsymbol{H}$, that figures in the trajectory equations. After solving the boundary-value problem for the magnetic field, we are only interested in the source-free vacuum field. We may therefore introduce a new potential $W(\boldsymbol{r})$, writing

$$\boldsymbol{B} = -\operatorname{grad} W, \quad \nabla^2 W = 0, \quad W = \mu_0 \chi \tag{6.12a}$$

We shall use this representation whenever this is possible and causes no confusion.

6.3 Variational Principles

In connection with the finite-element method, the subject of Chapter 12, it is of importance that the partial differential equations given above can also be derived from a variational principle, the integrand being the stored field energy. This principle takes different forms for scalar and vector potentials.

The general form of variational principle for m coupled functions of position $y_1(\boldsymbol{r}) \ldots y_m(\boldsymbol{r})$ in a three-dimensional domain S is given by

$$\delta W = \delta \int_S \Lambda(\boldsymbol{r}, y_1 \ldots y_m, \nabla y_1 \ldots \nabla y_m)\, d^3 r = 0 \tag{6.13}$$

The boundary ∂S and the boundary values $y_1 \ldots y_m$ on it must not vary. The corresponding Euler equations are

$$\frac{\partial \Lambda}{\partial y_i} = \sum_{k=1}^{3} \frac{\partial}{\partial x_k} \left[\frac{\partial \Lambda}{\partial(\partial y_i / \partial x_k)} \right] = \nabla \cdot \frac{\partial \Lambda}{\partial(\nabla y_i)} \tag{6.14}$$

The Cartesian components of the gradients are now treated as independent variables.

In applications involving electrostatic fields, we have $m = 1$, $y_1(\boldsymbol{r}) = \Phi(\boldsymbol{r})$ being the electrostatic potential. The corresponding Lagrange density Λ is given by

$$\Lambda = \frac{\varepsilon}{2} \boldsymbol{E}^2 - \rho\Phi, \quad (\boldsymbol{E} = -\nabla\Phi) \tag{6.15}$$

Applying Eq. (6.14), we obtain Eq. (6.5).

When source-free magnetic field domains are to be studied, we may choose scalar potential representations; as before $m = 1$ and now $y_1(\boldsymbol{r}) = \chi(\boldsymbol{r})$. The Lagrange density Λ is now the familiar energy density:

$$\Lambda = \frac{\mu}{2} \boldsymbol{H}^2, \quad (\boldsymbol{H} = -\nabla\chi) \tag{6.16}$$

The corresponding Euler equation is (6.12).

In the case of general magnetostatic fields, we have to choose $m = 3$, and $y_i(\boldsymbol{r}) = \boldsymbol{A}_i(\boldsymbol{r})$, $i = 1, 2, 3$, are then the three Cartesian components of the vector potential $\boldsymbol{A}$. We shall need the absolute flux density

$$B = |\boldsymbol{B}| = |\operatorname{curl} \boldsymbol{A}| \tag{6.17}$$

The source-free term of the Lagrange density Λ is then a function of the form

$$U(B) = \int_0^B H(B')dB' \tag{6.18}$$

provided that any magnetic media are isotropic, that is, their permeability may vary with $\boldsymbol{B}$ but not with the direction of $\boldsymbol{B}$; μ is now a scalar, not a tensor. For a nonlinear medium with saturation effects, but without hysteresis effects, a graph of such a function is shown in Fig. 6.2. Differentiation of Eq. (6.18) with respect to $\boldsymbol{B}$ gives

$$\boldsymbol{H}(\boldsymbol{B}) = \frac{\partial U}{\partial \boldsymbol{B}} = \frac{\boldsymbol{B}}{B} \frac{dU(B)}{dB} =: \nu(B)\boldsymbol{B} \tag{6.19a}$$

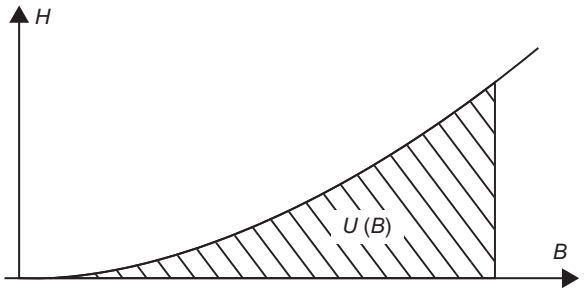


Figure 6.2
A branch of the magnetization curve, in which $U(B)$ denotes the area under the curve. As $B \to \infty$, the gradient tends asymptotically to a constant value.

in accordance with Eq. (6.3) and hence

$$\nu(B) = \frac{1}{B}\frac{dU}{dB} \tag{6.19b}$$

The complete Lagrange density is now given by

$$\Lambda = U(B) - \boldsymbol{j}\cdot\boldsymbol{A} \tag{6.20}$$

Evaluation of Eq. (6.14) for the three Cartesian components of $\boldsymbol{A}$ and expression of the result in vector notation leads to Eq. (6.8). These tedious but elementary calculations are not reproduced here.

In the case of linear (unsaturated) media, Eq. (6.18) simplifies to the familiar energy density

$$U(B) = \frac{1}{2\mu}B^2 = \frac{1}{2\mu}(\nabla \times \boldsymbol{A})^2 \tag{6.21}$$

$\mu = 1/\nu$ now being a constant. With the additional constraint (6.9), the Euler equations reduce to (6.10).

In all the cases considered here, the Lagrange density Λ has the physical meaning of an energy density, which generally contains an additional interaction term. The functional $F = \int \Lambda \, d^3r$ is thus an energy. Since this has no upper bound, the concrete evaluation of the variational principle always results in a *minimum* of F. This is of importance for the finite-element method. If the boundary ∂S is extended to infinity, the functional F must remain finite. This implies that any electrodynamic quantities must satisfy the 'natural' boundary conditions, which means that as $|r| \to \infty$ they must converge to zero in such a way that all the integrals involved remain finite.

6.4 Rotationally Symmetric Fields

Rotationally symmetric fields are of particular interest in electron optics, as the most common electron lenses are round, by which we mean they are built up from rotationally symmetric fields. Rotationally symmetric electrostatic fields are a simple special case of the Fourier series expansion treated in Chapter 7, Series Expansions, and will therefore not be considered here. Magnetic fields, however, require special attention, as we now explain.

For rotational symmetry it is advantageous to introduce cylindrical coordinates (z, r, φ), as defined in Section 2.4. It is necessary to assume that the current density $\boldsymbol{j}(\boldsymbol{r})$ is circular and it is further convenient, though not necessary, to assume that the vector potential $\boldsymbol{A}(\boldsymbol{r})$ is likewise circular. This means that both vector functions have only azimuthal components:

$$\boldsymbol{j}(\boldsymbol{r}) = j(z,\ r)\boldsymbol{i}_{\varphi} \tag{6.22}$$

$$\boldsymbol{A}(\boldsymbol{r}) = A(z, r)\boldsymbol{i}_{\varphi} \tag{6.23}$$

This already implies that $\operatorname{div} \boldsymbol{j} = 0$ and $\operatorname{div} \boldsymbol{A} = 0$. The cylindrical components of $\boldsymbol{B} = \operatorname{curl} \boldsymbol{A}$ are given by

$$B_z(z, r) = \frac{1}{r}\frac{\partial}{\partial r}(rA), \quad B_r = -\frac{\partial A}{\partial z}, \quad B_{\varphi} = 0 \tag{6.24}$$

which represent a magnetic field, the direction of which always lies in a *meridional* plane. The opposite case of a circular magnetic field produced by a meridional current distribution is of little interest in electron optics but is of importance in plasma physics. This case will not be treated here.

It is advantageous to introduce the magnetic flux function $\Psi(z, r)$, as in Section 2.5:

$$\Psi(z, r) = 2\pi \int_0^r r' B_z(z, r')dr' \tag{6.25}$$

This is the static special case of Eq. (2.32). An immediate consequence is that $B_z(r, z) = (2\pi r)^{-1}\partial\Psi/\partial r$ is the static special case of Eq. (2.33). The condition $\operatorname{div} \boldsymbol{B} = 0$, expressed in cylindrical coordinates, now becomes

$$\operatorname{div} B = \frac{\partial B_z}{\partial z} + \frac{1}{r}\frac{\partial}{\partial r}(rB_r) = 0 \tag{6.26}$$

in which we have used $B_{\varphi} \equiv 0$ (6.24).

Introducing the expression obtained above for B_z into (6.26), we find

$$\frac{\partial}{\partial r}(rB_r) = -\frac{1}{2\pi}\frac{\partial^2 \Psi}{\partial z \partial r} = \frac{\partial}{\partial r}\left(-\frac{1}{2\pi}\frac{\partial \Psi}{\partial z}\right)$$

Integration with respect to r yields first

$$rB_r = -\frac{1}{2\pi}\frac{\partial \Psi}{\partial z} + C(z)$$

$C(z)$ being an arbitrary differentiable function of z. At the optic axis ($r = 0$), B_r must remain finite. Since Ψ and $\partial\Psi/\partial z$ are proportional to r^2 in the vicinity of the optic axis, we can conclude that $C(z) \equiv 0$. Since the restriction to static fields is irrelevant in this context, we thus obtain Eq. (2.34).

By applying Stokes's integral theorem to the circle $z = \text{const}$, $r = \text{const}$, $r' \leq r$ shown in Fig. 2.1 we find immediately

$$\Psi(z, r) = 2\pi r A(z, r) \tag{6.27}$$

A corresponding relation has already been used in Section 4.2. In connection with the paraxial properties of magnetic round lenses, it is helpful to introduce an auxiliary potential function

$$\Pi(z,\ r) = \frac{2}{r}A(z,\ r) \tag{6.28}$$

We can now write the different representations of B_z and B_r as follows:

$$B_z(z,\ r) = \frac{\partial A}{\partial r} + \frac{A}{r} = \Pi + \frac{r}{2}\cdot\frac{\partial \Pi}{\partial r} = \frac{1}{2\pi r}\cdot\frac{\partial \Psi}{\partial r} \tag{6.29}$$

$$B_r(z,\ r) = -\frac{\partial A}{\partial z} = -\frac{r}{2}\cdot\frac{\partial \Pi}{\partial z} = -\frac{1}{2\pi r}\cdot\frac{\partial \Psi}{\partial z} \tag{6.30}$$

Comparing the different representations and noting that $\Pi(z,\ r)$ must remain finite as $r \to 0$, we see that

$$B_z(z, 0) = \Pi(z, 0) \tag{6.31}$$

Furthermore, we see that, in the immediate vicinity of the optic axis, B_r and A are proportional to r. These facts are of great importance in the physics of magnetic lenses.

When we come to discuss the boundary conditions, it will be useful to write Eqs. (6.29) and (6.30) in vector form:

$$\boldsymbol{B}(\boldsymbol{r}) = \frac{1}{2\pi r}\boldsymbol{i}_\varphi \times \text{grad}\ \Psi(z, r) \tag{6.32}$$

This is easily verified by writing out Eq. (6.32) in cylindrical components.

The partial differential equation to be satisfied by $A(z, r)$ is most easily obtained by substituting Eq. (6.24) into Eq. (6.8). In cylindrical coordinates, only the azimuthal component fails to vanish:

$$\frac{\partial}{\partial z}\left(\nu\frac{\partial A}{\partial z}\right)+\frac{\partial}{\partial r}\left(\nu\frac{\partial A}{\partial r}+\nu\frac{A}{r}\right)=-j(z,r) \tag{6.33}$$

In unsaturated materials $\mu=\nu^{-1}$ is a constant and Eq. (6.33) then simplifies to

$$\frac{\partial^2 A}{\partial z^2}+\frac{\partial^2 A}{\partial r^2}+\frac{1}{r}\frac{\partial A}{\partial r}-\frac{A}{r^2}=-\mu j \tag{6.34}$$

Introducing Eqs. (6.27) and (6.28) into this equation, more convenient partial differential equations, having no term in r^{-2}, are obtained:

$$\frac{\partial^2 \Pi}{\partial z^2}+\frac{\partial^2 \Pi}{\partial r^2}+\frac{3}{r}\frac{\partial \Pi}{\partial r}=-\frac{2\mu j}{r} \tag{6.35}$$

$$\frac{\partial^2 \Psi}{\partial z^2}+\frac{\partial^2 \Psi}{\partial r^2}-\frac{1}{r}\frac{\partial \Psi}{\partial r}=-2\pi\mu r j \tag{6.36}$$

In Chapter 7, Series Expansions and Chapter 8, Boundary-Value Problems, these equations will be encountered again as formal special cases of a more general differential equation.

6.5 Planar Fields

Planar fields are such that the field components are independent of one of the three Cartesian coordinates. In practice, they are idealizations of three-dimensional fields, obtained by neglecting the fringe-field domains in one direction. Fields of this type are approximately realized in such devices as electrostatic deflection units, slit lenses, secondary emission multipliers and deflection magnets with plane surfaces.

Without loss of generality, we may choose the Cartesian coordinate system in such a manner that the field does not depend on the coordinate y. In the z–x plane, we introduce the complex variable

$$w=z+\mathrm{i}x=r\mathrm{e}^{\mathrm{i}\varphi} \tag{6.37}$$

Any analytic function

$$f(w)=u(z,x)+\mathrm{i}v(z,x) \tag{6.38}$$

is a solution of Laplace's equation. This is a consequence of differentiability in the complex plane. As is well-known from the theory of analytic functions, u and v satisfy the Cauchy−Riemann equations

$$\frac{\partial}{\partial z}u(z,x) = \frac{\partial}{\partial x}v(z,x), \quad \frac{\partial}{\partial z}v(z,x) = -\frac{\partial}{\partial x}u(z,x) \tag{6.39}$$

From these, the orthogonality relation

$$\frac{\partial u}{\partial z}\frac{\partial v}{\partial z} + \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} \equiv \text{grad } u \cdot \text{grad } v = 0, \tag{6.40}$$

the conformity relation

$$\left|\text{grad } u\right| = \left|\text{grad } v\right| = \left|f'(w)\right| \tag{6.41}$$

(with $f'(w) \coloneqq df/dw$) and the Laplace equations

$$\nabla^2 u = 0, \quad \nabla^2 v = 0, \quad \left(\nabla^2 = \frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial x^2}\right) \tag{6.42}$$

can be derived. Eqs (6.40) and (6.41) are characteristic of a conformal mapping, as shown in Fig. 6.3. The function $f(w)$ defined by Eq. (6.38) can be interpreted as a transform from the square-shaped map in the z−x-plane (Fig. 6.3A) to a curvilinear map in the u−v-plane (Fig. 6.3B). As the size of the curved cells is decreased, these cells collapse to squares.

A standard method of calculating planar fields consists in finding the inverse function $[f(w)]^{-1}$ that transforms the given boundary in the z−x-plane into a pair of lines $u = \text{const}$. All lines $u(z, x) = \text{const}$ are then equipotentials of the field and all lines $v(z, x) = \text{const}$ are orthogonal flux lines. Such a transform is again a conformal mapping. This method of field calculation, commonplace in the mathematical literature, will not be outlined here since it is

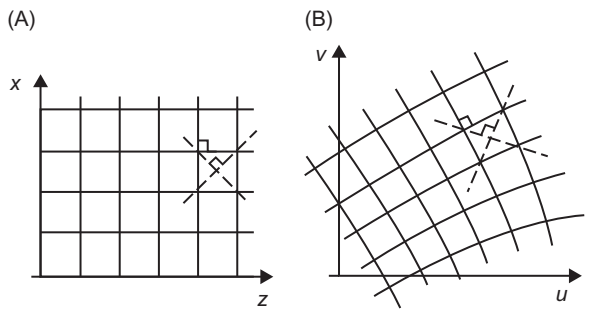


Figure 6.3

(A) Square grid in the (z, x)-plane. (B) Conformal map in the (u, v)-plane; the mapping is such that angles are conserved.

more complicated than the standard numerical techniques described later. It was regularly employed in the analysis of mass spectrometers (Boerboom and Chen, 1984; Wallington, 1970, 1971). An extremely detailed account is given by Durand (1966).

Some interesting analytical functions and the corresponding potential fields are the following:

1. Arbitrary convergent power series expansions:

$$f(w) = \sum_m a_m w^m = \sum_m a_m r^m \mathrm{e}^{\mathrm{i}m\varphi} \tag{6.43}$$

Nonnegative integral exponents m lead to regular Taylor series expansions, while negative integers generate singular Laurent series expansions. The potential functions obtained from Taylor series expansions are approximately realized in the central zones of multipole devices, see Chapter 7, Series Expansions. Nonintegral values of m give rise to potential fields in the vicinity of sharp edges. The potential of the sharp edge shown in Fig. 6.4, for example, is given by

$$u(r, \phi) = u_0 + \sum_{n=1}^{\infty} A_n r^{n\mu} \sin n\mu\varphi, \quad \mu = \frac{\pi}{2\pi - \gamma} \tag{6.44}$$

Here n is a positive integer. It is easy to verify that $u = u_0 = \text{const}$ on the lines $\varphi = 0$ and $\varphi = 2\pi - \gamma$ for all sets of coefficients $A_1, A_2, \ldots$

2. Complex Fourier integral expansions:

$$f(w) = \int_0^{\infty} A(k)\exp(\mathrm{i}kw)\ dk \qquad (k \text{ real}, \ \geq 0) \tag{6.45}$$

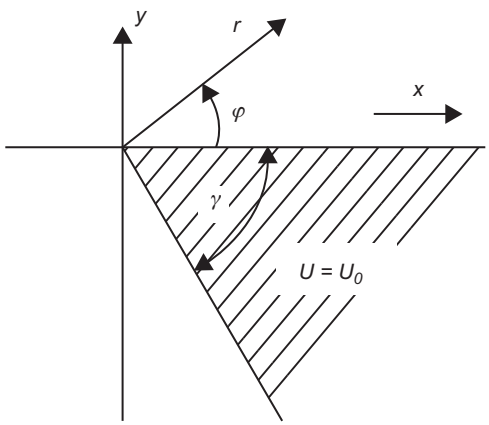


Figure 6.4
Coordinate system adapted to a sharp edge.

With $A(k) = A'(k) - iB'(k)$ and $A(0) = 0$ we obtain a potential distribution that satisfies $u(z, x) \to 0$ as $x \to \infty$:

$$u(z,x) = \int_0^\infty e^{-kx}\{A'(k)\cos kz + B'(k)\sin kz\}\ dk \tag{6.46}$$

This has been found useful in some investigations on electron mirrors.

3. Logarithmic singularity

$$f(w) = \log Cw, \quad u = \log Cr, \quad \upsilon = \varphi \tag{6.47}$$

(C is a positive constant.) This is of great importance in the practical application of the integral equation method, see Chapter 10, The Boundary-Element Method. The inverse function

$$w = z + ix = r\exp(i\varphi) = \frac{1}{C}\exp(u + i\upsilon) \tag{6.48a}$$

$$r = \frac{1}{C}\exp(u), \quad \varphi \equiv \upsilon \tag{6.48b}$$

describes the transform from quadratic to polar grids, see also Section 11.4.2.

CHAPTER 7

Series Expansions

As has been mentioned in Section 6.1, the electromagnetic field in the extended paraxial domain is of paramount interest in electron optics. In this domain, the field will be represented by series expansions, the general structures of which can already be derived without explicit solution of the corresponding boundary-value problem. As will be obvious later, the field in the vicinity of the optic axis can be obtained by analytic continuation of the axial distribution.

In this chapter we shall assume that the optic axis is straight. Series expansions adapted to curvilinear axes are required in the treatment of devices with sector fields and will be derived in that context (see Part X). We shall further assume that the paraxial domain is usually source-free. This is certainly true for all current distributions, since the windings of coils are always far from the optic axis. Electric space charge may accumulate in the extended paraxial domain. Its distribution is, however, so inhomogeneous that the corresponding series expansions are of little practical value. Here, therefore, we shall mostly exclude them; they will be considered in more detail when we come to treat electron guns (see Part IX).

7.1 Azimuthal Fourier Series Expansions

The aim of the following considerations is to decompose a three-dimensional field into a sequence of uncoupled two-dimensional fields. This is advantageous, since two-dimensional fields are far more easily calculated than three-dimensional ones. The required decoupling is obtained by expanding the field as a series of complete orthogonal functions, most favourably as a Fourier series (Glaser, 1952, Section 35). Since the present discussion is quite general, source distributions are not excluded.

7.1.1 Scalar Potentials

Let us now consider a general Poisson equation

$$\nabla^2 V(\boldsymbol{r}) = -S(\boldsymbol{r}) \tag{7.1}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00007-9

© 2018 Elsevier Ltd. All rights reserved.

regardless of the special meanings of the functions $V(\boldsymbol{r})$ and $S(\boldsymbol{r})$. In cylindrical coordinates (z, r, φ) this Poisson equation takes the form

$$\frac{\partial^2 V}{\partial z^2} + \frac{\partial^2 V}{\partial r^2} + \frac{1}{r} \cdot \frac{\partial V}{\partial r} + \frac{1}{r^2} \cdot \frac{\partial^2 V}{\partial \varphi^2} = -S(z, r, \varphi) \tag{7.2}$$

Both functions V and S are periodic with respect to the azimuth φ, the period being 2π, which suggests that we should introduce their Fourier series expansions. These can be represented in different equivalent forms. One essential requirement is that it should be easy to transform these series expansions and the corresponding ones for the gradient into *regular* power series expansions with respect to x and y. With these transforms in mind, we introduce the variables

$$w \coloneqq x + \mathrm{i}y = r \exp(\mathrm{i}\varphi) \tag{7.3}$$

$$s \coloneqq x^2 + y^2 = r^2 = ww^* \tag{7.4}$$

the asterisk denoting complex conjugation. The series expansions in question then have the form

$$V(z, r\ \varphi) = \sum_{m=0}^{\infty} r^m \Re\left\{ U_m(z, s) \mathrm{e}^{\mathrm{i}m\varphi} \right\} \tag{7.5}$$

$$S(z, r\ \varphi) = \sum_{m=0}^{\infty} r^m \Re\left\{ g_m(z, s) \mathrm{e}^{\mathrm{i}m\varphi} \right\} \tag{7.6}$$

and in the Cartesian representation

$$V(\boldsymbol{r}) = \sum_{m=0}^{\infty} \Re\left\{ U_m(z, s) w^m \right\} \tag{7.7}$$

with a corresponding expression for $S(\boldsymbol{r})$. Differentiation of Eq. (7.7) with respect to z immediately results in:

$$\frac{\partial V}{\partial z} = \sum_{m=0}^{\infty} \Re\left\{ w^m \frac{\partial U_m(z, s)}{\partial z} \right\} \tag{7.8}$$

The remaining derivatives are most favourably expressed in complex form:

$$\frac{\partial V}{\partial x} + \mathrm{i}\frac{\partial V}{\partial y} = \sum_{m=0}^{\infty} \left\{ w^{m+1} \frac{\partial U_m}{\partial s} + (w^*)^{m-1} \left(s \frac{\partial U_m^*}{\partial s} + m U_m^* \right) \right\} \tag{7.9}$$

For $m = 0$ this series expansion contains *no* singularity, since $(w^*)^{-1}\ s = w$. Recalling that U_0 must be a real function, we obtain the expression $2w\partial U_0/\partial s$ for the zero-order term of this series expansion.

Introducing Eqs (7.5) and (7.6) into (7.2), we first find a sequence of uncoupled differential equations

$$\frac{\partial^2 U_m}{\partial z^2} + \frac{\partial^2 U_m}{\partial r^2} + \frac{2m+1}{r}\frac{\partial U_m}{\partial r} = -g_m(z,r) \quad m = 0, 1, 2, \ldots \tag{7.10}$$

It is possible to introduce the variable $s = r^2$ (Glaser, 1952, Section 35); this is, however, unfavourable with respect to the numerical solution, see Chapter 11, The Finite-Difference Method (FDM).

In practice, in order to evaluate infinite series expansions numerically, we have to truncate them in a suitable manner. From Eq. (7.5) we can get an idea of how this is to be done. Since U_m and g_m may be regarded as slowly varying functions within the extended paraxial domain, the factor r^m is of greatest importance. Let R be some characteristic bore radius of the device in question. Then the exponent $m = M$ at which the series is truncated should be chosen in such a way that all terms $(r/R)^m$ with $m > M$ can be neglected. In practice $M = 5$ is usually sufficient, since $r/R \lesssim 10^{-1}$. In order to avoid any misunderstanding, we emphasize that in concrete calculations, Eq. (7.10) (with $0 < m < M$) are to be solved as a sequence of boundary-value problems, which implies that r may extend far beyond the paraxial domain. Only the values of the solutions obtained inside the extended paraxial domain are to be retained.

In Eq. (7.10), the coefficients of the partial derivatives are simple real factors. A further simplification is therefore possible by considering only real source terms $g_m(z, s)$ and real solutions $U_m(z, s)$. Complex solutions can easily be obtained by forming linear combinations with appropriate complex factors. In Section 7.2, we shall thus assume that the functions $U_m(z, s)$ are *real*.

7.1.2 Vector Potentials

In Chapter 4, Variational Principles and Chapter 5, Hamiltonian Optics, we have seen that the vector potential contributes to the canonical momentum and to the characteristic function. Though the vector potential itself is not an observable quantity, it still plays an important role in theoretical considerations. We now derive appropriate series expansions for it. Eqs (6.9) and (6.10) must be satisfied and furthermore we impose the natural boundary condition $\boldsymbol{A}(\boldsymbol{r}) \to 0$ for $|\boldsymbol{r}| \to \infty$. Then $\boldsymbol{A}(\boldsymbol{r})$ is uniquely defined. For technical reasons we have $j_x = j_y = 0$ on the optic axis, which implies that $A_x = A_y = 0$ for $x = y = 0$.

The desired series expansions will be similar in form to Eq. (7.7). Here it is more convenient to consider the imaginary parts instead of the real ones, which is only a minor difference. Furthermore it is convenient to introduce complex transverse components, thus

$$A_T := A_x + \mathrm{i}A_y = \frac{\mathrm{i}}{2}\sum_{m=1}^{\infty} C_m(z,s)w^m \tag{7.11}$$

$$A_z := -\Im\left(\sum_{m=0}^{\infty} D_m(z,s)w^m\right) \tag{7.12}$$

In a similar way the current density $\boldsymbol{j}(\boldsymbol{r})$ is represented by

$$j_T := j_x + \mathrm{i}j_y = \frac{\mathrm{i}}{2}\sum_{m=1}^{\infty} J_m(z,s)w^m \tag{7.13}$$

$$j_z := -\Im\left(\sum_{m=0}^{\infty} L_m(z,s)w^m\right) \tag{7.14}$$

In the functions introduced on the right-hand side the variable s is again defined by Eq. (7.4).

A straightforward differentiation in Cartesian coordinates results in

$$\operatorname{div}\boldsymbol{A} = -\Im\left\{\sum_{m=1}^{\infty}(D_{m-1\,|z} + sC_{m|s} + mC_m)w^{m-1}\right\} = 0$$

Here the subscripts behind vertical bars denote partial differentiations with respect to the corresponding variable (see Section 2.4). Since different powers w^m of w must be linearly independent, this condition $\operatorname{div}\boldsymbol{A} = 0$ can be satisfied only if

$$D_{m-1|z} + sC_{m\,s} + mC_m = 0 \quad (m \geq 1) \tag{7.15}$$

is valid for all integers m. Similarly the condition $\operatorname{div}\boldsymbol{j} = 0$ leads to

$$L_{m-1\,|\,z} + sJ_{m\,|\,s} + mJ_m = 0 \quad (m \geq 1) \tag{7.16}$$

The determination of $\nabla^2\boldsymbol{A}$ in Cartesian coordinates is also straightforward; we obtain

$$\nabla^2 A_T = \frac{\mathrm{i}}{2}\sum_{m=1}^{\infty}(C_{m\,|\,zz} + 4sC_{m\,|\,ss} + 4(m+1)C_{m|s})w^m$$

$$\nabla^2 A_z = -\Im\left\{\sum_{m=0}^{\infty}(D_{m\,|zz} + 4sD_{m|ss} + 4(m+1)D_{m|s})w^m\right\}$$

These series expansions are to be matched to Eqs (7.13) and (7.14). Using the linear independence of different powers of w and the vector Poisson equation $\nabla^2\boldsymbol{A} = -\mu\,\boldsymbol{j}$ we obtain

$$C_{m|\,zz} + 4sC_{m\,|ss} + 4(m+1)C_{m|\,s} = -\mu J_m \tag{7.17}$$

$$D_{m|zz} + 4sD_{m|ss} + 4(m+1)D_{m|s} = -\mu L_m \tag{7.18}$$

With the transformations

$$\frac{\partial}{\partial s} = \frac{1}{2r}\frac{\partial}{\partial r}, \quad \frac{\partial^2}{\partial s^2} = \frac{1}{4s}\frac{\partial^2}{\partial r^2} - \frac{1}{4sr}\frac{\partial}{\partial r}$$

we obtain the more suitable differential equations

$$C_{m|zz} + C_{m|rr} + \frac{2m+1}{r}C_{m|r} = -\mu J_m(z,s) \tag{7.19}$$

$$D_{m|zz} + D_{m|rr} + \frac{2m+1}{r}D_{m|r} = -\mu L_m(z,s) \tag{7.20}$$

which have the same mathematical structure as Eq. (7.10). Here, however, their solutions are additionally coupled by Eq. (7.15).

The magnetic flux density is obtained by differentiation, $\boldsymbol{B} = \nabla \times \boldsymbol{A}$. After some elementary calculations we find:

$$\begin{aligned} B_T &:= B_x + \mathrm{i}B_y \\ &= -\sum_{m=0}^{\infty}\left(sD^*_{m|s} + mD^*_m\right)(w^*)^{m-1} - \sum_{m=0}^{\infty}\left(\frac{1}{2}C_{m+1|z} - D_{m|s}\right)w^{m+1} \end{aligned} \tag{7.21}$$

In *source-free domains*, where $\boldsymbol{j} = 0$, this expression can be further simplified. In order to show this, we differentiate Eq. (7.15) with respect to s and make use of Eq. (7.17) with $J_m = 0$, obtaining

$$-D_{m-1|sz} = sC_{m|ss} + (m+1)C_{m|s} = -\frac{1}{4}C_{m|zz}$$

This can be integrated with respect to z. Since $\boldsymbol{A}$ has to vanish asymptotically, no additive constant can appear, and so

$$4D_{m|s} = C_{m+1|z} \tag{7.22}$$

Introducing this into Eq. (7.21) and eliminating $C_{m+1|z}$, we find now

$$B_T = -\sum_{m=0}^{\infty}\left\{\left(sD^*_{m|s} + mD^*_m\right)(w^*)^{m-1} + D_{m|s}w^{m+1}\right\}$$

This equation has the same structure as Eq. (7.9). Therefore with $\mu = \mu_0$ *in vacuo* we can introduce a scalar potential $W(\boldsymbol{r})$ by writing

$$W(\boldsymbol{r}) = \Re\sum_{m=0}^{\infty}(D_m(z,s)w^m) \tag{7.23}$$

Calculation of the axial component B_z confirms this result. A straightforward differentiation yields

$$B_z = A_{y|x} - A_{x|y} = \Re \sum_{m=1}^{\infty} (sC_{m|s} + mC_m) w^{m-1}$$

On the right-hand side the expression $D_{m|z}$ can be introduced by means of Eq. (7.15) and we then find

$$B_z = -\frac{\partial}{\partial z} \Re \left(\sum_{m=0}^{\infty} D_m(z, s) w^m \right) = -W_{|z}$$

since the permeability μ is constant. The results obtained are in agreement with Eqs. (6.11) and (6.12), as they should be.

Comparison of Eqs (7.12) and (7.23) shows that W and A_z are interrelated as the real and imaginary parts of a complex function:

$$W - \mathrm{i}A_z = \sum_{m=0}^{\infty} D_m(z, s) w^m \tag{7.24}$$

This relation holds, of course, only in source-free domains.

The representation of the vector potential given above is a generalization of Sturrock's (1951) formula. Other gauges, which do not satisfy div $\boldsymbol{A} = 0$, have been introduced by Glaser (1952, Section 36) and by Schwertfeger and Kasper (1974). The present procedure is convenient, since we always arrive at the same class of partial differential equations.

7.2 Radial Series Expansions

7.2.1 Scalar Potentials

In accordance with the assumption that the paraxial domain is source-free, we shall now investigate solutions of (7.10) with source terms vanishing for sufficiently small values of r. From Eqs (7.5) and (7.7), it can be seen that a power series expansion with respect to $s = r^2$ will be the most suitable. We therefore introduce

$$U_m(z, s) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} c_n(z, m) s^n \tag{7.25}$$

into Eq. (7.10), the factor $(-1)^n/(2n)!$ being included for reasons of convenience. The coefficients c_n, unknown at the moment, are related by recurrence formulae:

$$\left(1+\frac{2m+1}{2n+1}\right)c_{n+1}(z,m)=c_n''(z,m),\quad n=0,1,2,\ldots$$

where primes denote differentiation with respect to z. The coefficient $c_0(z, m)$ can be chosen arbitrarily provided that all its derivatives remain finite for all values of z ($|z|\to\infty$ included). This function is the *axial* value of U_m:

$$c_0(z,m)=U_m(z,0)=:u_m(z) \tag{7.26}$$

the subscript 0 being omitted to lighten the notation. In terms of r^2 the required power series expansion is now given by

$$\begin{aligned}U_m(z,r^2)&=\sum_{n=0}^{\infty}\frac{m!}{n!(m+n)!}\left(-\frac{r^2}{4}\right)^m u_m^{(2n)}(z)\\&=u_m(z)-\frac{r^2u_m''(z)}{4(m+1)}+\left(\frac{r^2}{4}\right)^2\frac{u_m^{(4)}(z)}{2!(m+1)(m+2)}+O(r^6)\end{aligned} \tag{7.27}$$

Apart from the notation, this radial series expansion is identical with that given by Glaser (1952); it is equivalent to that given by Kasper (1982) if α is identified with $2m+1$. Eq. (7.27) shows that each function $U_m(z, s)$ is already uniquely determined by its axial distribution $u_m(z)$ (called the axial harmonic), provided that the series expansion converges. This property of the solution is very similar to that of analytic functions in the complex plane and by analogy we shall refer to Eq. (7.27) as an analytic continuation of the axial values.

In the most general case, the convergence of Eq. (7.27) cannot be proven but must be assumed, at least in the extended paraxial domain. In this context there may arise problems since in fact Eq. (7.27) does *not* converge for all values of r. Unfortunately even reliable estimates for the radius of convergence are not known (apart from some special examples).

In principle, it should be possible to compute the field in an entire device by analytic continuation of the appropriate axial harmonics $u_m(z)$, since the solution as a boundary-value problem shows clearly that the singularities must be located outside or at the boundaries. But in practice this is impossible for various reasons. A first problem is that the analytical continuation is numerically unstable. Even if one starts with the correct functions $u_m(z)$, one will *not* obtain the correct boundary values of $U_m(z, r^2)$, since rounding and truncation errors, initially very small, may increase dramatically. Another serious difficulty is that it will be impossible to obtain reasonable shapes of electrodes or polepieces if the functions $u_m(z)$ are only slightly different from those corresponding to a realistic field.

This problem is further discussed in Chapter 34, Numerical Calculation of Trajectories, Paraxial Properties and Aberrations.

In a realistic field computation, the first step is the solution of a boundary-value problem. In this way the axial harmonics $u_m(z)$ can be determined uniquely. These have then to be differentiated numerically. Finally Eq. (7.27) can be evaluated for arbitrary values of z and sufficiently small values of r ($r/R \lesssim 10^{-1}$, see Section 7.1).

A very important practical application of the radial series expansion arises in general theoretical calculations, where no concrete numerical evaluations are required. The purpose of such calculations is the derivation of general rules for focusing properties and aberration coefficients in classes of devices. In this context, it is often helpful to use simple analytic models of the axial field distributions. These models must contain some free parameters with which a fit to a correct numerical solution for the $u_m(z)$ is possible.

7.2.2 Vector Potentials

Since the functions $C_m(z, s)$ and $D_m(z, s)$, introduced in Section 7.1.2, satisfy differential equations of the same basic type as Eq. (7.10), their radial series expansions must be similar to Eq. (7.27), always assuming that the domain of solution is source-free. Here we have to consider *two* series expansions, one for each of the two functions $C_m(z, s)$ and $D_m(z, s)$, but Eqs. (7.15) and (7.22) show that these are linearly dependent. The consequence is that, though the three components of $\boldsymbol{A}(\boldsymbol{r})$ are different functions of position, only *one* axial harmonic $\Pi_m(z)$ can be introduced independently for each Fourier component. Since the scalar potential W itself, given by Eq. (7.23), has little significance, we define the axial harmonics by means of the relations

$$\Pi_m(z) \coloneqq -D_{m|z}(z, 0), \quad (m \geq 0) \tag{7.28}$$

and hence

$$D_m(z, 0) = -\int_{-\infty}^{z} \Pi_m(z')dz' \tag{7.29}$$

Applying Eq. (7.20) (with $L_m(z, s) = 0$), the radial series expansion of $D_m(z, s)$ is then given by

$$\begin{aligned} D_m(z, r^2) &= -\sum_0^\infty \frac{m!}{n!(m+n)!}\left(-\frac{r^2}{4}\right)^n \Pi_m^{(2n-1)}(z) \\ &= -\int_{-\infty}^{z} \Pi_m(z')dz' + \frac{r^2\Pi_m'(z)}{4(m+1)} - \left(\frac{r^2}{4}\right)^2 \frac{\Pi_m'''(z)}{2!(m+1)(m+2)} + O\left(r^6\right) \end{aligned} \tag{7.30}$$

The axial values of the function $C_m(z, r^2)$ can be determined from Eq. (7.15). On the optic axis, the term $sC_m|_s$ vanishes and using Eq. (7.28), we find

$$C_m(z,0) = \Pi_{m-1}(z)/m \quad (m \geq 1) \tag{7.31}$$

The series expansion for $C_m(z, r^2)$ is given by

$$C_m(z,r^2) = \sum_{n=0}^{\infty} \frac{m!}{n!(m+n)!} \left(-\frac{r^2}{4}\right)^n C_m^{(2n)}(z,0)$$

which must be similar to Eq. (7.27), since in the homogeneous case the corresponding differential equations (7.10) and (7.17) have the same formal structure. With the aid of Eq. (7.31), we find

$$\begin{aligned} C_m(z,r^2) &= \sum_{n=0}^{\infty} \frac{(m-1)!}{n!(m+n)!} \left(-\tfrac{r^2}{4}\right)^n \Pi_{m-1}^{(2n)}(z) \\ &= \frac{\Pi_{m-1}(z)}{m} - \frac{r^2 \Pi''_{m-1}(z)}{4m(m+1)} + \left(\tfrac{r^2}{4}\right)^2 \frac{\Pi_{m-1}^{(4)}(z)}{2!m(m+1)(m+2)} - O\left(r^6\right) \end{aligned} \tag{7.32}$$

It is straightforward to prove, by carrying out the necessary differentiations, that Eqs (7.15) and (7.22) are satisfied.

In order to obtain the complete series expansions of the vector potential $\boldsymbol{A}$, the expansions, derived above, have to be substituted in Eqs (7.11) and (7.12). Practical expressions will be given in the next sections. Apart from the different notation, our results are identical with those given by Sturrock (1951). They differ from the formulae of Schwertfeger and Kasper (1974), which do not satisfy div $\boldsymbol{A} = 0$, and also from those of Glaser (1952), in which $A_z = 0$ is assumed and which do not necessarily satisfy $\boldsymbol{A} \to 0$ for $z \to \infty$. The particular gauge that does satisfy the natural boundary conditions is most convenient in practical applications.

7.2.3 Explicit Representations

In many practical applications, it is quite sufficient to truncate the power series expansions of the potentials after the terms of fourth order in x and y, and consequently those of the field strength after the third order. These series expansions play an important role in the theory of electron optical aberrations; in instruments in which the primary aberrations have been corrected, the next higher order terms are required

In order to introduce a comparatively simple and easily remembered *real* representation of the series expansions, we define real axial harmonics $p_{\mathrm{m}}(\mathrm{z})$, $q_m(z)$, $P_m(z)$ and $Q_{\mathrm{m}}(z)$ as follows:

$$U_m(z,0) =: \frac{(-1)^m}{m!}(p_m(z) - \mathrm{i}q_m(z)) \quad m = 0, 1, 2, \ldots \tag{7.33}$$

$$D_m(z,0) =: \frac{(-1)^m}{m!}(P_m(z) - \mathrm{i}Q_m(z)) \quad m = 0, 1, 2, \ldots \tag{7.34}$$

Eq. (7.33) will be used exclusively for the expansion of the electrostatic potential $\Phi(\boldsymbol{r})$, while Eq. (7.34) will refer exclusively to the vector potential $\boldsymbol{A}(\boldsymbol{r})$ and the flux density $\boldsymbol{B}(\boldsymbol{r})$. The Fourier coefficients with $m = 0$ and $m = 1$ have a special meaning which will be encountered frequently. We therefore introduce a special notation for these coefficients apart from q_0, which is identically zero:

$$p_0(z) =: \phi(z), \quad p_1(z) =: F_1(z), \quad q_1(z) =: F_2(z) \tag{7.35}$$

Here $\phi(z) = \Phi(z, 0, 0)$ is the familiar axial potential, $F_1(z) = -\Phi_{\mathrm{lx}}(z, 0, 0)$ and $F_2(z) = -\Phi_{\mathrm{ly}}(z, 0, 0)$ are the transverse components of the field strength $\boldsymbol{E}$ on the optic axis. The electric potential is then given by

$$\begin{aligned}
\Phi(r) = {} & \phi(z) - \frac{1}{4}(x^2+y^2)\phi''(z) + \frac{1}{64}(x^2+y^2)^2\phi^{(4)}(z) - \frac{1}{2304}(x^2+y^2)^3\phi^{(6)} \\
& - xF_1(z) - yF_2(z) + \frac{1}{8}(x^2+y^2)(xF_1'' + yF_2'') - \frac{1}{192}(x^2+y^2)^2(xF_1^{(4)} + yF_2^{(4)}) \\
& + \frac{1}{2}(x^2-y^2)p_2(z) + xyq_2(z) - \frac{1}{24}(x^2+y^2)\left\{(x^2-y^2)p_2'' + 2xyq_2''\right\} \\
& + \frac{1}{768}(x^2+y^2)^2\left\{(x^2-y^2)p_2^{(4)} + 2xyq_2^{(4)}\right\} \\
& - \frac{1}{6}p_3(z)x(x^2-3y^2) + \frac{1}{6}q_3(z)y(y^2-3x^2) + \frac{1}{96}(x^2+y^2)\left\{x(x^2-3y^2)p_3'' + y(3x^2-y^2)q_3''\right\} \\
& + \frac{1}{24}p_4(z)(x^4 - 6x^2y^2 + y^4) + \frac{1}{6}q_4(z)x(x^2y - y^3) \\
& - \frac{1}{480}(x^2+y^2)\left\{(x^4 - 6x^2y^2 + y^4)p_4'' + 4xy(x^2-y^2)q_4''\right\} \\
& - \frac{1}{120}\left\{x(x^4 - 10x^2y^2 + 5y^4)p_5 + y(5x^4 - 10x^2y^2 + y^4)q_5\right\} \\
& + \frac{1}{720}\left\{(x^2-y^2)(x^4 - 14x^2y^2 + y^4)p_6 + 2xy(3x^2-y^2)(x^2-3y^2)q_6\right\}
\end{aligned} \tag{7.36}$$

The corresponding series expansion in cylindrical coordinates reveals the structure better:

$$\begin{aligned}\Phi(\mathbf{r}) = {} & \phi - \frac{r^2}{4}\phi'' + \frac{r^4}{64}\phi^{(4)} - \frac{r^6}{2304}\phi^{(6)} \\ & - r(F_1\cos\varphi + F_2\sin\varphi) + \frac{r^3}{8}(F_1''\cos\varphi + F_2''\sin\varphi) - \frac{r^5}{192}(F_1^{(4)}\cos\varphi + F_2^{(4)}\sin\varphi) \\ & + \frac{r^2}{2}(p_2\cos 2\varphi + q_2\sin 2\varphi) - \frac{r^4}{24}(p_2''\cos 2\varphi + q_2''\sin 2\varphi) + \frac{r^6}{768}(p_2^{(4)}\cos 2\varphi + q_2^{(4)}\sin 2\varphi) \\ & - \frac{r^3}{6}(p_3\cos 3\varphi + q_3\sin 3\varphi) + \frac{r^5}{96}(p_3''\cos 3\varphi + q_3''\sin 3\varphi) \\ & + \frac{r^4}{24}(p_4\cos 4\varphi + q_4\sin 4\varphi) - \frac{r^6}{480}(p_4''\cos 4\varphi + q_4''\sin 4\varphi) \\ & - \frac{r^5}{120}(p_5\cos 5\varphi + q_5\sin 5\varphi) + \frac{r^6}{720}(p_6\cos 6\varphi + q_6\sin 6\varphi)\end{aligned} \tag{7.37}$$

For magnetic fields, we introduce the following special notation for the coefficients:

$$\Pi_0(z) =: B(z), \quad P_0(z) = -\int_{-\infty}^{z} B(z)dz, \quad Q_0 = 0 \tag{7.38}$$

$$P_1(z) =: B_1(z), \quad Q_1(z) =: B_2(z) \tag{7.39}$$

Again P_0 is essentially a scalar axial potential, but here this has little physical meaning. The functions $B(z)$, $B_1(z)$ and $B_2(z)$ are, however, very important, since they represent the axial value of $\boldsymbol{B}(\boldsymbol{r})$:

$$\boldsymbol{B}(0,0,z) = \boldsymbol{i}_x B_1(z) + \boldsymbol{i}_y B_2(z) + \boldsymbol{i}_z B(z) \tag{7.40}$$

The *scalar magnetic potential* is then given by:

$$\begin{aligned}W(\boldsymbol{r}) = {} & -\int B\,dz + \frac{1}{4}(x^2+y^2)B'(z) - \frac{1}{64}(x^2+y^2)^2B'''(z) + \frac{1}{2304}(x^2+y^2)^3B^{(5)} \\ & - xB_1(z) - yB_2(z) + \frac{1}{8}(x^2+y^2)(xB_1'' + yB_2'') - \frac{1}{192}(x^2+y^2)^2(xB_1^{(4)} + yB_2^{(4)}) \\ & + \frac{1}{2}(x^2-y^2)P_2(z) + xyQ_2(z) - \frac{1}{24}(x^4-y^4)P_2'' - \frac{1}{12}(x^3y+xy^3)Q_2'' \\ & + \frac{1}{768}(x^2+y^2)^2\left\{(x^2-y^2)P_2^{(4)} + 2xyQ_2^{(4)}\right\} - \frac{1}{6}x(x^2-3y^2)P_3(z) + \frac{1}{6}y(y^2-3x^2)Q_3(z)\end{aligned}$$

$$+\frac{1}{96}(x^2+y^2)\left\{x(x^2-3y^2)P_3''+y(3x^2-y^2)Q_3''\right\}+\frac{1}{24}(x^4-6x^2y^2+y^4)P_4(z)+\frac{1}{6}(x^3y-xy^3)Q_4(z)$$
$$-\frac{1}{480}(x^2+y^2)\left\{(x^4-6x^2y^2+y^4)P_4''+4xy(x^2-y^2)Q_4''\right\}$$
$$-\frac{1}{120}\left\{x(x^4-10x^2y^2+5y^4)P_5+y(5x^4-10x^2y^2+y^4)Q_5\right\}$$
$$+\frac{1}{720}\left\{(x^2-y^2)(x^4-14x^2y^2+y^4)P_6+2xy(3x^2-y^2)(x^2-3y^2)Q_6\right\} \tag{7.41}$$

or

$$W=-\int B\,dz+\frac{r^2}{4}B'(z)-\frac{r^4}{64}B'''(z)+\frac{r^6}{2304}B^{(5)}$$
$$-r(B_1\cos\varphi+B_2\sin\varphi)+\frac{r^3}{8}(B_1''\cos\varphi+B_2''\sin\varphi)-\frac{r^5}{192}(B_1^{(4)}\cos\varphi+B_2^{(4)}\sin\varphi)$$
$$+\frac{r^2}{2}(P_2\cos 2\varphi+Q_2\sin 2\varphi)-\frac{r^4}{24}(P_2''\cos 2\varphi+Q_2''\sin 2\varphi)+\frac{r^6}{768}(P_2^{(4)}\cos 2\varphi+Q_2^{(4)}\sin 2\varphi)$$
$$-\frac{r^3}{6}(P_3\cos 3\varphi+Q_3\sin 3\varphi)+\frac{r^5}{96}(P_3''\cos 3\varphi+Q_3''\sin 3\varphi)$$
$$+\frac{r^4}{24}(P_4\cos 4\varphi+Q_4\sin 4\varphi)-\frac{r^6}{480}(P_4''\cos 4\varphi+Q_4''\sin 4\varphi)$$
$$-\frac{r^5}{120}(P_5\cos 5\varphi+Q_5\sin 5\varphi)+\frac{r^6}{720}(P_6\cos 6\varphi+Q_6\sin 6\varphi) \tag{7.42}$$

The *vector potential* is given by

$$A_x=-\frac{y}{2}(B-\frac{1}{8}(x^2+y^2)B''+\frac{1}{192}(x^2+y^2)^2B^{(4)})$$
$$-\frac{xy}{2}B_1'+\frac{1}{24}xy(x^2+y^2)B'''-\frac{1}{768}xy(x^2+y^2)^2B_1^{(5)}$$
$$+\frac{1}{4}(x^2-y^2)B_2'-\frac{1}{48}(x^4-y^4)B_2'''+\frac{1}{1536}(x^2+y^2)(x^4-y^4)B_2^{(5)}$$
$$-\frac{1}{12}y(y^2-3x^2)P_2'-\frac{1}{12}x(x^2-3y^2)Q_2'+\frac{1}{192}(x^2+y^2)\left\{y(y^2-3x^2)P_2'''+x(x^2-3y^2)Q_2'''\right\}$$

$$
\begin{aligned}
&+\frac{(x^2+y^2)^2}{7680}\left\{(3x^2-y^2)yP_2^{(5)}-(x^2-3y^2)xQ_2^{(5)}\right\}\\
&-\frac{1}{12}xy(x^2-y^2)P_3'+\frac{1}{48}(x^4-6x^2y^2+y^4)Q_3'\\
&+\frac{1}{960}(x^2+y^2)\left\{4xy(x^2-y^2)P_3'''-(x^4-6x^2y^2+y^4)Q_3'''\right\}\\
&+\frac{1}{240}\left\{y(5x^4-10x^2y^2+y^4)P_4'-x(x^4-10x^2y^2+5y^4)Q_4'\right\}\\
&-\frac{(x^2+y^2)}{5760}\left\{(5x^4-10x^2y^2+y^4)yP_4'''-(x^4-10x^2y^2+5y^4)xQ_4'''\right\}\\
&-\frac{1}{1440}\left\{2xy(3x^4-10x^2y^2+3y^4)P_5'-(x^2-y^2)(x^4-14x^2y^2+y^4)Q_5'\right\}
\end{aligned}
\tag{7.43}
$$

$$
\begin{aligned}
A_y=&\frac{x}{2}[B-\frac{1}{8}(x^2+y^2)B''+\frac{1}{192}(x^2+y^2)^2B^{(4)}]\\
&+\frac{1}{4}(x^2-y^2)B_1'-\frac{1}{48}(x^4-y^4)B_1'''+\frac{1}{1536}(x^2+y^2)(x^4-y^4)B_1^{(5)}\\
&+\frac{xy}{2}B_2'-\frac{xy}{24}(x^2+y^2)B_2'''+\frac{1}{768}xy(x^2+y^2)^2B_2^{(5)}\\
&-\frac{x}{12}(x^2-3y^2)P_2'-\frac{y}{12}(3x^2-y^2)Q_2'+\frac{1}{192}(x^2+y^2)\left\{x(x^2-3y^2)P_2'''+y(3x^2-y^2)Q_2'''\right.\\
&-\frac{(x^2+y^2)^2}{7680}\left\{(x^2-3y^2)xP_2^{(5)}+(3x^2-y^2)yQ_2^{(5)}\right\}\\
&+\frac{1}{48}(x^4-6x^2y^2+y^4)P_3'+\frac{xy}{12}(x^2-y^2)Q_3'\\
&-\frac{1}{960}(x^2+y^2)\left\{(x^4-6x^2y^2+y^4)P_3'''+4xy(x^2-y^2)Q_3'''\right\}\\
&-\frac{x}{240}(x^4-10x^2y^2+5y^4)P_4'-\frac{y}{240}(5x^4-10x^2y^2+y^4)Q_4'\\
&+\frac{(x^2+y^2)}{5760}\left\{(x^4-10x^2y^2+5y^4)xP_4'''+(5x^4-10x^2y^2+y^4)yQ_4'''\right\}\\
&+\frac{1}{1440}(x^2-y^2)(x^4-14x^2y^2+y^4)P_5'+\frac{xy}{720}(3x^4-10x^2y^2+3y^4)Q_5'
\end{aligned}
\tag{7.44}
$$

$$
\begin{aligned}
A_z = &- xB_2(z) + yB_1(z) + \frac{1}{8}(x^2+y^2)(xB_2'' - yB_1'') - \frac{1}{192}(x^2+y^2)^2(xB_2^{(4)} - yB_1^{(4)}) \\
&+ \frac{1}{2}(x^2-y^2)Q_2 - xyP_2 - \frac{1}{24}(x^4-y^4)Q_2'' + \frac{xy}{12}(x^2+y^2)P_2'' \\
&+ \frac{1}{768}(x^2+y^2)^2\left\{(x^2-y^2)Q_2^{(4)} - 2xyP_2^{(4)}\right\} - \frac{x}{6}(x^2-3y^2)Q_3 - \frac{y}{6}(y^2-3x^2)P_3 \\
&+ \frac{1}{96}(x^2+y^2)\left\{x(x^2-3y^2)Q_3'' + y(y^2-3x^2)P_3''\right\} + \frac{1}{24}(x^4-6x^2y^2+y^4)Q_4 - \frac{xy}{6}(x^2-y^2)P_4 \\
&- \frac{1}{480}(x^2+y^2)\left\{(x^4-6x^2y^2+y^4)Q_4'' - 4xy(x^2-y^2)P_4''\right\} \\
&- \frac{1}{120}\left\{x(x^4-10x^2y^2+5y^4)Q_5 - y(5x^4-10x^2y^2+y^4)P_5\right\} \\
&+ \frac{1}{720}\left\{(x^2-y^2)(x^4-14x^2y^2+y^4)Q_6 - 2xy(3x^2-y^2)(x^2-3y^2)P_6\right\}
\end{aligned}
\tag{7.45}
$$

Differentiation using $\boldsymbol{B} = -\nabla W = \nabla \times \boldsymbol{A}$ and truncation of the resulting expressions beyond the fifth-order terms results in

$$
\begin{aligned}
B_x = &-\frac{x}{2}B'(z) + \frac{x}{16}(x^2+y^2)B''' - \frac{x}{384}(x^2+y^2)^2B^{(5)} \\
&+ B_1(z) - \frac{1}{8}(3x^2+y^2)B_1'' + \frac{1}{192}(x^2+y^2)(5x^2+y^2)B_1^{(4)} - \frac{1}{4}xyB_2'' + \frac{1}{48}(x^2+y^2)xyB_2^{(4)} \\
&- xP_2 - yQ_2 + \frac{1}{6}x^3P_2'' + \frac{y}{12}(3x^2+y^2)Q_2'' - \frac{x}{384}(x^2+y^2)(3x^2-y^2)P_2^{(4)} - \frac{y}{384}(x^2+y^2)(5x^2+y^2)Q_2^{(4)} \\
&+ \frac{1}{2}(x^2-y^2)P_3 + xyQ_3 - \frac{1}{96}(5x^4-6x^2y^2-3y^4)P_3'' - \frac{xy}{24}(3x^2+y^2)Q_3'' \\
&- \frac{x}{6}(x^2-3y^2)P_4 + \frac{y}{6}(y^2-3x^2)Q_4 + \frac{x}{240}(3x^4-10x^2y^2-5y^4)P_4'' + \frac{y}{120}(5x^4-y^4)Q_4'' \\
&+ \frac{1}{24}(x^4-6x^2y^2+y^4)P_5 + \frac{xy}{6}(x^2-y^2)Q_5 \\
&- \frac{x}{120}(x^4-10x^2y^2+5y^4)P_6 - \frac{y}{120}(5x^4-10x^2y^2+y^4)Q_6
\end{aligned}
\tag{7.46}
$$

$$
\begin{aligned}
B_y = &-\frac{y}{2}B'(z)+\frac{y}{16}(x^2+y^2)B''' -\frac{y}{384}(x^2+y^2)^2B^{(5)}\\
&+B_2(z)-\frac{1}{8}(3y^2+x^2)B_2''-\frac{1}{4}xyB_1''+\frac{1}{192}(x^2+y^2)(x^2+5y^2)B_2^{(4)}+\frac{xy}{48}(x^2+y^2)B_1^{(4)}\\
&-xQ_2+yP_2-\frac{1}{6}y^3P_2''+\frac{x}{12}(x^2+3y^2)Q_2''-\frac{y}{384}(x^2+y^2)(x^2-3y^2)P_2^{(4)}\\
&-\frac{x}{384}(x^2+y^2)(5y^2+x^2)Q_2^{(4)}\\
&+\frac{1}{2}(x^2-y^2)Q_3-xyP_3-\frac{1}{96}(3x^4+6x^2y^2-5y^4)Q_3''+\frac{xy}{24}(x^2+3y^2)P_3''\\
&-\frac{x}{6}(x^2-3y^2)Q_4-\frac{y}{6}(y^2-3x^2)P_4+\frac{x}{120}(x^4-5y^4)Q_4''-\frac{y}{240}(5x^4+10x^2y^2-3y^4)P_4''\\
&+\frac{1}{24}(x^4-6x^2y^2+y^4)Q_5-\frac{xy}{6}(x^2-y^2)P_5-\frac{x}{120}(x^4-10x^2y^2+5y^4)Q_6+\frac{y}{120}(5x^4-10x^2y^2+y^4)P_6
\end{aligned}
\tag{7.47}
$$

$$
\begin{aligned}
B_z = &B-\frac{1}{4}(x^2+y^2)B''+\frac{1}{64}(x^2+y^2)^2B^{(4)}-\frac{1}{2304}(x^2+y^2)^3B^{(6)}\\
&+xB_1'+yB_2'-\frac{1}{8}(x^2+y^2)(xB_1'''+yB_2''')+\frac{1}{192}(x^2+y^2)^2(xB_1^{(5)}+yB_2^{(5)})\\
&-\frac{1}{2}(x^2-y^2)P_2'-xyQ_2'+\frac{1}{24}(x^2+y^2)\left\{(x^2-y^2)P_2'''+2xyQ_2'''\right\}\\
&-\frac{1}{768}(x^2+y^2)^2\left\{(x^2-y^2)P_2^{(5)}+2xyQ_2^{(5)}\right\}\\
&+\frac{1}{6}x(x^2-3y^2)P_3'-\frac{1}{6}y(y^2-3x^2)Q_3'-\frac{1}{96}(x^2+y^2)\left\{x(x^2-3y^2)P_3''''-y(y^2-3x^2)Q_3'''\right\}\\
&-\frac{1}{24}(x^4-6x^2y^2+y^4)P_4'-\frac{1}{6}xy(x^2-y^2)Q_4'+\frac{1}{480}(x^2+y^2)(x^4-6x^2y^2+y^4)P_4'''\\
&+\frac{1}{120}xy(x^4-y^4)Q_4'''+\frac{1}{120}\left\{x(x^4-10x^2y^2+5y^4)P_5'+y(5x^4-10x^2y^2+y^4)Q_5'\right\}\\
&-\frac{1}{720}\left\{(x^2-y^2)(x^4-14x^2y^2+y^4)P_6'+2xy(3x^2-y^2)(x^2-3y^2)Q_6'\right\}
\end{aligned}
\tag{7.48}
$$

In these formulae, general Cartesian components B_x, B_y, and B_z are clearly distinguished from their axial values B_1, B_2, B by the notation of the subscripts. The expressions for the electric field strength are obtained by interchanging the symbols as follows:

$$\begin{aligned} &B_{1,2} \to F_{1,2}, \quad B \to -\phi', \quad B_{x,y,z} \to E_{x,y,z} \\ &P_j \to p_j, \quad Q_j \to q_j \end{aligned} \tag{7.49}$$

The resulting formulae are not given explicitly here.

7.3 Rotationally Symmetric Fields

In Sections 7.3–7.5, we give the paraxial series expansions for the most important applications to be studied in later chapters in the notation that will be used there.

7.3.1 Electrostatic Fields

Since it describes round lenses, the rotationally symmetric scalar potential field is the most important special case. For electrostatic fields, the following identifications are necessary in Sections 7.1.1 and 7.2.1:

$$m = 0, \quad U_0(z, s) = \Phi(\boldsymbol{r}) =: \Phi(z, r), \quad u_0(z) = \Phi(z, 0) =: \phi(z)$$

With this notation and setting $\phi^{(n)} := d^n\phi/dz^n$, we obtain

$$\begin{aligned} \Phi(z, r) &= \sum_{n=0}^{\infty} \frac{1}{(n!)^2} \left(-\frac{r^2}{4}\right)^n \phi^{(2n)}(z) \\ &= \phi(z) - \frac{r^2}{4}\phi''(z) + \frac{r^4}{64}\phi^{(4)}(z) - \frac{r^6}{2304}\phi^{(6)}(z) + O(r^8) \end{aligned} \tag{7.50}$$

The Cartesian components of $\boldsymbol{E} = -\text{grad}\ \Phi$ are most rapidly obtained by direct differentiation of (7.50)

$$E_z(z, r) = -\phi'(z) + \frac{r^2}{4}\phi^{(3)}(z) - \frac{r^4}{64}\phi^{(5)}(z) + \frac{r^6}{2304}\phi^{(7)}(z) - O(r^8) \tag{7.51}$$

$$E_r = -\frac{r}{2}R_E, \quad E_x = -\frac{x}{2}R_E, \quad E_y = -\frac{y}{2}R_E \tag{7.52a}$$

with

$$R_E(z, r^2) := \phi''(z) - \frac{r^2}{8}\phi^{(4)}(z) + \frac{r^4}{192}\phi^{(6)}(z) - \frac{r^6}{9216}\phi^{(8)}(z) + O(r^8) \tag{7.52b}$$

It is possible to express Eq. (7.50) in closed form as a complex integral (Glaser, 1952)

$$\Phi(z,r)=\frac{1}{2\pi}\int_0^{2\pi}\phi(z+\mathrm{i}r\cos\alpha)d\alpha$$

This has, however, little practical value and does not circumvent the difficulties described above associated with the analytic continuation.

In some electron optical devices, the space charge of the beam is important. We therefore present here the corresponding series expansions. It is convenient to expand the space charge density $\rho(z, r)$ as in Eq. (7.50):

$$\begin{aligned}\varepsilon_0^{-1}\rho(z,r)&=\sum_{n=0}^{\infty}\frac{1}{(n!)^2}\left(-\frac{r^2}{4}\right)^n a_n(z)\\&=a_0(z)-\frac{r^2}{4}a_1(z)+\frac{r^4}{64}a_2(z)-\cdots\end{aligned}\tag{7.53}$$

The coefficients a_n are here *independent* functions of z. Substituting Eq. (7.53) into the axisymmetric form of Poisson's equation,

$$\frac{\partial^2\Phi}{\partial z^2}+\frac{\partial^2\Phi}{\partial r^2}+\frac{1}{r}\frac{\partial\Phi}{\partial r}=-\frac{\rho}{\varepsilon_0}$$

and introducing for Φ a series expansion similar to Eq. (7.50) but with unknown coefficients still to be determined, we find after some elementary calculations

$$\begin{aligned}\Phi(z,r)\quad&=\phi(z)-\frac{r^2}{4}\left\{\phi''+a_0(z)\right\}+\frac{r^4}{64}\left\{\phi^{(4)}+a_0''+a_1(z)\right\}\\&\quad-\frac{r^6}{2304}\left\{\phi^{(6)}+a_0^{(4)}+a_1''+a_2(z)\right\}+\cdots\end{aligned}\tag{7.54}$$

The coefficient $a_0(z)$, representing the most important space charge term, is related to the axial space charge density by $\rho(z, 0) = \varepsilon_0 a_0(z)$.

7.3.2 Magnetic Fields

The series expansions for round magnetic fields are obtained by recognizing that the formulae of Section 6.4 are a special case of those given in Sections 7.1.2 and 7.2.2. We need to consider here only fields in source-free vacuum domains, where a scalar potential W can be employed.

Eqs (6.23), (6.28) and (7.11), (7.12) describe the *same* field in different ways if

$$\Pi(z,r) \equiv C_1(z,s), \quad \Im(D_0(z,s)) = 0$$

so that recalling Eq. (6.31) and setting $m = 1$ in (7.31), we have

$$\Pi(z,0) \equiv C_1(z,0) = \Pi_0(z) = B(z)$$

$B(z)$ being the axial flux density. From Eq. (7.32) we now have the series expansion

$$\begin{aligned}\Pi(z,r) &= \sum_{n=0}^{\infty} \frac{1}{n!(n+1)!}\left(-\frac{r^2}{4}\right)^n B^{(2n)}(z) \\ &= B(z) - \frac{r^2}{8}B''(z) + \frac{r^4}{192}B^{(4)}(z) - \frac{r^6}{9216}B^{(6)}(z) + O(r^8)\end{aligned} \tag{7.55a}$$

Using this, we obtain series expansions for the various components of A and related quantities:

$$A_\varphi = A = \frac{r}{2}\Pi(z,r) \quad \Psi = \pi r^2 \Pi(z,r) \tag{7.55b}$$

$$A_x = -\frac{y}{2}\Pi(z,r), \quad A_y = \frac{x}{2}\Pi(z,r), \quad A_z = 0 \tag{7.55c}$$

The components of $\boldsymbol{B}$ can be obtained in two different but equivalent ways, from $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ and from $\boldsymbol{B} = -\nabla W$. In the latter case we have to start from Eq. (7.30) with $W = D_0(z, r^2)$. In both cases, we arrive at

$$B_z(z,r) = B(z) - \frac{r^2}{4}B''(z) + \frac{r^4}{64}B^{(4)}(z) - \frac{r^6}{2304}B^{(6)}(z) + O(r^8) \tag{7.56a}$$

$$B_r = -\frac{r}{2}R_M, \quad B_x = -\frac{x}{2}R_M, \quad B_y = -\frac{y}{2}R_M \tag{7.56b}$$

with $R_M = \partial\Pi/\partial z$ given by

$$R_M(z,r^2) = B'(z) - \frac{r^2}{8}B'''(z) + \frac{r^4}{192}B^{(5)}(z) - \frac{r^6}{9216}B^{(7)}(z) + O(r^8) \tag{7.56c}$$

These expressions are of particular interest in connection with the physics of round lenses.

For completeness, we state the series expansions of the scalar potential:

$$W(z,r) = -\int B\,dz + \frac{r^2}{4}B' - \frac{r^4}{64}B''' + \frac{r^6}{2304}B^{(5)} - O(r^8) \tag{7.57}$$

On comparing all these series expansions, there are seen to be only two sets of denominators, which appear frequently in different contexts.

7.4 Multipole Fields

In electron optics the meaning of this term is slightly different from that familiar in electrodynamics. In the present context, we do not consider series expansions of fields in terms of spherical harmonics but only those in terms of the azimuth in cylindrical coordinates, as given by Eqs. (7.5) and (7.9). Multipole fields are then those that have well-defined symmetry properties with respect to the azimuth φ, as is illustrated schematically in Fig. 7.1.

In practice, such fields are often created by a suitable configuration of electrodes and polepieces, their major axes being parallel to the optic axis, see Fig. 7.2. Since the extent of these elements must be finite, fringe fields are inevitable. It is thus impossible to create 'pure' multipole fields in the sense that their dependence on the azimuth φ corresponds to a single harmonic (or finite number of them). In practice this is of no consequence; the only essential requirement is that a well-defined symmetry exists. Consequently each physical multipole field consists of a superposition of different harmonics having the same symmetry

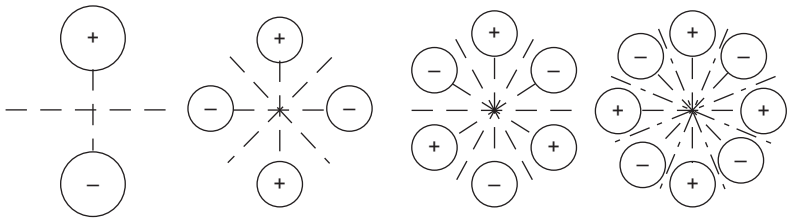


Figure 7.1
From left to right, multipoles of order $m = 1$ (dipole), $m = 2$ (quadrupole), $m = 3$ (sextupole) and $m = 4$ (octopole). The optic axis is always perpendicular to the plane of the diagram.

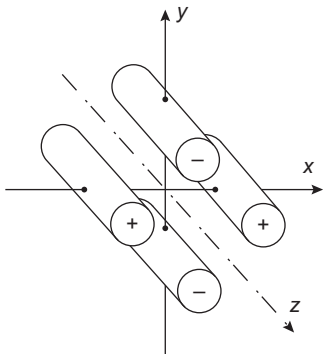


Figure 7.2
Simplified diagram of a real quadrupole (see also Fig. 19.1).

properties. The field is then classified by its lowest order harmonic component. Here we state explicitly the electrostatic multipole potentials of the lowest orders $m = 1$ and $m = 2$.

The *dipole field* ($m = 1$) is characterized by having only one plane of even symmetry and a perpendicular one of odd symmetry. Commonly, two such fields, rotated at 90° with respect to each other, are superimposed; the potential is then given up to terms in r^5 by

$$\begin{aligned} -\Phi_D = {} & r\cos\varphi\left\{F_1(z) - \frac{r^2}{8}F_1''(z) + \frac{r^4}{192}F_1^{(4)}(z)\right\} + \frac{1}{6}r^3\cos 3\varphi\left\{p_3(z) - \frac{r^2}{16}p_3''(z)\right\} \\ & + \frac{1}{120}r^5\cos 5\varphi\, p_5(z) + O(r^7) + r\sin\varphi\left\{F_2(z) - \frac{r^2}{8}F_2''(z) + \frac{r^4}{192}F_2^{(4)}(z)\right\} \\ & + \frac{1}{6}r^3\sin 3\varphi\left\{q_3(z) - \frac{r^2}{16}q_3''(z)\right\} + \frac{1}{120}r^5\sin 5\varphi\, q_5(z) + O(r^7) \end{aligned} \tag{7.58}$$

Such fields are employed in *deflection units*, see Chapter 32, Paraxial Properties of Deflection Systems and Chapter 40 of Volume 2.

The *quadrupole field* ($m = 2$) is characterized by two perpendicular planes of even symmetry and two planes of odd symmetry, inclined at 45° relative to the former. More generally, two such fields, inclined at 45° with respect to each other, may be superimposed. The electrostatic potential is then given up to terms in r^6 by

$$\begin{aligned} \Phi_Q = {} & \frac{1}{2}r^2\cos 2\varphi\left\{p_2(z) - \frac{r^2}{12}p_2''(z) + \frac{r^4}{384}p_2^{(4)}(z)\right\} + \frac{1}{2}r^2\sin 2\varphi\left\{q_2(z) - \frac{r^2}{12}q_2''(z) + \frac{r^4}{384}q_2{}^{(4)}(z)\right\} \\ & + \frac{1}{720}r^6\left\{p^6(z)\cos 6\varphi + q_6(z)\sin 6\varphi\right\} + O(r^8) \end{aligned} \tag{7.59}$$

Fields of this type occur in multiplets of quadrupole lenses (Chapter 29, The Aberrations of Quadrupole Lenses and Octopoles) and in stigmators (Chapter 32, Paraxial Properties of Deflection Systems).

The radial series expansions of the multipole fields of higher multiplicity ($m > 2$) can be terminated after the nonvanishing term of lowest order. The reason for this is that, within the paraxial domain, these fields represent only weak perturbations or corrections. The potential is then given by

$$\Phi_m(r) = \frac{(-r)^m}{m!}\left\{p_m(z)\cos m\varphi + q_m(z)\sin m\varphi\right\} \tag{7.60}$$

The most important of these in practice are the sextupole and octopole fields. So far we have given explicit expressions only for the electrostatic potential Φ. The corresponding magnetostatic potential W is obtained by replacing the symbols as follows:

$$\Phi \to W, \quad F_1 \to B_1, \quad F_2 \to B_2, \quad p_j \to P_j, \quad q_j \to Q_j \tag{7.61}$$

for each subscript j.

7.5 Planar Fields

In Section 6.5 we introduced planar solutions of Laplace's equation as analytic functions of a complex variable $z + ix$. This slightly unorthodox choice was adopted for the purposes of electron optics, where the z-axis is almost always made to coincide with the optic axis. We now reconsider planar fields in the (z, x)-plane, confining the discussion to fields with well-defined symmetry properties.

The power series expansion of potentials with *odd* mirror symmetry with respect to the plane $x = 0$ is given by

$$\begin{aligned} -\Phi(z,x) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} F^{(2n)}(z) \\ &= x\left\{F(z) - \frac{x^2}{6}F''(z) + \frac{x^4}{120}F^{(4)}(z) + O(x^6)\right\} \end{aligned} \tag{7.62}$$

It is easily verified that this is a solution of Laplace's equation. The function $F(z) = -\partial\Phi/\partial x$, $(x = 0)$, has the meaning of an axial field strength.

Fields of this type are approximately realized in the electric deflection units of oscillographs (Chapter 32, Paraxial Properties of Deflection Systems) and (as the analogous magnetic potential) in the fringe-field domains of sector magnets with plane fronts (see Chapter 52 of Volume 2). Eq. (7.62) can be transformed into a special case of (7.58), the nonzero axial harmonics then being

$$p_1(z) \equiv F(z), \quad p_3(z) = -\frac{1}{4}F''(z), \quad p_5(z) = \frac{1}{16}F^{(4)}(z) \tag{7.63}$$

This shows that the planar deflection field is a special type of dipole field.

The paraxial series expansion of potentials with *even* mirror symmetry with respect to the plane $x = 0$ is given by

$$\Phi(z,x) = \phi(z) - \frac{x^2}{2}\phi''(z) + \frac{x^4}{24}\phi^{(4)}(z) - \frac{x^6}{720}\phi^{(6)}(z) + O(x^8) \tag{7.64}$$

$\varphi(z)$ being the potential in the symmetry plane. These planar fields are a special case of multipole fields with $p_2 = -\phi''/2$, $p_4 = \phi^{(4)}/8$. They occur in electrostatic slit lenses.

7.6 Fourier–Bessel Series Expansions

In Section 7.1 we introduced azimuthal Fourier series expansions with coefficient functions depending on z and r. The evaluation of the paraxial series expansions, derived in the subsequent sections, is not the only way of calculating the coefficient functions. An alternative procedure is to separate the general scalar potential $V(\boldsymbol{r})$ into two functions, one in z only, the other in r only:

$$V_m(z, r) = r^m U_m(z, r^2) =: Z_m(z) R_m(r) \tag{7.65}$$

When this is introduced—together with a factor $\exp(\mathrm{i}m\varphi)$—into Laplace's equation, ordinary differential equations are obtained:

$$Z_m''(z) + k^2 Z_m(z) = 0 \tag{7.66a}$$

$$R_m''(z) + \frac{1}{r} R_m'(z) - (k^2 + \frac{m^2}{r^2}) R_m(r) = 0 \tag{7.66b}$$

The separation constant here has an arbitrary positive value k^2. The general solution of (7.66a) is

$$Z_m(z) = C_m(k) \mathrm{e}^{\mathrm{i}kz}, \quad (-\infty < k < \infty) \tag{7.67a}$$

$C_m(k)$ being any regular function of k. Eq. (7.66b) is the differential equation for modified Bessel functions, its regular solution being given by

$$R_m(r) = I_m(kr) \tag{7.67b}$$

Putting all this together, we obtain a solution of Laplace's equation in the form of a Fourier-Bessel series expansion:

$$V(z, r, \varphi) = \sum_{m=0}^{\infty} \Re \left\{ \mathrm{e}^{\mathrm{i}m\varphi} \int_{-\infty}^{\infty} C_m(k) \mathrm{e}^{\mathrm{i}kz} I_m(kr) dk \right\} \tag{7.68}$$

The paraxial series expansion can now be obtained by introducing the well-known Taylor series expansion

$$I_m(x) = \left(\frac{x}{2}\right)^m \sum_{n=0}^{\infty} \frac{(x^2/4)^n}{n!(m+n)!}, \quad (m = 0, 1, 2, \ldots) \tag{7.69}$$

with $x = kr$ into Eq. (7.68). The resulting expression will not be given here. The main difference between it and (7.27) is that repeated differentiations are replaced by Fourier integrals.

The most important special case is that of rotational symmetry, $m = 0$, Eq. (7.68) then simplifying to

$$V(z, r) = \int_{-\infty}^{\infty} C_0(k) \mathrm{e}^{\mathrm{i}kz} I_0(kr) dk \tag{7.70}$$

The reality of this expression is guaranteed by requiring that

$$C_0^*(-k) = C_0(k) \tag{7.71}$$

This Fourier coefficient $C_0(k)$ is the Fourier transform of the axial potential $u_0(z)$. The relation between differentiations and Fourier transforms is very simple here:

$$u_0^{(\nu)}(z) = \int_{-\infty}^{\infty} (\mathrm{i}k)^{\nu} C_0(k) \mathrm{e}^{\mathrm{i}kz}\, dk \tag{7.72}$$

Furthermore, it is easy to determine $C_0(k)$ from the boundary values $V(z, a)$ of the potential on the surface $r = a$ of an infinitely long cylinder. Applying the inverse Fourier transform to Eq. (7.70) with $r = a$, we obtain rapidly

$$C_0(k) = \left\{ 2\pi I_0(ka) \right\}^{-1} \int_{-\infty}^{\infty} V(z, a) \mathrm{e}^{-\mathrm{i}kz} dz \tag{7.73}$$

This expression satisfies (7.71).

Historically, Eqs (7.70) and (7.73) have played an important role in the development of simple analytic field models for round electron lenses. A few details are given in Chapters 35 and 36 of Volume 2. Nowadays, interest in these models has dwindled, for it is as easy to calculate field distributions exactly as to match parameters to a model. One or two models remain useful for teaching purposes and to gain a rapid qualitative understanding of the dependence of the properties of some device on various parameters. Nevertheless, the Fourier–Bessel series expansion is still of some interest. For instance, van der Merwe (1978a–c, 1979, 1980) used it to calculate rotationally symmetric lenses and Franzen (1984) applied it to quadrupole lenses in cathode-ray tubes.

CHAPTER 8

Boundary-Value Problems

Hitherto we have concentrated on fields in the extended paraxial domain. We are in a position to carry out the field calculation once a sequence of axial harmonics $u_m(z)$, $m = 0, 1, 2, \ldots$ is known, but these functions are so far unspecified. The potential inside a domain of solution is specified by its boundary values at the surfaces of this domain and by its source distribution. We elaborate on this in the following sections.

8.1 Boundary-Value Problems in Electrostatics

In electron optics, the electric fields inside insulators and in current-carrying metal conductors are of very little interest and will not be considered here. The domain of solution is the vacuum part of the device in question. This may be multiply connected but it always contains the optic axis. Its boundary is formed by the surfaces of all surrounding metallic electrodes or at least by relevant parts of these. It may prove to be convenient to assume that parts of the boundary are located in the vacuum and even to extend these to infinity, though this is clearly an idealization.

In almost all cases of practical interest, the electric field exhibits simplifying symmetry properties, since a completely unsymmetric field serves no practical purpose. Imperfections in the machining of the electrodes will not be considered here; this topic is treated in Section 9.4.6 and Chapter 31, Parasitic Aberrations. Any symmetry properties of the field can be exploited to reduce the relevant domain Γ of solution; the field obtained is subsequently completed by means of symmetry operations.

These remarks are illustrated in the example shown in Fig. 8.1. The appropriate choice of the domain Γ and its boundary $\partial\Gamma$ does, of course, depend strongly on the particular properties of the device in question.

Whenever it is sufficient to consider the field in a planar axial section through the device, we shall adopt the notation presented in Fig. 8.2. The vectors $\boldsymbol{n}$ and $\boldsymbol{t}$ are unit vectors. The surface normal $\boldsymbol{n}$ is directed outwards from medium 1, even in the general three-dimensional case. The contour of the boundary in the axial section will always be oriented positively in the sense shown in Figs 8.1 and 8.2. This choice will be adopted throughout this Part. In the case of electrostatic fields, the medium 1 will be identical with the domain Γ of the desired solution.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00008-0
© 2018 Elsevier Ltd. All rights reserved.

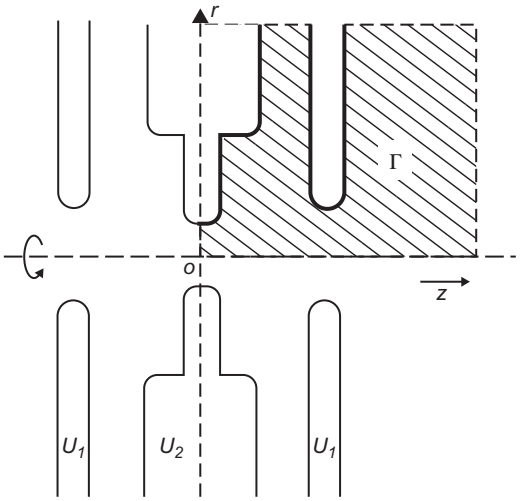


Figure 8.1
Round symmetric electrostatic lens. The domain Γ within which the solution is sought can be confined to the vacuum region for which $z \geq 0$, $r \geq 0$. Whether the domain Γ must be closed or can be extended to infinity depends on the method of calculation.

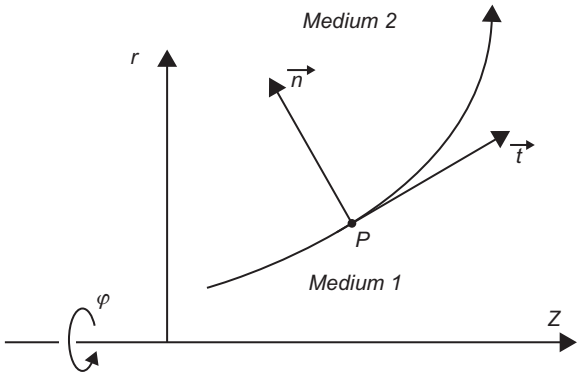


Figure 8.2
Unit vectors normal to the surface ($\boldsymbol{n}$) and tangent to it in the meridional section ($\boldsymbol{t}$) at an arbitrary point P of the boundary. The unit vector $\boldsymbol{i}_\varphi$ is perpendicular to the meridional section shown and satisfies $\boldsymbol{i}_\varphi = \boldsymbol{t} \times \boldsymbol{n}$.

The boundary-value problem for the electrostatic field is now defined in the following manner:

1. At all electrode surfaces the electrostatic potentials $\Phi(r)$ must have a constant value equal to the known potential of the corresponding electrode.

2. If the boundary consists of several separate parts located in the vacuum, the corresponding surface potential is generally not constant and is to be defined reasonably, for instance by means of linear, quadratic or logarithmic interpolations. This situation often arises in narrow gaps between electrodes.
3. If a plane of negative mirror symmetry of the potential forms a part of the boundary, the potential over this plane is constant, usually zero.
4. At all infinite parts of the boundary, the potential is constant. These constant values must be chosen consistently.
5. At all planes or at an axis of positive mirror symmetry, the normal component of the field strength (normal derivative of the potential) vanishes. The optic axis in every rotationally symmetric device is certainly such an axis.

The boundary-value problem, specified in this manner, has a unique solution, and later we shall describe computational methods for obtaining this solution. In this context, the following relations are very useful. Since all electrode surfaces must be equipotentials, the field strength on their vacuum side is given by

$$\boldsymbol{E}(\boldsymbol{r}) = -\frac{1}{\varepsilon_0}\sigma(\boldsymbol{r})\boldsymbol{n}(\boldsymbol{r}) \tag{8.1}$$

Its magnitude is then

$$-\boldsymbol{E}\cdot\boldsymbol{n} = \frac{\partial\Phi}{\partial n} = \sigma(\boldsymbol{r})/\varepsilon_0 \tag{8.2}$$

Throughout this Part the symbol $\partial/\partial n = \boldsymbol{n}\cdot\nabla$ denotes the familiar normal derivative. The function $\sigma(\boldsymbol{r})$, defined for all metallic surfaces, is the surface charge density. Initially, this function is unknown, but once the boundary-value problem has been solved it may be used with advantage in the ensuing field computations, see Chapter 9, Integral Equations.

8.2 Boundary Conditions in Magnetostatics

Whereas the material properties of the metallic electrodes are unimportant in electrostatics, since the electrostatic field vanishes inside any conductor, the situation in magnetostatics is far more complicated. Apart from the case of perfect superconductors, the magnetic field inside polepieces does not vanish. In consequence, it is not always possible to confine the domain of solution to the vacuum part of the field. It is of course this region that is of greatest interest for calculating the optical properties, but a knowledge of the field distribution in the yoke is often needed when the shape of the latter is being designed. In the most general case, the field computation will become extremely complicated; instead of treating this, therefore, we shall consider some classes of important field configurations

with simplifying properties. First, however, we shall formulate the general boundary conditions, which must always be satisfied.

We shall use the notation explained in Fig. 8.2; here medium 1 is the vacuum while medium 2 is any ferromagnetic or superconducting material. In the general case, at all surfaces of materials, the *interface conditions*

$$\boldsymbol{n}\cdot(\boldsymbol{B}_2-\boldsymbol{B}_1)=0 \tag{8.3}$$

$$\boldsymbol{n}\times(\boldsymbol{H}_2-\boldsymbol{H}_1)=\boldsymbol{\omega}(\boldsymbol{r}) \tag{8.4}$$

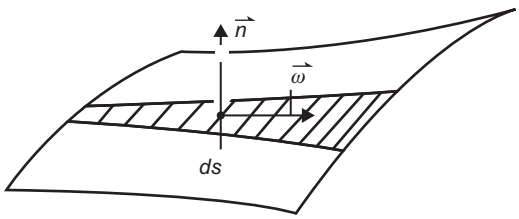
must be satisfied, the subscript referring to the material in which the field vectors are defined. The function $\boldsymbol{\omega}(\boldsymbol{r})$ is the *surface current density*. This is a vector function defined only on surfaces. It must always have the same direction as the local tangent: $\boldsymbol{\omega}(\boldsymbol{r})=\omega(\boldsymbol{r})\boldsymbol{t}_c(\boldsymbol{r})$, $\boldsymbol{t}_c(\boldsymbol{r})$ being a normalized tangential vector which may differ from the vector $\boldsymbol{t}$ introduced earlier. The physical meaning is as follows: $dI=\omega(\boldsymbol{r})ds$ is the electric current flowing through a surface line-element ds oriented perpendicularly to $\boldsymbol{\omega}(\boldsymbol{r})$, see Fig. 8.3. The whole distribution of surface currents must, of course, satisfy the requirements for the conservation of electric current.

Such surface current distributions arise in superconducting devices. They are caused by induction effects, when the field in the vacuum domain is switched on. Furthermore, surface current distributions offer a convenient way of describing flat layers of current-conducting windings located in the vicinity of the surfaces of magnetic shielding tubes. Further examples are given below.

In very many cases, the function $\boldsymbol{\omega}(\boldsymbol{r})$ vanishes identically, Eq. (8.4) then simplifying to

$$\boldsymbol{n}\times(\boldsymbol{H}_2-\boldsymbol{H}_1)=0 \tag{8.5}$$

Eq. (8.3) expresses the continuity of the normal component of $\boldsymbol{B}$, while Eq. (8.5) implies that the tangential component of $\boldsymbol{H}$ is continuous. Even when saturation effects occur, we



Figuer 8.3

Representation of a surface current distribution; the direction of the vector $\boldsymbol{\omega}$ is the same as that of the tangent. The current increment $dI=\boldsymbol{\omega}\cdot\boldsymbol{ds}$ is constant along the stripe indicated.

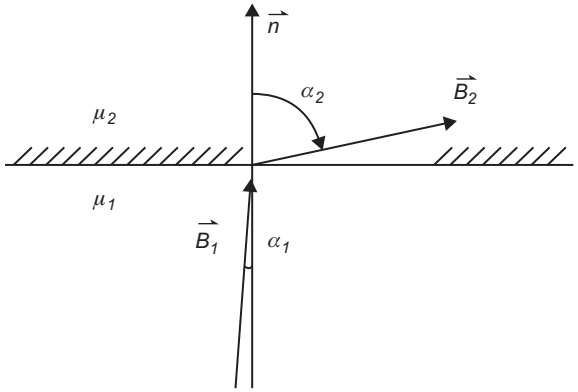


Figure 8.4
Refraction of lines of magnetic flux for $\mu_2/\mu_1 = 50$.

may introduce material coefficients (Eq. 6.3 or 6.19). Writing $B_j = \mu_j H_j$, $\mu_j = 1/\nu_j$ for $j = 1$, 2, we can now derive the familiar law for the refraction of flux lines (Fig. 8.4):

$$\frac{\tan \alpha_2}{\tan \alpha_1} = \frac{\mu_2}{\mu_1} \tag{8.6}$$

The continuity laws break down at sharp edges, where no local surface normal $\boldsymbol{n}$ can be defined. We shall therefore assume that such edges are slightly rounded off, as is the case in all practical devices.

These interface conditions are very simple but refer to vector fields. The computation of vector fields is possible in principle but is usually complicated. It is therefore advantageous to use *scalar* potentials from which the field can be determined by differentiation. Unfortunately, the scalar potentials $\chi(\boldsymbol{r})$ and $\Pi(\boldsymbol{r})$, introduced in Eqs (6.11 and 6.12) and (6.28) respectively, are of only very limited applicability. In order to circumvent this difficulty, it is usual to separate the magnetic field strength $\boldsymbol{H}(\boldsymbol{r})$ into the contribution $\boldsymbol{H}_0(\boldsymbol{r})$ of the isolated coils in vacuum and the contribution $\boldsymbol{H}_M(\boldsymbol{r})$ of the ferromagnetic parts,

$$\boldsymbol{H}(\boldsymbol{r}) = \boldsymbol{H}_0(\boldsymbol{r}) + \boldsymbol{H}_M(\boldsymbol{r}) \tag{8.7}$$

By definition the following conditions are to be satisfied in the whole space:

$$\operatorname{div} \boldsymbol{H}_0(\boldsymbol{r}) = 0, \quad \operatorname{curl} \boldsymbol{H}_0(\boldsymbol{r}) = \boldsymbol{j}(\boldsymbol{r}) \tag{8.8}$$

Together with the natural boundary conditions at infinity, this is already sufficient to calculate $\boldsymbol{H}_0(\boldsymbol{r})$ uniquely by means of Biot–Savart's law:

$$\boldsymbol{H}_0(\boldsymbol{r}) = \int \frac{(\boldsymbol{r}' - \boldsymbol{r}) \times \boldsymbol{j}(\boldsymbol{r}')}{4\pi |\boldsymbol{r}' - \boldsymbol{r}|^3} d^3 r' \tag{8.9}$$

This function can be obtained by differentiation of potentials. In the general case, the vector potential is appropriate:

$$\boldsymbol{H}_0 = \frac{1}{\mu_0}\,\mathrm{curl}\,\boldsymbol{A}_0, \quad \boldsymbol{A}_0(\boldsymbol{r}) = \frac{\mu_0}{4\pi}\int \frac{\boldsymbol{j}(\boldsymbol{r}')d^3r'}{|\boldsymbol{r}'-\boldsymbol{r}|} \tag{8.10}$$

In simply connected domains excluding any sources, we may also use Eqs (6.11 and 6.12). For a single closed winding carrying an electric current I, this representation takes the familiar form

$$\boldsymbol{H}_0 = -\,\mathrm{grad}\,\chi_0, \quad \chi_0(\boldsymbol{r}) = \frac{I}{4\pi}\int_S \frac{(\boldsymbol{r}-\boldsymbol{r}')\cdot d\boldsymbol{a}'}{|\boldsymbol{r}-\boldsymbol{r}'|^3} \tag{8.11}$$

The two-dimensional surface integral is to be evaluated over any surface S enclosed by the windings. An example of this is shown in Fig. 8.5. Since the integral in Eq. (8.11) is equal to the solid angle Ω under which the winding would be seen from the point $\boldsymbol{r}$ (see Fig. 8.5), $\chi_0(\boldsymbol{r})$ is known as the 'solid angle potential'. As the point $\boldsymbol{r}$ passes through the surface S, $\chi_0(\boldsymbol{r})$ varies discontinously, the jump being $\pm I$. In the case of several closed windings the contributions of all the windings are to be summed up appropriately. Important practical applications of this integral are given in Chapter 40 of Volume 2.

The representation (8.11) is most convenient in the extended paraxial domain, since it is always possible to define the surfaces S of integration in such a way that Eq. (8.11) is unique in this domain. Apart from some special cases, Eq. (8.10) is less convenient. The evaluation of Eq. (8.9) is always possible. In the case of surface current distributions,

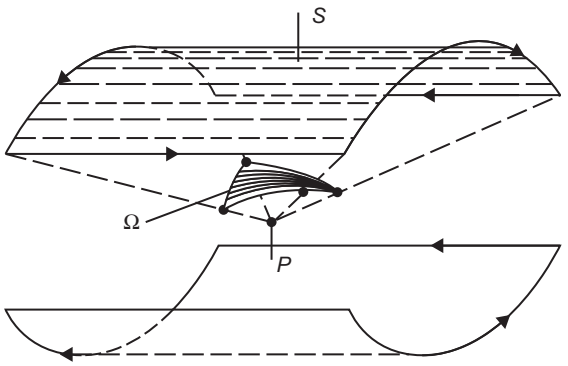


Figure 8.5

Pair of saddle coils. The arrows indicate the local direction of the electric current. The solid angle subtended by the upper surface at some arbitrary point P is denoted by Ω. The solid angle subtended at P by the lower surface can likewise be found and is to be subtracted from Ω.

the three-dimensional integration collapses to a surface integration, which means that $\boldsymbol{j}(\boldsymbol{r}')d^3r'$ is to be replaced by $\omega(\boldsymbol{r}')da'$. In every case, the resulting function $\boldsymbol{H}_0(\boldsymbol{r})$ is unique in the whole space. In the following considerations we shall assume that these integrations can be carried out for all points of reference $\boldsymbol{r}$ that are to be considered.

Let us now focus our attention on the second term $\boldsymbol{H}_M(\boldsymbol{r})$ in Eq. (8.7). Since the whole current density $\boldsymbol{j}$ is already associated with $\boldsymbol{H}_0$, the field $\boldsymbol{H}_M$ must satisfy curl $\boldsymbol{H}_M = 0$, and hence

$$\boldsymbol{H}_M(\boldsymbol{r}) = -\text{grad}\ \chi_M(\boldsymbol{r}) \tag{8.12}$$

$\chi_M(\boldsymbol{r})$ being a *unique* scalar potential in the whole space and defined as the *reduced* magnetic scalar potential. In every unsaturated medium ($\mu = \text{const}$) it can be concluded from div $\boldsymbol{B} = 0$, div $\boldsymbol{H}_0 = 0$ and div $\boldsymbol{H} = 0$ that div $\boldsymbol{H}_M = 0$ so that

$$\nabla^2\chi_M(\boldsymbol{r}) = 0 \quad \text{for} \quad \mu = \text{const} \tag{8.13}$$

This is invalid at the material surfaces, where formal scalar surface charge distributions must be introduced; these are the analogue of electrostatic surface charges. Magnetic surface charges, however, have no physical meaning, but are only a convenience in calculating, as will become obvious in Section 9.2.

The interface conditions (8.3), (8.4) and (8.5) are considerably simplified by the separation in Eq. (8.7) combined with (8.12). Since $\chi_M(\boldsymbol{r})$ must be a unique function in the whole space, this potential itself and the tangential components of its gradient must be continuous at all material surfaces, while the normal component will be discontinuous. The field contribution $\boldsymbol{H}_0$ has the opposite behaviour: its normal component is continuous at interfaces, while the tangential components are discontinuous, the corresponding jump being obtained from Eq. (8.4):

$$\boldsymbol{n} \times (\boldsymbol{H}_0)_2 - (\boldsymbol{H}_0)_1 = \boldsymbol{\omega} \tag{8.14}$$

This jump is already considered in the Biot–Savart integration over the surface currents and hence Eqs (8.4) and (8.5) contain no further information.

Introducing Eqs (8.7) and (8.12) into (8.13) and using the fact that

$$\boldsymbol{B}_j = \mu_j\boldsymbol{H}_j = \mu_j(\boldsymbol{H}_0 - \nabla\chi_M)_j \quad (j = 1, 2)$$

on both sides of the corresponding interface, we obtain the fundamental interface condition

$$\mu_2\left(\frac{\partial\chi_M}{\partial n}\right)_2 - \mu_1\left(\frac{\partial\chi_M}{\partial n}\right)_1 = (\mu_2 - \mu_1)\,\boldsymbol{n}\cdot\boldsymbol{H}_0 \tag{8.15}$$

The implementation of this condition will be examined in Section 9.2.

8.3 Examples of Boundary-Value Problems in Magnetostatics

In this section we shall confine our considerations to important classes of boundary-value problems in the proper sense, by which we mean cases in which it is possible to confine the necessary field calculation entirely to *one* medium (1). The subscript 1 will then be omitted when this does not cause confusion, and we set $\mu_1 = \mu_0$ when medium (1) is the vacuum.

8.3.1 Devices with Superconducting Yokes

Owing to the Meissner–Ochsenfeld effect, the magnetic field is completely expelled from the interior of any superconductor. On the vacuum side of its surface the magnetic field must have a locally tangential direction, see Fig. 8.6. This condition can be satisfied only by the presence of appropriate surface current distributions $\boldsymbol{\omega}(\boldsymbol{r})$, which are *unknown* prior to the solution of the corresponding boundary-value problem. Since $\boldsymbol{\omega}(\boldsymbol{r})$ must be known in order to calculate $\boldsymbol{H}_0(\boldsymbol{r})$, the separation (8.7) is unhelpful in this context.

In rotationally symmetric devices, the necessary boundary conditions are most simply satisfied by use of the flux potential Ψ. From Eq. (6.32) it is obvious that $\boldsymbol{B}$ is tangential if the boundary contour C is a line $\Psi = \text{const}$. The Dirichlet problem for $\Psi(z, r)$ is then very simple:

$$\begin{aligned} &\Psi(z,r) = \Psi_B = \text{const}, && (z,r) \in C \\ &\Psi(z,0) = 0, && -\infty < z < \infty \end{aligned} \tag{8.16a}$$

For the field in the paraxial domain, the potential $\Pi(z, r)$ is more suitable. From Eqs (6.27), (6.28) and from the regularity requirements at the optic axis, the boundary conditions

$$\begin{aligned} &\Pi(z,r) = \Psi_B/\pi r^2 && (z,r) \in C \\ &\partial\Pi/\partial r = 0 && \text{for } r = 0 \end{aligned} \tag{8.16b}$$

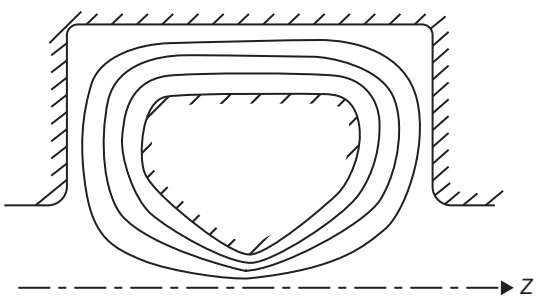


Figure 8.6
Magnetic flux lines in a superconducting device; the upper half of an axial section is shown.

can be derived. The solution of the corresponding boundary-value problem for Π is not unduly complicated. The constant Ψ_B has the physical meaning of the total flux through the bore of the superconducting polepiece.

An analogous Dirichlet problem can be formulated for planar fields but this will not be treated here since the simplification to planar field structures is generally not satisfactory in superconducting devices.

8.3.2 Conventional Round Magnetic Lenses

Fig. 8.7 shows an axial section through a typical magnetic lens, and Fig. 8.8 the relevant vacuum domain of the magnetic field. The contours of the casing are schematically simplified. It is only approximately possible to confine the field calculation to the domain Γ; for this the following assumptions must be made:

1. The permeability of the casing material must be extremely high, $\mu_2 \geq 10^4 \mu_0$, and saturation effects must nowhere occur.
2. The cross-section of the casing must be large enough to ensure that practically all the magnetic flux flows through the gap.
3. The gap has to be long (in the radial direction) and narrow, so that the field between the pole faces may be regarded as practically homogeneous.

These assumptions cannot, of course, be satisfied precisely in a mathematical sense. The following considerations represent a technical simplification.

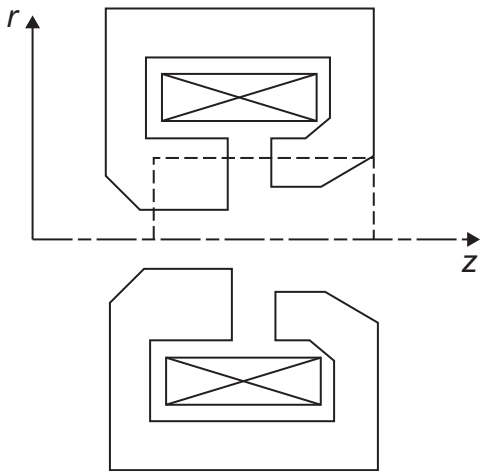


Figure 8.7
Highly simplified axial section through a conventional magnetic lens.

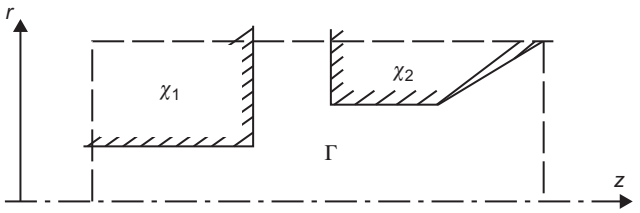


Figure 8.8
Enlarged view of the region of Fig. 8.7 enclosed in dashed lines; the scalar potential $\chi(\boldsymbol{r})$ is defined within the domain Γ.

Since the currents in the coils are located completely outside the domain Γ, we may use the total scalar potential $\chi(\boldsymbol{r})$. From Eq. (8.6) we can conclude that the angle α_1 on the vacuum side of the casing surfaces must be extremely small. Thus the flux lines intersect these surfaces practically *orthogonally*. This implies that the surfaces may be regarded as equipotentials $\chi(\boldsymbol{r}) = \text{const}$. One surface potential χ_1 may be chosen arbitrarily, for instance $\chi_1 = 0$. The other is determined by Ampère's law

$$\oint \boldsymbol{H} \cdot d\boldsymbol{r} = NI = \chi_1 - \chi_2 \tag{8.17}$$

In this relation *NI* is the total number of ampère-turns of the coil; the integration loop must enclose all the windings and must pass through the gap.

The boundary-value problem to be solved now takes the following form:

1. Inside Γ the potential $\chi(z, r)$ is to be calculated by solving $\nabla^2\chi = 0$.
2. At the surfaces of the polepieces and in asymptotic regions of the bores the potential is constant, χ_1 or χ_2 respectively.
3. At the upper part of the boundary, inside the gap, the potential is to be interpolated linearly.
4. On the optic axis, $\partial\chi/\partial r = 0$ must be satisfied.

The simplifying assumptions reduce this boundary-value problem to the analogue of an electrostatic problem and it can be solved by means of corresponding techniques. As in Eqs (8.1) and (8.2), it will be convenient to introduce formal magnetic surface charge densities, satisfying

$$\boldsymbol{H}(\boldsymbol{r}) = -\sigma_M(\boldsymbol{r})\boldsymbol{n}(\boldsymbol{r}) \tag{8.18}$$

$$-\boldsymbol{H} \cdot \boldsymbol{n} = \partial\chi/\partial n = \sigma_M(\boldsymbol{r}) \tag{8.19}$$

Another problem that can be solved by employing the potential $\chi(\boldsymbol{r})$ is the fringe field of sector magnets with screening plates (Part X of Volume 2).

8.3.3 Unconventional Round Magnetic Lenses

The approximations described above are of limited application and may break down even when comparatively simple configurations are considered. A typical example is shown in Fig. 8.9. The 'gap' in the shield of this lens is so wide that a simple linear interpolation is unreasonable and the distribution of the electric current in the coil is now of importance. This is clearly a case for the flux potential $\Psi(z, r)$, since the interface conditions are then strongly simplified.

The continuity of the magnetic flux requires that $\Psi(z, r)$ be continuous at any material surface not conducting surface currents. Consequently the tangential component of $\nabla\Psi$ is also continuous. Hence Eq. (8.3) is already satisfied: introducing Eq. (6.32) into (8.3) and recalling that $\boldsymbol{n} \times \boldsymbol{i}_\varphi = \boldsymbol{t}$ for the tangential vector $\boldsymbol{t}$ (Fig. 8.2), we do indeed obtain

$$(\boldsymbol{t} \cdot \nabla\Psi)_2 = (\boldsymbol{t} \cdot \nabla\Psi)_1$$

The second interface condition is obtained by introducing Eq. (6.32) into (8.5). After some elementary calculations, we find

$$\nu_2 \left(\frac{\partial \Psi}{\partial n} \right)_2 = \nu_1 \left(\frac{\partial \Psi}{\partial n} \right)_1 \tag{8.20}$$

In the case of an unsaturated casing with extremely high permeability, $\mu_2 \gg \mu_1$, $\nu_2 \ll \nu_1$, it is reasonable to make the approximation $\nu_2 \to 0$, Eq. (8.20) then simplifying to

$$\left(\frac{\partial \Psi}{\partial n} \right)_1 = 0 \tag{8.21}$$

We now have to solve the following boundary-value problem: the domain Γ of solution is the whole of space excluding all ferromagnetic parts. Inside this domain Eq. (6.36) is to be solved. At the optic axis and at infinity, Ψ must vanish, while at the iron surfaces the Neumann condition (8.21) must be satisfied.

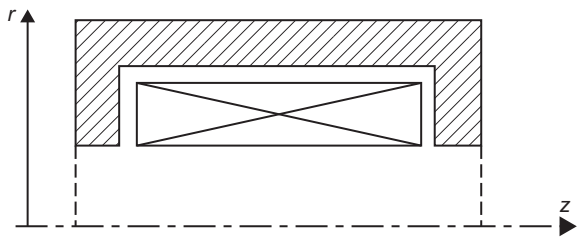


Figure 8.9
Axial section through a very simple unconventional magnetic lens.

The simplifications introduced above are invalid if parts of the iron become seriously saturated. In cases where this is liable to occur, another approximate field calculation is to be used, which is described in Chapter 12, The Finite-Element Method (FEM).

8.3.4 Toroidal Magnetic Deflection Systems

Such systems are frequently employed as scanning units, for instance in old television tubes and scanning electron microscopes. A simplified diagram is given in Fig. 8.10. Two pairs of coils, rotated at 90° with respect to each other, are wound round a rotationally symmetric ferrite shield in such a way that each winding remains in a meridional plane.

It is convenient to regard the coils as surface current distributions. Following Schwertfeger and Kasper (1974), the $\boldsymbol{H}$-field inside the yoke may be neglected, since the permeability is very high. The deflection currents are never strong enough to cause saturation effects and we can therefore confine the following discussion to the vacuum domain Γ of the device. We now omit the subscript 1. In view of these assumptions, Eq. (8.4) simplifies to

$$\boldsymbol{H}(\boldsymbol{r}) \times \boldsymbol{n}(\boldsymbol{r}) = \boldsymbol{\omega}(\boldsymbol{r}) \quad (\boldsymbol{r} \in \partial\Gamma) \tag{8.22}$$

$\partial\Gamma$ denoting the surfaces of the coils on their vacuum side. Since any integration contour which remains completely in the vacuum never encloses any current lines, $\oint \boldsymbol{H} \cdot d\boldsymbol{r} = 0$ is always valid and hence the scalar potential χ is unique in the whole vacuum domain. Forming the vector product of Eq. (8.22) with $\boldsymbol{n}$ and introducing $\boldsymbol{H} = -\nabla\chi$ we obtain

$$\text{grad}\ \chi - (\boldsymbol{n} \cdot \text{grad}\ \chi)\boldsymbol{n} = \boldsymbol{\omega} \times \boldsymbol{n} \quad \text{on} \quad \partial\Gamma \tag{8.23}$$

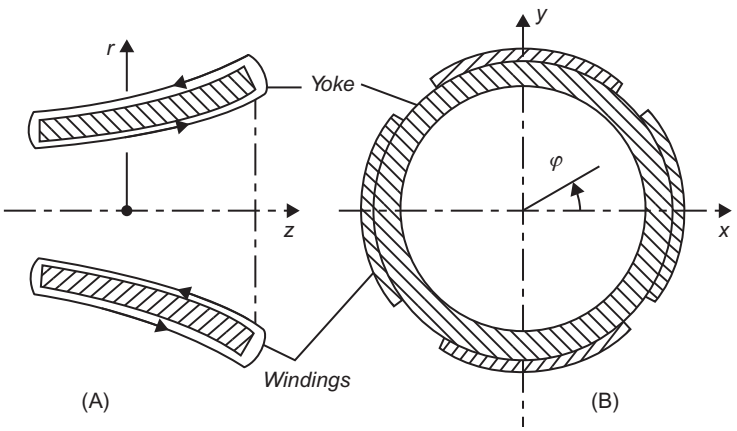


Figure 8.10
Simplified representation of a toroidal deflection system. (A) Axial section, (B) cross-section.

The expression on the left-hand side is the *tangential* component of $\nabla\chi$; this now has a given value for all surface points.

So far these considerations are quite general and are also applicable to deflection systems with saddle coils. The characteristic feature of toroidal systems is that the direction of the vector function $\boldsymbol{\omega}$ is meridional. This implies that $\boldsymbol{\omega}(\boldsymbol{r})$ may be written

$$\boldsymbol{\omega}(\boldsymbol{r}) = J(\varphi)\boldsymbol{t}(\boldsymbol{r})/r \tag{8.24}$$

φ being the azimuth with respect to the optic axis and $\boldsymbol{t}(\boldsymbol{r})$ the local tangential vector in the meridional direction; $J(\varphi)$ is the azimuthal current distribution function, which means $J(\varphi)d\varphi$ is the total electric current flowing through the windings located between φ and $\varphi + d\varphi$.

It is now possible to integrate Eq. (8.23), since Eq. (8.23) is consistent with the assumption that the boundary values of χ are only dependent on φ. Introducing $\nabla\chi = r^{-1}\,\chi'(\varphi)\boldsymbol{i}_\varphi$ and Eq. (8.24) into (8.23), and recalling that $\boldsymbol{t} \times \boldsymbol{n} = \boldsymbol{i}_\varphi$, we obtain $\chi'(\varphi) = J(\varphi)$ and hence

$$\chi(\varphi) = \chi_0 + \int_0^\varphi J(\alpha)d\alpha \quad \text{on} \quad \partial\Gamma \tag{8.25}$$

This is essentially the same as the formula of Schwertfeger and Kasper (1974); here the derivation is more general, since Eq. (8.23) may also be applied to more general types of deflection systems, see Section 9.4.4.

The boundary-value problem to be solved is now comparatively simple: at the surface $\partial\Gamma$ of a rotationally symmetric shield the potential $\chi(\boldsymbol{r})$ has uniquely determined boundary values which are *not* rotationally symmetric. At infinity the potential must vanish, and in the vacuum domain Γ, Laplace's equation is satisfied.

This presentation of the important classes of boundary-value problems is by no means complete. We cannot devote more space to them here but we hope that the reader has some impression of the complexity of the problems to be considered.

CHAPTER 9

Integral Equations

As is well-known in classical electrodynamics, it is possible to reduce the problem of solving a boundary-value problem to that of solving an integral equation. This is very advantageous since methods of field calculation based upon integral equations have gained great importance. In this chapter we shall present the general theory; details of numerical procedures are given in Chapter 10, The Boundary-Element Method.

9.1 Integral Equations for Scalar Potentials

In the following account, we consider a domain Γ in three-dimensional space and its boundary $\partial\Gamma$. Inside the domain Γ we attempt to solve a uniquely specified boundary-value problem for Poisson's equation (7.1)

$$\nabla^2 V(\boldsymbol{r}) = -S(\boldsymbol{r}) \tag{9.1}$$

9.1.1 General Theory

In order to obtain an integral equation, we start from Green's theorem for a modified domain Γ' with boundary $\partial\Gamma'$ and for a variable of integration $\boldsymbol{r}'$

$$\int_{\Gamma'} (G\nabla'^2 V - V\nabla'^2 G) d^3 r' = \int_{\partial\Gamma'} \left(G\frac{\partial V}{\partial n'} - V\frac{\partial G}{\partial n'} \right) da' \tag{9.2}$$

valid for any differentiable functions $G(\boldsymbol{r}')$ and $V(\boldsymbol{r}')$, regardless of their special meanings. The operator $\partial/\partial n' = \boldsymbol{n}' \cdot \nabla'$ is the so-called normal derivative, the derivative in the direction of the *outward* oriented surface normal $\boldsymbol{n}'$ on $\partial\Gamma'$. The boundary $\partial\Gamma'$ itself may consist of several distinct closed surfaces and the integral on the right-hand side of Eq. (9.2) is then the sum of the contributions arising from the different surfaces; this summation is implicit in the notation.

The function G in Eq. (9.2) can be chosen arbitrarily; the most suitable choice is the free-space Green's function, defined by

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00009-2
© 2018 Elsevier Ltd. All rights reserved.

$$G(\boldsymbol{r},\boldsymbol{r}') = (4\pi|\boldsymbol{r}-\boldsymbol{r}'|)^{-1} \tag{9.3}$$

and satisfying the differential equation

$$\nabla^2 G = \nabla'^2 G' = -\delta(\boldsymbol{r}-\boldsymbol{r}')$$

where $\delta(\boldsymbol{r}-\boldsymbol{r}')$ denotes Dirac's distribution. The use of the latter is familiar but not always favourable. Difficulties arise if the reference point $\boldsymbol{r}$ is located on the boundary. In order to circumvent these difficulties, we do *not* employ the δ-function formalism but instead, we modify the given domain Γ as sketched in Fig. 9.1. An internal point $\boldsymbol{r}$ is completely enclosed by a small sphere of solid angle $\Omega_1 = 4\pi$. Around a regular boundary point, a small hemisphere with solid angle $\Omega_2 = 2\pi$ is excluded; at a sharp line-edge of intersection angle α, a spherical segment with solid angle $\Omega_3 = 2\alpha$ is removed; finally, an external point $\boldsymbol{r}$ needs no exclusion surface and hence $\Omega_4 = 0$. In the domain Γ', obtained after excluding the immediate neighbourhood of the reference point, Eq. (9.2) is valid.

In order to simplify the notation, we introduce the abbreviation

$$\varepsilon(\boldsymbol{r}) \coloneqq \frac{1}{4\pi}\Omega(\boldsymbol{r}) = \begin{array}{ll} 1 & \text{for } \boldsymbol{r} \in \Gamma, \\ \alpha/2\pi & \text{for } \boldsymbol{r} \in \partial\Gamma, \\ 0 & \text{for } \boldsymbol{r} \notin \Gamma, \end{array} \tag{9.4}$$

which implies that $\alpha = \pi$, $\varepsilon = 1/2$ for *regular* boundary points. We also write

$$\sigma(\boldsymbol{r}') \coloneqq \frac{\partial V}{\partial n'} \equiv \boldsymbol{n}' \cdot \nabla' V(\boldsymbol{r}') \tag{9.5}$$

$$P(\boldsymbol{r},\boldsymbol{r}') \coloneqq \frac{\partial G}{\partial n'} \equiv \frac{\boldsymbol{n}' \cdot (\boldsymbol{r}-\boldsymbol{r}')}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|^3} \tag{9.6}$$

At the boundary these normal derivatives will be discontinuous; they are then defined as the limits obtained on approaching from the interior of Γ'.

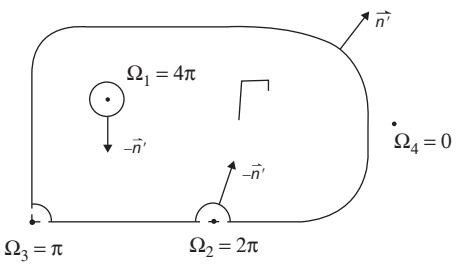


Figure 9.1
Two-dimensional section through a three-dimensional domain Γ, showing various positions of reference points and the associated solid angles.

Introducing all this into Eq. (9.2), we soon notice that only the terms involving P are critical so far as the limit of vanishing radius ρ of the sphere or sector of exclusion is concerned. We may then approximate the slowly varying function $V(\boldsymbol{r}')$ by $V(\boldsymbol{r})$. Recalling that $|\boldsymbol{r}-\boldsymbol{r}'| = \rho = \boldsymbol{n}' \cdot (\boldsymbol{r}-\boldsymbol{r}')$ on the spherical surfaces S, we obtain

$$\int_{\partial\Gamma'} V(\boldsymbol{r}')\frac{\partial G}{\partial n'}da' \to \oint_{\partial\Gamma} V(\boldsymbol{r}')P(\boldsymbol{r},\boldsymbol{r}')da' + \lim_{\rho\to 0} V(\boldsymbol{r})\int_S \frac{da'}{4\pi\rho^2}$$

The final surface integral is just the expression $\varepsilon(\boldsymbol{r})$ defined by Eq. (9.4). The symbol $\oint$ denotes the principal value of the corresponding integral, defined as the value obtained by proceeding to the limit $\rho \to 0$ in the integration over the not-excluded parts of $\partial\Gamma$. For reference points $\boldsymbol{r}$ outside $\partial\Gamma$ this is straightforward; for $\boldsymbol{r} \in \partial\Gamma$ this limit exists, since in this case we have $\boldsymbol{n}' \cdot (\boldsymbol{r} - \boldsymbol{r}') \to 0$ because of the orthogonality between tangents and surface normals. In the subsequent presentation we shall not indicate the principal value explicitly since all improper surface integrals are to be evaluated in this way.

Putting all this together, we arrive finally at

$$\begin{aligned}\varepsilon(\boldsymbol{r})V(\boldsymbol{r}) = &\int_\Gamma G(\boldsymbol{r},\boldsymbol{r}')S(\boldsymbol{r}')d^3r' \\ &+ \int_{\partial\Gamma}\{G(\boldsymbol{r},\boldsymbol{r}')\sigma(\boldsymbol{r}') - P(\boldsymbol{r},\boldsymbol{r}')V(\boldsymbol{r}')\}da'\end{aligned} \tag{9.7}$$

The expressions on the right-hand side can be interpreted in the following way: the first is a space-source term, the second a surface-source term and the last is a surface-polarization term. Since the boundary functions $V(\boldsymbol{r}')$ and $\sigma(\boldsymbol{r}')$ are still independent, Eq. (9.7) alone does not suffice to determine the potential distribution uniquely. We may prescribe an additional boundary condition

$$a(\boldsymbol{r}')V(\boldsymbol{r}') + b(\boldsymbol{r}')\sigma(\boldsymbol{r}') = c(\boldsymbol{r}'), \quad \boldsymbol{r}' \in \partial\Gamma \tag{9.8}$$

the surface functions $a(\boldsymbol{r}')$, $b(\boldsymbol{r}')$ and $\boldsymbol{c}(\boldsymbol{r}')$ being known with $a^2 + b^2 > 0$. It is then possible to solve Eq. (9.7). We now discuss the two most familiar special cases.

9.1.2 Dirichlet Problems

Here the boundary values $\overline{V}(\boldsymbol{r}')$ are specified uniquely, while $\sigma(\boldsymbol{r}')$ is unknown. In Eq. (9.8) we may choose $a \equiv 1$, $b \equiv 0$; $c(\boldsymbol{r}') \equiv \overline{V}(\boldsymbol{r}')$ are then the given boundary values. Eq. (9.7) is now a two-dimensional integral equation of Fredholm's first kind for the unknown σ. After solving it, the same Eq. (9.7) can be used to evaluate the potential at any point $\boldsymbol{r}$ in Γ.

Although this procedure is perfectly correct, it is rather inconvenient for numerical solutions since the factor $\varepsilon(\boldsymbol{r})$ is discontinuous and the polarization term requires careful

handling in order to obtain the appropriate principal value. It is therefore highly desirable to find alternative forms of the integral equation that do not contain these terms.

If the boundary $\partial\Gamma$ consists of a number of *closed* surfaces with *constant* boundary values $\overline{V}$ on them, the difficult terms can be eliminated completely. For simplicity, we consider only one such surface $\partial\Gamma$, the domain Γ being its *exterior*. The second boundary is an infinite sphere; since the field is required to satisfy the natural boundary condition, this sphere need not be considered here. We now have to evaluate the integral term

$$I := \int_{\partial\Gamma} P(\boldsymbol{r},\boldsymbol{r}')V(\boldsymbol{r}')da' = V\int_{\partial\Gamma} P(\boldsymbol{r},\boldsymbol{r}')da'$$

Since Eq. (9.7) is quite generally valid, we are at liberty to set V equal to a constant; for the moment, therefore, we introduce

$$V \equiv 1, \quad \sigma \equiv 0, \quad S \equiv 0$$

and recalling that Γ is now the exterior, we obtain the mathematical identity

$$\int_{\partial\Gamma} P(\boldsymbol{r},\boldsymbol{r}')da' = 1 - \varepsilon(\boldsymbol{r}) \tag{9.9}$$

The integral expression I now simplifies to

$$I = \{1 - \varepsilon(\boldsymbol{r})\}V$$

Introducing this into Eq. (9.7) we find

$$V(\boldsymbol{r}) = \int_{\Gamma} G(\boldsymbol{r},\boldsymbol{r}')S(\boldsymbol{r}')d^3r' + \int_{\partial\Gamma} G(\boldsymbol{r},\boldsymbol{r}')\sigma(\boldsymbol{r}')da' \tag{9.10}$$

This Fredholm equation for σ is now quite generally applicable without exception. When the solution is still unknown, the reference point $\boldsymbol{r}$ must be located at the boundary and $\boldsymbol{V(r)}$ on the left-hand side is then the corresponding boundary value. After obtaining the solution for σ, the same equation (9.10) may be employed to compute the potential at any point $\boldsymbol{r}$ in the space, even on $\partial\Gamma$ or outside Γ. The potential itself is *continuous* if the reference point crosses the boundary.

We can generalize Eq. (9.10) to include configurations with a boundary consisting of several closed surfaces and even boundary values of V that are *not* constant on these surfaces. The basic form of Eq. (9.10) remains unaltered, but now the surface charge density σ is the difference between the normal derivatives of V on each side of the corresponding surface. Physically, Eq. (9.10) can be interpreted as a Coulomb integral over

space charges and over surface charges as a degenerate case of the former. When it comes to numerical evaluation, Eq. (9.10) is far more convenient than Eq. (9.7).

9.1.3 Neumann Problems

It is now the boundary values of $\partial V/\partial n$ that are uniquely specified while those of $\boldsymbol{V(r)}$ are unknown. This boundary value problem has a solution only if $\oint\sigma(\mathbf{r}')da'$ vanishes on $\partial\Gamma$. In Eq. (9.8) we may specify $a \equiv 0$, $b \equiv 1$, so that $\boldsymbol{c(r')} = \boldsymbol{\sigma(r')}$ are given boundary values. Eq. (9.7) is now an integral equation of Fredholm's *second* kind for the boundary values of $\boldsymbol{V(r)}$. Apart from an unimportant additive constant, this integral equation has a unique solution. After this has been found, Eq. (9.7) can be used to compute the potential $\boldsymbol{V(r)}$ at arbitrary points inside Γ.

The polarization term cannot be eliminated from Eq. (9.7). Since this term contains a strong singularity and is discontinuous at the boundary $\partial\Gamma$, great care must be taken in numerical computations. Thus the concrete evaluation of such expressions should be avoided whenever possible by appropriate transformation of the boundary-value problem.

9.2 Problems with Interface Conditions

The theory outlined in Section 9.1.1 is quite standard in classical electrodynamics and can be applied to boundary-value problems in electrostatics and magnetostatics. The integral equation given below is less familiar. In connection with electron optical applications it has been mentioned by Kasper (1982) and explicitly derived by Scherle (1983), who also demonstrated that it can be applied in practical numerical computations. Alternative formulations will be given at the end of this section.

Since interface conditions are most important in *magnetostatic* problems, we shall confine the discussion to these, although it would be no problem to establish an integral equation for electric fields. It is necessary to assume unsaturated (linear) media. For simplicity, we consider here only two different domains, the vacuum domain Γ_1 and a ferromagnetic shield Γ_2, $\Gamma_1 \cup \Gamma_2$ being the whole space. The convention concerning the choice of the surface normal, represented in Fig. 8.2, then holds. Generalizations to more than two different domains have been worked out (Scherle, 1983); apart from the introduction of an iterative solution technique, they contain nothing essentially new; some results will be given in Chapter 10, The Boundary-Element Method.

We start from Eq. (9.7). The appropriate potential is here $\chi_M(\boldsymbol{r})$, introduced in Eqs (8.7) and (8.12). From Eq. (8.13), we see that there is no space-source term. On the boundary $\partial\Gamma = \Gamma_1 \cap \Gamma_2$ we have $\varepsilon = 1/2$ and hence Eq. (9.7) specializes to

$$\frac{\chi_M(\boldsymbol{r})}{2} = (-1)^j \oint_{\partial\Gamma} \left\{ \chi_M(\boldsymbol{r}')P(\boldsymbol{r},\boldsymbol{r}') - G(\boldsymbol{r},\boldsymbol{r}')\sigma_j(\boldsymbol{r}') \right\} da' \quad (j = 1, 2, \quad \boldsymbol{r} \in \partial\Gamma) \tag{9.11}$$

The index j indicates the domain from which the surface $\partial\Gamma$ is approached; we recall that in the polarization term the principal value of the integral is to be taken. Since neither the potential itself nor its normal derivatives σ_j are known, a second relation is needed, namely Eq. (8.15). This enables us to eliminate the normal derivative from Eq. (9.11) by forming appropriate linear combinations. In this context we need to use Eq. (9.5) with $\chi_M \equiv V$. The result of these elementary calculations is as follows:

$$-\lambda\chi_M(\boldsymbol{r}) + \oint_{\partial\Gamma} P(\boldsymbol{r},\boldsymbol{r}')\chi_M(\boldsymbol{r}')da' = \oint_{\partial\Gamma} G(\boldsymbol{r},\boldsymbol{r}')\boldsymbol{n}'\cdot\boldsymbol{H}_0(\boldsymbol{r}')da' \tag{9.12}$$

with $\boldsymbol{r} \in \partial\Gamma$ and

$$\lambda := \frac{1}{2}\,\frac{\mu_2 + \mu_1}{\mu_2 - \mu_1} \tag{9.13}$$

This is an integral equation of Fredholm's second kind for the surface values of the potential. Once it has been solved, the problem reduces to an ordinary Dirichlet problem. We can hence introduce the calculated surface values of χ_M in the left-hand side of Eq. (9.10) and solve this equation (with vanishing space-source S) for σ. It can be shown that $\sigma = \sigma_1 - \sigma_2$, but this is of little use, since the solution of Eq. (9.10) gives σ directly.

The derivation presented here differs from Scherle's method but is equivalent to it. In the literature on magnetic field computation, many other forms of integral equation are derived, which are essentially equivalent to Eq. (9.12), but not always so suitable for numerical evaluation. Many technical points were elucidated in the proceedings of COMPUMAG (1976) and the subject has been reviewed by Iselin (1981). Besides Scherle's thesis, the publications of Lucas (1976) and Kuroda (1983) are particularly concerned with field calculation for electron optical designs by means of integral equations. Scherle's method has the advantage that only one scalar integral equation for a potential is needed instead of three coupled ones for a vector field and that the singularity of the integral kernel $P(\boldsymbol{r}, \boldsymbol{r}')$ is the weakest possible. The field can be evaluated everywhere in space.

9.3 Reduction of the Dimensions

The derivation of a two-dimensional integral equation means that three-dimensional unknown functions have already been reduced to two dimensions. Very often the integral equations obtained are soluble only numerically. In electron optics, however, there is an important class of configurations that can be treated by means of one-dimensional integral

equations. This is the class of all devices the electrodes or polepieces of which have rotationally symmetric surfaces. It is not necessary to assume that the boundary values are also rotationally symmetric. In Chapter 7, Series Expansions, we have shown that the three-dimensional Poisson equation can be reduced to a sequence of *uncoupled* two-dimensional equations by means of Fourier series expansions. Here we shall show that integral equations can also be simplified in an analogous manner. For Dirichlet problems, an approximate theory has been developed by Kasper and Scherle (1982) and by Kasper (1984a,b); the theory of a method of evaluating Eq. (9.12) has been developed by Scherle, who has also demonstrated that it can be applied in practice.

9.3.1 Dirichlet Problems

The space-source term in Eq. (9.10) can now be omitted without loss of generality. This term alone produces a particular solution $\boldsymbol{V_s(r)}$. If $\boldsymbol{V_s} \neq 0$, then the subsequent reasoning is valid for $\boldsymbol{V} - \boldsymbol{V_s}$ instead of $\boldsymbol{V}$. Since this adds nothing new, we assume that $S(\boldsymbol{r}) \equiv 0$, Eq. (9.10) then simplifying to a pure surface integral

$$V(\boldsymbol{r}) = \frac{1}{4\pi}\int_{\partial\Gamma} \frac{\sigma(\boldsymbol{r}')da'}{|\boldsymbol{r} - \boldsymbol{r}'|} \tag{9.14}$$

There is no advantage to be gained by separating the factor r^m in Eq. (7.5) and we thus introduce the notation

$$V(\boldsymbol{r}) = \sum_{m=0}^{\infty} \Re\{V_m(z, r)\mathrm{e}^{\mathrm{i}m\varphi}\} \tag{9.15}$$

The boundary $\partial\Gamma$ is suitably represented in parametric form in terms of the azimuth φ and the arc-length s along the meridional line C passing through the reference point $\boldsymbol{r}$ (see Fig. 9.2). Since $\partial\Gamma$ is assumed to be rotationally symmetric, this line C can be represented

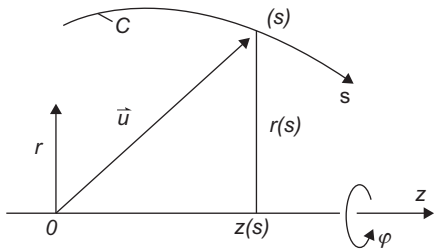


Figure 9.2
Meridional section through a rotationally symmetric boundary with contour C, along which the arc-length s is adopted as parameter.

by $z = z(s)$, $r = r(s)$. Whenever this causes no confusion, we shall omit the argument s and introduce the simplifying notation $z' = z(s')$, $r' = r(s')$. When z, z' and r, r' appear as arguments of functions, the vectors $\boldsymbol{u} = (z, r)$, $\boldsymbol{u}' = (z', r')$ will be used.

The values of the functions V_m, occurring in Eq. (9.15), on the boundary are given by the inverse Fourier transform of $V(\boldsymbol{u}(s), \varphi)$ and hence depend on s:

$$\upsilon_m(s) \coloneqq V_m(\boldsymbol{u}(s)) = \frac{\kappa_m}{2\pi}\int_0^{2\pi} V(\boldsymbol{u}(s), \varphi)\mathrm{e}^{-\mathrm{i}m\varphi}d\varphi \tag{9.16}$$

with

$$\kappa_m = 2 - \delta_{m,0} = \begin{cases} 1 & \text{for } m = 0 \\ 2 & \text{for } m \neq 0 \end{cases} \tag{9.16a}$$

It is helpful to expand the surface charge density σ as a Fourier series with respect to φ:

$$\sigma(\boldsymbol{r}) = \sum_{m=0}^{\infty} \Re\left\{\sigma_m(s)\mathrm{e}^{\mathrm{i}m\varphi}\right\} \tag{9.17}$$

Introducing Eqs (9.17) and (9.15) with the coefficients $\upsilon_m(s)$ into (9.14), we obtain

$$\sum_{m=0}^{\infty} \Re\left\{\upsilon_m(s)\mathrm{e}^{\mathrm{i}m\varphi}\right\} = \sum_{m=0}^{\infty} \Re\int_{\partial\Gamma} \frac{\sigma_m(s')\mathrm{e}^{\mathrm{i}m\varphi'}}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}da'$$

The distance $D \coloneqq |\mathbf{r} - \boldsymbol{r}'|$ is explicitly given by

$$D(\boldsymbol{u}, \boldsymbol{u}', \alpha) = \left\{(z - z') + r^2 + r'^2 - 2rr'\cos\alpha\right\}^{1/2} \tag{9.18}$$

$\alpha = \varphi' - \varphi$ being the difference of azimuth. The element of surface area can be written as $da' = r'\, ds'\, d\alpha$. Eliminating φ', we find

$$\sum_{m=0}^{\infty} \Re\left\{\upsilon_m(s)\mathrm{e}^{\mathrm{i}m\varphi}\right\} = \sum_{m=0}^{\infty} \Re\left\{\mathrm{e}^{\mathrm{i}m\varphi}\int_C G_m(\boldsymbol{u}, \boldsymbol{u}')r'\sigma_m(s')ds'\right\}$$

where the abbreviation

$$G_m(\boldsymbol{u}, \boldsymbol{u}') = \frac{1}{4\pi}\int_0^{2\pi} \frac{\mathrm{e}^{\mathrm{i}m\alpha}d\alpha}{D(\boldsymbol{u}, \boldsymbol{u}', \alpha)}$$

has been introduced. This expression is a *real* function of its arguments, since it can be rewritten as

$$G_m(\boldsymbol{u}, \boldsymbol{u}') = \frac{1}{2\pi}\int_0^{\pi} \frac{\cos(m\alpha)d\alpha}{D(\boldsymbol{u}, \boldsymbol{u}', \alpha)} \tag{9.19}$$

From Eq. (9.18) the *symmetry* relation

$$G_m(\boldsymbol{u}, \boldsymbol{u}') = G_m(\boldsymbol{u}', \boldsymbol{u}) \tag{9.20}$$

is obvious. These functions are essentially the Fourier coefficients of the free-space Green's function $G(\boldsymbol{r}, \boldsymbol{r}')$.

From the uniqueness of Fourier series expansions, we now obtain a sequence of *uncoupled* one-dimensional Fredholm equations, given in explicit notation by

$$\upsilon_m(s) = \int_C G_m(\boldsymbol{u}(s)\, \boldsymbol{u}(s'))r(s')\sigma_m(s')ds' \quad \text{for} \quad m = 0, 1, 2, \ldots \tag{9.21}$$

These have the formal structure

$$\upsilon_m(s) = \int_C K_m(s, s')\sigma_m(s')ds' \tag{9.22}$$

In Chapter 10, The Boundary-Element Method, we shall show that the kernel functions G_m of Eq. (9.19) can be evaluated analytically and that the resulting expressions contain complete elliptic integrals. We shall further show that there are convenient techniques for numerical solution of (9.22). Thus the solution of Dirichlet problems with rotationally symmetric boundaries can be regarded as a standard technique.

9.3.2 Interface Conditions

In the case of rotationally symmetric boundaries, Scherle's integral equation (9.12) can again be decomposed into a sequence of uncoupled integral equations for the Fourier potentials $V_m(z, r)$. As before, we start from Eq. (9.15), with $V \equiv \chi_M$. It is now necessary to introduce a Fourier series expansion for the component $\boldsymbol{n} \cdot \boldsymbol{H}_0$ appearing on the right-hand side of Eq. (9.12):

$$\boldsymbol{n}(\boldsymbol{r}) \cdot \boldsymbol{H}_0(\boldsymbol{r}) = \Re \sum_{m=0}^{\infty} N_m(z, r)\mathrm{e}^{\mathrm{i}m\varphi} \tag{9.23}$$

Then, by arguments similar to those described above, we obtain first

$$\oint_{\partial\Gamma} G(\boldsymbol{r}, \boldsymbol{r}')\boldsymbol{n}' \cdot \boldsymbol{H}_0(\boldsymbol{r}')da' = \sum_{m=0}^{\infty} \Re\left\{q_m(s)\mathrm{e}^{\mathrm{i}m\phi}\right\} \tag{9.24}$$

with

$$q_m(s) \coloneqq \oint_C G_m(\boldsymbol{u}, \boldsymbol{u}')r'N_m(\boldsymbol{u}')ds' \tag{9.25}$$

This is a known function of s which can be evaluated by numerical integration over s'.

In order to evaluate the integral expression on the left-hand side of Eq. (9.12), we need the Fourier coefficients of the function $P(\boldsymbol{r}, \boldsymbol{r}')$ defined by Eq. (9.6). Since the normal derivative involves only the coordinates z' and r', the operator $\partial/\partial n'$ can be applied *after* the integration over φ' has been carried out. We thus obtain

$$\begin{aligned} P_m(\boldsymbol{u}, \boldsymbol{u}') &:= \int_0^{2\pi} \frac{\partial}{\partial n'} G(\boldsymbol{r}, \boldsymbol{r}') \mathrm{e}^{\mathrm{i}m(\varphi - \varphi')}\, d\varphi' \\ &= \frac{\partial}{\partial n'} G_m(\boldsymbol{u}, \boldsymbol{u}') \end{aligned} \tag{9.26}$$

Like the integral expressions (9.19), these functions take only *real* values but the symmetry properties do not hold.

Introducing the Fourier series expansions of $\boldsymbol{V}(\boldsymbol{r})$ into the left-hand side of Eq. (9.12) and considering Eq. (9.26), we finally obtain a sequence of uncoupled Fredholm equations of the second kind:

$$\begin{gathered} -\lambda v_m(s) + \oint_C P_m(\boldsymbol{u}(s), \boldsymbol{u}(s')) r(s') v_m(s') ds' = q_m(s) \\ m = 0, 1, 2, \ldots \end{gathered} \tag{9.27}$$

Since G_m, P_m and λ are real, it is sufficient to investigate solution techniques for *real* integral equations. Complex solutions can easily be obtained by forming linear combinations of real solutions with constant complex factors. We have thus achieved a major simplification of the original boundary value problem.

9.3.3 Planar Fields

In this case the integral equations are already familiar. We shall only examine briefly the Dirichlet problem in the (z, x) plane. Let C be a boundary line represented parametrically in terms of the arc-length s: $z = z(s)$, $x = x(s)$. This line may consist of several distinct curves. Let $v(s) = V(z(s), x(s))$ be given boundary values. Then the integral equation takes the form

$$v(s) = \int_C q(s') \ln \left[a \left\{ (z - z')^2 + (x - x')^2 \right\}^{-1/2} \right] ds' \tag{9.28}$$

where $z' = z(s')$, $x = x(s')$ has been introduced; a is an arbitrary positive normalization constant and $q(s')$ the source distribution function on C. Since the logarithmic kernel function is singular at infinity, either $\int q(s')\, ds' = 0$ or the entire field must be enclosed within a closed loop on which V is zero.

9.4 Important Special Cases

In this section, we shall examine some field calculation problems that can be solved by numerical evaluation of the integral equations derived above. Technical details are discussed in Chapter 10, The Boundary-Element Method.

9.4.1 Rotationally Symmetric Scalar Potentials

Here it is not necessary to carry out the Fourier series expansions; the required results are obtained directly for $m = 0$. The function $G_0(\boldsymbol{u}, \boldsymbol{u}')$ is the scalar potential of a uniformly charged ring of radius r' or r and is treated in detail in Chapter 10, The Boundary-Element Method. Eq. (9.21) with $m = 0$ was the starting point for the development of the boundary-element method (BEM) of field calculation, first introduced by Cruise (1963) and extensively studied by Harrington (Harrington, 1967, 1968; Harrington et al., 1969; Mautz and Harrington, 1970; Kuno and Uchikawa, 1985; Kuno et al., 1988; Tsuboi et al., 1990a,b, 1992, 1998, 1999; Watanabe et al., 1990; Binns et al., 1992; Tanaka et al., 1992; Read and Bowring, 2011). This version of the BEM can be directly applied to electrostatic round lenses of various shapes (Lewis, 1966; Singer and Braun, 1970; Read et al., 1971; Adams and Read, 1972a,b; Harting and Read, 1976; Read, 2015b), electron guns with arbitrary rotationally symmetric cathodes (Rauh, 1971; Kuroda and Suzuki, 1972), round electron mirrors and conventional magnetic lenses.

9.4.2 Rotationally Symmetric Vector Potentials

Though the general theory was developed for scalar potentials, it can easily be modified to make it applicable to vector potentials. Since each Cartesian component of $\boldsymbol{A}$ separately satisfies a Poisson equation like Eq. (9.1) with $S \rightarrow \mu_0 j_k$ ($k = 1, 2, 3$), the general conclusions leading to Eq. (9.10) must hold for each component of $\boldsymbol{A}$ separately. Collecting the three integral equations into a single vector expression, we now find

$$\boldsymbol{A}(\boldsymbol{r}) = \mu_0 \int_{\Gamma} G(\boldsymbol{r}, \boldsymbol{r}')\, \boldsymbol{j}(\boldsymbol{r}') d^3 r' + \mu_0 \int_{\partial\Gamma} G(\boldsymbol{r}, \boldsymbol{r}') \boldsymbol{\omega}(\boldsymbol{r}') da' \tag{9.29}$$

The condition $\operatorname{div} \boldsymbol{A} = 0$ is satisfied if the surface currents are conserved, as must be the case for physical reasons.

In rotationally symmetric devices, the vectors $\boldsymbol{A}, \boldsymbol{j}$ and $\boldsymbol{\omega}$ have only azimuthal components (see also Section 6.4). In the case of superconducting round lenses, the space currents $\boldsymbol{j}$ are absent, Eq. (9.29) then simplifying to

$$A(\boldsymbol{u}) = \mu_0 \oint_C \int_0^{2\pi} \frac{\omega(\boldsymbol{u}') \boldsymbol{i}_{\varphi} \cdot \boldsymbol{i}_{\varphi'} r' ds' d\alpha}{4\pi D(\boldsymbol{u}, \boldsymbol{u}', \alpha)}$$

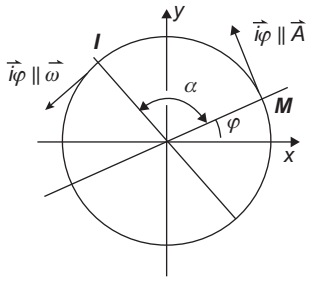


Figure 9.3
The relation between vectors and angles in rotationally symmetric vector-potential fields. ***M*** is a fixed point and ***I*** is the integration point.

where the notation introduced in Section 9.3 ($u = (z, r)$ etc.) has been used. The factor $\cos\alpha = \boldsymbol{i}_\varphi \cdot \boldsymbol{i}_{\varphi'}$ arises from the projection of $\boldsymbol{\omega}$ on the direction of $\boldsymbol{A}$, see Fig. 9.3. Using Eq. (9.19) with $m = 1$, Eq. (6.27) and the boundary conditions Eq. (8.16), we finally obtain the integral equation:

$$2\pi r\mu_0 \oint G_1(\boldsymbol{u}, \boldsymbol{u}')r'\omega(\boldsymbol{u}')ds' = \Psi_B = \text{const} \tag{9.30}$$

which may formally be considered as a special case of Eq. (9.21) with $\upsilon(s) = \Psi_B/2\pi\mu_0 r(s)$.

It is also of some interest that the azimuthal vector potential of a rotationally symmetric coil in the absence of magnetic materials can be represented by

$$A_0(\boldsymbol{u}) = \mu_0 \int\int_F G_1(\boldsymbol{u}, \boldsymbol{u}')r'\, j(\boldsymbol{u}')dr'dz' \tag{9.31}$$

F being the domain $r' > 0$ of the axial section through the coil. This is a special case of Eq. (8.10).

9.4.3 Unconventional Magnetic Lenses

In Section 8.3.3 we have indicated that such lenses can be treated by obtaining the solution of a Neumann problem for the rotationally symmetric flux potential Ψ. Alternatively, they can be calculated as a special case of Eqs (9.25) and (9.27) with $m = 0$. In this case, no Fourier series expansions are necessary, the quantities of interest being the zero-order Fourier coefficients themselves. In this way really complicated devices with very open structures like those investigated by Mulvey (1982) can be calculated, provided that saturation effects need not be considered.

9.4.4 Magnetic Deflection Coils

The following considerations (Kasper, 1984a) are valid for both toroidal coils and saddle coils. If the windings are so close to the shield surface that the surface current distribution approximation is reasonable, the total scalar potential $\chi(\boldsymbol{r})$ may be used. The boundary conditions on $\chi(\boldsymbol{r})$ are given by Eq. (8.23). In the case of toroidal systems, we could integrate this equation, but in more complicated cases the corresponding integration becomes very complicated. The application of Fourier series expansions, however, makes this integration unnecessary, as we now show.

We again represent each surface function as a function of the azimuth φ and the arc-length s along the axial contour C. The meridional tangent vector is $\boldsymbol{t} = \partial \boldsymbol{r}/\partial s$. The surface potential and its gradient are then given by

$$\chi = \chi(\varphi, s), \quad \nabla\chi = \frac{1}{r}\frac{\partial \chi}{\partial \varphi}\boldsymbol{i}_\varphi + \frac{\partial \chi}{\partial s}\boldsymbol{t} + \frac{\partial \chi}{\partial n}\boldsymbol{n} \tag{9.32}$$

The surface current distribution may be written

$$\boldsymbol{\omega}(\varphi, s) = \omega_\varphi(\varphi, s)\boldsymbol{i}_\varphi + \omega_s(\varphi, s)\boldsymbol{t} \tag{9.33}$$

Introducing Eqs (9.33) and (9.32) into (8.23) and recalling that $\boldsymbol{t} \times \boldsymbol{n} = \boldsymbol{i}_\varphi$ and $i_\varphi \times \boldsymbol{n} = -\boldsymbol{t}$, we obtain

$$\frac{\partial \chi}{\partial \varphi} = r\omega_s, \quad \frac{\partial \chi}{\partial s} = -\omega_\varphi \tag{9.34}$$

The condition obtained by evaluating $\partial^2\chi/\partial\varphi\partial s$ from each of these,

$$\frac{\partial^2 \chi}{\partial \varphi \partial s} = \frac{\partial}{\partial s}(r\omega_s) = -\frac{\partial}{\partial \varphi}\omega_\varphi \tag{9.35}$$

is identical with the continuity equation for surface currents and thus imposes a restriction on the choice of surface current distributions, which is automatically satisfied by any real distribution of wires.

For a single pair of deflection coils, we can choose the origin of φ in such a way that the appropriate Fourier series expansions can be written as

$$\omega_s(\varphi, s) = \sum_m M_m(s) \cos(m\varphi) \tag{9.36}$$

$$\omega_\varphi(\varphi, s) = -\sum_m A_m(s) \sin(m\varphi) \tag{9.37}$$

$$\chi(\varphi, s) = \sum_m \upsilon_m(s) \sin(m\varphi) \tag{9.38}$$

m always being an *odd* integer. Introduction of these series expansions into (9.34) and (9.35) results in

$$\upsilon_m(s) = \frac{1}{m} r(s) M_m(s), \quad m = 1, 3, 5, \ldots \tag{9.39}$$

and

$$\upsilon'_m(s) = \frac{d}{ds}\left(\frac{r}{m} M_m\right) = A_m(s), \quad m = 1, 3, 5, \ldots \tag{9.40}$$

This is an important simplification. The boundary values $\upsilon_m(s)$ of the Fourier potentials are already uniquely determined by the Fourier coefficients of the meridional component, while the azimuthal components A_m can be obtained by mere differentiation and are not needed explicitly for the solution of the boundary-value problem: the only major computations needed are the Fourier transformation of $\omega_s(\varphi, s)$ and the subsequent numerical solution of Eq. (9.21). Thereafter the integral expression in Eq. (9.21) can be evaluated to give $V_m(z, r)$ at any point (z, r) outside the surface.

In the case of a purely toroidal system of coils the surface current density is given by

$$\omega_\varphi \equiv 0, \quad \omega_s = IW(\varphi)/r(s) \tag{9.41}$$

I being the electric current through the windings and $W(\varphi)$ the winding density, by which we mean that $W(\varphi)d\varphi$ is the number of windings between φ and $\varphi + d\varphi$. In this case Eq. (9.39) simplifies to

$$\upsilon_m = \frac{4I}{\pi m} \int_0^{\pi/2} W(\varphi) \cos(m\varphi) d\varphi, \quad m = 1, 3, 5, \ldots \tag{9.42}$$

(Schwertfeger and Kasper, 1974). We cannot expect any further simplification, as we now have to solve Dirichlet problems with constant boundary values.

If the windings are not close to the shield surface, as is frequently the case in systems of saddle coils, the one-dimensional integral equations arising from Scherle's equation, namely Eqs (9.25) and (9.27), must be evaluated. Then, of course, the necessary computation is considerably greater but the results will be very accurate. The details are discussed in Chapter 40 of Volume 2.

9.4.5 Multipole Systems

Electric or magnetic multipole systems are commonly used as stigmators or as elements of aberration correctors, where quadrupoles, sextupoles, octopoles and dodecapoles are needed.

Nowadays such systems are also used as deflectors in scanning devices. Since these systems are in no sense rotationally symmetric, the theory of one-dimensional integral equations cannot be properly applied to them. If, however, parts of the surfaces form a rotationally symmetric face and the gaps between adjacent poles are comparatively narrow, as shown in Figs 9.4 and 9.5, the theory is approximately applicable. Magnetic systems like the one shown in Fig. 9.4 were in use in the devices developed by the Darmstadt group (see Rose, 1971; Rose and Plies, 1973) to compensate the third-order spherical aberration and the axial chromatic aberration in an electron microscope of very high resolution. Electric systems of the kind shown in Fig. 9.5 have been investigated by Munro and Chu (1982a,b) and Chu and Munro (1982a,b) and are used as deflection units in electron lithography devices.

The approximation underlying the treatment of these systems is that in the gaps, it must be possible to make a reasonable interpolation for the potential with respect to the azimuth φ, usually by a linear expression. Then, on a rotationally symmetrical surface — consisting of the cylindrical bore, the gaps, the ring-shaped parts of the end-planes and sometimes parts

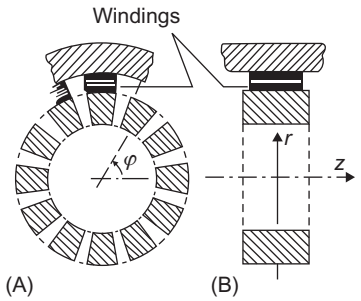


Figure 9.4
Simplified representation of a magnetic multipole system. (A) Cross-section, (B) meridional section. Only parts of the coils and the outer screening ring are shown.

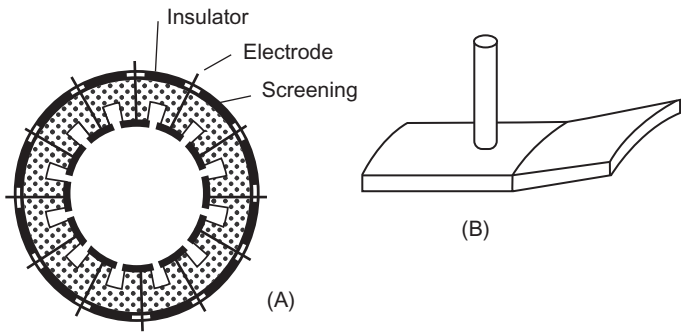


Figure 9.5
Simplified view of an electric multipole device. (A) Cross-section, (B) perspective view of a single electrode.

of the outer surfaces of the poles and a screening surface enclosing the whole systems as well — reasonable values of the potential can be defined. These are then to be introduced into Eq. (9.16) after which the uncoupled one-dimensional integral equations (9.21) can be solved. The errors caused by the incorrect boundary values do not seriously influence the field in the paraxial domain. This approximation is therefore often satisfactory.

9.4.6 Small Perturbations of the Rotational Symmetry

Since it is impossible to build rotationally symmetric devices perfectly, the effects of small perturbations such as shifts, tilts or ellipticity of the polepieces or electrodes on the field in the paraxial domain are of interest. Such effects determine the tolerance limits for the machining of electron optical devices, and they have attracted considerable attention (see Chapter 31, Parasitic Aberrations). The first successful attempts to calculate them numerically were made by Janse (1971). Here, we describe how his method can be used to calculate the field in an imperfect round magnetic lens.

Janse's method is a perturbation calculus. First, as a zero-order approximation, the ideal round system is calculated; along the contour C in the axial section through the system the boundary values $\boldsymbol{H}(s)$ of the field strength are then determined (s being the arc-length along C). This is illustrated in Fig. 9.6. Owing to the rotational symmetry of the perfect system, the field strength $\boldsymbol{H}(\boldsymbol{r}) = \boldsymbol{H}(\varphi, s)$ at any surface point $\boldsymbol{r}$ with coordinates φ and s is then known.

In the next step a deviation function $\boldsymbol{d}(\boldsymbol{r})$ is defined as the local shift from the ideal surface to the real one. In the parametrization adopted here, this is a function $\boldsymbol{d} = \boldsymbol{d}(\varphi, s)$. The shift may be in any direction but must be very small as shown in Fig. 9.6.

Since the real distorted pole surface coincides with the equipotentials $\chi(\boldsymbol{r} + \boldsymbol{d}) = \text{const}$, the potential at the ideal round surface is perturbed by a quantity

$$\delta\chi(\boldsymbol{r}) = \chi(\boldsymbol{r}) - \chi(\boldsymbol{r} + \boldsymbol{d}) \approx -\boldsymbol{d} \cdot \text{grad}\, \chi = \boldsymbol{d}(\boldsymbol{r}) \cdot \boldsymbol{H}(\boldsymbol{r}) \tag{9.43a}$$

or

$$\delta\chi(\varphi, s) = \boldsymbol{d}(\varphi, s) \cdot \boldsymbol{H}(\varphi, s) \tag{9.43b}$$

This function represents the boundary values of a perturbation potential and is to be introduced into Eq. (9.16). The integral equations derived above can then be applied to this problem.

Janse did not in fact use integral equation techniques but solved the Dirichlet problems corresponding to Eq. (7.10) with vanishing source terms by means of the finite-difference method. By solving Eq. (9.12) with $V \equiv \chi$, however, the required normal derivative

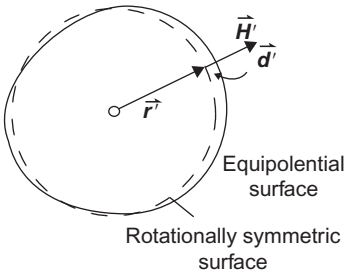


Figure 9.6
Simplified cross-section through an imperfect lens, showing the real and the ideal contour of a polepiece. The vectors $\boldsymbol{r}'$, $\boldsymbol{d}'$ and $\boldsymbol{H}'$ are the projections of the position vector $\boldsymbol{r}$, the shift $\boldsymbol{d}(\boldsymbol{r})$ and the field strength $\boldsymbol{H}(\boldsymbol{r})$, respectively.

$\boldsymbol{H} = -\boldsymbol{n}\sigma$ at the equipotential surfaces is obtained directly, thus saving subsequent differentiation. These simplifications are, of course, not possible if $\sigma \neq \partial\chi/\partial n$.

After solution of the corresponding boundary-value problems, the rotationally symmetric field and the perturbation field are to be superimposed. It is sufficient to limit this to the paraxial domain, whereupon the influence of various kinds of perturbations on the electron trajectories can be studied. The publications of Greenfield and Monastyrskii (2004), Ivanov (2015), Ivanov and Brezhnev (2004), Lencová (2002c), Munro (1988), Murata et al. (1996, 1997, 2001, 2004a,b), Ozaki et al. (1981), Read (1996, 2000, 2004, 2015a), Read and Bowring (1996, 2011), Uchikawa et al. (1981), Watcharotone et al., (2008), Yavor (1993, 1996) and Yavor and Berdnikov (1995) are relevant here.

9.5 Résumé

Without entering into the details of concrete numerical calculations, we have developed a general theory of field calculation in systems having a straight optic axis. Two basic ideas, the introduction of azimuthal Fourier series expansions and the formulation of integral equations, have been worked out in some detail, since these are particularly well adapted to the needs of electron optical field calculations. The use of Fourier series expansions results in a sequence of uncoupled mathematical structures of lower dimensions. Since we shall finally be interested in the field in the paraxial domain, we may terminate the calculation of the Fourier coefficients (axial harmonics) after the first few orders, which are of most importance. This is thus a very economic technique. The use of integral equations rather than partial differential equations further reduces the number of dimensions, since parts of the necessary integrations have already been carried out. In all cases in which the material properties of the polepieces or electrodes are constant, integral equation methods have proved to be very powerful and efficient. There are, of course, problems that cannot be solved in this way. These will be treated in the context in which they arise.

CHAPTER 10

The Boundary-Element Method

In Chapter 9, Integral Equations, we derived various types of such equations. We now turn to their concrete numerical solution. The corresponding procedure, called the integral-equation method (IEM) or the boundary-element method (BEM), has proved to be very powerful. Since the early investigations (Cruise, 1963; Lewis, 1966; Harrington, 1967, 1968; Singer and Braun, 1970), many presentations have been published, which often differ only in minor details. It was introduced into electron optics by Read et al. (1978) and a short survey by Read (2016) summarizes the progress that has been made. We draw attention to some key publications: Kuno and Uchikawa (1985), Costabel (1987), Ströer (1987, 1988), Kuno et al. (1988), Kasper and Ströer (1990), Tsuboi et al. (1990a, b, 1992, 1998, 1999), Watanabe et al. (1990), Binns et al. (1992), Tanaka et al. (1992), Read and Bowring (2011). The list of references given in the bibliography makes no pretence at completeness. We recall that the physical idea behind this method is to calculate the charge density distribution (in the electrostatic case) corresponding to the voltages applied to the electrodes and then calculate the potential distribution in space created by this charge distribution. The magnetic counterpart of this has been fully explored by Murata et al. (1996, 1997, 2001, 2004a,b) and extended to cover saturation by Murata et al. (2016) and Murata and Shimoyama (2017). The complexity of the method in some practical situations is a consequence of the complicated nature of the boundary conditions (Desbruslais and Munro, 1987). We first consider one-dimensional integral equations.

10.1 Evaluation of the Fourier Integral Kernels

The numerical solution of Eqs (9.21) and (9.27) requires the evaluation of the Fourier integral kernels G_m, defined by Eqs (9.18) and (9.19), and of their partial derivatives. The special case $m = 0$ is already familiar in classical electrodynamics, since $G_0(z, r; z', r')$ is the potential at (z, r) when a uniformly charged ring is located at (z', r'). Cases for which $m \neq 0$ have been investigated by Kasper and Scherle (1982).

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00010-9

© 2018 Elsevier Ltd. All rights reserved.

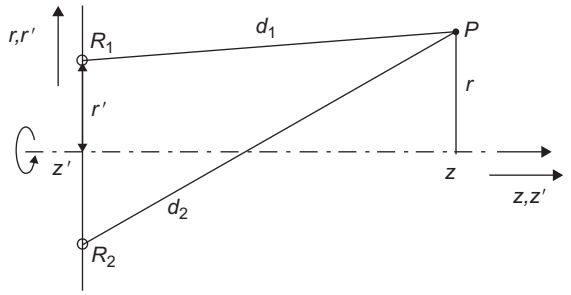


Figure 10.1
In an axial section through an axisymmetric configuration, a single ring is seen as the circles R_1 and R_2. P is an arbitrary reference point.

10.1.1 Introduction of Moduli

The meaning of some of the geometric variables appearing in the subsequent theory is presented in Fig. 10.1. We assume that a ring of radius r' is located in the plane $z' = \text{const}$. The coordinates (z, r) define an arbitrary point in the axial section through the field. It is convenient to introduce the distances

$$d_{1,2} := \sqrt{(z-z')^2 + (r \mp r')^2} \tag{10.1}$$

and the quantities

$$p := d_1 d_2 = [\{(z-z')^2 + r^2 + r'^2\}^2 - 4r^2 r'^2]^{1/2} \tag{10.2}$$

and

$$S := d_1 + d_2 = \sqrt{2}\{p + (z-z')^2 + r^2 + r'^2\}^{1/2} \tag{10.3}$$

These quantities have a very simple geometric meaning: d_1 and d_2 are the minimal and the maximal distance from the ring, respectively, as shown in Fig. 10.1, and S is the sum of these distances. The surfaces $S = \text{const}$ are oblate spheroids with the ring as their common focal line. It is helpful to introduce dimensionless moduli:

$$k := \frac{2\sqrt{rr'}}{d_2}, \quad k' := \sqrt{1-k^2} = \frac{d_1}{d_2} \tag{10.4}$$

$$\rho := \frac{4rr'}{S^2} = \frac{d_2 - d_1}{d_2 + d_1} = \frac{1-k'}{1+k'} \tag{10.5}$$

These moduli are defined only in the interval [0,1]. It is of great importance that the relation between k and ρ is a Landen transformation of the arguments of elliptic integrals (Whittaker and Watson, 1927; Section 22.42), see Section 10.1.3.

10.1.2 Radial Series Expansions

Again using the abbreviations $\boldsymbol{u} = (z, r)$, $\boldsymbol{u}' = (z', r')$, we obtain from Eqs (9.18) and (9.21) the explicit integral representation

$$G_m(\boldsymbol{u};\boldsymbol{u}') = \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos m\alpha \, d\alpha}{\{(z-z')^2 + r^2 + r'^2 - 2rr' \cos\alpha\}^{1/2}} \tag{10.6}$$

Introducing the variables S and ρ, defined by Eqs (10.3) and (10.5), we soon find the simpler representation

$$G_m(\boldsymbol{u};\boldsymbol{u}') = \frac{1}{\pi S}\int_0^{2\pi} \frac{\cos m\alpha \, d\alpha}{(1 - 2\rho\cos\alpha + \rho^2)^{1/2}} \tag{10.7}$$

The required radial series expansion is now most easily obtained by expanding the denominator in terms of Legendre polynomials $P_l(\cos\alpha)$:

$$(1 - 2\,\rho\cos\alpha + \rho^2)^{-1/2} = \sum_{l=0}^{\infty} \rho^l P_l(\cos\alpha)$$

(Whitaker and Watson, 1927, Chapter XV). Integration of each term of the resulting sum yields

$$G_m(\boldsymbol{u};\boldsymbol{u}') = \frac{1}{S}\sum_{j=0}^{\infty} a_{mj}\rho^{m+2j} \tag{10.8}$$

the coefficients being given by

$$a_{mj} = \frac{(2j)!(2m+2j)!}{2^{2m+4j}(j!)^2\{(j+m)!\}^2} \tag{10.9}$$

Since ρ is proportional to r as long as $r \ll r'$, (10.8) essentially represents a series expansion with respect to r. This agrees with our former results that the m^{th} Fourier coefficient of the potential must be proportional to r^m if r is very small (see Section 7.1).

The series expansion Eq. (10.8) converges for all values of ρ in the interval $0 \le \rho < 1$. In practice it can be evaluated from a reasonable number of terms in the interval $0 \le \rho \le 0.5$. The number of terms necessary depends on the acceptable error limit and on the actual value of ρ. With $j_{max} = 20$, excellent accuracy is obtained. Since $\rho \le 0.5$ usually covers the domain occupied by the electron trajectories, Eq. (10.8) is very useful in practical field computations.

10.1.3 Recurrence Relations

The solution of the integral equation derived in Chapter 9, Integral Equations, requires the Fourier kernels G_m to be evaluated in the interval $0 \le \rho < 1$. In the interval $0.5 < \rho < 1$ this

can be performed by means of appropriate recurrence relations. The latter can take different forms, which have been investigated by Kasper and Scherle (1982). For the sake of brevity, we shall treat here only one of these forms.

Introducing the more familiar modulus k Eq. (10.4) into Eq. (10.6) and substituting $\beta = \alpha/2$, we find the representation

$$G_m(\boldsymbol{u};\boldsymbol{u}') = \frac{1}{\pi d_2}\int_0^{\pi/2} \frac{\cos 2m\beta \, d\beta}{(1-k^2\cos^2\beta)^{1/2}}$$

which is clearly a generalization of the familiar complete elliptic integral of the first kind. In fact, the special case $m = 0$ is the well-known formula for the potential of a uniformly charged ring.

For the subsequent discussion, we introduce the complete elliptic integrals

$$\left.\begin{aligned} K(k) &:= \int_0^{\pi/2} (1-k^2\sin^2\beta)^{-1/2}\, d\beta \\ E(k) &:= \int_0^{\pi/2} (1-k^2\sin^2\beta)^{1/2}\, d\beta \\ D(k) &:= \int_0^{\pi/2} \sin^2\beta\,(1-k^2\sin\beta)^{-1/2}\, d\beta \equiv (K-E)k^{-2} \end{aligned}\right\} \tag{10.10}$$

For all values $0 < k < 1$, these can be computed very accurately by means of the algorithm of repeated algebraic and geometric means (see Abramowitz and Stegun, 1965, p. 598). In terms of Eq. (10.10), the lowest order Fourier kernels may be written

$$G_0 = \frac{1}{\pi d_2}K(k), \quad G_1 = \frac{2D(k)-K(k)}{\pi d_2}$$

As it is necessary to introduce the modulus ρ for the evaluation of the radial series expansion, it is convenient to introduce ρ into the integral expression as well; this implies that we have to establish the functions $I_m(\rho)$:

$$G_m(\boldsymbol{u};\boldsymbol{u}') =: S^{-1}I_m(\rho), \quad 0 \le \rho < 1 \tag{10.11}$$

As far as G_0 and G_1 are concerned, this is easily done by means of the Landen transformation

$$K(k) = (1+\rho)K(\rho), \quad E(k) = \frac{2}{1+\rho}E(\rho) - (1-\rho)K(\rho)$$

After some elementary calculations, it is easy to confirm that Eq. (10.11) is satisfied for $m = 0$ and $m = 1$ by

$$I_0(\rho) = \frac{2}{\pi}K(\rho), \quad I_1(\rho) = \frac{2}{\pi}\rho D(\rho) \tag{10.12}$$

The higher order integral expressions are given by the linear recurrence relation ($m \geq 1$)

$$(2m+1)I_{m+1}(\rho) = 2m\left(\rho + \frac{1}{\rho}\right)I_m(\rho) - (2m-1)I_{m-1}(\rho) \tag{10.13}$$

This differs from the expressions given by Kasper and Scherle: here the more convenient normalization $I_0(0) = 1$ has been chosen. Eq. (10.13) can be easily evaluated in ascending sequence. This recursive procedure is slightly unstable. For practical purposes it is stable enough if $\rho \geq 0.5$ and $m \leq 12$. There are ways of extending the region of allowable orders m, but these will not be treated here since $m = 12$ is quite sufficient. Combined with the paraxial series expansion Eq. (10.8) for $\rho \leq 0.5$, the recursive procedure defined by Eq. (10.12) and (10.13) provides a convenient way of computing generalized elliptic integrals.

An interesting complete integral representation of the function $I_m(\rho)$ is the expression

$$I_m(\rho) = \frac{2\rho^m}{\pi}\int_0^{\pi/2} \sin^{2m}\beta\,(1-\rho^2\sin^2\beta)^{-1/2}\,d\beta$$

Obviously Eq. (10.12) are satisfied. By expanding the integrand as a power series and performing the integrations before the summation, it is possible to verify Eqs (10.8) and (10.9). Finally, Eq. (10.13) can be proved by induction.

10.1.4 Analytic Differentiation

It is of great importance in connection with field computation that the derivatives with respect to z and r can also be calculated quite easily. Differentiation of the expressions for S and ρ gives

$$\frac{\partial S}{\partial z} = -\frac{\partial S}{\partial z'} = S(z-z')/p \tag{10.14}$$

$$\frac{\partial \rho}{\partial z} = -\frac{\partial \rho}{\partial z'} = -2\rho(z-z')/p \tag{10.15}$$

$$\frac{\partial S}{\partial r} = \frac{r}{pS}(S^2 - 4r'^2) \tag{10.16}$$

$$\frac{\partial \rho}{\partial r} = \frac{4r'}{S^2}\left(1 - \frac{2r^2}{p} + \frac{8r^2r'^2}{pS^2}\right) \tag{10.17}$$

The derivatives with respect to r' are obtained by interchanging r and r'. On the optic axis ($r = 0$ or $r' = 0$) all these derivatives remain regular. The differentiation of Eq. (10.8) is now a straightforward procedure. For $\rho < 0.5$ the derivatives of Eqs (10.12) and (10.13)

are to be used. From the standard formulae for the derivatives of complete elliptic integrals we have

$$I_0'(\rho) = \frac{\rho I_0(\rho) - I_1(\rho)}{1-\rho^2}, \quad I_1'(\rho) = I_0'(\rho)/\rho \tag{10.18}$$

The differentiation of Eq. (10.13) results immediately in

$$(2m+1)I_{m+1}'(\rho) = 2m\left(\rho + \frac{1}{\rho}\right)I_m'(\rho) - (2m-1)I_{m-1}'(\rho) - 2m\left(\frac{1}{\rho^2} - 1\right)I_m(\rho) \tag{10.19}$$

For $m \geq 1$ this is an inhomogeneous linear recurrence relation, which can easily be solved after evaluating Eq. (10.13).

The derivatives of $G_m(\boldsymbol{u}, \boldsymbol{u}')$ can now be put into a very convenient form. From Eqs (10.14) and (10.15) we find easily

$$\frac{\partial G_m}{\partial z} = -\frac{\partial G_m}{\partial z'} = \frac{z'-z}{pS} Q_m(\rho), \quad Q_m(\rho) := I_m + 2\rho I_m'(\rho) \tag{10.20a}$$

After some elementary calculations we obtain from Eqs (10.16) and (10.17) the result

$$\frac{\partial G_m}{\partial r} = \frac{1}{pS}\left\{\frac{\rho}{r} I_m'(\rho) + (r'\rho - r)Q_m(\rho)\right\} \tag{10.20b}$$

As $r \to 0$ this formula gives a finite result since ρ/r remains finite.

Altogether we obtain a comparatively simple numerical procedure for the computation of the kernel functions $G_m(\boldsymbol{u}, \boldsymbol{u}')$ and their derivatives. This procedure can be carried out for arbitrary values of the arguments with the exception of singular combinations, the latter being defined by $d_1 = 0$ or $k = \rho = 1$.

Computer programs embodying this procedure have been tested and its practical applicability is now well established. In the literature it is usually the special case of rotationally symmetric fields ($m = 0$) that is investigated. Apart from the fact that there is no need to evaluate the recurrence formulae, this specialization brings no major simplification.

10.2 Numerical Solution of One-Dimensional Integral Equations

We now discuss the solution of integral equations such as Eqs (9.21) and (9.27). These are special cases of a Fredholm equation of the second kind:

$$\int_C K(s,s')Y(s')ds' = U(s) + \lambda Y(s) \tag{10.21}$$

Depending on the particular application, the variables appearing in Eq. (10.21) have different physical meanings. In applications to Dirichlet problems (9.21), Y is a surface charge density, U is a surface potential, while λ vanishes. In applications to problems with interface conditions (9.27), Y is a surface potential, U arises from the boundary values of an external field (9.23) and (9.25), while λ is given by (9.14). The variables s and s' are usually arc-lengths in an axial section through the device, but may also be other parameters, such as angles, if these are more favourable. In most applications the contour line C of the boundary consists of two or more separated loops. As a result of using any symmetry properties of the device, these loops may be open or closed, as is shown in Fig. 10.2. In the transformation to a linear scale for s and s', the contours are mapped onto a sequence of disjoint intervals, as is shown in Fig. 10.3. On closed loops the functions appearing in Eq. (10.21) must satisfy cyclic conditions, since they must be unique.

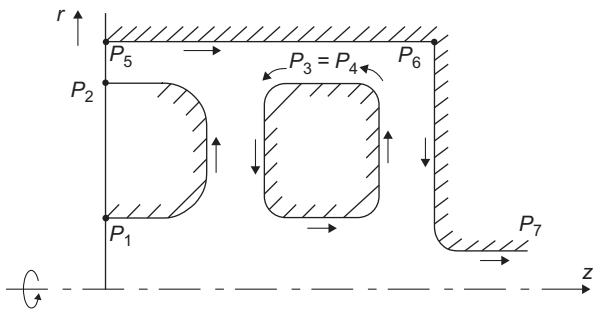


Figure 10.2
Upper-right quadrant of an axial section through a device with axial and mirror symmetry. The contour ***C*** now consists of three separate curves, one closed, the other open. The orientation has been chosen in such a way that the field domain is always on the right as one follows the contour; this makes it easy to distinguish between interior and exterior domains.

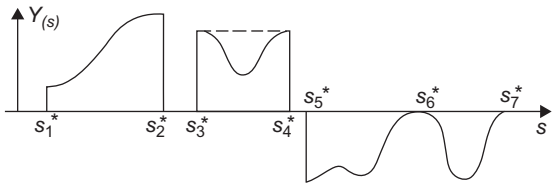


Figure 10.3
The intervals of s and the surface charge distribution corresponding to Fig. 10.2. The three branches of ***C*** now correspond to three separated intervals. The surface charge density $Y(s)$ must satisfy $Y(s_3^*) = Y(s_4^*)$ (cyclicity) and $Y'(s_1^*) = Y'(s_2^*) = Y'(s_5^*) = 0$ (even or mirror symmetry).

10.2.1 Conventional Solution Techniques

The structure of the kernel function $K(s, s')$, is so complicated that Eq. (10.21) can only be solved numerically. In every case, this requires a suitable discretization of the boundary, as is shown in Figs 10.3 and 10.4. The solution technique is generally analogous to that used in quantum mechanics: the unknown function $Y(s)$ is expanded in terms of so-called trial functions $T_k(s)$, $(k = 1 \ldots M)$:

$$Y(s) = \sum_{k=1}^{M} Q_k T_k(s) \tag{10.22}$$

with initially unknown coefficients $Q_1 \ldots Q_M$. This number M must not be smaller than the number N of boundary elements.

The simplest and generally preferred technique is known as *collocation*. Eq. (10.22) is required to satisfy the integral equation (10.21) at the endpoints $s_1 \ldots s_N$ of the boundary elements. This leads to a system of linear equations for the coefficients $Q_1 \ldots Q_N$ with $M = N$:

$$\sum_{k=1}^{N} A_{jk} Q_k = U_j = U(s_j) \qquad (1 \leq j \leq N) \tag{10.23a}$$

$$A_{jk} = \int K(s_j, s') T_k(s') ds' - \lambda T_k(s_j) \tag{10.23b}$$

The integration is to be carried out over the whole boundary C. Usually the trial functions are nonvanishing only in a smaller set of intervals in C. The integration is then limited to these intervals.

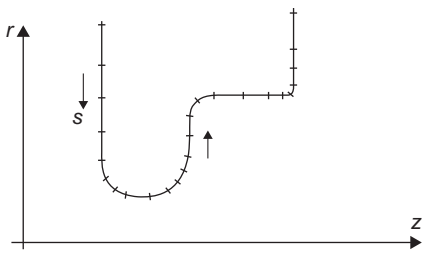


Figure 10.4

Part of a boundary and the appropriate discretization. The line elements must be short in the vicinity of edges and in regions where the boundary curvature is large. Abrupt large changes in the length of the elements should be avoided.

A more advanced approximation is the *Galerkin method*. Now the integral equation is to be satisfied not at the nodes but for the *projections* to the trial functions. Hence Eq. (10.21) is now multiplied in turn by the functions $T_i\,(s)$ and also integrated over the variable s. In this context it is useful to introduce the notation

$$K_{ik} = <T_i|K|T_k> \;\coloneqq \iint T_i(s)K(s,s')T_k(s')\,ds\,ds' \tag{10.24a}$$

$$U_i = <T_i|U> \coloneqq \int T_i(s)U(s)ds \tag{10.24b}$$

$$T_{ik} = <T_i|T_k> \;\coloneqq \int T_i(s)T_k(s)\,ds \tag{10.24c}$$

We then arrive at the following system of linear equations:

$$\sum_{k=1}^{M}(K_{jk} - \lambda T_{jk})Q_k = U_j \qquad (1 \le j \le M) \tag{10.25}$$

This method has some advantages. The most important is that it satisfies the integral equation not at the nodes but *on average,* which implies that the error is minimized. Evidently the matrix in Eq. (10.23a) is asymmetric, while the matrix in Eq. (10.24a) is *symmetric* if this holds for the kernel function $K(s, s')$. This is true for the kernels G_m in Eq. (10.6). Moreover the matrix Eq. (10.24a) is then *positive definite*, at least for $\lambda = 0$, so that the system Eq. (10.25) can be solved by means of the well-known *Cholesky* algorithm. The solution comes close to that of the minimization of a corresponding functional.

Another advantage is the possibility of using more than the minimum number N of trial functions and variables Q_k. A favourable choice is, for instance, the use of function values $Y(s_n)$ and derivatives $Y'(s_n)$ at the nodes, hence $M = 2\,N$. In each boundary element the function $Y(s)$ can then be approximated by a cubic Hermite polynomial, which is fairly accurate. Each node has two degrees of freedom, which gives a better result than the simple one-degree approximation.

The main disadvantage of the Galerkin method is the need to carry out *double* integrations over logarithmic singularities, while single integrations are already difficult enough. This problem was solved by Ströer (1987) and reported by Kasper (2001, Section 6.3.3). For reasons of conciseness we cannot deal with this topic here and refer to these publications.

Analogous techniques exist for the boundary-element method in three dimensions. All the variables then become functions of two surface coordinates and hence all the integrals are double. Full flexibility is possible only if the surface elements are general triangles. This is beyond the scope of the volume.

10.2.2 The Charge Simulation Method

In order to circumvent the complicated evaluation of the improper integrations in the diagonal elements H_{ii}, Hoch et al. (1978) have replaced the surface charge distribution by a sequence of charged rings located *inside* the electrodes, as shown in Fig. 10.5. In addition, they introduced charged thin planar apertures. We postpone consideration of these to Section 10.3. The rings are characterized by coordinates $z' =: z_k^0$, $r' =: r_k^0$ and charges Q_k ($k = 1 \ldots N$). At the boundary C an equal number of control points with coordinates $z =: z_i$, $r =: r_i$ and boundary potentials U_i are introduced. Apart from an unimportant constant factor, the conditions that the rotationally symmetric potential $V(z, r)$ has the prescribed boundary values are of the form

$$\sum_{k=1}^{N} Q_k G_0(\boldsymbol{u}_i, \boldsymbol{u}_k^0) = U_i, \quad i = 1 \ldots N \tag{10.26}$$

(we again use $\boldsymbol{u} = (z, r)$, $\boldsymbol{u_i} = (z_i, r_i)$, $\boldsymbol{u}_k^0 = (z_k^0, r_k^0)$). Once the solution of this linear system of equations is known, the potential and its gradient are simply given by

$$V(\boldsymbol{u}) = \sum_{k=1}^{N} Q_k G_0(\boldsymbol{u}; \boldsymbol{u}_k^0) \tag{10.27a}$$

$$\nabla V(\boldsymbol{u}) = \sum_{k=1}^{N} Q_k \nabla G_0(\boldsymbol{u}; \boldsymbol{u}_k^0) \tag{10.27b}$$

This is certainly the simplest version of the boundary element method. Kasper and Scherle (1982) have generalized it for Fourier series expansions like Eq. (9.15). In many practical applications to conventional round lenses and to deflection units this method has worked quite satisfactorily, the relative error in the field in its paraxial domain being roughly 10^{-3} to 5×10^{-3}. Difficulties may arise in the vicinity of sharp edges and especially at the

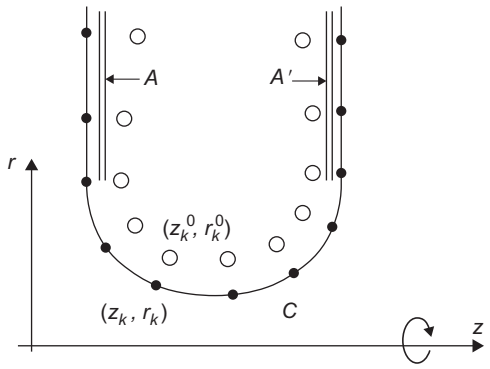


Figure 10.5

Part of an electrode, showing the boundary curve C and a possible configuration of control points (z_k, r_k) and ring positions (z_k^0, r_k^0). A and A' are the upper halves of meridional sections through thin charged plane apertures.

surfaces of cathode tips in electron guns. Depending on the shape of the boundaries and on the prescribed boundary values, the equipotential over which the expression Eq. (10.27a) assumes a constant electrode potential U has a wavy structure, as shown in Fig. 10.6. This causes a deviation of the local field strength from its correct direction parallel to the local surface normal of the electrode. In order to avoid this weakness, the charge can be distributed over an inner surface, as shown in Fig. 10.7.

It is sufficient to assume constant charge density on each surface element. Replacing the summation in Eq. (10.26) by the appropriate integration, one obtains quite an accurate solution. Weyßer (1983) has applied this method to electron guns and found that it works reasonably well.

In spite of its wide field of applications, the refined method of Hoch et al. is not quite satisfactory in every respect because it cannot be employed to solve problems with interface conditions, since the singularities must then be located at the interfaces themselves. In this case, the corresponding integration over the singularities cannot be circumvented. In the next section we shall present a better solution.

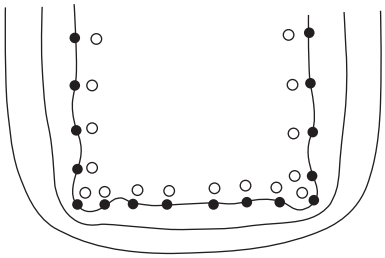


Figure 10.6
Wavy structure of equipotential lines produced by disjoint rings. This effect, exaggerated here for clarity, decreases rapidly with increasing distance from the surface.

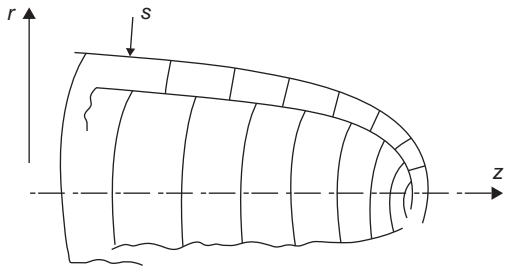


Figure 10.7
Simplified perspective drawing of the discretization adopted for the cathode in an electron gun. Each control point at the physical surface S is associated with an interior conical mantle over which the corresponding part of the surface charge is to be distributed uniformly or, better, continuously.

The cases of simplified field calculation given here are examples of a more general method, the charge simulation method. Any distribution of charges or currents, the field of which can be calculated analytically, can be considered if the singularities are located far enough outside the domain of particle trajectories. Hence not only charged rings but also point charges, charged rods, plates or discs can be used to simulate the field in the particle domain. The method can also be improved by a least-squares fit, which minimizes the approximation error (see e.g., Kasper (2001, Ch. 7.3)). Such a method is necessary in the treatment of *electron guns*, as it is practically impossible to start electron trajectories accurately close to singularities, if the cathode surface is a part of the charge distribution determined by the BEM.

10.2.3 Combination with Interpolation Kernels

Apart from the special case of electron guns, it is favourable to locate the singularities on the material surfaces, and for the calculation of magnetic fields this is usually essential. In order to obtain sufficient accuracy, it is then necessary to use a better than piecewise linear approximation for the surface sources, which can be expressed only by piecewise polynomial surface functions. Since the physical source distribution is continuous at the boundary nodes, the same should hold also for the simulated function and its derivative. This renders the calculation more complicated. There are different ways of achieving this. One method has already been mentioned in Section 10.2.1: the use of piecewise cubic Hermite polynomials. The rank of the system matrix is then $2N$. An alternative is to eliminate the derivatives by means of finite-difference formulae, so that the system has the lowest possible rank N. This approach was chosen by Kasper (2001, Section 3.3) who used so-called *modified interpolation kernels*. These are a sequence of polynomials, defined over four or more intervals and joined smoothly together at the nodes. We shall present here only the third order, the lowest nontrivial one. The method can be extended to higher orders.

10.2.3.1 General formalism

Consider four successive pairs (x_n, y_n), $(1 \le n \le 4)$ of sampling data of a function $y(x)$ to be interpolated, as is sketched in Fig. 10.8. The reference abscissa x should be located in the *central interval*, $x_2 \le x \le x_3$. We define the interval lengths

$$h_n = x_{n+1} - x_n \qquad (n = 1, 2, 3) \tag{10.28a}$$

and the relative abscissa

$$t = (x - x_2)/h_2 \quad (0 \le t \le 1).$$

We now start from the well-known cubic Hermite polynomial

$$y(x) = y_2 p_1(t) + y_3 p_2(t) + h_2 y'_2 q_1(t) - h_2 y'_3 q_2(t) \tag{10.28b}$$

with the four normalized cubic polynomials

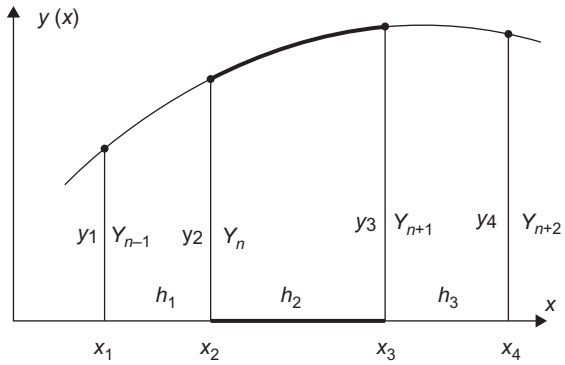


Figure 10.8
Configuration of the variables for the Hermite interpolation (10.28). The bold line indicates the interval of interpolation. The notations on the function values indicate their local and global numbering.

$$\begin{aligned} p_1(t) &= (1+2t)(1-t)^2, \qquad p_2(t) = t^2(3-2t) = 1 - p_1(t), \\ q_1(t) &= t(1-t)^2, \qquad q_2(t) = t^2(1-t) = t(1-t) - q_1(t) \end{aligned} \tag{10.29}$$

It is straightforward to verify that this is correct for any boundary values y_2, y_3, y_2' and y_3'. These functions are to be used also in other contexts in which Hermite interpolations are required, for instance in the case mentioned in Section 10.2.1.

The unknown derivatives y_2' and y_3' are now approximated by the central three-point differences, valid for a locally parabolic function:

$$y_n' = \{h_{n-1}^2(y_{n-1} - y_n) + h_n^2(y_n - y_{n-1})\}/\{h_{n-1}(h_{n-1} + h_n)h_n\}, \quad (n = 2, 3)$$

After introducing this into Eqs (10.28) and collecting all terms which refer to the same sampling point, we obtain

$$y(x) = \sum_{n=1}^{4} y_n f_n(t) \tag{10.30}$$

with the kernel functions

$$\begin{aligned} f_1(t) &= -\frac{h_2^2 q_1(t)}{h_1(h_1 + h_2)} \\ f_4(t) &= -\frac{h_2^2 q_2(t)}{h_3(h_3 + h_2)} \\ f_2(t) &= p_1(t) + q_1(t)\left(\frac{h_2}{h_1} - 1\right) + \frac{q_2(t)h_3}{h_2 + h_3} \\ f_3(t) &= p_2(t) + q_2(t)\left(\frac{h_2}{h_3} - 1\right) + \frac{q_1(t)h_1}{h_1 + h_2} \end{aligned} \tag{10.31}$$

The resulting function $y(x)$ has continuous derivatives of first order, as is evident from its construction. It is exact for any quadratic polynomial and fairly accurate for cubic ones, as near the midpoint the error changes its sign, so that it largely cancels out in integrals over the interval.

This algorithm can be executed for all inner intervals, when the local labels, used here, are adjusted to the global ones, moving from A to B. In order to avoid confusion, we denote local data by x and y, as above, while global ones are denoted by X and Y, hence $x_n = X_{k+n-2}$, $y_n = Y_{k+n-2}$, $(n = 1\ldots4)$ for the application to the interval (X_k, X_{k+1}), $(A < k < B-1)$ (see Fig. 10.8). However, special formulae are necessary for positions near the global endpoints X_A and X_B.

10.2.3.2 Marginal positions

If all functions are *periodic* (also with a possible leap), then values that would be out of range can be removed by setting

$$Y_{A-k} = Y_{B-k} + Y_A - Y_B, \quad Y_{B+m} = Y_{A+m} + Y_B - Y_A \tag{10.32a}$$

and analogously for X and any other periodic function.

If the function has a *local extremum* at the lower endpoint A, the value Y_{A-1} is missing from the global representation and $y_2' = 0$ in Eq. (10.28b). This leads to

$$q_1 = f_1 = 0 \tag{10.32b}$$

in all functions in Eq. (10.31). Likewise an extremum on the other margin leads to $y_3' = q_2 = f_4 = 0$. If neither of these conditions arises, then a *linear approximation* is reasonable. On the lower margin we then set $y_2' = (y_3 - y_2)/h_2$ in Eq. (10.28b), whereupon we find

$$\begin{aligned} f_1 &= 0 \\ f_2 &= p_1 - q_1 + \frac{q_2 h_3}{h_2 + h_3} \\ f_3 &= p_2 + q_1 + \frac{q_2(h_2 - h_3)}{h_3} \\ f_4 &= -q_2 \frac{h_2 - h_3}{h_3} \quad \text{as in} \quad (10.31) \end{aligned} \tag{10.32c}$$

Similar relations hold for a linear function on the upper end. With a suitable combination of these conditions quite general interpolation problems can be solved. We have thus obtained a general way to discretize functions, and this gives us the possibility of reducing the determination of an unknown function to the solution of a system of equations for its sampling values.

10.2.3.3 General properties

This interpolation simplifies considerably in the case of equidistant intervals, of length h. The kernels then become *symmetric* functions, $\boldsymbol{F}(-u) = \boldsymbol{F}(u)$. Moreover they are identical in shape apart from the necessary shift. The interpolation then takes the form $X_n = X_0 + nh$ and

$$y(x) = \sum_n \boldsymbol{F}\left(\frac{x - X_n}{h}\right) \tag{10.33}$$

With $u := |\,(x - X_n)/h\,|$ this kernel function is then given by

$$\begin{aligned} \boldsymbol{F}(u) &= 1 - 2.5u^2 + 1.5u^3 \qquad (0 \le u \le 1) \\ \boldsymbol{F}(u) &= 2 - 4u + 2.5u^2 - 0.5u^3 \qquad (1 \le u \le 2) \\ \boldsymbol{F}(u) &= 0 \quad \text{for } u \ge 2. \end{aligned} \tag{10.34}$$

Obviously the interpolation condition $\boldsymbol{F}(k) = \delta_{k,0}$ (k an integer) is satisfied, as it must be. It is of importance that the function $\boldsymbol{F}(u)$ has the following integral properties

$$\int_{-\infty}^{\infty} \boldsymbol{F}(u)u^n du = \delta_{n,0} \quad (n \le 3) \qquad \int_{-\infty}^{\infty} \boldsymbol{F}(u)u^4 du = -0.3 \tag{10.35a}$$

Hence for any cubic polynomial $P(x)$ we obtain exactly

$$\int_{-\infty}^{\infty} P(x)\boldsymbol{F}\left(\frac{x - X_k}{h}\right) dx = hP(X_k) \tag{10.35b}$$

This shows that for decreasing length h the function $h^{-1}\,\boldsymbol{F}(u)$ approaches Dirac's delta function.

Integration over piecewise cubic polynomials as given by Eq. (10.33) or (10.30) with constant h and the cyclic condition $Y_B = Y_A$ leads to the well-known Maclaurin integral formula

$$\int_A^B Y(X)dX = h\sum_{A+1}^{B} Y_n \tag{10.35c}$$

the accuracy of which is better than fourth order for a periodic function. This concept can be extended to polynomials of higher orders. With $2M$ intervals of definition for $\boldsymbol{F}(u)$ (here $M = 2$) interpolation kernels of order $2M - 1$ are obtained which are $(M - 1)$ times continuously differentiable. The first nonvanishing term in (10.35b) is then of order $2M$ instead of 4. This means that Eq. (10.35c) becomes more accurate, and the approximation to Dirac's function becomes better. The price for this gain is stronger oscillations in the outer intervals, as can be seen in Fig. 10.9. In order to define the kernel of next higher order $M = 3$ it is favourable to introduce the function

$$G(w) = w^3(7 - 12w + 5w^2)/24$$

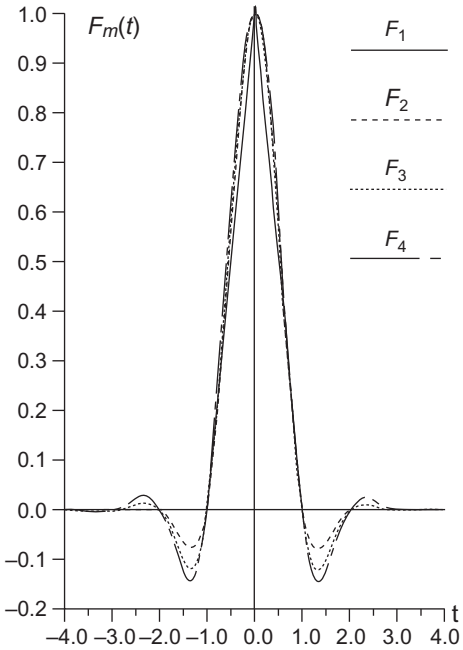


Figure 10.9
Modified interpolation kernels up to order $M = 4$ (after Kasper (2001, p. 89)).

The kernel, a system of polynomials of fifth power, is then given by

$$\begin{aligned}
&\boldsymbol{F}(u) = w(4 + 4w - w^2 - w^3)/6 + 10\,G(w), && (0 \le |u| \le 1, w = 1 - u)\\
&\boldsymbol{F}(u) = -\,w(2 + w - 2w^2 - w^3)/24 - 5\,G(w), && (1 \le |u| \le 2, w = 2 - u)\\
&\boldsymbol{F}(u) = G(w), && (2 \le |u| \le 3, w = 3 - u)\\
&\boldsymbol{F}(u) = 0 && (|u| \ge 3).
\end{aligned} \tag{10.36}$$

(By explicit differentiations and evaluations for the arguments $u = 0, 1, 2, 3$ and $w = 0, 1$ it is straightforward to verify that all continuity conditions are satisfied and the coefficients of five-point finite differences are assumed at $u = 0, 1, 2$.)

The use of this function will certainly provide more accurate results but its application requires the introduction of a parameter transform to a grid with *equal interval lengths*, which can become awkward. The use of asymmetric five-point finite differences for the derivatives is possible in principle but so very complicated that this is not recommended.

10.2.3.4 Solution of integral equations

We now return to the solution of the integral equation (10.21). The functions $T_k\,(s)$ in Eq. (10.22) are now to be identified with the functions $\boldsymbol{F}(u)$ with $u = s - s_n$. The unknown variables Q_k are here the sampling values Y_k of the function $Y(s)$ at the nodes, as is obvious.

For reasons of conciseness we shall present the method, as we did earlier, for only one coherent boundary, the labels being $A = 1$, $B = N$, $1 \leq k \leq N-1$. A generalization to more disjunct parts, as is usually necessary, provides no problem. The method, outlined below is a *collocation* technique. Formulation as a *Galerkin* method is possible.

Let us consider a boundary interval $s_k \leq s \leq s_{k+1}$ $(1 < k < N-1)$. Then, in order to apply the formulae derived above, the interval lengths

$$h_n = s_{k+n-1} - s_{k+n-2} \quad (n = 1, 2, 3)$$

(see Eq. 10.28a) are required. The function $Y(s)$ is then interpolated by

$$Y(s_k + h_2 t) = \sum_{n=1}^{4} Y_{n+k-2} f_n(t)$$

(see Eq. 10.30). Introduction into Eq (10.21) with Eqs (10.29) and (10.31) and evaluation at $s = s_i$ leads to

$$\int K(s_i, s')Y(s')ds' = U(s_i) + \lambda Y_i \qquad (i = 1, \ldots, N)$$

$$\sum_{k=1}^{N-1} \left\{ \int_0^1 \left\{ K(s_i, s_k + h_2 t) h_2 \sum_{n=1}^{4} Y_{n+k-2} f_n(t) dt \right\} \right. = U(s_i) + \lambda Y_i \qquad (i = 1, \ldots, N)$$

with $h_2 = s_{k+1} - s_k$. Each local variable depends on the integration interval with label k and must hence be recalculated for each such interval. The same is true for the sequence below. At the endpoints $k = 1$ or $k = N-1$, some terms are out of range; these terms are to be eliminated by use of the special boundary conditions Eq. (10.32). This is implicit also in the scheme outlined further below.

From the upper relation it turns out that a linear system of equations of the form Eq. (10.23) is obtained. The matrix elements are most favourably determined by *successive summation*, as is common in numerical calculation. In this presentation $\coloneqq$ means, as usual, replacement by the right hand side of a statement. The method proceeds as follows:

$$\begin{array}{lll}
 & \textit{for } (i = 1 \text{ to } N) & \\
\{ & \textit{for } (j = 1 \text{ to } N) \quad \{A_{ij} = -\lambda \delta_{ij}\} & \text{(initialization)} \\
 & \textit{for } (k = 1 \text{ to } N-1) & \\
 & \{ \textit{for } (n = 1, 2, 3) \quad \{h_n = s_{k+n-1} - s_{k+n-2}\} & \text{(Eq. 10.28a)} \\
 & \textit{for } (n = 1, 2, 3, 4) & \\
 & \{ j \coloneqq k + n - 2, \quad A_{ij} \coloneqq A_{ij} + h_2 \int_0^1 K(s_i, s_k + h_2 t)\ f_n(t)\ dt\} & \\
 & \} \text{ (end of loop over } k) & \\
\} & \text{(end of loop over } i) &
\end{array} \qquad (10.37)$$

In the loops over the label n it must be remembered that, on endpoints, the terms with $n = 1$ or $n = 4$ are missing and the functions $f_n\,(t)$ must then be modified according to the corresponding boundary conditions.

Thereafter the system Eq. (10.22) of linear equations is to be solved for $Y_1 \ldots Y_N$. We must assume here that λ is not an eigenvalue of the matrix, though the integration technique would still be valid. The interpolation technique explained above can then be applied to obtain a smooth function $Y(s)$.

This is the most general case, as it allows *unequal lengths* of the subintervals without a parameter transform of the variable s. Generally the necessary integrations are to be carried out numerically. Only in the case when the positions corresponding to the labels i and k are widely separated may simplifications be possible. The integration over an interval adjacent to a singularity has to be performed by asymmetric special Gauss-quadrature-like sum formulae. Their positions and weights are derived by Ströer (1987) and reprinted in Kasper (2001, p. 432) (see next section). The method furnishes acceptable results, provided that the discretization of the integration intervals is chosen reasonably. For reasons of conciseness we have tacitly assumed only one domain of integration. The generalization to several disjunct domains, as is sketched in Fig. 10.2, creates no problem.

10.2.3.5 Application to field calculations

Only the case of a rotationally symmetric electric potential $V(\boldsymbol{u}_0)$ with $\boldsymbol{u}_0 = (z_0, r_0)$ is dealt with here. We have then to consider the kernel function $K\,(s, s') \coloneqq G_o\,(\boldsymbol{u}_0, \boldsymbol{u}(s'))$, and $Y(s')$ has the physical meaning of a surface charge density $\sigma(s')$. The integral to be evaluated then becomes

$$V(\boldsymbol{u}_0) = \int_C G_o(\boldsymbol{u}_0, \boldsymbol{u}(s'))\sigma(s')ds' \tag{10.38}$$

with a corresponding formula for the gradient. The procedure is analogous to the scheme in Eq. (10.37) with the modifications that the loop over i has to be omitted and that the sampling values of the surface charge density have to be considered. The scheme becomes then

$$\begin{aligned}
&V(\boldsymbol{u}_0) = 0 \qquad \text{(initialization)}\\
&\textit{for}\ \ (k = 1\ \text{ to }\ N - 1)\\
\{\quad &\textit{for}\ \ (n = 1, 2, 3)\ \ \{h_n = s_{k+n-1} - s_{k+n-2}\} \quad \text{(Eq. (10.28}a\text{))}\\
&\textit{for}\ \ (n = 1, 2, 3, 4)\\
&\{V(\boldsymbol{u}_0)\ \coloneqq\ V(\boldsymbol{u}_0) + h_2 \int_0^1 G_0\,(\boldsymbol{u}_0, \boldsymbol{u}\,(s_k + h_2 t)\,)\ f_n(t)\ \ \sigma_{k+n-2} dt\}\\
\}\quad &\text{(end of loop over } k)
\end{aligned} \tag{10.39}$$

In the same manner the components of grad $V(\boldsymbol{u})$ can also be calculated, even together in the same loop. This method can be applied for the determination of the field at any position in space, even on the surfaces. In the latter case the necessary integration over singularities may become tedious but is always possible. Sufficiently far from all singularities the integration can approximately be reduced to mere summations, as in Eq. (10.27a), but now the points $\boldsymbol{u}_k$ are located at the surfaces. The method, outlined briefly here, can be improved in very many respects. These cannot be dealt with here for reasons of space. For more details we refer to Kasper (2001, pp. 86−96 and Chapter VI) where references to other work can be found.

10.2.4 Evaluation of Improper Integrals

A careful inspection of the kernel functions to be evaluated shows that these contain logarithmic singularities; the derivatives of the kernels may have singularities of first order. For the sake of brevity, we assume the abscissa of the singularity in question to be $x = 0$.

A singular function of the type

$$f_1(x) = \ln|\varphi_1(x)| \quad \text{with} \quad \varphi_1(0) = 0, \quad \varphi_1'(0) \neq 0$$

can be reduced to the form

$$f_1(x) = \ln|x| + \ln|\varphi_1(x)/x|$$

where the second term on the right is a *regular* function, which can be expanded as a Taylor series about $x = 0$. Likewise, a function of the type

$$f_2(x) = 1/\varphi_2(x) \quad \text{with} \quad \varphi_2(0) = 0, \quad \varphi_2'(0) \neq 0$$

can be reduced to

$$f_2(x) = \frac{1}{x\varphi_2'(0)} + \frac{x\varphi_2'(0) - \varphi_2(x)}{x\varphi_2'(0)\varphi_2(x)}$$

where again the second term is regular, since both numerator and denominator have a common zero of second order, which cancels out. It is therefore sufficient to consider functions of the form

$$F(x) = f_1(x)\ln|x| + f_2(x)/x + f_3(x) \tag{10.40}$$

$f_1(x)$, $f_2(x)$, and $f_3(x)$ being arbitrary *regular* functions.

Kasper (1983) has developed a simple method for integration over such functions. The corresponding formula is similar to a Gauss quadrature and takes the basic form

$$\int_{-h}^{h} F(x)dx = h\sum_{\mu=1}^{N} w_\mu\{F(p_\mu h) + F(-p_\mu h)\} + O(h^{2N+1}) \tag{10.41}$$

with positive abscissae p_μ and weight factors w_μ. Thanks to the symmetric arrangement of these, the contributions from the singularity in $f_2(x)/x$ and from all antisymmetric terms cancel out, and Eq. (10.41) automatically furnishes the principal value of the integral. The parameters p_μ and w_μ are obtained by numerical solution of the nonlinear equations

$$\left.\begin{aligned} \int_0^1 u^{2\nu}du &= \sum_{\mu=1}^{N} w_\mu p_\mu^{2\nu} = \frac{1}{2\nu+1} \\ \int_0^1 u^{2\nu}\ln\frac{1}{u}du &= \sum_{\mu=1}^{N} w_\mu p_\mu^{2\nu}\ln\frac{1}{p_\mu} = \frac{1}{(2\nu+1)^2} \end{aligned}\right\} \nu = 0, 1 \ldots N-1$$

The results for $N = 4$, which are adequate in most cases, are as follows:

$$\begin{aligned} p_1 &= 0.0399\ 4596\ 2203 & w_1 &= 0.1270\ 7679\ 2574 \\ p_2 &= 0.2801\ 7249\ 6204 & w_2 &= 0.3267\ 4417\ 6078 \\ p_3 &= 0.6361\ 2394\ 4954 & w_3 &= 0.3523\ 4912\ 8452 \\ p_4 &= 0.9223\ 6045\ 1138 & w_4 &= 0.1938\ 3290\ 3896 \end{aligned}$$

It is now easy to integrate all kernel functions $K(s, s')$ and their normal derivatives over arbitrary smooth boundary contours C. It is necessary neither to approximate the contour by a polygon, nor to choose the arc-length as the variable of integration. As Eq. (10.41) makes clear, there is no need to express the integrand explicitly in the form Eq. (10.40), which would be extremely tedious. The only knowledge required is the *implicit* singular character of the integrand. The user of this method need only set up a program that supplies correctly the value of the integrand for arbitrary arguments.

The symmetric quadrature formulae are favourable because antisymmetric terms cancel out automatically. But they are not useful in every case, for instance, not for the calculation of the integrals appearing in Eq. (10.37) or (10.39). In order to evaluate such integrals, Ströer (1987) has compiled quadrature formulae for asymmetric integrals. Here we shall reproduce them only in the lowest order. For functions of the general form

$$F(x) = A(x) + B(x)\ln(x) \quad (0 < x \leq 1)$$

with *regular* coefficients $A(x)$ and $B(x)$ we have

$$\int_0^1 F(x)dx = \sum_{n=1}^{6} g_n F(x_n)$$

The abscissae x_n and weights g_n appearing in this formula are given in the following table:

n	x_n	g_n
1	0.003025 80213 75463	0.011351 33881 72726
2	0.040978 25415 59506	0.075241 06995 49165
3	0.170863 29552 68773	0.188790 04161 54163
4	0.413255 70884 47932	0.285820 72182 72273
5	0.709095 14679 06286	0.284486 42789 14088
6	0.938239 59037 71671	0.154310 39989 37584

Analogous formulae for 9 and 15 points have also been compiled by Ströer and are reproduced in Kasper (2001, p. 432). After a suitable transformation of the variables, these data can also be used for calculation of improper integrals with a singularity at $x = 1$.

In conclusion, it emerges that the BEM, at least in its one-dimensional version, is a highly attractive procedure. Practical examples demonstrating this are given in Section 10.5.

10.3 Superposition of Aperture Fields

Though the general method described above can be applied to any axisymmetric Dirichlet problem, this is not always the best way of obtaining the solution. Problems may arise in systems with very narrow apertures in electrodes with plane fronts, as shown in the example presented in Fig. 10.10. Here many integration subintervals are necessary in order to obtain the electric field correctly in its asymptotic domains. The computation of surface charge distributions can, however, be confined to the principal inhomogeneous domains of the field if the asymptotic field is represented correctly by appropriate superpositions of the fields of thin plane circular apertures. A combination of the BEM with such fields has been proposed by Hoch et al. (1978). Pure superposition of aperture fields had been investigated earlier by Regenstreif (1951), Lenz (1956) and Dommaschk (1965). In the following

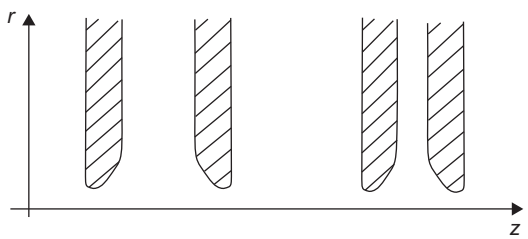


Figure 10.10
Upper half of the axial section through an accelerator.

analysis, we shall first study the field of one single aperture and then superpositions of such fields. For a detailed study, see Becker et al. (2004).

10.3.1 Electric Field of a Single Aperture

We consider now a thin plane circular aperture with radius $r' = R$. Without loss of generality, we can choose the coordinate system in such a manner that the aperture plane is given by $z' = 0$. It is appropriate to introduce oblate spheroidal coordinates (u, v, φ,), φ being the usual azimuth and u, v defined by the transformation

$$\begin{gathered} z = Ruv, \quad r = R\{(1+u^2)(1-v^2)\}^{1/2} \\ (-\infty < u < \infty, \quad 0 \le v \le 1) \end{gathered} \tag{10.42}$$

These are shown in Fig. 10.11. The surfaces $u = \text{const}$ are confocal oblate spheroids, the surfaces $v = \text{const}$ confocal orthogonal hyperboloids. Among the latter, the optic axis ($v = 1$) and the surface of the aperture itself ($v = 0$) are degenerate special cases.

In these new coordinates, Laplace's equation takes the form

$$\frac{\partial}{\partial u}\left\{(1+u^2)\frac{\partial \Phi}{\partial u} + \frac{\partial}{\partial v}(1-v^2)\frac{\partial \Phi}{\partial v}\right\} = 0$$

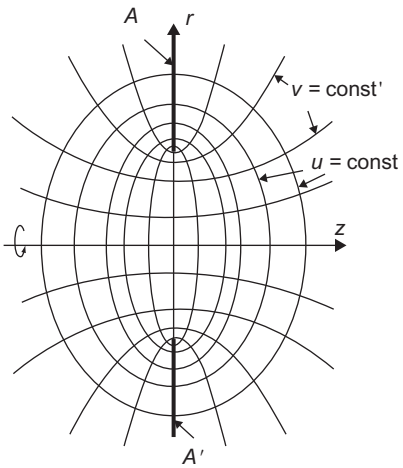


Figure 10.11
Coordinate lines $u = \text{const}$ and $v = \text{const}'$ in a system of oblate spheroidal coordinates (u, v, φ). The azimuth φ is constant in this axial section. A and A' denote the closed parts of the aperture plane; these are singularities of the coordinate system.

This can be solved by separation of variables,

$$\Phi(u, \upsilon) = F(u)P(\upsilon)$$

Writing the separation constant in the familiar form $l(l+1)$, we obtain

$$\frac{d}{d\upsilon}\{(1-\upsilon^2)P'(\upsilon)\} = -\,l(l+1)P(\upsilon) \tag{10.43}$$

$$\frac{d}{du}\{(1+u^2)F'(u)\} = -\,l(l+1)F(u) \tag{10.44}$$

Eq. (10.43) is Legendre's differential equation, its *regular* solutions being the well-known Legendre polynomials $P_l(\upsilon)$ for *integral values* of the subscript l. Eq. (10.44) reduces to Eq. (10.43) if we write $u' = \mathrm{i}u$ and is hence solved by general Legendre functions with imaginary argument. For physical reasons the resulting solution for Φ must correspond asymptotically to a homogeneous field, which means that $\Phi \sim z = Ru\upsilon$ for $|u| \gg 1$. With this constraint, the general solution is

$$\Phi(u, \upsilon) = Au\upsilon + B\upsilon(1 + \upsilon \arctan u) + C$$

The coefficients A, B and C are uniquely specified by the conditions $\Phi(u, 0) = \Phi_0$, $\partial\Phi/\partial z = -E_l$ for $z \to -\infty$ and $\partial\Phi/\partial z = -E_r$ for $z \to +\infty$, the result being

$$\Phi(u, \upsilon) = \Phi_0 + R\upsilon\left\{-\frac{E_l + E_r}{2}u + \frac{E_l - E_r}{\pi}(1 + u \arctan u)\right\} \tag{10.45}$$

The constants Φ_0, E_l and E_r have the physical meaning of the aperture potential and the asymptotic field strengths, respectively. Examples of such solutions, differing in the choice of the constants, are shown in Fig. 10.12A and B.

In order to represent the potential and the asymptotic field strength in cylindrical coordinates, we need the inverse transformation corresponding to Eq. (10.42). This can be expressed in terms of the distances d_1 and d_2, defined by Eq. (10.1) and shown in Fig. 10.1; in this context, the singular ring is the edge of the aperture, $z' = 0$, $r' = R$. We then have

$$u^2 = \frac{(d_1 + d_2)^2}{4R^2} - 1, \quad \upsilon^2 = 1 - \frac{(d_2 - d_1)^2}{4R^2} \tag{10.46}$$

For the later computations, it is more convenient to express the transform in *relative* coordinates $\bar{z} = z/R$, $\bar{r} = r/R$; we find

$$\xi := \bar{z}^2 + \bar{r}^2 - 1 \equiv u^2 - \upsilon^2 \tag{10.47a}$$

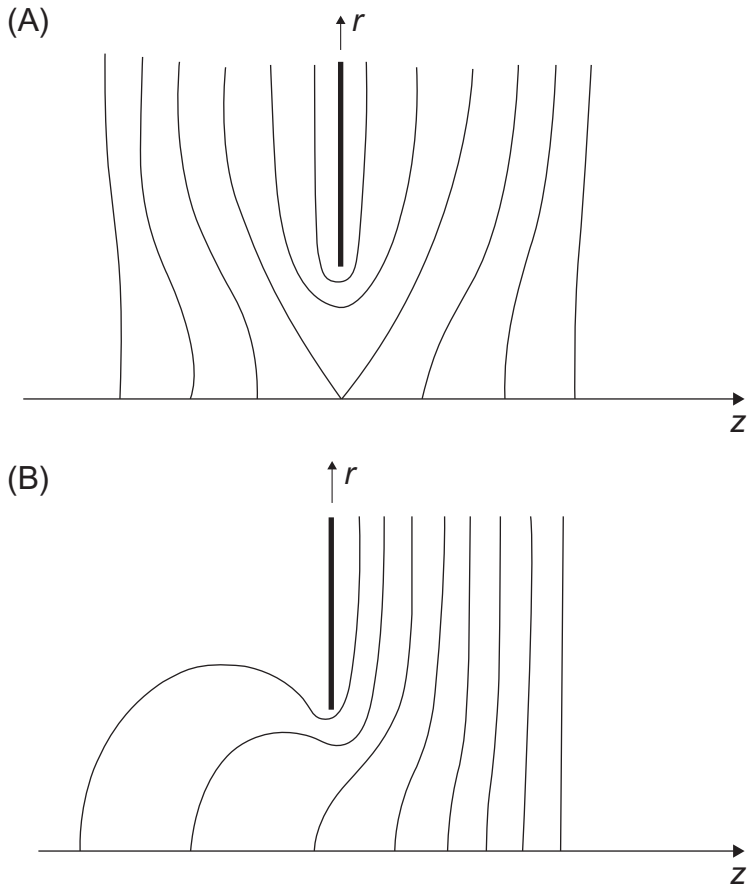


Figure 10.12
Equipotentials in the upper half of an axial section through the field of a thin charged aperture. (A) Symmetric field, $\boldsymbol{E_1 = -E_2}$; (B) $\boldsymbol{E_1} = 0$, the field vanishes asymptotically as $z \to -\infty$.

$$\eta := \sqrt{\xi^2 + 4\overline{z}^2} \equiv u^2 + \upsilon^2 \tag{10.47b}$$

$$\upsilon = \begin{cases} \sqrt{(|\xi| + \eta)/2} & \text{for } \xi < 0 \\ |\overline{z}|/\sqrt{(|\xi| + \eta)/2} & \text{for } \xi \geq 0 \end{cases} \tag{10.47c}$$

$$u = \overline{z}/\upsilon \tag{10.47d}$$

The matrix of the partial derivatives can then be computed very easily:

$$w := (R\eta)^{-1} \equiv \{R(u^2 + \upsilon^2)\}^{-1} \tag{10.48a}$$

$$\frac{\partial u}{\partial z} = wv(1+u^2), \quad \frac{\partial u}{\partial r} = wu\bar{r} \tag{10.48b}$$

$$\frac{\partial v}{\partial z} = wu(1-v^2), \quad \frac{\partial v}{\partial r} = -wv\bar{r} \tag{10.48c}$$

By means of these relations it is easy to compute $\Phi(z, r)$ and its derivatives $\partial\Phi/\partial z$, $\partial\Phi/\partial r$ or even derivatives of higher orders. The corresponding expressions are not given here.

10.3.2 Superposition Procedure

We now consider configurations of coaxial thin apertures and their fields. Any such configuration of N apertures is uniquely specified by the aperture positions z_i, the radii R_i and the physical surface potentials U_i, $i = 1, 2, \ldots N$, and also by the asymptotic field strengths E_l for $z \to -\infty$ and E_r for $z \to \infty$. The linear superposition of the corresponding single-aperture fields can be represented in many different but equivalent ways. Here we choose a superposition of N *symmetric* single-aperture potentials and the potential Φ_H of one homogeneous field:

$$\Phi_A(z, r) = \sum_{i=1}^{N} C_i v_i (1 + u_i \arctan u_i) + A + Bz \tag{10.49}$$

the quantities u_i and v_i being spheroidal coordinates referring to the aperture with subscript i. This representation is the most simple. The coefficients $C_1 \ldots C_N$ are to be determined from the asymptotic field for large values of r:

$$C_i v_i (1 + u_i \arctan u_i) \to \frac{\pi C_i}{2R_i} |z - z_i| \quad \text{for} \quad r \gg R_i \tag{10.50}$$

Since the asymptotic field strength must be

$$F_i := \frac{U_i - U_{i+1}}{z_{i+1} - z_i} \quad \text{for } z_i < z < z_{i+1}, \quad K = 1 \ldots N-1 \tag{10.51a}$$

$$F_0 := E_l \text{ for } z \to -\infty, \quad F_N := E_r \text{ for } z \to \infty \tag{10.51b}$$

at large off-axis distances r, we find

$$C_k = (F_{k-1} - F_k) R_k / \pi, \quad K = 1 \ldots N \tag{10.52}$$

The total contribution of the homogeneous field can be represented as

$$A + Bz = \frac{1}{2}\{U_1 - E_l(z - z_1) + U_N - E_r(z - z_N)\} \tag{10.53}$$

The algorithm corresponding to these formulae is very easy to program.

10.3.3 Combination with the BEM

In earlier investigations, prior to the publication of Hoch et al. (1978), mere superposition of aperture fields was used to approximate the electric field in devices with plane electrodes containing circular bores. This is entirely reasonable for very thin electrodes separated by large distances, $|z_{i+1} - z_i| \gg R_i + R_{i+1}$. Generally speaking, this method fails when applied to devices with thick electrodes although Regenstreif modelled the latter by two thin electrodes at the same potential. A typical example is shown in Fig. 10.13A, which represents equipotentials in an axial section through an electron gun with a thick wehnelt electrode.

In such a situation we suggest the following procedure. As a first step the appropriate aperture parameters are determined, so that the potentials Φ_A, given by Eq. (10.49), can be computed for any point in space. The chosen apertures may be located in the front planes of the electrodes, or inside the electrodes as suggested by Hoch et al. (1978). In the latter case the appropriate aperture potentials are to be obtained by linear extrapolation.

The contribution Φ_A alone, of course, will not give the full result. We therefore consider additional surface charge distributions. These can be confined to the vicinity of the bores, where the surface values of Φ_A differ strongly from the prescribed boundary values. Thus, as a second step, we introduce a reasonable discretization of those parts B of the boundaries where such surface charges are to be applied. We can then solve numerically the integral equation

$$\int_B \frac{1}{\varepsilon_0} G_0(z, r; z', r')\sigma(s')ds' = U(s) - \Phi_A(z, r) \tag{10.54}$$

where $z = z(s)$, $z' = z(s')$, . . ., are parametric representations of surface points and $U(s)$ is the given boundary value function.

After solving Eq. (10.54), the potential $\Phi(z, r)$ and the field strength $\boldsymbol{E} = -\nabla\Phi$ at any point of reference can be computed by numerically superposing Φ_A and $-\nabla\Phi_A$ and the corresponding terms arising from the surface Coulomb integrals. The construction of this field is such that it satisfies the boundary conditions with a high degree of accuracy. The results for the chosen example are presented in Fig. 10.13B.

The consideration of aperture fields is only one of many possible ways of extending the BEM. Another possibility is the superposition of the fields produced by axial charge distributions. This can be helpful for field calculation in systems with pointed cathodes and will therefore be dealt with in Chapter 45 of Volume 2.

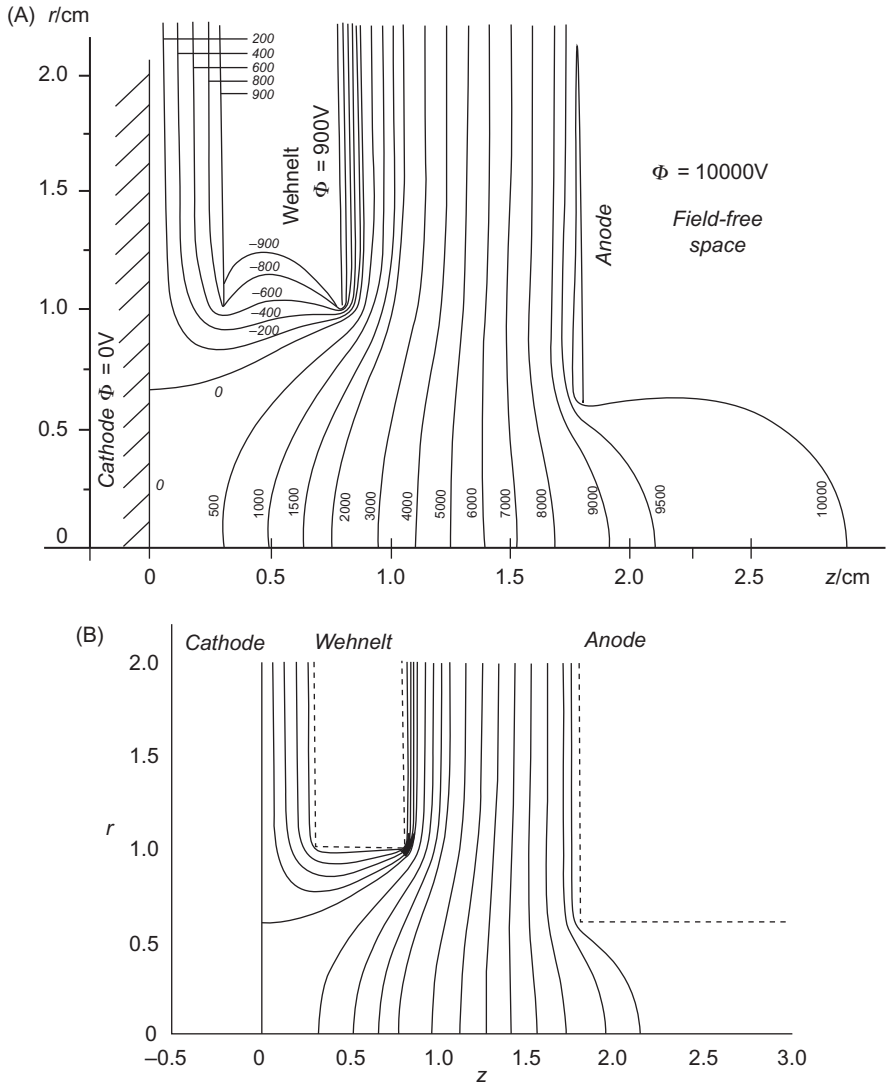


Figure 10.13

Equipotentials $\Phi(z, r) = \text{const}$ in the upper half of an axial section through an electron gun with a plane cathode, a thick wehnelt and a thin anode. (A) Attempt to solve the Dirichlet problem by mere superposition of aperture potentials. The equipotential $\Phi(z, r) = -900$ V does not fit the cylindrical bore of radius 1 cm of the wehnelt at all well. This example makes it very clear that the potentials of additional surface charge distributions are indispensable. (B) Equipotentials $\Phi(z, r) = \text{const}$ in the upper half of an axial section through an electron gun with a plane cathode, a thick wehnelt and a thin anode. The improvement achieved by the introduction of a suitable surface charge distribution.

10.3.4 Extrapolation of the Number of Segments

The accuracy of the results obtained is inevitably affected by the number of subdivisions of the boundary. A procedure for extrapolating from a number N of subdivisions towards an infinite number has been proposed by Read (2000), who observed that the asymptotic dependence of the discretization error of the boundary-element method is not well understood (see Sloan, 1992, which contains an excellent list of related publications). Read found that, for electrostatic lenses at least, it is preferable to choose the segments in such a way that the charge carried by each is approximately the same. Suppose that some quantity x, such as the focal length of a lens, is calculated for several values of N. A linear fit of these values to $1/N^p$ can be obtained for some value of p, which has to be determined. Read finds empirically that p is typically equal to 2 but its value depends on the quantity x being calculated. Once p is known, the quantity in question can be extrapolated to 'infinity' from

$$x_\infty = \frac{\dfrac{x_2}{N_1^p} - \dfrac{x_1}{N_2^p}}{\dfrac{1}{N_1^p} - \dfrac{1}{N_2^p}}$$

in which x_i denoted the value of x obtained with N_i segments.

10.4 Three-Dimensional Dirichlet Problems

There are three-dimensional Dirichlet problems that cannot be reduced to a sequence of two-dimensional ones. Such problems arise in most situations with nonrotationally symmetric boundaries. A typical example is the field in the vicinity of the hairpin of a thermionic electron gun, see Figs 10.19 and 10.20A and B. This case has been investigated by Eupper (1985) in order to estimate the influence of the electric field perturbation on the astigmatism in the electron beam.

The extreme complexity of general three-dimensional boundary-value problems renders their concrete numerical solution much more complicated than that of two-dimensional problems. Here we can deal only with one family of problems, the three-dimensional Dirichlet problem for Laplace's equation $\nabla^2 V(\boldsymbol{r}) = 0$.

The general method of solving such problems may appear to be straightforward. The boundary $\partial\Gamma$ is dissected into a set of N sufficiently small finite surface elements Δ_k, the centroids of which are located at $\boldsymbol{r}_k$ ($k = 1 \ldots N$). In each such element, the surface charge density σ_k is assumed to be constant. Then Eq. (9.14) is approximated by a linear system of equations:

$$V(\boldsymbol{r}_j) = \frac{1}{4\pi\varepsilon_0} \sum_{k=1}^{N} \sigma_k \int_{\Delta_k} \frac{da'}{|\boldsymbol{r}_j - \boldsymbol{r}'|}, \quad j = 1 \ldots N \tag{10.55}$$

This system can be solved for $\sigma_1 \ldots \sigma_N$, after which the surface Coulomb integral can be evaluated for any position $\boldsymbol{r}$ of reference.

This is, in fact, the usual way of solving such problems. For instance, Munro and Chu (1982a,b) have applied this method to an electrostatic deflection unit. Such devices are still very simple, since all the surface elements Δ_k can be chosen to be rectangles. In more general cases, when triangular surface elements cannot be avoided, the method may become very tedious. It is certainly possible to calculate the potential of any triangular surface element with a uniform or even a linear charge distribution in a completely analytical manner (Durand, 1966; Eupper, 1985), but this is very laborious. Although the corresponding expression for $V(\boldsymbol{r})$ can be built up entirely from elementary functions, its evaluation is very slow. Since the number of surface elements must be large in order to achieve good accuracy, the whole procedure is extremely inefficient. The problem of saving unnecessary operations is far more important than in the case of two-dimensional field calculations.

With a view to improving the efficiency, Eupper (1982, 1985) has made an unconventional proposal. In order to avoid the evaluation of improper integrals, the charges are assumed to be located on surfaces $\partial\Gamma'$ chosen close to the corresponding electrode surfaces $\partial\Gamma$, but in the *interior* of the corresponding electrode as is shown in Fig. 10.14A and B. These surfaces are to be chosen reasonably, in the sense that the distance between parallel parts of $\partial\Gamma'$ must be larger than the distance to the corresponding material surface $\partial\Gamma$. Difficulties thus arise with this method for very thin electrodes and in the vicinity of sharp edges.

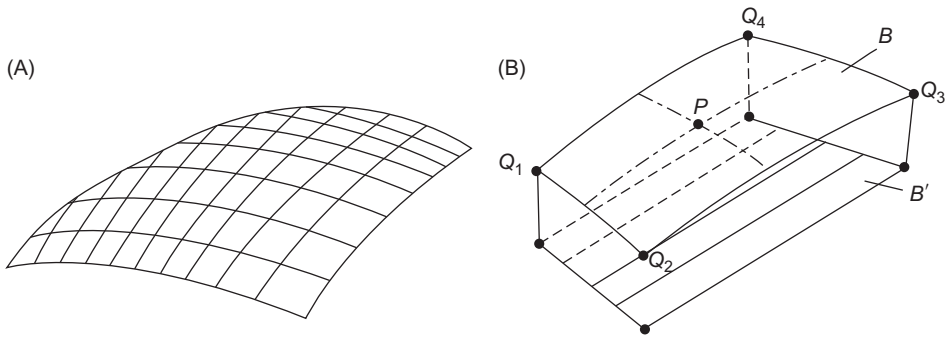


Figure 10.14

(A) Dissection of a curved surface into surface elements by means of two families of surface curves. In at least one of these families, neighbouring curves are locally parallel (provided we consider small enough regions). (B) Perspective view of a single surface element and the associated trapezoid. The lines in the latter indicate the charged bars from which the surface charge element is built up. P and $Q_1, \ldots Q_4$ are control points for the potential.

The interior surfaces $\partial\Gamma'$ are now dissected into general trapezoidal elements (with parallelograms and triangles as special cases). Each such element Δ_k is associated with one control point $\boldsymbol{r}_k$ located on the true electrode surface $\partial\Gamma$. The conditions that the potentials $V(\boldsymbol{r}_1)$, $V(\boldsymbol{r}_2)\ldots V(\boldsymbol{r}_N)$ assume their prescribed values are now set up; this is straightforward. The new idea here is that one part of the necessary integrations, that in the longitudinal direction of each trapezium, can be carried out analytically, resulting in the potential of a charged bar; the corresponding expression will be given below. The remaining integration in the transverse direction is then carried out numerically. This procedure is comparatively simple and yet much faster than entirely analytic integration. The discretization by trapezia is so flexible that even complicated problems like those shown in Figs 10.19 and 10.20 can be solved satisfactorily.

Let us now consider a charged bar of length $2a$, the direction of which is indicated by a unit vector $\boldsymbol{t}$. Let the charge per length unit be $q(s)$ for $-a \le s \le a$. The origin of the coordinate system may be chosen to coincide with the centroid of the bar. The potential is then given by

$$V(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int_{-a}^{a}\frac{q(s)ds}{|\boldsymbol{r}-\boldsymbol{t}s|} \tag{10.56}$$

There are two ways of carrying out the integration, direct analytic integration and integration after an appropriate series expansion. In the most important case of a linear charge distribution $q(s) = q_0 + q_1 s$, direct integration results in

$$4\pi\varepsilon_0 V(\boldsymbol{r}) = (q_0 + q_1\xi)\ln\left(\frac{d_1+d_2+2a}{d_1+d_2-2a}\right) + q_1(d_1-d_2) \tag{10.57}$$

with

$$d_{1,2} = |\boldsymbol{r} \mp a\boldsymbol{t}|, \quad \xi = \boldsymbol{t}\cdot\boldsymbol{r} \equiv (d_2^2 - d_1^2)/4a \tag{10.57a)s}$$

The quantities d_1 and d_2 are the distances of the point $\boldsymbol{r}$ from the endpoints of the bar, as shown in Fig. 10.15.

Alternatively, we may use asymptotic multipole series expansion. After introducing spherical coordinates R, ϑ, defined by

$$R \coloneqq |\boldsymbol{r}|, \quad \mu \coloneqq \cos\vartheta \coloneqq \xi/R \tag{10.58}$$

we first write down the series expansion

$$|\boldsymbol{r}-\boldsymbol{t}s|^{-1} = (R^2 - 2R\mu s + s^2)^{-1/2} = R^{-1}\sum_{l=0}^{\infty}\left(\frac{s}{R}\right)^l P_l(\mu), \quad s < R$$

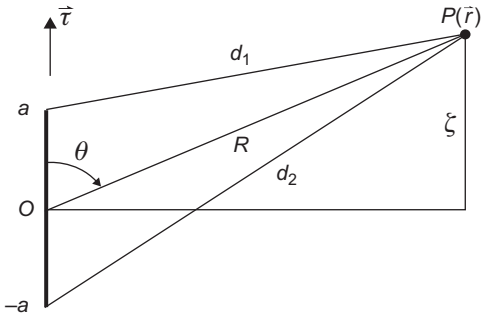


Figure 10.15
Notation employed to characterize the position of an arbitrary point ***P*** relative to a bar of length 2*a*.

where the $P_l(\mu)$ are Legendre polynomials. Introducing this into Eq. (10.56) and defining moments of the charge distribution by

$$M_l := a^{-(l+1)} \int_{-a}^{a} q(s) s^l ds \tag{10.59}$$

we obtain

$$4\pi\varepsilon_0 V(\boldsymbol{r}) = \sum_{l=0}^{\infty} \left(\frac{a}{R}\right)^{l+1} M_l P_l(\mu) \tag{10.60}$$

Using the relation $\mu P_l'(\mu) + (l+1)P_l(\mu) = P_{l+1}'(\mu)$, the gradient of this potential can be written

$$4\pi\varepsilon_0 \nabla V = \frac{1}{a} \sum_{l=0}^{\infty} \left(\frac{a}{R}\right)^{l+2} M_l \left\{ \boldsymbol{t} P_l'(\mu) - \frac{\boldsymbol{r}}{R} P_{l+1}'(\mu) \right\} \tag{10.61}$$

In the case of a linear charge distribution the moments M_l are given by

$$M_l \begin{cases} q_0/(l+1/2), & l \text{ even} \\ q_1 a/(1+l/2), & l \text{ odd} \end{cases}$$

The series expansions (10.60) and (10.61) have a very simple structure and are hence attractive if they can be terminated after a very few terms. This is the case when $R/a \geq 5$. In conclusion, the calculation of such a potential field can be made fast enough to be useful as a basic routine in the numerical solution of three-dimensional Dirichlet problems.

Practical tests have shown that it is quite sufficient to assume constant surface charge density in each trapezoidal element if the field is needed only in domains far from all

boundaries, as is the case in deflection units. When the method is applied to asymmetric electron guns, higher accuracy—especially of the field strength—can be achieved by choosing the charge distribution in each surface element as a bilinear function of the corresponding coordinates, so that the whole distribution becomes continuous (in the vicinity of the cathode). The structure of the linear system of equations for the surface charges does, of course, become more complicated. Another was of extending the BEM to three dimensions was suggested by Tsuboi et al. (1999).

Alternative techniques for the solution of three-dimensional Dirichlet problems are the three-dimensional versions of the finite-difference and finite-element methods. These are generally so dependent on the problem in question that they will not be treated in any detail here. For an example, see the detailed description of the use of the three-dimensional FDM to calculate electrostatic focusing fields in cathode ray tubes given by Franzen (1984) and the review of Rouse (1994).

10.5 Examples of Applications of the Boundary-Element Method

The rotationally symmetric solution of (9.11), corresponding to the order $m = 0$ in (9.27), is useful for calculating the magnetic field of a round lens with a very wide gap (Scherle, 1983) as is shown in Fig. 10.16. For clarity, only the equipotentials of the reduced scalar potential $\chi_M(\boldsymbol{r})$ are presented. A typical example of field calculation in deflection systems (Scherle, 1983) is shown in Fig. 10.17A–C. Again, only the equipotentials of $\chi_M(\boldsymbol{r})$ in one section through the field are shown. The case in which there are three different domains

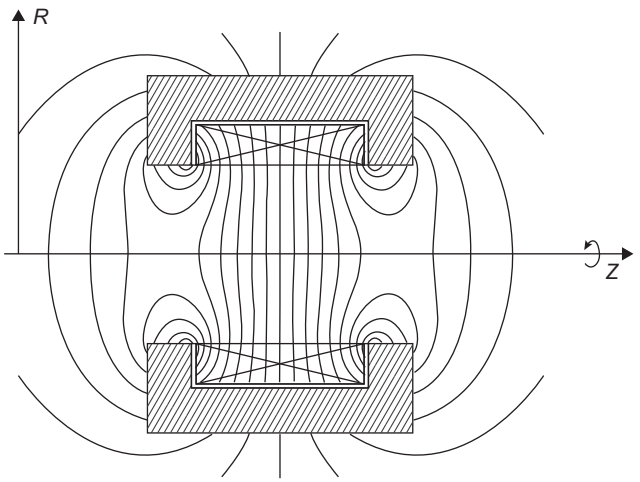


Figure 10.16
Equipotentials of the reduced magnetic scalar potential $\chi_M(\boldsymbol{r})$ of a magnetic lens with a very wide gap. *Courtesy of W. Scherle (1983).*

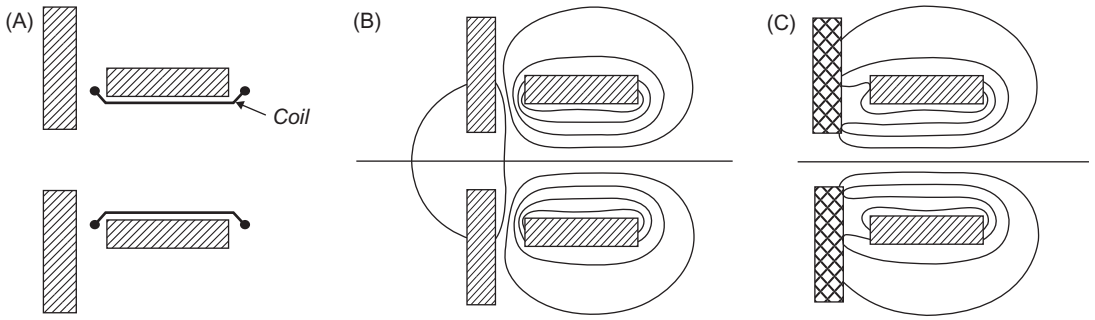


Figure 10.17

Equipotentials of the reduced magnetic scalar potential $\chi_m(r)$ in an axial section through a system of saddle coils, a ferromagnetic yoke and a pierced shielding plate. (A) The position of the saddle coils (B) Ferromagnetic shielding plate (C) Superconducting shielding plate. (Note: the coils and their field $H_0(r)$ are omitted from (B) and (C) to prevent confusion.). *Courtesy of W. Scherle (1983).*

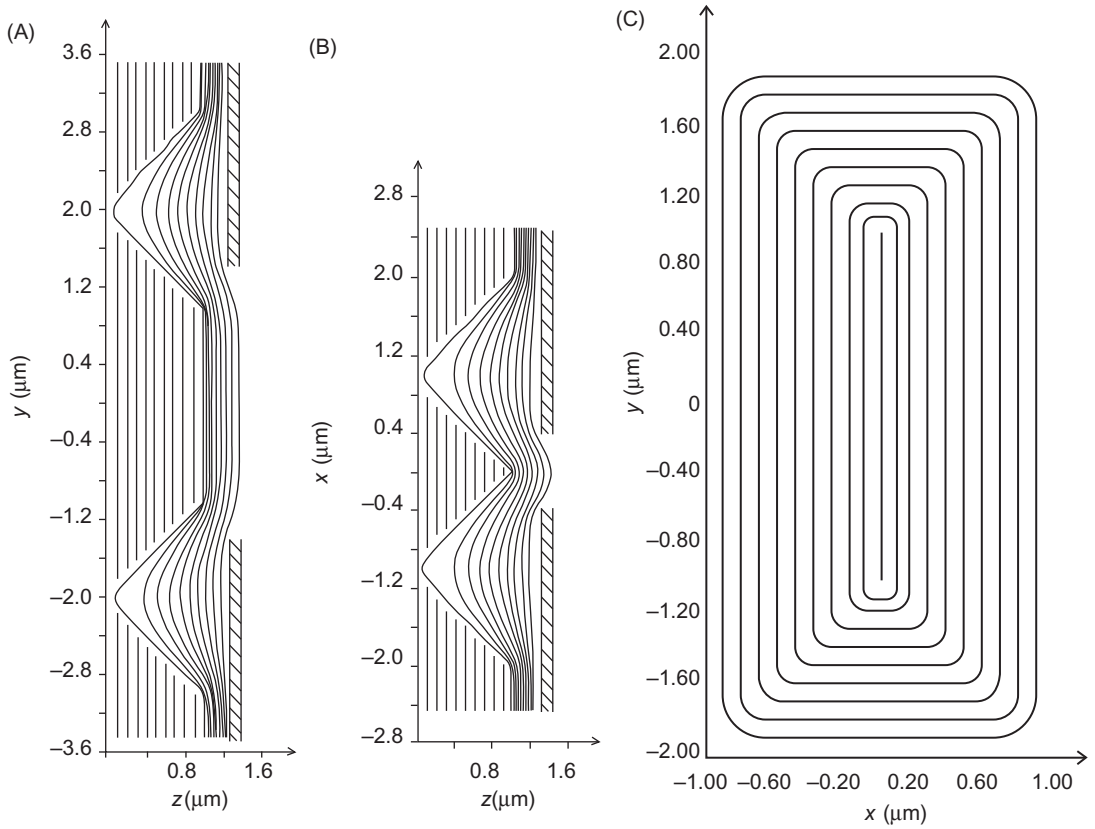


Figure 10.18

(A)–(C) Equipotentials in different sections through a field-electron emission source consisting of a cathode with a hipped roof and an anode with a rectangular bore. *Courtesy of M. Eupper (1982).*

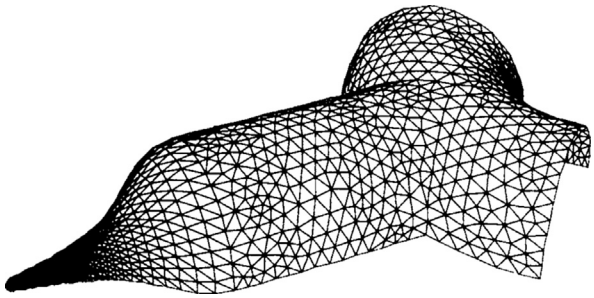


Figure 10.19
Perspective view of a pointed cathode welded on a hairpin support; Only parts of the surface discretization are shown for reasons of clarity. *Courtesy of M. Eupper (1983).*

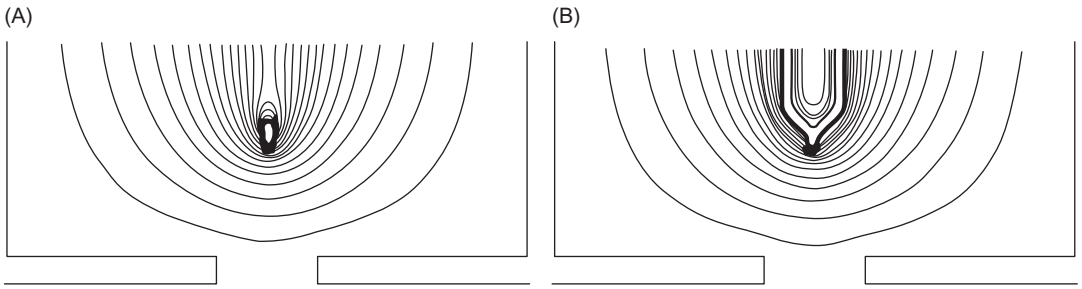


Figure 10.20
(A and B) Equipotentials of the electrostatic potential in two perpendicular sections through the system shown in Fig. 10.19. The position of the anode is indicated. The total width is 10 mm and the anode–cathode distance is 2 mm.

of solution can be analysed by iterative solution of two coupled integral equations of the form (9.11) (Scherle, 1983). The magnetic form of the method has been used to compute magnetic fields in the presence of saturated magnetic materials in a very impressive publication by Murata et al. (2016).

The three-dimensional BEM becomes a very powerful tool when problems with very large differences in the geometrical dimensions are to be solved. Some typical examples are field-electron emission sources with a hipped-roof cathode (Fig. 10.18A–C; Eupper, 1982) and with a rotationally symmetric tip welded on a hairpin-shaped support (Figs 10.19 and 10.20A and B). More details of field calculations in electron sources are given in Chapter 45 of Volume 2.

CHAPTER 11

The Finite-Difference Method (FDM)

The finite-difference method (FDM), usually combined with an iterative technique to solve the corresponding linear system of equations, is a standard procedure for field computation. It was introduced by H. Liebmann as early as 1918 and is thus often called 'Liebmann's method'. The associated mathematical theory is exhaustively studied in the literature, for instance by Varga (1962), Forsythe and Wasow (1960), Ames (1969) and Jacobs (1977). Survey articles on the application of the FDM to electron optical problems have been published by Weber (1967), Bonjour (1980) and Kasper (1982), who has devoted an entire volume to this method and those treated in Chapter 12, The Finite-Element Method (FEM) and Chapter 13, Field-Interpolation Techniques (Kasper, 2001); see also Munro (1988, 1997). In recent years, however, more powerful techniques have been developed for solving boundary-value problems, and the original form of the FDM has lost some of its earlier importance; we shall therefore discuss it only briefly. Nevertheless, it is used in the popular program suite SIMION, which is continually being improved (Dahl et al., 1990 and for a good account of the history of SIMION, Dahl, 2000; www.simion.com). A paper by Dowsett (2015) shows how devices in which very different length scales are present can be modelled by using several overlapping potential arrays with different mesh densities (see also Dowsett, 2011). Multiregion FDM has also been studied by Edwards (2011). For three-dimensional elements such as quadrupoles, the FDM is easier to programme than the finite-element method (Rouse, 1994).

11.1 The Choice of Grid

The basic idea of the FDM is to cover the entire domain of solution of a boundary-value problem by a finite rectangular grid. In order to obtain the greatest possible simplification, it is usual to specialize to square-shaped grids. We have to distinguish between regular internal points (A), irregular internal points (B), regular axial points (C, D), irregular axial points (E) and boundary points (F), as shown in Fig. 11.1. In the practical organization of a FDM program, each point has to be assigned to one of the classes and handled accordingly, which complicates the actual application of the FDM. Since the boundary of the domain is in general curved, these complications cannot be avoided.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00011-0
© 2018 Elsevier Ltd. All rights reserved.

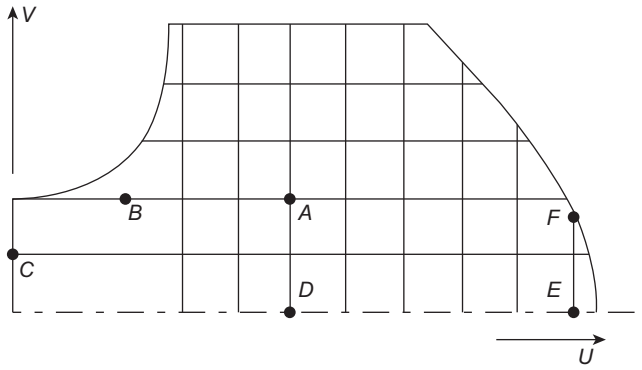


Figure 11.1
Grid with square meshes and distinction between different types of node.

In the subsequent presentation we shall consider a two-dimensional Dirichlet problem associated with a general elliptic differential equation:

$$\begin{aligned} A(u,v)V_{|uu} + B(u,v)V_{|vv} + a(u,v)V_{|u} + b(u,v)V_{|v} \\ = C(u,v)V(u,v) + G(u,v) \end{aligned} \tag{11.1}$$

(For an explanation of the notation, see Section 2.4.) The coordinate system (u, v) may be curvilinear, but is almost invariably chosen to be *orthogonal*. A term with $V_{|uv}$ cannot then appear and is hence omitted from Eq. (11.1).

There are essentially two different ways of deriving discrete formulae, the Taylor series method and the integral method. Both are in very widespread use and equivalent in the sense that they differ only in higher order terms of the discretization errors.

11.2 The Taylor Series Method

Since irregular configurations are to be treated in the vicinity of the boundary at least, we now consider a general five-point configuration, as shown in Fig. 11.2A. In order to establish a discrete form of Eq. (11.1), we expand $V(u, v)$ as a Taylor series with respect to the coordinate differences $u-u_0$ and $v-v_0$. In applications to five-point configurations, we have to truncate this after the second-order terms. This implies that, along the lines $u = u_0 = \text{const}$ and $v = v_0 = \text{const}$, we can approximate the potential by Lagrange interpolation parabolae. For instance, the parabola that fits the potential at the points P_3, P_0 and P_1 (see Fig. 11.2A) is given by

$$V(u, v_0) = V_0 + (u - u_0)V_{|u} + \frac{1}{2}(u-u_0)^2 V_{|uu}$$

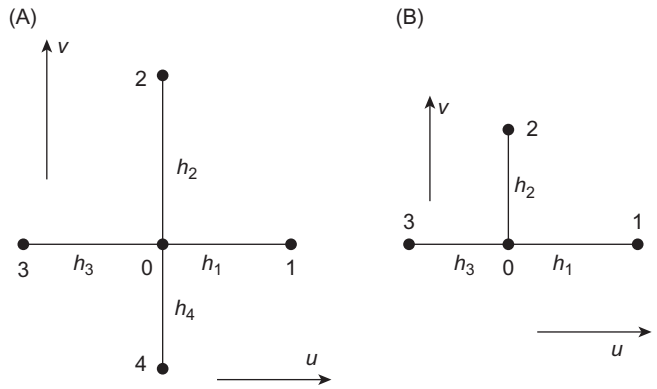


Figure 11.2
(A) Five-point configuration for internal nodes 0. (B) Four-point configuration to be used when the node 0 lies on the axis.

The derivatives $V_{|u}$ and $V_{|uu}$ refer to the central point P_0 and are given by

$$\begin{aligned} V_{|u} &= \frac{1}{h_1 h_3 (h_1 + h_3)} \left\{ h_3^2 V_1 - h_1^2 V_3 + \left(h_1^2 - h_3^2\right) V_0 \right\} \\ V_{|uu} &= \frac{2}{h_1 h_3 (h_1 + h_3)} \left\{ h_3 V_1 + h_1 V_3 - (h_1 + h_3) V_0 \right\} \end{aligned}$$

Similar expressions are obtained for the derivatives $V_{|v}$ and $V_{|vv}$. Introducing all these into Eq. (11.1), we obtain a finite-difference approximation for the latter. This is to be solved for the value V_0 at the grid point P_0 in question. The resulting linear relation has the general form

$$V_0 = \beta_0 + \beta_1 V_1 + \beta_2 V_2 + \beta_3 V_3 + \beta_4 V_4 \tag{11.2a}$$

the coefficients $\beta_0 \ldots \beta_4$ depending on the position of the grid point in question and being given by

$$\begin{aligned} \beta_0 &= G_0 / N \\ \beta_1 &= \frac{2A_0 + a_0 h_3}{h_1 (h_1 + h_3) N}, \quad \beta_2 = \frac{2B_0 + b_0 h_4}{h_2 (h_2 + h_4) N} \\ \beta_3 &= \frac{2A_0 - a_0 h_1}{h_3 (h_1 + h_3) N}, \quad \beta_4 = \frac{2B_0 - b_0 h_2}{h_4 (h_2 + h_4) N} \\ N &= C_0 + \frac{2A_0 + a_0 (h_3 - h_1)}{h_1 h_3} + \frac{2B_0 + b_0 (h_4 - h_2)}{h_2 h_4} \end{aligned} \tag{11.2b}$$

A relation of this kind holds for all regular and irregular internal grid points. On the axis of symmetry, however, only four-point configurations can be evaluated, as is shown in Fig. 11.2B. In such a case the symmetry condition $V(-u, v) = V(u, v)$ leads to

$$V_{|v} = 0, \qquad \lim_{v \to 0} \left[\frac{V_v(u, v)}{v} \right] = V_{|vv}(u, 0)$$

The second-order derivative is then given by

$$V_{|vv}(u, 0) = 2 \frac{V_2 - V_0}{h_2^2}$$

Introducing these approximations into Eq. (11.1) we obtain the four-point formula. It is necessary to assume that $b(u, v) = \tilde{b}(u, v)/v$, where $\tilde{b}(u, v)$ is an even function with respect to v and may also vanish. The coefficients are then given by

$$\begin{aligned} &\beta_4 = 0, \quad \beta_2 = 2(B_0 + \tilde{b}_0)/Nh_2^2 \\ &N = C_0 + \frac{2A_0 + a_0(h_3 - h_1)}{h_1 h_3} + \frac{2(B_0 + \tilde{b}_0)}{h_2^2} \end{aligned} \tag{11.2c}$$

while β_0, β_1 and β_3 remain the same as in Eq. (11.2b), apart from the different normalization factor N.

As a comparatively simple example, we shall now consider the differential equation

$$V_{|zz} + V_{|rr} + \frac{\alpha}{r} V_{|r} = -g(z, r) \tag{11.3}$$

which includes Eq. (7.10) as a special case with $\alpha = 2m + 1$. We limit the discussion to regular grid points, for which the finite-difference approximations can be given easily in explicit notation. With

$$V_{i,k} \coloneqq V(ih, kh) \quad (i, k \quad \text{integers}) \tag{11.4}$$

and a similar notation for $g(z, r)$ we obtain

$$\begin{aligned} V_{i,k} = &\frac{1}{4}(V_{i+1,k} + V_{i-1,k} + V_{i,k+1} + V_{i,k-1} + h^2 g_{i,k}) \\ &+ \frac{\alpha}{8k}(V_{i,k+1} - V_{i,k-1}), \qquad k > 1 \end{aligned} \tag{11.5a}$$

for internal mesh points and

$$V_{i,0} = \frac{1}{2(\alpha + 2)} \left\{ V_{i+1,0} + V_{i-1,0} + 2(\alpha + 1)V_{i,1} + h^2 g_{i,0} \right\} \tag{11.5b}$$

for axial mesh points (if $\alpha \neq -2$). The resulting discretization error is of fourth order in the mesh-length h. The set of relations Eq. (11.5a,b) forms a linear system of equations specified by two subscripts i and k.

11.3 The Integration Method

In the subsequent presentation, we assume that the differential equation to be solved is self-adjoint:

$$\frac{\partial}{\partial u}(PV_{|u}) + \frac{\partial}{\partial \upsilon}(PV_{|\upsilon}) = QV + S \tag{11.6}$$

the coefficients P, Q and S being regular functions of u and υ. Applying Gauss's integral theorem to any domain R and its closed boundary C we obtain

$$\oint_C P\frac{\partial V}{\partial n}\,ds = \iint_R (QV + S)\,du\,d\upsilon \tag{11.7}$$

$\partial V/\partial n$ denoting the normal derivative of V in the outward direction. This integral relation is exact. Its practical evaluation, however, requires several simplifying assumptions. For instance, when applying Eq. (11.7) to the configuration shown in Fig. 11.3A, we can make the approximation $\partial V/\partial n = (V_i - V_0)/h_i$ $(i = 1\ldots4)$ on the corresponding side of the rectangular contour. Furthermore we make the simplification $V \approx V_0$ under the double integral; we then obtain the finite-difference equation

$$\sum_{i=1}^{4} \frac{V_i - V_0}{h_i} \int_{s_i} P\,ds = \iint_R (QV_0 + S)\;\,du\,d\upsilon \tag{11.8a}$$

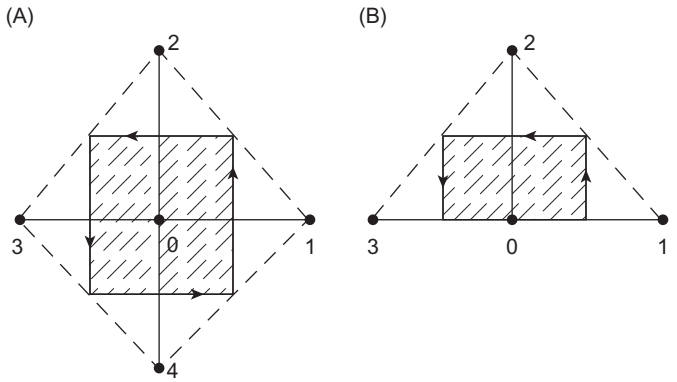


Figure 11.3
Rectangular loops and areas of integration (A) for internal nodes and (B) for axial nodes. In the latter case, both the loop and the area can be doubled by exploiting the mirror symmetry.

Provided that the integrands are sufficiently simple analytic functions, it is possible to perform the remaining integrations analytically, but this brings no essential gain in accuracy, since the approximation made for $\partial V/\partial n$ is then too inaccurate. Thus, in order to obtain a practical form of the discretization, we assume $Q \to Q_0 = Q(u_0, v_0)$, $S \to S_0 = S(u_0, v_0)$ on the right-hand side of Eq. (11.8a). On the left-hand side, we assume that the integrand P is piecewise constant, for instance $P(u, v) \to P(u_0 + h_1/2, v_0) =: \overline{P}_1$ on the right-hand side of the rectangular integration path shown in Fig. 11.3A, the other parts of this path being treated analogously. We then find immediately

$$\begin{aligned}
&(V_1 - V_0)\frac{h_2 + h_4}{2h_1}\overline{P}_1 + (V_2 - V_0)\frac{h_1 + h_3}{2h_2}\overline{P}_2 \\
&+ (V_3 - V_0)\frac{h_2 + h_4}{2h_3}\overline{P}_3 + (V_4 - V_0)\frac{h_1 + h_3}{2h_4}\overline{P}_4 \qquad (11.8b) \\
&= \frac{1}{4}(V_0 + Q_0 S_0)(h_1 + h_3)(h_2 + h_4)
\end{aligned}$$

Solving this for V_0, we obtain a linear relation which differs from Eq. (11.2) only in the values of the coefficients $\beta_0 \ldots \beta_4$. The two discretizations are equivalent in the sense that they differ only in discretization errors of third or fourth order in the mesh-length. These error terms are to be neglected in any case. For axial nodes O, some special considerations are necessary, which are not given here. For the most important special case mentioned below, the reader will find them in Janse (1971) and Kasper (1976).

The above considerations can be applied to Eq. (11.3), since this differential equation can be rewritten as

$$\frac{\partial}{\partial z}(r^\alpha V_{|z}) + \frac{\partial}{\partial r}(r^\alpha V_{|r}) = -r^\alpha g(z, r) \qquad (11.9)$$

so that in Eq. (11.6) we have $P = r^\alpha$, $Q = 0$, $S = -r^a g$. The corresponding discretization formulae have been published by Janse (1971) and Kasper (1976, 1982) and will not be repeated here. The discretization differs from Eq. (11.5a) essentially in the fact that here the integrations in Eq. (11.8a) are carried out analytically and all the coefficients remain strictly positive, whereas in Eq. (11.5a), the coefficient of $V_{i,k-1}$ becomes negative for $\alpha > 2k$. Positive coefficients mean increased stability of the entire system of equations when these are solved by iterative techniques (see Section 11.5), but the final accuracy of the solution obtained is not better than that given by Eq. (11.5a). It is interesting to note that the discretization formula obtained by integration over the configuration shown in Fig. 11.3B is identical with Eq. (11.5b).

The integral method may be further generalized to discretization in general triangular grids. The corresponding algorithms have been derived by Colonias (1974) and Winslow (1967). These will not be treated here, since they seem to be less favourable than the finite-element method presented in Chapter 12, The Finite-Element Method (FEM). Some other refinements of the FDM are possible; see, e.g., Lenz (1973) and Kasper (1982) for further details.

11.4 Nine-Point Formulae

The accuracy of the FDM can be improved considerably by the use of nine-point formulae (Durand, 1966; Kasper, 1976, 1984a,b, 2001). These are advantageous in the case of a regular grid; the solution of the problem that arises for irregular grid points is given below.

We reconsider the differential equation of the general form Eq. (11.6). More particularly we assume that the coefficient functions have a common factor υ^α with $\upsilon \geq 0$, $\alpha \geq -1$, so that

$$\frac{\partial}{\partial u}(p^2\upsilon^\alpha V_{|u}) + \frac{\partial}{\partial \upsilon}(p^2\upsilon^\alpha V_{|\upsilon}) + p^2\upsilon^\alpha(\hat{q}V + s) = 0 \tag{11.10}$$

$p(u, \upsilon)$, $\hat{q}$ (u, υ) and $s(u, \upsilon)$ being *finite* analytical functions of their variables and $p > 0$. There is a wide class of differential equations that fit Eq. (11.10). One practical example is Eq. (11.9) with $p \equiv 1$, $\hat{q} \equiv 0$, $\upsilon = r$, $s = g$; we shall meet others below. The same type of discretization can be applied to all these equations, as will be obvious from the following considerations.

We first note that by writing

$$V(u, \upsilon) =: \frac{U(u, \upsilon)}{p(u, \upsilon)} \tag{11.11}$$

Eq. (11.10) collapses to the simpler form

$$\Delta_\alpha U = -g(u, \upsilon) := -(q(u, \upsilon)U + ps) \tag{11.12}$$

with

$$\Delta_\alpha := \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial \upsilon^2} + \frac{\alpha}{\upsilon}\frac{\partial}{\partial \upsilon} \tag{11.13}$$

and

$$q(u, \upsilon) := \hat{q}(u, \upsilon) - \frac{\Delta_\alpha p}{p} \tag{11.14}$$

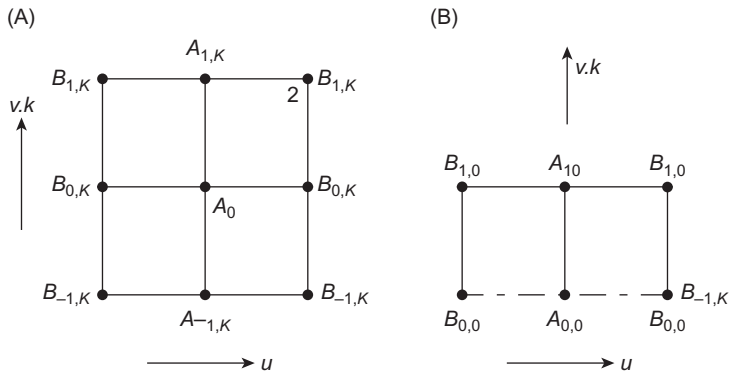


Figure 11.4
Notation employed for the points and coefficients in nine-point configurations, (A) General case and (B) on-axis node.

If we consider cylindrical coordinates, $u = z$ and $\upsilon = r$, all the differential equations derived in Chapter 7, Series Expansions, are seen to be special cases of Eq. (11.12). There are, however, important examples of representations in other coordinate systems, as we shall see later.

Just as for the five-point discretization, we have to distinguish between on-axis formulae ($\upsilon = 0$) and off-axis formulae ($\upsilon > 0$), and again these can take different forms. Here we shall present only the results derived by Kasper (1984a), since these seem to be most favourable for practical applications. For reasons of space we cannot reproduce here their lengthy derivation, to be found in the corresponding publication.

We again use a notation with two subscripts $i = u/h$, $k = \upsilon/h$; the notation for the coefficients is explained in Figs 11.4A and B. The discretization formulae are most conveniently given by an *implicit* representation in terms of a new array W, defined by

$$W_{i.k} \coloneqq U_{i,k} + \frac{h^2}{12} g_{i,k} \equiv p_{i,k} \left\{ V_{i,k} + \frac{h^2}{12} (q_{i,k} V_{i,k} + s_{i,k}) \right\} \tag{11.15}$$

The *off-axis* discretization ($k \neq 0$) is found to be

$$\begin{aligned} W_{i,k} = {} & A_{-1,k} W_{i,k-1} + A_{1,k} W_{i,k+1} + A_{0,k} h^2 g_{i,k} \\ & + \sum_{j=-1}^{1} B_{j,k} (W_{i-1,k+j} + W_{i+1,k+j}) + O(h^6) \end{aligned} \tag{11.16}$$

The coefficients are given by

$$\begin{aligned} \gamma_k &:= \frac{\alpha(\alpha-2)}{12k^2-3} \qquad N_k := \frac{1}{20+8\gamma_k} \\ A_{\pm 1,k} &= 4B_{\pm 1,k} \qquad B_{\pm 1,k} = N_k\left(1+\gamma_k \pm \frac{\alpha-\gamma_k}{2k}\right) \\ A_{0,k} &= 6N_k, \qquad\qquad B_{0,k} = N_k(4-2\gamma_k) \end{aligned} \tag{11.17}$$

(Kasper, 2001, Section 4.4.1). This set of coefficients is obviously independent of the mesh-length h. It is sensible to compute it at the beginning of the program and leave it in store.

On the optic axis ($v = 0$, $k = 0$) a slightly different discretization is necessary. Equation (11.15) remains valid but in the discretization not only are the coefficients different but some other terms appear:

$$\begin{aligned} W_{i,0} = A_{1,0}W_{i,1} + \sum_{j=0}^{1} B_{j,0}(W_{i-1,\,j} + W_{i+1,\,j}) \\ + h^2\left\{A_{0,0}g_{i,0} + C(g_{i,0} - g_{i,1})\right\} + O(h^6) \end{aligned} \tag{11.18}$$

the set of coefficients being given by

$$\begin{aligned} \beta &:= \frac{(1+\alpha)(6+\alpha)}{6(3+\alpha)}, \qquad \gamma_0 := \frac{1}{2(2+\alpha-\beta)} \\ A_{1,0} &= 2\gamma_0(1+\alpha-\beta) \\ B_{0,0} &= \gamma_0(1-\beta), \qquad\qquad B_{1,0} = \gamma_0\beta \\ A_{0,0} &= \gamma_0, \qquad\qquad\qquad C = \gamma_0\frac{\alpha(1+\alpha)}{6(3+\alpha)} \end{aligned} \tag{11.19}$$

For $\alpha < 0$ ($\alpha = -1$ for flux fields), the axial discretization fails, but then we have simply $W_{i,0} = U_{i,0} = g_{i,0} = s_{i,0} = 0$, so that Eq. (11.18) is no longer needed.

In order to reduce the necessary amount of calculations and storage locations, we rewrite Eq. (11.15) in the form

$$h^2 g_{i,k} = \frac{(q_{i,k}W_{i,k} + p_{i,k}s_{i,k})}{\left(\frac{1}{h^2} + \frac{1}{12}q_{i,k}\right)} = C_{i,k}W_{i,k} + S_{i,k} \tag{11.20a}$$

from which the field V has been eliminated. The sets of coefficients $C_{i,k}$ and $S_{i,k}$ are calculated once in the beginning and stored. Next, the boundary values of W are determined

from Eq. (11.15) and stored. The boundary-value problem for the array W can now be solved using Eqs (11.16) and (11.18) and with Eq. (11.20a) as source terms. In this major calculation only three arrays, C, S and W, are needed simultaneously. Finally the required function V is obtained by solving Eq. (11.15) for $V_{i,k}$ or equivalently from

$$V_{i,k} = p_{i,k}^{-1}\left\{W_{i,k} - \frac{1}{12}(C_{i,k}W_{i,k} + S_{i,k})\right\} \tag{11.20b}$$

There is no iteration and the coefficients C, S and W are no longer needed once the expressions on the right-hand side have been evaluated.

11.5 The Finite-Difference Method in Three Dimensions

The three-dimensional form of the method has been studied in great detail by Rouse (1994); here we simply give the basic structure, referring to Rouse's article for details. We now consider six points in the neighbourhood of a central node (Fig. 11.5A). Laplace's equation leads to

$$\alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 + \alpha_4 V_4 + \alpha_5 V_5 + \alpha_6 V_6 = \alpha_0 V_0$$

(cf. 11.2a), in which

$$\alpha_1 = \frac{2}{h_1(h_1 + h_2)}, \quad \alpha_2 = \frac{2}{h_2(h_1 + h_2)}$$
$$\alpha_3 = \frac{2}{h_3(h_3 + h_4)}, \quad \alpha_4 = \frac{2}{h_4(h_3 + h_4)}$$
$$\alpha_5 = \frac{2}{h_5(h_5 + h_6)}, \quad \alpha_6 = \frac{2}{h_6(h_5 + h_6)}$$
$$\alpha_0 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6$$

If an electrode intersects a mesh line away from a node, a small modification usually gives a satisfactory solution and requires very little modification of the set of linear equations to be solved. Suppose that the electrode (potential V_e) intersects the mesh between node zero and node 1 (Fig. 11.5B) The length h_1 is now replaced by the distance from the node zero to the electrode and the potential is set equal to V_e.

Rouse extends this basic theory in two important ways: the theory is modified to permit dielectric material (notably, insulators) to be included and a simple way of adapting it to calculate the magnetic fields of magnetic materials and coils is described.

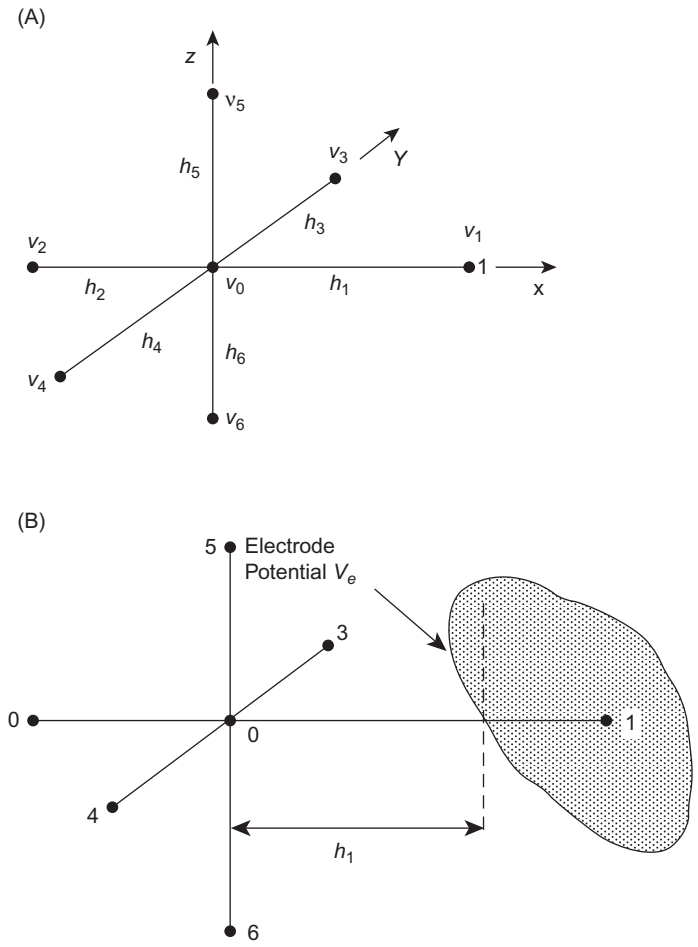


Figure 11.5

(A) The six nearest neighbours of a central node (0) at distances h_1, h_2,. . .h_6 from the central node. (B) An electrode at potential V_e intersects the arm joining the central node and node 1 at a distance h_{1e} from the central node.

For solution of the finite-difference equations, Rouse prefers successive over-relaxation (Section 11.7). Gaussian elimination is extremely slow; the incomplete Cholesky conjugate gradient method is faster but cumbersome.

11.6 Other Aspects of the Method

11.6.1 Expanding Spherical-Mesh Grid

The formalism developed above fits very many different forms of the equations of Laplace, Poisson and Helmholtz. In electron optics, they are most frequently expressed in cylindrical

coordinates but this is not always advantageous. In order to calculate the field in field-emission electron guns (see Part IX), Kang et al. (1981, 1983) introduced an expanding spherical grid which they called SCWIM (spherical coordinates with increasing mesh-width). This can be regarded as a conformal mapping of an originally cylindrical grid with square-shaped meshes:

$$z + \mathrm{i}r = R_0 \exp(u + \mathrm{i}\vartheta) = R(\cos\vartheta + \mathrm{i}\sin\vartheta)$$

with an arbitrary positive constant R_0, and hence

$$R = R_0 \exp u, \quad \vartheta \equiv \upsilon \tag{11.21}$$

If the mesh-length h in the coordinates (u, υ) is constant, the corresponding grid in the real space expands exponentially, as shown in Fig. 11.6. (This interpretation is not mentioned by Kang et al.).

Introducing Eq. (11.21) into the rotationally symmetric Poisson equation in spherical coordinates (R, ϑ):

$$\frac{\partial^2 V}{\partial R^2} + \frac{2}{R}\frac{\partial V}{\partial R} + \frac{1}{R^2}\frac{\partial^2 V}{\partial \vartheta^2} + \frac{\cot\vartheta}{R^2}\frac{\partial V}{\partial \vartheta} = -\frac{\rho(R,\vartheta)}{\varepsilon_0} \tag{11.22}$$

we find after some elementary calculations

$$\frac{\partial}{\partial u}\left(\mathrm{e}^u \sin\upsilon \frac{\partial V}{\partial u}\right) + \frac{\partial}{\partial \upsilon}\left(\mathrm{e}^u \sin\upsilon \frac{\partial V}{\partial \upsilon}\right) = -\frac{R_0^2 \mathrm{e}^{3u}\sin\upsilon}{\varepsilon_0}\rho(R_0\mathrm{e}^u, \upsilon) \tag{11.23}$$

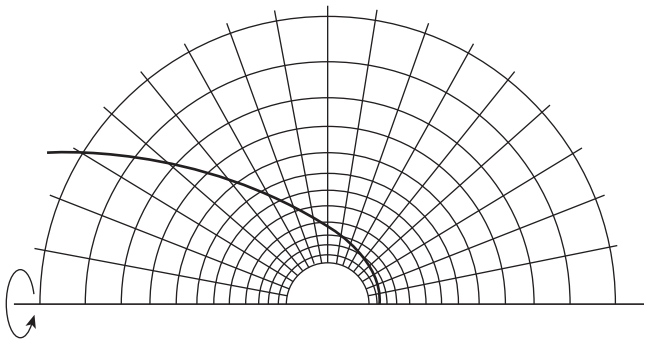


Figure 11.6
An expanding spherical-mesh grid and part of a curved boundary that does not fit this grid ($h = \pi/20$).

This can be brought into the form Eq. (11.10) with

$$\alpha = 1, \quad p^2 = \mathrm{e}^u \frac{\sin \upsilon}{\upsilon}, \quad \hat{q} \equiv 0, \quad (0 \le \upsilon < \pi) \tag{11.24a}$$

$$s = \frac{R_0^2 \mathrm{e}^{2u}}{\varepsilon_0} \rho(R_0 \mathrm{e}^u, \upsilon) = \frac{R^2}{\varepsilon_0} \rho(R, \vartheta) \tag{11.24b}$$

The evaluation of Eq. (11.14) results in

$$q = \frac{1}{4}\left(\frac{1}{\sin^2 \upsilon} - \frac{1}{\upsilon^2}\right) = \frac{1}{12}\left(1 + \frac{\upsilon^2}{5}\right) + O(\upsilon^4) \tag{11.24c}$$

The derivation of the corresponding nine-point discretization is now a straightforward matter. It has been proposed by Kasper (1984a) and worked out by Killes (1985). In comparison with a five-point discretization with equal meshes, the gain in accuracy is considerable, so that it is certainly worthwhile to use the nine-point discretization whenever this is possible. All that is required is the determination of the coefficients in Eq. (11.20) from Eq. (11.24), the result being

$$p_{j,k} = \exp\left(\frac{1}{2} jh\right)\left(\frac{\sin hk}{hk}\right)^{1/2}$$

$$q_{j,k} =: q_k = \frac{1}{4}\left(\frac{1}{\sin^2 hk} - \frac{1}{h^2 k^2}\right)$$

$$C_{j,k} =: C_k = \frac{12 q_k}{q_k + 12h^{-2}}$$

$$S_{j,k} = \frac{12 p_{j,k} s_{j,k}}{q_k + 12h^{-2}}$$

where j and k are integers. Equations (11.16) and (11.18) with (11.20a) and finally (11.20b) can then be solved.

As Killes (1985) pointed out, this method is useful for $0 \le \vartheta = hk \le \pi/2$. For larger values of ϑ, it is better to discretize the variable $\vartheta' := \pi - \vartheta$ and to join the fields in the two domains together smoothly at $\vartheta = \vartheta' = \pi/2$. The concept of conformal mapping can be generalized further but we shall not pursue this here.

11.6.2 Extrapolation on Multiple Grids

In Chapter 10, The Boundary-Element Method, we have seen that extrapolation can be beneficial by reducing the size of the matrix to be inverted. Here, the extrapolation is

designed to improve the accuracy of results obtained with the five-point FDM without increasing the number of nodes. A potential problem is first solved using a small number of nodes ($N \times N$), then again after doubling the number of nodes in both directions ($2N \times 2N$) and yet again after doubling the number a last time ($4N \times 4N$). Three values are thus obtained at each node of the coarse mesh and these are extrapolated to give an improved estimate. The resulting values at the $N \times N$ nodes are then interpolated to furnish an improved $4N \times 4N$ mesh. In the example examined by Becker (2008), where $N = 32$, an improvement of more than an order of magnitude was obtained, which would have required a 407×407 grid without interpolation. However, no such improvement was found when Kasper's nine-point discretization was employed.

11.6.3 Combination with the BEM

It frequently happens that a boundary-value problem is to be solved in which the boundary contours do not fit the grid. For every irregular internal node the general five-point formulae then have to be applied. This complicates the FDM very considerably and diminishes its accuracy. This disadvantage can be circumvented in different ways. In the case treated above and for the solution of other Dirichlet problems for Poisson's equation $\nabla^2 V = -\rho(\boldsymbol{r})/\varepsilon_0$, combination with the boundary-element method (BEM) is effective (Kasper, 1984a,b; Killes, 1985). First, the regular grid is extended beyond the boundaries so that only *regular* points are obtained, as is demonstrated in Figs 11.6 and 11.7. In this grid, Poisson's equation is solved with arbitrary reasonable boundary values. The values at the true curved boundary are then determined by interpolation. This can be done very accurately since interpolation in regular grids raises no problems (Killes, 1985). The values obtained are now subtracted from the prescribed boundary values, and with these differences the Dirichlet problem $\nabla^2 V' = 0$ is solved using the BEM. The required total solution is then obtained by superposition. The advantage lies in the fact that the solution of

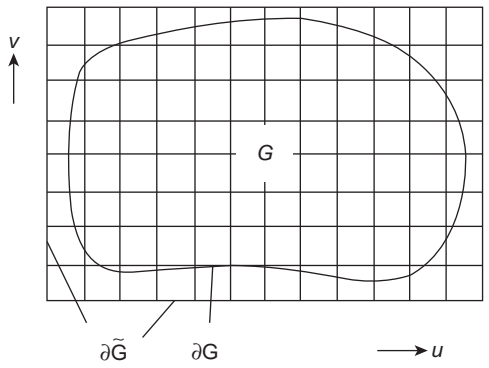


Figure 11.7
Extension of a regular grid beyond the curved boundary ∂G of a given domain G.

Poisson's equation by means of the FDM is much faster than the evaluation of Coulomb integrals, while the BEM can be applied easily to configurations with curved boundaries.

11.7 Iterative Solution Techniques

These will be briefly treated here; the same techniques are also used to solve the equations resulting from finite-element approximations. In order to obtain a highly accurate final solution, the mesh-length must be sufficiently small. Then, of course, the rank $\boldsymbol{N}$ of the system of equations for the potential at the internal nodes becomes very large, typically $\boldsymbol{N} \sim 10^4$. The solution can be obtained directly but a solution can also be found iteratively, since the coefficient matrix of the system is sparse.

Numerous iterative techniques for solving large linear systems of equations have been developed. The corresponding mathematical literature is very extensive; some of the major works are cited in the bibliographic listing.

The first step in the application of any of these methods is the choice of an appropriate numbering of the grid points. All the internal grid points must be counted exactly once in a one-dimensional sequence and no boundary point must be counted as the whole domain of solution is scanned. For instance, an appropriate numbering of the internal points with indices $I_1 \le i \le I_2$, $K_1 \le k \le K_2$ is given by

$$\mu := i - I_i + 1 + (k - K_1)(I_2 - I_1 + 1)$$
$$1 \le \mu \le N := (K_2 - K_1 + 1)(I_2 - I_1 + 1)$$

but permutations of this sequence are also allowed. With this linear sequence of numbering, the whole system of finite-difference equations can be brought into the general form

$$V_\mu = \sum_{\nu=1}^{N}{}' \, C_{\mu\nu} V_\nu + Q_\nu, \quad \mu = 1 \ldots N \tag{11.25}$$

the prime indicating that the case $\mu = \nu$ is to be excluded. The matrix on the right-hand side is large but sparse, its nonzero elements can be easily calculated and it is therefore not necessary to store them. The inhomogeneous terms Q_ν arise from the boundary values of the potential and from the source terms.

The standard iterative technique for solving linear systems of equations like Eq. (11.25) is the successive over-relaxation method (SOR). The corresponding procedure is defined by

$$S_\mu^{(n+1)} := \sum_{\nu=1}^{\mu-1} C_{\mu\nu} V_\nu^{(n+1)} + \sum_{\nu=\mu+1}^{N} C_{\mu\nu} V_\nu^{(n)} + Q_\nu \tag{11.26a}$$

$$V_\mu^{(n+1)} = V_\mu^{(n)} + \omega (S_\mu^{(n+1)} - V_\mu^{(n)}), \quad \mu = 1 \ldots N \tag{11.26b}$$

Here the superscript in parentheses denotes the iteration number. The starting values $V_1^{(0)} \ldots V_N^{(0)}$ can be chosen arbitrarily; a sensible guess for these is quite sufficient. The constant ω, the over-relaxation parameter, must satisfy $1 < \omega < 2$. It is very important for the convergence of the iteration procedure to choose it suitably, as will be discussed below. The practical meaning of Eq. (11.26a,b) is as follows. The quantity $S_\mu^{(n+1)}$ is an approximation for the expression on the right-hand side of Eq. (11.25). This Gauss−Seidel value is calculated from the newest values V_ν ($\nu = 1 \ldots N$); it is, however, not accepted as the next approximation, but instead the difference between it and the preceding value is magnified by a factor ω, as is obvious from Eq. (11.26b).

In the practical setup of a SOR program, each successive approximation $V_\mu^{(l)}$, $l = 0, 1, 2, \ldots$ for the same variable V_μ is stored in a single location assigned to V_μ, and for S only one storage location is necessary. Since the array $\{Q_\nu, \nu = 1 \ldots N\}$ is usually sparse, the necessary number of storage locations is in practice not much greater than N.

The theory of the convergence of the SOR is investigated exhaustively in the mathematical literature; see Ames (1969), Varga (1962) and Weber (1967), for example. A sufficient and often necessary criterion for the convergence of SOR is that

$$\sum_{\nu=1}^{N}{}' \left| C_{\mu\nu} \right| \leq 1, \quad \mu = 1 \ldots N \tag{11.27}$$

At least one of these sums must be less than 1. For $\alpha > 1$, criterion Eq. (11.27) is violated by Eq. (11.5a), so that SOR cannot be applied to Eq. (11.5a). When the process does converge, its rate depends essentially on the choice of ω, as is shown in Fig. 11.8. The number of iterations N_{it} needed to reduce the iteration error Δ below a given error limit ε does of course depend on the definition of Δ, the choice of ε and the initialization $V_\mu^{(0)}$,

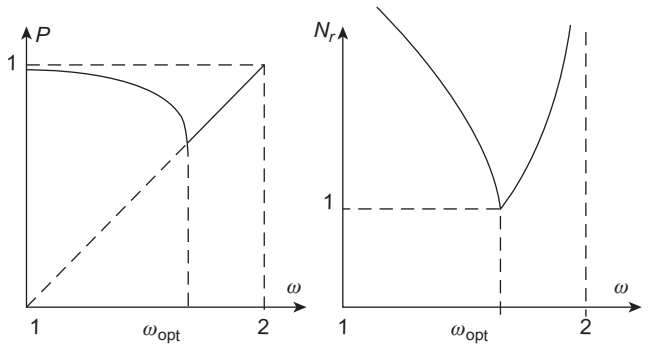


Figure 11.8
Asymptotic behaviour of the damping factor p and relative iteration number N_r as functions of the relaxation parameter ω.

$\mu = 1 \ldots N$. But the value of ω_{opt}, where N_{it} has its sharp minimum, and the relative iteration number, $N_r = N_{it}/N_{\min}$, do *not* depend on these quantities.

According to the general theory of SOR, the optimum value of ω is given by

$$\omega_{\text{opt}} = \frac{2}{1 + \sqrt{1 - \lambda^2}}, \quad |\lambda| < 1 \tag{11.28}$$

λ being the largest (real) eigenvalue of the matrix C in Eq. (11.25). For any value of ω this quantity λ is related to the asymptotic damping factor p, $(p < 1)$ by

$$\lambda = \frac{(p + \omega - 1)^2}{\omega^2 p} \tag{11.29}$$

Since λ and p are very difficult to calculate exactly, Carré (1961) has proposed a method of estimating these quantities, and hence ω_{opt}, from the actual rates of convergence during the computations. Winslow (1967) has modified this method. Both versions work adequately in many applications but not in every case. For further information the reader is referred to the original papers and Eiermann and Varga (1993).

A further refinement of the SOR is the familiar successive line over-relaxation (SLOR) method. Here SOR is combined with Gaussian elimination for tridiagonal subsystems. For instance, in applications to Eq. (11.5a,b), these equations are first solved directly along each radial row of the grid, the values in the neighbouring rows being regarded as known for the moment; thereafter the values obtained are modified by over-relaxation and the algorithm proceeds to the next row. The whole grid is scanned repeatedly in this way until sufficient convergence has been achieved.

The main advantage of the SLOR method is that it removes the instabilities of the simple SOR when Eq. (11.27) is not satisfied. This has been reported by Kasper and Lenz (1980), who applied the SLOR to Eq. (11.5a,b).

Still more refined techniques for solving large but sparse systems of linear equations are the alternating direction implicit methods (ADI: Peaceman and Rachford, 1955; Varga, 1962; Jacobs, 1977), the strongly implicit methods (Stone, 1968) and the cyclic reduction methods (Buneman, 1971, 1973a,b). With the vast increase of memory capacity since those methods were introduced, direct solution techniques have become attractive. Special procedures that order the corresponding matrix in such a way that only its nonzero part is stored then become desirable. Munro (1971, 1973) has applied such methods to systems of equations arising in electron optical field computations.

CHAPTER 12

The Finite-Element Method (FEM)

Though the basic ideas and equations of the finite-difference method (FDM) are very simple, the practical application of this method to boundary-value problems can become extremely tedious if the boundaries are of an irregular shape. In the finite-element method (FEM), this difficulty is removed by the use of general triangular grids, as is shown in Fig. 12.1. Such grids can be fitted to any shape of boundary, once the latter has been represented approximately by a polygon. Since the numerical differentiation now becomes very complicated, partial differential equations are not considered here. Instead, the equations governing the values of the potential at the nodes of the grid are derived directly from an appropriate variational principle (see Section 6.3).

The FEM was proposed by Courant (1943). It came into practical use with the development of modern computers and has found widespread application in mechanical and electrical engineering. Typical examples are problems in fluid dynamics and aerodynamics, elasticity, heat conduction and magnetic field computations for electric machines (Chari and Silvester, 1980). For further details we refer to the books of Zienkiewicz (1967, 1971, 2013). Some mathematical problems associated with the FEM were treated in detail by Norrie and de Vries (1973). In electron optics, the FEM was first used by Munro (1971, 1973), who applied it to the computation of magnetic fields in round lenses. Since this is of especial interest in electron optics, we concentrate on this application without neglecting electrostatic fields, including those in electron guns. A section is devoted to deflection units and multipoles. The presentation of the FEM given below differs from Munro's version in the introduction of form functions and in the unification of the methods for saturated and unsaturated lenses. For further background information, see Munro (1980, 1987a,b), Kasper (1987a,b,c), Morton (1987) and Reid (1987). Many aspects of the method are developed in more detail in a book by Kasper (2001); a monograph on *The Finite-Element Method in Charged Particle Optics* has been written by Khursheed (1999).

12.1 Formulation for Round Magnetic Lenses

The appropriate variational principle for magnetic field calculation is Eq. (6.13) in combination with Eqs (6.17), (6.18) and (6.29). Though magnetization curves are usually

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00012-2
© 2018 Elsevier Ltd. All rights reserved.

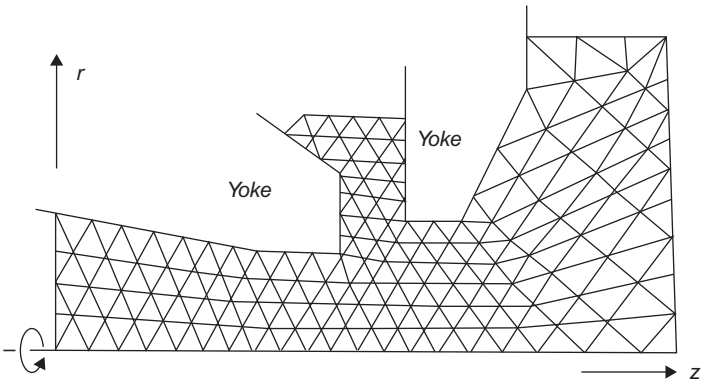


Figure 12.1
Finite-element discretization; this example concerns the vacuum part of a meridional section through a round magnetic lens.

presented in the form $B = B(H)$ and thus Eq. (6.18) is quite familiar, it is more convenient for the FEM to write

$$\beta \coloneqq B^2(z, r), \quad V(\beta) \coloneqq U(B) \tag{12.1}$$

Then Eq. (6.19) is in agreement with

$$\frac{1}{\mu} = \nu(\beta) = 2V'(\beta) \tag{12.2}$$

We now introduce cylindrical coordinates (z, r, φ) and perform the integration over φ. The variational principle Eq. (6.13) then takes the form

$$W = 2\pi \iint_S \Big\{ V(\beta) - j(z, r)A(z, r) \Big\} r\, dr\, dz = \text{minimum} \tag{12.3}$$

S being the domain of integration in the axial section and

$$\beta = (\nabla \times \boldsymbol{A})^2 = \left(\frac{\partial A}{\partial z} \right)^2 + \left(\frac{\partial A}{\partial r} + \frac{A}{r} \right)^2 \tag{12.4}$$

In order to solve this variational equation, some simplifying assumptions are necessary.

The expression (12.3) must be minimized with respect to any permitted variations of the vector potential $A(z, r)$. In the FEM these are variations of the values $A_j = A(z_j, r_j)$, $j = 1 \ldots N$, assumed at the *internal* nodes of a triangular grid, N being the total number of such nodes. In this context it is convenient to introduce dimensionless form functions $f_j(z, r)$

associated with the node with the corresponding subscript; a detailed definition will be given below.

We start now from a series expansion

$$A(z,r) = \sum_{k=1}^{N} A_k f_k(z,r) \tag{12.5}$$

and with Eq. (12.4) in mind, we introduce the abbreviation

$$\begin{aligned} F_{l,k}(z,r) = F_{k,l}(z,r) &= f_{l|z} f_{k|z} + (f_{l|r} + r^{-1} f_l)(f_{k|r} + r^{-1} f_k) \\ &\equiv \frac{\partial f_l}{\partial z}\frac{\partial f_k}{\partial z} + \left(\frac{\partial f_l}{\partial r} + \frac{f_l}{r}\right)\left(\frac{\partial f_k}{\partial r} + \frac{f_k}{r}\right) \end{aligned} \tag{12.6}$$

The function β may now be written as the quadratic form

$$\beta(z,r) = \sum_{l=1}^{N}\sum_{k=1}^{N} F_{l,k}(z,r) A_l A_k \tag{12.7}$$

Introducing Eqs (12.5) and (12.7) into (12.3) we obtain a discretization of this functional. The minimization condition now takes the form

$$\frac{\partial W}{\partial A_i} = 2\pi \iint_S \left\{ V'(\beta)\frac{\partial \beta}{\partial A_i} - f_i(z,r) j(z,r) \right\} r\, dr\, dz = 0, \quad (i = 1 \ldots N) \tag{12.8}$$

Evaluating this expression and recalling Eq. (12.2), we soon notice that it is favourable to introduce the matrix elements

$$L_{i,k} = 2\pi \iint_S \nu(\beta) F_{i,k}(z,r)\, r\, dr\, dz = L_{k,i} \tag{12.9}$$

$$M_i = 2\pi \iint_S j(z,r) f_i(z,r)\, r\, dr\, dz \tag{12.10}$$

Eq. (12.8) then takes the concise form

$$\sum_{k=1}^{N} L_{i,k}(\beta) A_k = M_i, \quad i = 1 \ldots N \tag{12.11}$$

which represents a *nonlinear* system of equations in the general case, since in saturated media ν is a function of β and hence depends implicitly on $A_1 \ldots A_N$.

So far, the discussion has been quite general. The choice of the form functions $f_j(z, r)$ is quite arbitrary, except that these must remain linearly independent so that the matrix L in

Eq. (12.9) is invertible for fixed values of ν. In order to perform the necessary numerical integrations in Eqs (12.9) and (12.10), however, a reasonably simple choice for the form functions is necessary. In the simplest choice, equivalent to Munro's original version of the FEM, they are piecewise linear functions which are joined together continuously at the nodes. Even the first-order derivatives are then discontinuous on the grid lines, however. With an alteration of the numbering, the form function corresponding to an arbitrary internal node 0 is sketched in Fig. 12.2A and B. It is nonzero only in the configuration shown, consisting of $n = 6$ triangles with the common node 0, and there it is a pyramid of unit height. It is defined in one particular subdomain ($i = 1$), see Fig. 12.3, by

$$f_{012}(z, r) = \frac{z_1 r_2 - z_2 r_1 + z(r_1 - r_2) + r(z_2 - z_1)}{z_1 r_2 - z_2 r_1 + z_0(r_1 - r_2) + r_0(z_2 - z_1)} \tag{12.12a}$$

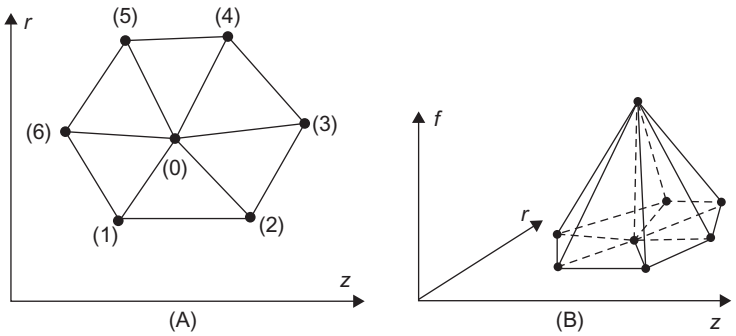


Figure 12.2
(A) Hexagonal configuration ($n = 6$) of nearest neighbours of an arbitrary internal node 0. (B) Perspective view of the corresponding linear form function; outside the hexagonal domain, this function vanishes.

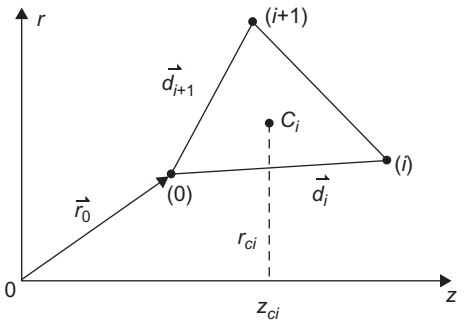


Figure 12.3
Notation for one of the triangular elements of which the hexagonal configuration of Fig. 12.2 is composed; $\boldsymbol{d}_j = \boldsymbol{r}_j - \boldsymbol{r}_0$ ($j = i, i + 1$) denote the side vectors and C_i the centroid. In the text, the area is denoted by a_i.

which assumes the nodal values

$$f_{012}(z_0, r_0) = 1, \quad f_{012}(z_1, r_1) = f_{012}(z_2, r_2) = 0 \tag{12.12b}$$

Linear functions of this type are said to be barycentric.

The remaining calculation is elementary but very lengthy. We have to differentiate the linear form functions, substitute these into Eqs (12.9) and (12.10) and then integrate. In this context some simplifying assumptions are necessary, essentially concerning the factors $\nu(\beta)$ and $j(z, r)$. In order to facilitate the integration, these factors are assumed to be piecewise linear or even piecewise constant functions. The results of these considerations will be given at the end of Section 12.2, since we need the notation introduced there.

12.2 Formulation for Self-adjoint Elliptic Equations

A variational principle is certainly a very common starting point but is not the only possible one. The following method is equivalent to it but can be made more general so that it remains applicable in cases where no variational principle is known.

We now regard (z, r) as quasi-Cartesian coordinates in a meridional plane and consider a self-adjoint elliptic differential equation of the form

$$\frac{\partial}{\partial z}\left(P\frac{\partial \Psi}{\partial z}\right) + \frac{\partial}{\partial r}\left(P\frac{\partial \Psi}{\partial r}\right) + Q(z, r) = 0 \tag{12.13}$$

$P = P(z, r)$ being a positive coefficient function. As in Eq. (12.5), we expand the potential Ψ in terms of form functions:

$$\Psi(z, r) = \sum_{k=1}^{M} \Psi_k f_k(z, r) \tag{12.14}$$

Here M is the total number of nodes; the numbering can always be chosen so that $i = 1 \ldots N$ refer to internal nodes, while $j = N + 1 \ldots M$ refer to boundary nodes. In Dirichlet problems, the boundary values $\Psi_{N+1} \ldots \Psi_M$ are kept fixed. Substituting this in Eq. (12.13), multiplying throughout by $f_i(z, r)$, $i = 1 \ldots N$, and integrating over the whole domain S of solution, we find

$$\iint_S \left[\sum_k \Psi_k f_i \left\{\frac{\partial}{\partial z}(P f_{k|z}) + \frac{\partial}{\partial r}(P f_{k|r})\right\} + Q f_i\right] dz\, dr = 0$$

With the necessary continuity conditions, integration by parts now leads to

$$\iint_S \left\{ -\sum_k P\Psi_k (f_{i|z} f_{k|z} + f_{i|r} f_{k|r}) + Q f_i \right\} dz\, dr + \oint_{\partial S} P\Psi_k f_i (n_z f_{k|z} + n_r f_{k|r})\, ds = 0, \quad i = 1 \ldots N$$

n_z, n_r denoting the cylindrical components of the outward-directed boundary normal. The contour integral vanishes, since the form functions referring to inner nodes must vanish at the boundary in the case of a Dirichlet problem. Introducing the arrays

$$G_{i,k}(z, r) = f_{i|z} f_{k|z} + f_{i|r} f_{k|r} = \nabla f_i \cdot \nabla f_k \tag{12.15a}$$

$$\overline{P}_{i,k} = \iint_S P(z, r) G_{i,k}(z, r)\, dz\, dr \tag{12.15b}$$

$$\overline{Q}_i = \iint_S Q(z, r) f_i(z, r)\, dz\, dr \tag{12.16}$$

we obtain the finite-element equations

$$\sum_{k=1}^{N} \overline{P}_{i,k} \Psi_k = \overline{Q}_i, \quad i = 1 \ldots N \tag{12.17}$$

which are identical with those obtained by evaluating the corresponding variational principle. From the latter, it might be concluded that Eq. (12.17) with (12.15b) remains valid even when the normal derivatives of the form functions are discontinuous at the grid lines, but this is not always true. In Eq. (12.17) the contributions of the boundary values to the inhomogeneity are incorporated on the left-hand side, as the summation covers *all* the nodes.

The matrix elements (12.15a) are considerably simpler than (12.6), as they are scalar products. Consequently, the results of the discretization using linear form functions can be cast into a fairly simple explicit form. Assuming that $P(z, r)$ and $Q(z, r)$ are constant and refer to the centroid C_i in each triangular element, we find, after some lengthy elementary calculations, for a configuration of n triangular elements with common node 0 like that shown in Fig. 12.2:

$$\Psi_0 \sum_{i=1}^{n} P_{ci} (\boldsymbol{d}_{i+1} - \boldsymbol{d}_i)^2 / a_i = \sum_{i=1}^{n} \frac{1}{a_i} \left[P_{ci} \left\{ d_i^2 \Psi_{i+1} + d_{i+1}^2 \Psi_i - \boldsymbol{d}_i \cdot \boldsymbol{d}_{i+1} (\Psi_i + \Psi_{i+1}) \right\} + \frac{4}{3} \sum_{i=1}^{n} a_i Q_{ci} \right] \tag{12.18}$$

Here the notation of Figs 12.2 and 12.3 has been adopted; it is cyclic in the sense that $\boldsymbol{d_{i+n}} = \boldsymbol{d_i}$, $\Psi_{i+n} = \Psi_i$; a_i is the area of the element with the side vectors $\boldsymbol{d_i}$ and $\boldsymbol{d_{i+1}}$.

This theory can be applied to the field in round magnetic lenses. The potential Ψ is then to be identified with the flux function $\Psi = 2\pi rA$, introduced in Section 6.4, and (12.13) must be identified with the flux equation arising from Eq. (6.33):

$$\frac{\partial}{\partial z}\left(\frac{\nu}{r}\frac{\partial \Psi}{\partial z}\right) + \frac{\partial}{\partial r}\left(\frac{\nu}{r}\frac{\partial \Psi}{\partial r}\right) = -2\pi j(z, r) \tag{12.19}$$

The coefficients are hence

$$P(z, r) = \nu(z, r)/r, \quad Q(z, r) = 2\pi\, j(z, r) \tag{12.20}$$

The differential equation and the corresponding finite-element discretization remain applicable even in the nonlinear case; we then have $\nu = \nu(\beta)$ with

$$\beta = |\boldsymbol{B}|^2 = \frac{1}{4\pi^2 r^2}\left(\Psi_{|z}^2 + \Psi_{|r}^2\right) = \frac{1}{4\pi^2 r^2}\sum_i \sum_k \Psi_i \Psi_k G_{i,k} \tag{12.21}$$

In the vicinity of the optic axis at least, quadratic form functions are necessary, since we know that $\Psi \propto r^2$ in the paraxial domain. Recalling that Ψ must vanish at the outer boundary and at the axis, we perceive that all the summations run only over the internal nodes ($i = 1 \ldots N$), as in Section 12.1.

We now state briefly the corresponding formulae for the vector potential, which result from the considerations in Section 12.1. Although not identical with Munro's formulae, they are equivalent to them.

Again adopting the notation introduced in Fig. 12.3, we find for the value of β at the centroid[1] C_i

$$\beta_{ci} = \frac{1}{4a_i^2}\left\{\boldsymbol{d}_{i+1}(A_0 - A_i) + \boldsymbol{d}_i(A_{i+1} - A_0) + \boldsymbol{s}_i(A_0 + A_i + A_{i+1})\right\}^2$$

with an additional shift

$$\boldsymbol{s}_i = \frac{2a_i}{3r_{ci}}\boldsymbol{i}_z$$

Then with $\nu_{ci} \coloneqq \nu(\beta_{ci})$ and $j_{ci} = j(z_{ci}, r_{ci})$, both referring to the centroid C_i of the element with label i, we obtain

[1]The centroid, also known as the centre-of-mass, is the point of intersection of the lines joining a vertex to the mid-point of the side facing the vertex.

$$A_0 \sum_{i=1}^{n} \frac{r_{ci}\nu_{ci}}{a_i}(\boldsymbol{d}_{i+1}-\boldsymbol{d}_i+\boldsymbol{s}_i)^2 - \frac{4}{3}\sum_{i=1}^{n} r_{ci}a_i j_{ci} = \sum_{i=1}^{n} \frac{r_{ci}\nu_{ci}}{a_i}\left\{A_i(\boldsymbol{d}_{i+1}-\boldsymbol{s}_i) - A_{i+1}(\boldsymbol{d}_i+\boldsymbol{s}_i)\right\}\cdot(\boldsymbol{d}_{i+1}-\boldsymbol{d}_i+\boldsymbol{s}_i) \tag{12.22}$$

Apart from different material coefficients, the essential difference between Eqs (12.22) and (12.18) lies in the appearance of the shift $\boldsymbol{s}_i$ which results from the term A/r in Eq. (12.4).

Another interesting application of the FEM to magnetic lenses is the calculation of magnetic circuits made of anisotropic material. In this case the reluctance $\nu(\beta)$ is to be replaced by a symmetric tensor. Its components depend directly on the position $\boldsymbol{r}$ as a consequence of the variable crystallographic orientation in the material and indirectly due to saturation effects. Such calculations are extremely complicated; nevertheless, magnetic circuits with anisotropic material can be advantageous. Balladore et al. (1981, 1984) have shown that the size and weight of the yoke can be appreciably reduced in this way. Other formulations of the FEM that are used to study electron lenses, deflection units and multipoles are described in Section 12.4. Determination of the field functions for both electron lenses and deflectors is described by Zhu et al. (1996).

12.3 Solution of the Finite-Element Equations

In the case of linear (unsaturated) media, the matrix elements $L_{i,k}$ in Eq. (12.11) and $\overline{P}_{ik}$ in Eq. (12.17) are constants; the corresponding systems of equations are therefore linear and can be solved by means of standard techniques. It is usual to employ direct solution techniques, in which case it is desirable to make use of an ordering that minimizes the bandwidth of the corresponding sparse matrix. These techniques cannot be outlined here; the reader is referred to the corresponding literature (Cuthill and McKee, 1969; Gibbs et al., 1976; Duff, 1977). A very fast iterative procedure using the preconditioned or incomplete Cholesky conjugate-gradient (ICCG) method (Meijerink and van der Vorst, 1977) has been developed by Lencová and Lenc (1984, 1986) and this is now the standard technique.

Eq. (12.18) has already been cast into a form which is suitable for iterative techniques such as SOR and SLOR. These are efficient if the coefficients in Eq. (12.18), referring to each internal node (0), are computed once and for all at the beginning and then stored. In the absence of source terms Q_{ci} and with $n = 6$, the total memory requirement is 7 N.

When saturation or other nonlinear effects are present, the situation becomes more complicated, as Eqs (12.11) and (12.17) are now nonlinear systems of equations: iterative procedures are unavoidable. Direct techniques must be combined with Newton's iterative procedure (Munro, 1973). When the SOR is employed, a quasi-linearization is necessary; Eqs (12.11) and (12.17) already have the appropriate form if

$\nu(\beta)$ is first treated as a linear coefficient during each iteration over the field and then recalculated before the next cycle according to Eq. (12.2) together with (12.7) or (12.21). A mesh designed for highly saturated lenses has been devised by Podbrdský and Krivanek (1988). The use of the FEM to study permanent-magnet lenses is discussed by Kamminga (1975).

12.4 Improvement of the Finite-Element Method

12.4.1 Introduction

The form of the FEM presented above is only the very simplest version. We have chosen it in order to display the basic ideas clearly. The FEM can, of course, be improved in many ways. The corresponding theory has been developed in so much detail that it is impossible to treat it adequately here. It is even impossible to present a complete list of references. We must confine our considerations to some essential points.

The linear form functions given by Eq. (12.12) are those of the lowest permissible order. For the azimuthal component of the vector potential $A(z, r)$ such a choice is reasonable, since in the most important, paraxial domain the function $A(z, r)$ is proportional to r. There are also presentations of the FEM in which a linear approximation is made for scalar potentials (Munro, 1973) and flux functions (Bonjour, 1980). This was clearly a locally very inaccurate approximation, criticized by Kasper and Lenz (1980), who showed that the FEM is then less accurate than the FDM, at least in the paraxial domain. This weakness is avoided by modifying the finite-element equations for elements close to the axis or by using higher order approximations for the potential. The corresponding theory has been presented by Konrad and Silvester (1973). It is then necessary to introduce additional points of reference along the mesh lines of the grid and also in the interior of the elements (see Fig. 12.4). The values of the potentials at these additional points are unknown. The rank of the system of finite-element equations and, of course, the complexity of its structure

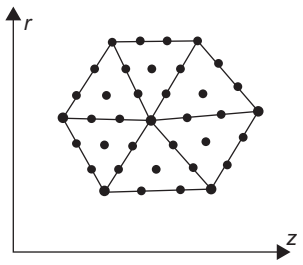


Figure 12.4
Configuration of finite elements and reference points to be used in a third-order approximation.

increase accordingly but a gain in accuracy may be achieved. The second-order finite-element method is presented in Section 12.4.3.

A second important aspect is the appropriate choice of the grid. In practice it is unreasonable to require the user of a program to choose every node individually. Suitable algorithms for the automatic generation of the grid have been proposed by Winslow (1967) and Munro (1973). Whether these algorithms work efficiently or not depends essentially on the shape of the boundary in question. More recently, Hermeline (1982) proposed a new method, which works well but is highly sophisticated. The user of a commercial finite-element program is supplied with a method for generating the grid, but setting up one's own program will generally be very laborious. The benefits of graded meshes are emphasized by Lencová (1995b). Khursheed has shown that 'conformal', nearly square meshes are superior to nearly rectangular meshes when very different dimensions are involved, in electron guns for example (1997a,b). This choice of mesh is incorporated in Khursheed's KEOS package. Adaptive mesh generation for gun studies is explained at length by Xia and Broers (1995).

A third aspect of the FEM − and indeed of the FDM − is the need to confine the spatial extension of the grid. Very often the fringe-fields of a particular configuration spread out over a much larger region than can be covered by the grid. In order to keep the error introduced by cutting off the field at the boundary of the grid sufficiently small, the size of the grid must often be very large. This drawback of the FEM can be removed by the introduction of additional infinite elements. The trial functions for the potential, to be evaluated in such elements, must be consistent with the asymptotic form of the real potential. Such infinite elements have been proposed by Bettess (1977) and their use in electron optics has been investigated by Lencová and Lenc (1982, 1984). However, it is now possible to use such a large number of meshes that such special precautions are often no longer necessary.

The problem of computing field strengths, which requires sophisticated interpolation routines, will be discussed in Chapter 13, Field-Interpolation Techniques; it also arises in connection with the use of differential algebra in Section 34.8.

12.4.2 Alternative Formulations

The foregoing general presentation is indispensable when a novel configuration is to be studied. We now examine in more detail its application to the common elements of electron optics: round lenses, multipoles and deflection units. We first reconsider the first-order finite-element method (FOFEM) for magnetic lenses. In Munro's original work, the value of the integrand Eq. (6.13 or 12.3) was taken at the centroid of each triangular element.

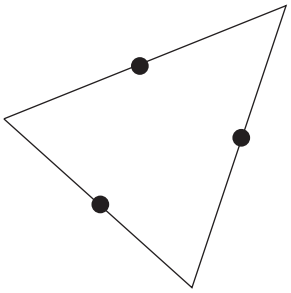


Figure 12.5
Midpoints of sides for a finite-element calculation.

However, other choices have been examined by Lencová and Lenc (1986), who found that the accuracy is improved by taking the average of the values at the midpoints of the sides of the triangle (Fig. 12.5). Related versions of the FOFEM also have advantages. Use of $P \coloneqq 2\,A/r$ instead of A in the energy functional (12.3) has the attraction that, close to the optic axis, $P(r) \approx B(z) - \frac{1}{8}r^2B''(z) + \ldots$ The functional can then be evaluated in terms of (r, z) or of (r^2, z). Another formulation was proposed by Lencová and Lenc (1992), designed to treat correctly the boundary between materials of different permeability, such as the yoke and polepieces of a magnetic lens. For the components of the magnetic flux B, they use

$$B_z = \frac{\sum_i c_i A_i}{\sum_i c_i r_i} + \frac{\sum_i |c_i| A_i}{\sum_i |c_i| r_i}$$

$$B_r = -\frac{\sum_i b_i A_i}{\sum_i b_i z_i} = -\frac{\sum_i b_i A_i}{\sum_i c_i r_i}$$

$$\sum_i b_i z_i = \sum_i c_i r_i = D$$

A further possibility is to replace A by the flux $\Psi = 2\pi r A$, which we met at the end of Section 12.2.

For all these slightly different approaches, the energy stored in an element of the mesh can be written as

$$\Delta W = \frac{\pi(k+1)}{6} \sum_{i=1}^{3} \left(-JF_k p_i + \frac{F_k}{2D_k^2 \mu} \sum_{j=1}^{3} q_{ij} V_j \right) V_i$$

where J denotes the current density in any triangle.

Table 12.1: Values of the coefficients of p_i and q_{ij} appearing in the expression for the stored energy

Method	p_i	q_{ij}
A (centroid)	$r_{s1}/3$	$b_ib_j + (c_i + d)(c_j + d)$
B (midpoints of sides)	$(r_{s1} + r_i)/4$	$b_ib_j + (c_i + d)(c_j + d) - d^2 + d^2 r_{s1}\{(1 - \delta_{ij})r_{ij} + \delta_{ij}(r_{il} + r_{jl})\}$
C	$(r_{s1} + r_i)/4$	$b_ib_j + (c_i + e_i)(c_j + e_j)$
P_1 (r, z)	$f_i/20$	$3r_{s3}(\xi b_ib_j + c_ic_j)/r_{s1} + 2d(c_if_j + c_jf_i) + 2dD_1(1 + \delta_{ij})(r_i + r_j + r_{s1})$
P_2 (r^2, z)	$r_i/6$	$[u_{s1}d_id_j + 2u_{s2}c_ic_j + \{(u_{s1} + u_i)c_j + (u_{s1} + u_j)c_i\}D_2 + (1 + \delta_{ij})D_2{}^2]/12$
Ψ	$1/6\pi r_j$	$\{(u_{12} + u_{23} + u_{13})d_id_j/6 + c_ic_j\}/\pi^2$

When (r, z) is used, $k = 1$; for (r^2, z), $k = 2$. $F_1 = r_{s1}D_1 \quad F_2 = D_2$. The geometrical quantities p_i and q_{ij} are listed in Table 12.1 for the various choices. The other quantities appearing in this Table are as follows:

$$r_{s1} = r_1 + r_2 + r_3$$

$$r_{s2} = r_1^2 + r_2^2 + r_3^2 + r_1r_2 + r_1r_3 + r_2r_3$$

$$r_{s3} = r_1^3 + r_2^3 + r_3^3 + r_1^2r_2 + r_1^2r_3 + r_1r_2^2 + r_1r_3^2 + r_2^2r_3 + r_2r_3^2 + r_1r_2r_3$$

$$u_i = r_i^2$$

$$u_{s1} = u_1 + u_2 + u_3$$

$$u_{s2} = u_1^2 + u_2^2 + u_3^2 + u_1u_2 + u_1u_3 + u_2u_3$$

$$b_1 = r_2 - r_3 \quad b_2 = r_3 - r_1 \quad b_3 = r_1 - r_2$$

$$c_1 = z_3 - z_2 \quad c_2 = z_1 - z_3 \quad c_3 = z_2 - z_1$$

$$d_1 = u_2 - u_3 \quad d_2 = u_3 - u_1 \quad d_3 = u_1 - u_2$$

$$D_1 = b_1c_2 - b_2c_1 \quad D_2 = d_1c_2 - d_2c_1 \quad d = D_1/r_{s1}$$

$$e_i = \frac{D_1|c_i|}{r_1|c_1| + r_2|c_2| + r_3|c_3|}$$

$$f_i = r_{s2} + r_i(r_i + r_{s1})$$

$$r_{ij} = \frac{1}{r_i + r_j} \quad \text{if} \quad r_i + r_j \neq 0, \quad \text{otherwise} \quad r_{ij} = 0$$

$$u_{ij} = \frac{1}{u_i + u_j} \quad \text{if} \quad u_i + u_j \neq 0, \quad \text{otherwise} \quad u_{ij} = 0$$

$$\xi = 5/8 \ \text{ if two vertices of the triangle are on the axis, otherwise } \xi = 0$$

The errors associated with each of these formulations are analysed by Lencová and Lenc (1996a,b) and Lencová (1999), who conclude that B (midpoints of sides) and P_1 give good results when $B(z)$ is required, especially when a fine graded mesh is adopted. For exact ray tracing, however, C or P_2 is better.

12.4.3 First- and Second-Order Finite-Element Methods (FOFEM and SOFEM)

The advantages of using second-order "isoparametric" finite elements with nine nodes instead the triangular elements with linear variation of potential have been assessed by Zhu and Munro (1989, 1995) and the second-order version is routinely used in some of the commercial program suites of Munro's Electron Beam Software (MEBS).

First, the region enclosed by electrodes or magnetic material is divided into large quadrilateral regions, which may have curved sides; some of the latter will coincide with the physical boundaries. These quadrilateral regions are regarded as Coons' patches (Fig. 12.6)

$$\boldsymbol{r}(p,q) = (1-q)\boldsymbol{R}_1(p) + q\boldsymbol{R}_3(p) + (1-p)\boldsymbol{R}_2(q) + p\boldsymbol{R}_4(q) \\ -(1-p)(1-q)\boldsymbol{R}_1(0) - (1-p)q\boldsymbol{R}_3(0) - p(1-q)\boldsymbol{R}_1(1) - pq\boldsymbol{R}_3(1)$$

where $\boldsymbol{r}(p,q)$ is the position vector at a point (p, q) inside or on the perimeter of the patch. A finer quadrilateral mesh is then generated by plotting lines of constant p and q, as shown in Fig. 12.7. The resulting quadrilaterals are now the finite elements, which have nine nodes and still have curved boundaries. This is not convenient for the subsequent calculations and the quadrilaterals are therefore mapped to rectangles (Fig. 12.8A)

$$z(u,v) = \sum_{i=0}^{2}\sum_{j=0}^{2} z_{ij}\alpha_i(u)\alpha_j(v)$$

$$r(u,v) = \sum_{i=0}^{2}\sum_{j=0}^{2} r_{ij}\alpha_i(u)\alpha_j(v)$$

$$\Phi(u,v) = \sum_{i=0}^{2}\sum_{j=0}^{2} \Phi_{ij}\alpha_i(u)\alpha_j(v)$$

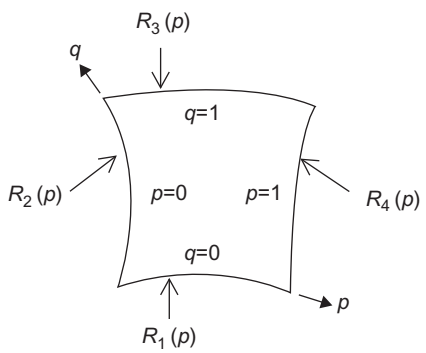


Figure 12.6
Notation for a linear Coons' patch.

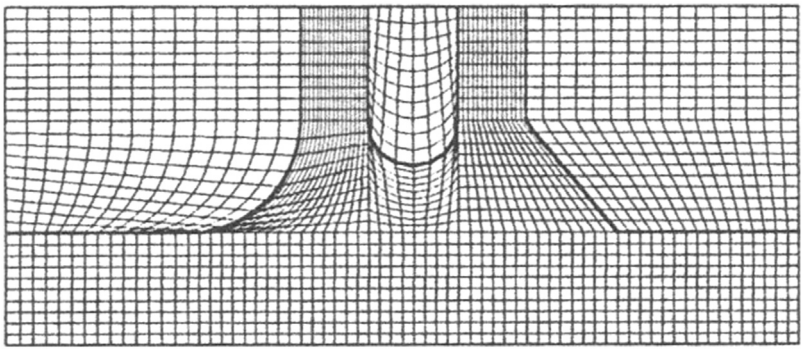


Figure 12.7
Fine quadrilateral mesh covering an electrostatic lens. Courtesy J. Rouse.

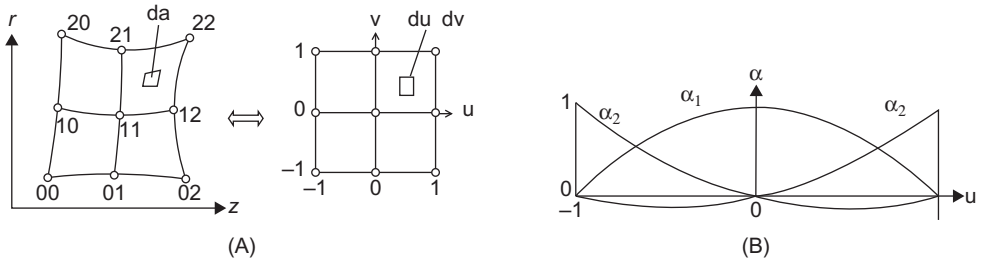


Figure 12.8
(A) Mapping of a curvilinear element into a square element. (B) The Lagrange functions. After Munro (1997), Courtesy: Taylor & Francis.

in which z_{ij} and r_{ij} denote the coordinates at the nine nodes and Φ_{ij} denotes the potential in the electrostatic case. The functions α_0, α_1, α_2 are the quadratic Lagrange functions

$$\alpha_0(x) = \frac{x(x-1)}{2}, \quad \alpha_1(x) = -(x-1)(x+1), \quad \alpha_2(x) = \frac{x(x+1)}{2}$$

in which x may represent u or v (Fig. 12.8B).

Minimization of the functional proceeds as before, yielding a set of linear equations for the values of the potential. These are then solved by Gaussian elimination or preferably by the incomplete Cholesky conjugate-gradient (ICCG) method. These equations are more complicated than in the first-order case, as nodes on the boundaries between two finite elements and those at the point of intersection of four finite elements require special attention.

We note that the relative merits of the first- and second-order finite-element methods have excited considerable discussion, Munro and colleagues arguing in favour of the SOFEM while Lencová prefers the first-order method. With the large storage capacity and greatly

increased speed of modern computers, the accuracy of the SOFEM can now be rivalled by increasing the number of elements employed in the FOFEM without any unacceptable increase in computing time.

12.5 Comparison and Combination of Different Methods

In Chapters 10–12, we have dealt with three major methods of calculating potentials, the BEM, the FDM and the FEM. The question that now arises is, which one should be preferred in a given case. The answer depends on the details of the particular problem to be solved.

In all cases in which a one-dimensional linear integral equation can be derived, the BEM is the most advantageous means of obtaining a solution. The necessary discretization can be easily fitted to arbitrary boundaries, regardless of whether these are curved or piecewise straight with sharp edges. Even extreme differences in the dimensions of boundaries, as in field-emission electron guns, for example, are no obstacle to this method. There is no need to cut off fringe-fields, as theoretically the domain of solution is the whole space. With a comparatively modest memory capacity, high accuracy can be achieved. The linear system of equations to be solved is well-conditioned and can hence be solved directly by means of a simple Gaussian elimination without any pivoting. After determining the appropriate surface-source distributions, the analytic expressions for the field strength can be evaluated at any point of reference. In principle, there is no need for additional interpolation and numerical differentiation techniques, though these may be helpful in some cases.

The same conclusions hold when the two-dimensional BEM is applied to three-dimensional boundary-value problems (Section 10.4). Of course, this method is then more complicated than the one-dimensional BEM, but this is an inevitable consequence of the greater complexity of the problem to be solved; alternative methods such as the FDM and the FEM will also become more complicated.

The FDM is suitable only when a regular grid fits the boundary, since the inclusion of irregularities, though quite elementary, is very tedious. Since highly regular domains of solution are very rare, the FDM is not advantageous, in electron optics at least, unless the improvements outlined in Sections 11.5.2 and 11.5.3 are incorporated. It is, as we have already noted, used in the program SIMION, where the use of multiple grids is current.

The FEM is theoretically applicable to any kind of boundary-value problem, even in three dimensions. This method is certainly very effective if a highly perfected and tested program package is available but if this is not the case, we should prefer the BEM, since the latter can be easily programmed by a single scientist, at least in the one-dimensional version. When saturation effects become important in ferromagnetic materials, however, the FEM is definitely the best choice.

Systematic comparisons of the FDM, FEM and BEM were made several years ago by Cubric et al. (1997, 2D problems; 1999, 3D problems). At that time, the BEM outperformed the other two methods but with later hardware developments, the FEM may well have narrowed or even closed the gap. Furthermore, the measures of performance used at that date would not necessarily be adopted today. The commercial software proposed by MEBS and SPOC rely largely on the finite-element method, with occasional use of the finite-difference method. The CPO programs (Read), on the other hand, use the boundary-element method. In the light of all this, no clear-cut preference emerges – the differences between the methods are insignificant with the performance of modern computers.

Several combinations of the different methods of field calculation have been investigated, and one has already been outlined in Section 11.4.3. In electron optics this possibility has been found valuable for field computation in field-emission electron guns (see Part IX). More generally, the combination of different methods for the solution of Dirichlet problems in electrostatics has been investigated by Schaefer (1982, 1983), who has proved quite generally that iterative solutions of Dirichlet problems in two and more overlapping domains converge. He has developed a suitable technique for the solution of such problems, which he calls Schwarz's alternating method. Though this method can be very powerful, we cannot devote space to it here. Unfortunately, this method does not work for problems with interface conditions or for nonlinear problems, where a suitable coupling of different methods is particularly interesting. In these cases a combination of the FEM with the BEM is possible, as has been proposed by McDonald and Wexler (1972) and McDonald et al. (1973) and by Lencová and Lenc (1982, 1984). This last paper gives very full details of the theory and programming of the procedure for a Mulvey lens similar to that shown in Fig. 12.9. A similar method has been proposed by Kasper (1984a,b); we now outline this briefly.

A typical example of the application of a hybrid method is presented in Fig. 12.9, which shows a half-axial section through an open magnetic lens. Owing to this open structure the magnetic fringe-field extends so far that it becomes somewhat impractical to apply the FEM in the vacuum domain V. On the other hand, the saturation effects in the polepieces make the application of Scherle's method impossible. We hence apply the FEM only to the polepieces and use the BEM in the outer domain.

The finite-element equations can be solved if the boundary values of $A(z, r)$ at the surface contour C are known. Then by means of suitable interpolation techniques the normal derivative $(\partial A/\partial n)_i$ on the inner side can be determined. When this has been done, we also know the normal derivative $(\partial A/\partial n)_v$ on the vacuum side from

$$\frac{1}{\mu_i}\left(\frac{\partial \hat{A}}{\partial n}\right)_i = \frac{1}{\mu_0}\left(\frac{\partial \hat{A}}{\partial n}\right)_v, \quad \hat{A} := rA \tag{12.23}$$

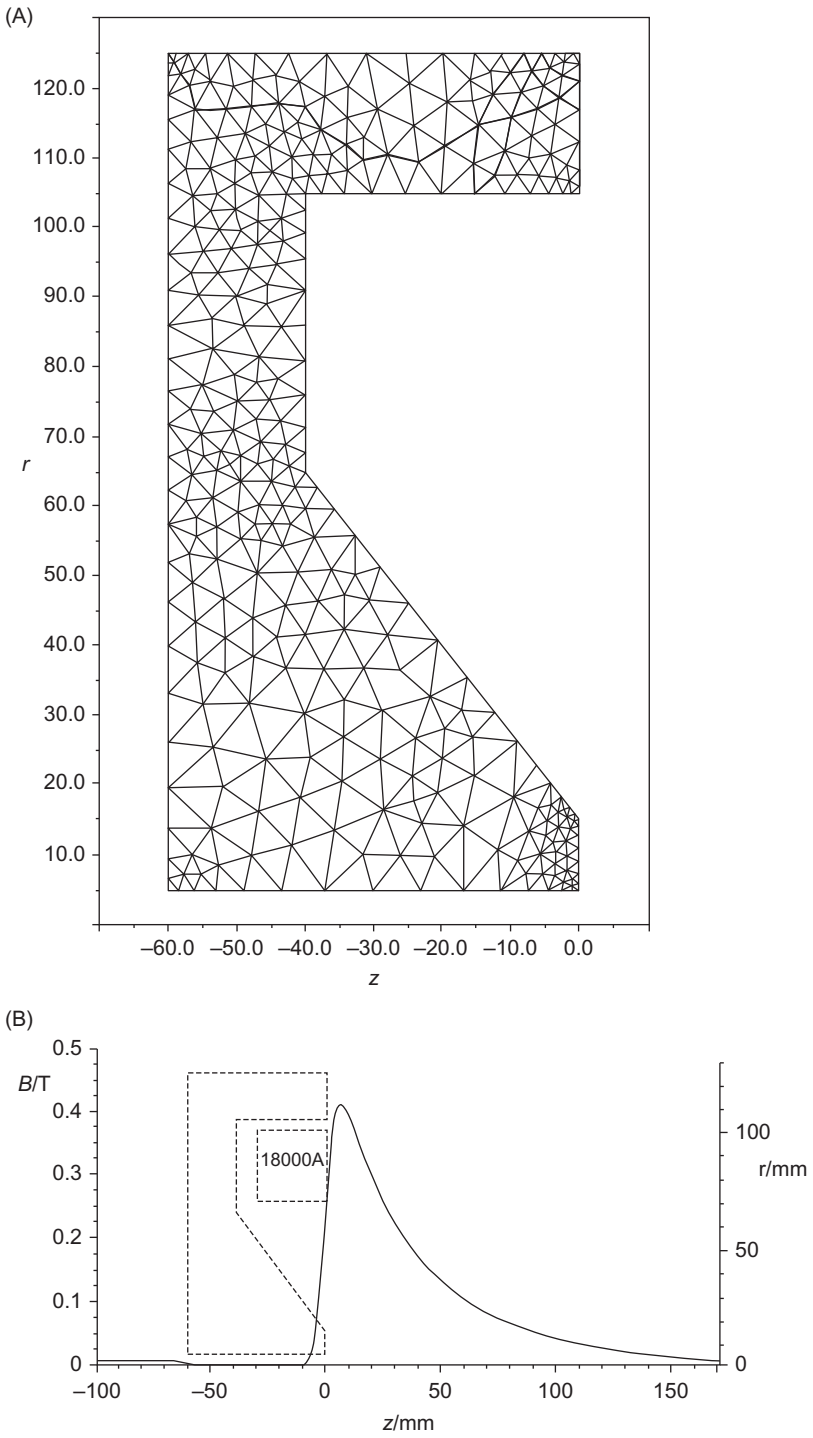


Figure 12.9

(A) Upper half of a meridional section through an open, round, magnetic lens with a ferromagnetic core and a rectangular distribution of windings. Only the interior of the yoke is discretized by a triangular-mesh grid. (B) Axial field strength for an excitation of 18000 A-turns.

We can now solve the integral equation

$$\frac{1}{2}A(\boldsymbol{u}) = \oint_C \left\{ A(\boldsymbol{u}')\frac{\partial G_1}{\partial n'} - G_1(\boldsymbol{u},\boldsymbol{u}')\left(\frac{\partial A}{\partial n'}\right)_v \right\} r'\,ds' + \mu_0 \int_V G_1(\boldsymbol{u},\boldsymbol{u}')j(\boldsymbol{u}')\,r'\,dr'\,dz' \qquad (12.24)$$

in which we have written $\boldsymbol{u} = (z, r)$, $\boldsymbol{u}' = (z', r')$ and G_1 is defined by (9.21) with $m = 1$.

When the whole system of coupled equations has been solved iteratively, we have the appropriate solution in the partly saturated iron together with an unbounded and smooth vacuum field. Thus the drawbacks of using each of the individual methods separately have been circumvented.

We repeat that the need for such a hybrid method has almost vanished, given the capabilities of modern computers. Nevertheless, Kubo et al. (2017) have used at once the finite-element method (EOD) and the finite-difference method (SIMION) to model the optics of the entire column of a transmission electron microscope, a Hitachi In-situ Interferometry TEM (I2TEM), which includes a biprism and a CEOS aberration corrector (See Chapter 41, Aberration Correction). EOD was needed to model the magnetic lens properties, where SIMION could not compete. An example of the results is illustrated in Fig. 12.10. The splitting convergent-beam mode (Houdellier et al., 2015), in which a biprism is employed, was also simulated as an example of the flexibility of the procedure.

12.6 Deflection Units and Multipoles

The finite-element method can be used to study these nonrotationally symmetric elements; some familiarity with the contents of Chapters 32 and 33, Paraxial Properties of Deflection Systems and The Aberrations of Deflection Systems, is assumed.

The finite-element method was first used to calculate the properties of deflectors by Munro and Chu (1982), who considered toroidal deflectors and saddle-coil deflectors on a cylindrical former. This work was extended by Lencová et al. (1989) who included the possibility of using a conical former and introduced a modified form of the energy functional. We seek a solution of the form

$$\boldsymbol{H} = \nabla\chi + \boldsymbol{F} \quad \boldsymbol{J} = \text{curl}\,\boldsymbol{F} \qquad (12.25)$$

where χ is the magnetostatic potential, $\chi = -W/\mu$ (7.41, 7.42) and $\boldsymbol{J}$ is the current density in the coil windings. We shall need to compute several harmonics of the field, which makes it preferable to replace χ by a new function Ψ:

$$\begin{aligned} \chi(r,z,\varphi) &= \sum_{m\ \text{odd}} \chi_m(r,z)\cos m\varphi \\ &=: \sum_{m\ \text{odd}} r^m \Psi_m \cos m\varphi \end{aligned} \qquad (12.26)$$

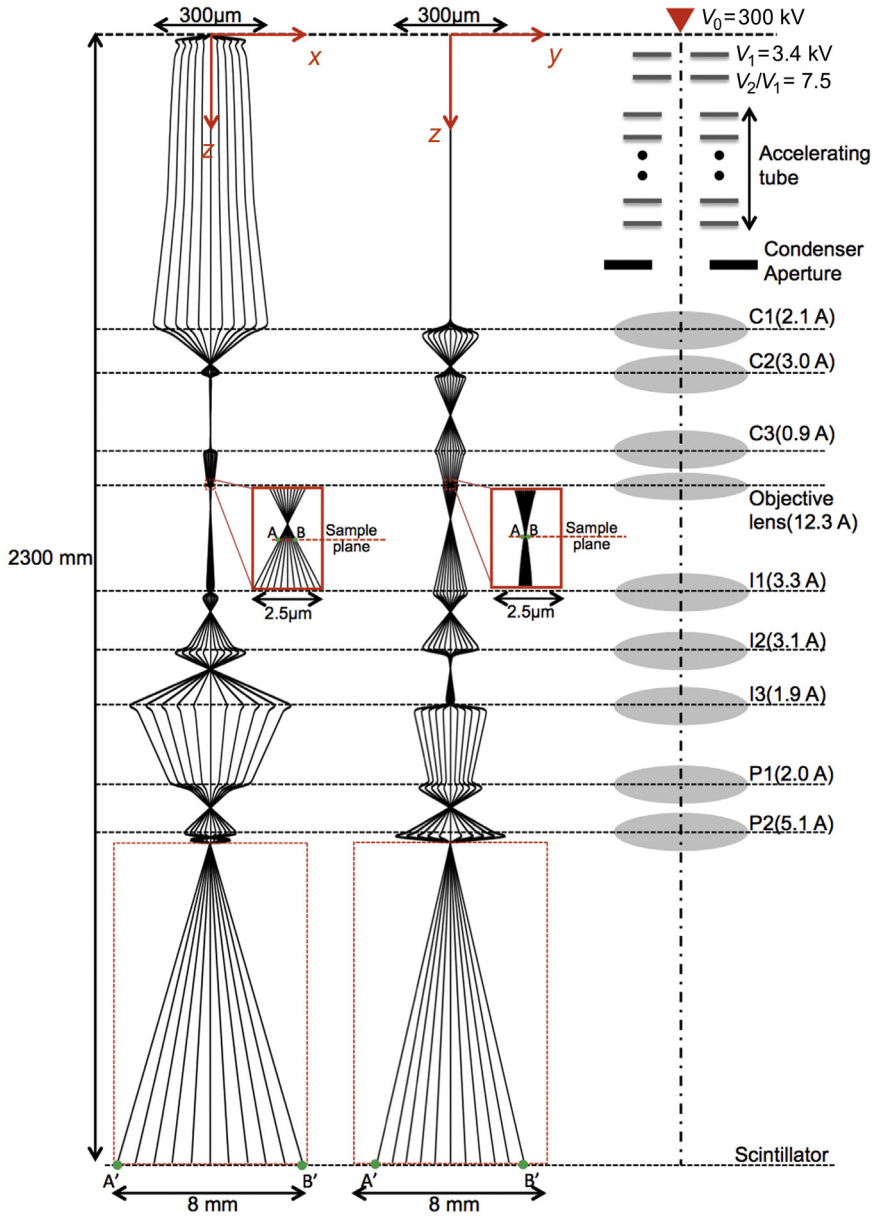


Figure 12.10

Electron trajectories in the I2TEM. The trajectories have been considerably expanded laterally. The aberration corrector, situated between the objective lens and the first intermediate lens (I1) is not included in this simulation. I: intermediate lens; P: projector lens.

The boundary condition on the optic axis, $\chi(0, z, \varphi) = 0$, is replaced by the Neumann boundary condition $\partial\Psi_m/\partial r = 0$.

The vector function $\boldsymbol{F}$ is zero everywhere outside the windings and can be chosen to be directed along the normal to the surface of the zone enclosed by the conducting wires. It has a slightly different form for the two coil configurations but a unified expression can be generated (Lencová et al., 1989). For toroidal coils, only the component F_φ is needed. This is first rewritten as

$$F_\varphi(r, z, \varphi) = \frac{g(r, z) f(\varphi)}{r} \tag{12.27}$$

after which the *loading function* $f(\varphi)$ is expanded as a Fourier series:

$$\begin{aligned} f(\varphi) &= \sum_{m \text{ odd}} f_m \sin m\varphi \\ &= \sum_{m \text{ odd}} \frac{4\pi NI}{\pi} \sin \varphi_c \sin m\varphi \end{aligned} \tag{12.28}$$

The function $g(r, z)$ characterizes the location of the windings:

$$\begin{aligned} g(r, z) &= 1 \quad \text{inside the windings} \\ g(r, z) &= 0 \quad \text{elsewhere} \end{aligned} \tag{12.29}$$

The angle between the planes containing the windings is denoted by φ_c.

For saddle coils, the loading function is expanded as

$$\begin{aligned} f(\varphi) &= \sum_{m \text{ odd}} f_m \cos m\varphi \\ &= \sum_{m \text{ odd}} \frac{4\pi NI}{\pi mt} \sin m\varphi_c \cos m\varphi \end{aligned} \tag{12.30}$$

where the current is assumed to flow through a small zone of thickness t. The angle φ_c is now defined by the shape of the coil (Fig. 12.11A). The components of $\boldsymbol{J}$ are given by

$$J_r = \frac{1}{r}\frac{\partial F_z}{\partial \varphi}, \quad J_z = -\frac{1}{r}\frac{\partial F_r}{\partial \varphi}, \quad J_\varphi = \frac{\partial F_r}{\partial z} - \frac{\partial F_z}{\partial r} \tag{12.31}$$

We find

$$F_r = g(r, z) f(\varphi) \cos \alpha \quad F_z = -\, g(r, z) f(\varphi) \sin \alpha \tag{12.32}$$

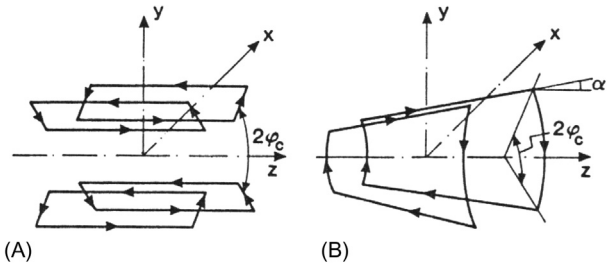


Figure 12.11
Geometry of toroidal (a) and tapered saddle (b) coils used in electron beam deflection.

in which $\alpha = 0$ for cylindrical coils and $\alpha = \pi/2$ for flat coils. (The case in which α varies with z can also be included but is not considered here.)

Before replacement of χ by Ψ, the components of the energy functional (here denoted by U to prevent confusion with the magnetic scalar potential) take the form

$$U_m = \frac{\pi\mu}{2}\iint\left\{\left(\frac{\partial\chi_m}{\partial r}+a_1 g f_m\right)^2+\left(\frac{\partial\chi_m}{\partial z}+a_2 g f_m\right)^2+\left(-\frac{m\chi_m}{r}+a_3 g f_m\right)^2\right\}r\,dr\,dz \quad (12.33)$$

Odd m, saddle coils: $f_m = 4NI\sin m\varphi_c/\pi mt$

Odd m, toroidal coils: $f_m = 4NI\sin m\varphi_c/\pi$

(For even values of m, the values of f_m are twice these.)

For saddle coils, $a_1 = \cos\alpha, \quad a_2 = -\sin\alpha, \quad a_3 = 0$

For toroidal coils, $a_1 = 0, \quad a_2 = 0, \quad a_3 = 1$

On replacing χ by Ψ, U_m becomes

$$U_m = \frac{\pi\mu}{2}(L - R + A) \quad (12.34)$$

in which

$$\begin{aligned} L &= \iint r^{2m}\left\{r\left(\frac{\partial\Psi_m}{\partial r}\right)^2+r\left(\frac{\partial\Psi_m}{\partial z}\right)^2+2m\Psi_m\frac{\partial\Psi_m}{\partial r}+2m^2\frac{\Psi_m^2}{r}\right\}dr\,dz \\ R &= -2gf_m\iint r^m\left\{a_1\left(r\frac{\partial\Psi_m}{\partial r}+m\Psi_m\right)+a_2 r\frac{\partial\Psi_m}{\partial z}-a_3 m\Psi_m\right\}dr\,dz \end{aligned} \quad (12.35)$$

(A has no effect on the subsequent theory). The integration is taken over the triangular elements (Fig. 12.11A and B).

The calculation now follows the same lines as that for round lenses. We give the essential steps without comment. Linear shape functions, $\psi(r, z) = f + gz + hr$, are employed, with

$$f = \frac{\sum d_i\psi_i}{D}, \quad g = \frac{\sum b_i\psi_i}{D}, \quad h = \frac{\sum c_i\psi_i}{D} \tag{12.36}$$

where ψ_i denote the values of Ψ_m at the vertices of each triangle, the area of which is $D/2$.

$$d_1 = z_2r_3 - z_3r_2, \quad b_1 = r_2 - r_3, \quad c_1 = z_3 - z_2, \quad \text{etc} \tag{12.37}$$

In order to evaluate L and R (Eq. 12.35), the following formulae, derived by Lencová et al, are required:

$$\begin{aligned}
\iint r^n dr\,dz &= \frac{D}{(n+1)(n+2)}R(n) \\
\iint r^n\psi dr\,dz &= \frac{D}{(n+1)(n+2)(n+3)}\sum_i T_i(n)\psi_i \\
\iint r^n\psi^2 dr\,dz &= \frac{D}{(n+1)(n+2)(n+3)(n+4)}T_{ij}(n)\psi_i\psi_j
\end{aligned} \tag{12.38}$$

in which the integration is again taken over each triangle and

$$\begin{aligned}
R(n) &= \sum_{k=0}^{n}\sum_{l=0}^{n-k} r_1^k r_2^l r_3^{n-k-l} \\
T_i(n) &= \sum_{k=0}^{n}\sum_{l=0}^{n-k}(k+1) r_i^k r_p^l r_q^{n-k-l}, \quad p \neq i, \quad q \neq i,p, \quad T_i(0) = 1 \\
T_{ij}(n) &= \sum_{k=0}^{n}\sum_{l=0}^{n-k}(k+1)(l+1) r_i^k r_j^l r_p^{n-k-1}, \quad j \neq i, \quad p \neq i,j \\
T_{ii}(n) &= \sum_{k=0}^{n}\sum_{l=0}^{n-k}(k+2)(l+1) r_i^k r_p^l r_q^{n-k-1}, \quad p \neq i, \quad q \neq i,p
\end{aligned} \tag{12.39}$$

Minimization of U_m leads to the following expression for each triangle:

$$f_{ij} = \frac{\mu}{D}(b_ib_j + c_ic_j)R(2m+1) + \mu\frac{m}{2m+1}c_iT_j(2m) + c_iT_j(2m) + DT_{ij}(2m-1) \tag{12.40}$$

and these enable us to build the matrix characterizing the set of linear equations for ψ. On the right-hand side, we have

$$\begin{aligned}
g_i = &-2\mu fgm\frac{2m+3}{(m+2)(m+3)}\big[a_1\{(m+1)c_iR(m+1) + mDT_i(m)\} \\
&+ a_2(m+1)b_iR(m+1) - a_3D(m+3)T_i(m-1)\big]
\end{aligned} \tag{12.41}$$

$R(n)$, $T_i(n)$ and $T_{ij}(n)$ can be easily evaluated with the aid of the following recursion rules:

$$\begin{aligned} R(n) &= r_1 R(n-1) + S(n), & R(1) &= r_1 + r_2 + r_3 \\ T_i(n) &= r_i T_i(n-1) + R(n), & T_i(1) &= r_i + r_1 + r_2 + r_3 \\ T_{ii}(n) &= r_i T_{ii}(n-1) + 2T_i(n), & T_{ii}(1) &= 4r_i + 2(r_1 + r_2 + r_3) \\ T_{ij}(n) &= r_j T_{ij}(n-1) + T_i(n), & T_{ij}(1) &= r_i + r_j + r_1 + r_2 + r_3 \end{aligned} \tag{12.42a}$$

where

$$S(n) = r_2 S(n-1) + r_3^n, \quad S(1) = r_2 + r_3 \tag{12.42b}$$

The fact that the expansion of Ψ_m has no linear term renders the linear shape function unsuitable. For triangles with two vertices on the axis, f_{ij} should be replaced by

$$\bar{f}_{ij} = f_{ij} - \frac{\mu}{D}\frac{2m+1}{4(m+1)} b_i b_j R(2m+1) \tag{12.43}$$

(Lencová et al, 1989; Lencová and Lenc, 1996b).

We shall not devote a separate section to multipoles. The essential point to remember is that the appropriate power of r should be removed before calculating the harmonics.

The imperfections of construction or alignment that give rise to parasitic aberrations have also been studied in depth. Since more than one method has been employed, we postpone discussion of this to Section 31.3.

12.7 Related Work

We have cited only a small selection of the many papers on the finite-element method. The following are also of direct relevance: Aiming and Khursheed (1999), Barth et al. (1990), Edgcombe (1997, 1999), Elster et al. (2008), Hodkinson and Tahir (1995), Horák and Zlámal (2015), Jánský et al. (2008, 2009), Khursheed (1994, 1996, 1997a,b), Khursheed and Dinnis (1989), Khursheed and Pei (1996), Lenc and Lencová (1997), Lencová (1975, 1980, 1988a,b, 1994, 1995a,b, 1996, 1998, 2002a,b. 2003, 2004a,b), Lencová and Lenc (2004), Lencová et al. (1996), Mulvey and Nasr (1981), Munro (1993), Park et al. (2008), Radlička (2008), Tahir (1985), Tahir and Mulvey (1990) and Tahir et al. (1993).

CHAPTER 13

Field-Interpolation Techniques

The finite-difference and the finite-element methods yield the values of a potential at the nodes of a discrete grid. This is only the first step in a full field calculation, since the computation of electron trajectories requires a knowledge of the field strength at arbitrary points in the field. This implies that suitable techniques for interpolation and numerical differentiation will be required.

The application of predictor-corrector methods to the computation of trajectories requires that the field strength be a smooth function, especially on the grid lines separating two adjacent meshes. With respect to the FDM, this problem has been solved satisfactorily. Calculation of field strengths with the accuracy needed for ray tracing requires an elaborate interpolation procedure. An example of a particularly difficult case is described by Kang et al. (2007, 2009, 2010) in their work on differential algebra.

The boundary-element method has the advantage that the field strength can be computed as a continuous superposition of analytic functions once the surface-source distributions have been determined.

13.1 One-Dimensional Differentiation and Interpolation

Numerical differentiation and interpolation in one dimension are the basis for all the corresponding procedures in two and three dimensions. Moreover they are of immediate importance in electron optics, since a knowledge of the axial potential and of its derivatives is sufficient for the determination of most electron optical properties.

We consider here an arbitrary analytical function $Y(z)$, which may be an axial potential, an axial deflection field strength or any other function of interest. Let us assume now that only the discrete values

$$Y_i \coloneqq Y(z_i), \quad i = 0 \ldots N, \quad (z_i > z_{i-1}) \tag{13.1}$$

are known. They may, for instance, result from a field calculation program using the FDM or the FEM. We now wish to calculate $Y(z)$ and some of its derivatives for arbitrary values of z with $z_0 \leq z \leq z_N$.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00013-4
© 2018 Elsevier Ltd. All rights reserved.

This is a standard problem in numerical analysis and a wide variety of methods has been devised to solve it but not all of these are suitable. If problems are not to arise in ray-tracing programs (see Chapter 34: Numerical Calculation of Trajectories, Paraxial Properties and Aberrations), all the required derivatives of $Y(z)$ must be continuous at $z_0 \ldots z_N$. Most of the familiar interpolation techniques do not satisfy this requirement, and the best technique proves to be Hermite interpolation, which will now be outlined.

13.1.1 Hermite Interpolation

We assume for the moment that the derivatives Y' and Y'' at $z_0 \ldots z_N$ are known; their determination will be discussed further below. We can then apply cubic or quintic Hermite interpolation. In each interval $z_{i-1} \le z \le z_i$ $(i = 1 \ldots N)$, *cubic* interpolation is described by

$$h_i := z_i - z_{i-1}, \quad t := (2z - z_i - z_{i-1})/h_i, \quad |t| \le 1 \tag{13.2}$$

$$f_{1,2}(t) = \frac{1}{4}(2 \mp 3t \pm t^3), \quad g_{1,2}(t) = \frac{1}{8}(t^3 \mp t^2 - t \pm 1) \tag{13.3}$$

$$Y(z) = Y_{i-1} f_1(t) + Y_i f_2(t) + h_i Y'_{i-1} g_1(t) + Y'_i g_2(t) \tag{13.4}$$

while *quintic* interpolation is given by Eqs (13.2) and (13.3) in combination with

$$\begin{aligned} F_{1,2}(t) &= \frac{1}{2}(1 \mp t) \mp \frac{t}{16}(7 - 10t^2 + 3t^4) \\ G_{1,2}(t) &= -\frac{t}{32}(7 - 10t^2 + 3t^4) \pm \frac{1}{32}(5 - 6t^2 + t^4) \end{aligned} \tag{13.5}$$

$$\begin{aligned} H_{1,2}(t) &= \frac{1}{64}(1 - t^2)^2(1 \mp t) \\ Y(z) &= Y_{i-1} F_1(t) + Y_i F_2(t) + h_i \left\{ Y'_{i-1} G_1(t) + Y'_i G_2(t) \right\} \\ &\quad + h_i^2 \left\{ Y''_{i-1} H_1(t) + Y''_i H_2(t) \right\} \end{aligned} \tag{13.6}$$

The *form functions* $f_{1,2}(t)$, $g_{1,2}(t)$ or $F_{1,2}(t)$, $G_{1,2}(t)$, $H_{1,2}(t)$, respectively, are defined in such a way that $Y(z)$ and certain of its derivatives assume the prescribed values at z_i and z_{i-1}. Since each internal endpoint z_i is common to the intervals $[z_{i-1}, z_i]$ and $[z_i, z_{i+1}]$, $Y(z)$ and $Y'(z)$ are continuous in the cubic Hermite procedure, while the quintic procedure also ensures the continuity of $Y''(z)$.

13.1.2 Cubic Splines

The formulae given above require that the derivatives at $z_0 \ldots z_N$ have been calculated and stored prior to the actual interpolation. We now discuss the determination of these

derivatives. The *cubic spline* technique (13.3−13.4) is very convenient and is in widespread use. Cubic spline functions are Hermite interpolation functions (13.4) that remain continuous after *two* differentiations. This requirement imposes conditions on $Y'_0 \ldots Y'_N$, which can be cast into the form of a tridiagonal system of equations. With the abbreviation $k_i = h_i^{-1}$, this is given by

$$\begin{aligned} &k_i Y'_{i-1} + 2(k_i + k_{i+1})Y'_i + k_{i+1}Y'_{i+1} \\ &= 3k_i^2(Y_i - Y_{i-1}) + 3k_{i+1}^2(Y_{i+1} - Y_i) \\ &i = 1 \ldots N-1 \end{aligned} \tag{13.7}$$

The terminal values Y'_0 and Y'_N can be chosen independently, provided that they are not determined uniquely by such constraints as symmetries or periodicity. If there is apparently no reasonable way of determining Y'_0 and Y'_N, the linear equations

$$\begin{aligned} k_1 Y'_0 + (k_2 + k_1)Y'_1 &= 2D_1 + \frac{k_1}{k_1 + k_2}(D_1 + D_2) \\ (3k_{N-1} - k_N)Y'_{N-1} + k_N Y'_N &= 2D_N + \frac{k_N}{k_N + k_{N-1}}(D_N + D_{N-1}) \\ D_\nu &:= (Y_\nu - Y_{\nu-1})k_\nu^2, \quad \nu = 1, 2 \ \ldots \ N-1, N \end{aligned} \tag{13.8}$$

can be combined with Eq. (13.7); these equations are obtained if $Y'''(z)$ is assumed to be continuous at $z = z_1$ and $z = z_{N-1}$. The complete tridiagonal system of equations can be solved directly by means of the Gauss algorithm without pivoting.

In cubic splines, the second derivative is only a piecewise linear function and hence not very accurate; one should thus not use cubic splines if $Y''(z)$ is needed explicitly. In order to obtain high accuracy in such cases, an improved differentiation technique is necessary and quintic Hermite interpolation should then be used.

13.1.3 Differentiation Using Difference Schemes

Among the many ways of performing numerical differentiations, the technique outlined below has proved very effective. The explicit use of the unequal spacing of the abscissae $z_0 \ldots z_N$ makes the formulae cumbersome and should be avoided. This can be achieved in the following way.

We choose a parametric representation of the function in question, the parameter being denoted by x:

$$z = z(x), \quad Y = Y(x) \tag{13.9a}$$

$$x = ih, \quad z_i = z(ih), \quad Y_i = Y(ih), \quad i = 0, 1, \ \ldots N \tag{13.9b}$$

Without loss of generality we can choose $h = 1$, as we do below. Since the two functions $z(x)$ and $Y(x)$ are to be treated in the same manner, it is sufficient to deal only with the differentiation of $Y(x)$; the corresponding derivatives will be denoted by dots.

We now introduce finite differences

$$\Delta Y_i = Y_{i+1} - Y_i \tag{13.10a}$$

$$\delta^2 Y_i = Y_{i+1} - 2Y_i + Y_{i-1} = \Delta Y_i - \Delta Y_{i-1} \tag{13.10b}$$

$$\delta^{2n+2} Y_i = \delta^{2n} Y_{i+1} - 2\delta^{2n} Y_i + \delta^{2n} Y_{i-1}, \quad n \geq 1 \tag{13.10c}$$

Derivatives with respect to x are then given by

$$\begin{aligned} S_j &:= \frac{1}{2} Y_i - \frac{1}{12}\delta^2 Y_j + \frac{1}{60}\delta^4 Y_j - \frac{1}{280}\delta^6 Y_j + \frac{1}{1260}\delta^8 Y_j - \dots \\ \dot{Y}_i &= S_{i+1} - S_{i-1} \\ \ddot{Y}_i &= \delta^2 Y_i - \frac{1}{12}\delta^2 Y_i + \frac{1}{90}\delta^6 Y_i - \frac{1}{560}\delta^8 Y_i + \frac{1}{3150}\delta^{10} Y_i - \dots \end{aligned} \tag{13.11}$$

If the highest order is chosen reasonably, these formulae give accurate results, since they are highly symmetric. In the vicinity of the margins they are not directly applicable. In order to avoid special asymmetric formulae, it is preferable to extrapolate the function $Y(x)$ a certain distance beyond the interval in which the derivatives are actually needed. This can be done with the aid of symmetries, periodicities or well-known asymptotic properties. If none of these is applicable, a polynomial extrapolation can be made. For a polynomial of degree n, this extrapolation takes the simple form

$$Y_{j+1} = \sum_{k=0}^{n} \binom{n+1}{k+1} (-1)^k Y_{j-k} \tag{13.12}$$

Analogous formulae with correspondingly lower degree hold for the differences ΔY_{j+1} and $\boldsymbol{\delta^{2m} Y_{j+1}}$.

Sometimes the first-order increments $\Delta Y_1 \dots \Delta Y_k$ are given directly, for instance in the procedure outlined in Section 10.2.3. It is then possible to set up the differentiation procedure directly in terms of these increments. This provides additional numerical stability, as the subtraction of large Y-values is avoided. The corresponding elementary manipulations are not given here.

Finally the required derivatives with respect to the coordinate z are given by

$$Y'_i = \dot{Y}_i / \dot{z}_i, \quad Y''_i = \left(\ddot{Y}_i \dot{z}_i - \dot{Y}_i \ddot{z}_i\right) / \dot{z}_i^3, \quad (i = 0 \dots N) \tag{13.13}$$

Derivatives of higher orders can be computed easily by applying this procedure to these sets of derivatives instead of to the function itself. If the abscissae $z_0 \ldots z_N$ are equidistant, the differentiations of $z(x)$ can be omitted since we have simply

$$\dot{z}_i = z_{i+1} - z_i = h = \text{const}, \quad \ddot{z}_i = 0$$

The procedure is thus very economic.

For the interpolation of derivatives $Y^{(n)}$, the following procedure is efficient: Eqs (13.2), (13.5) and (13.6) are completed by

$$\begin{aligned} Y^{(n)}(z) = {} & Y_{i-1}^{(n)} F_1(t) + Y_i^{(n)} F_2(t) + h_i \left\{ Y_{i-1}^{(n+1)} G_1(t) + Y_i^{(n+1)} G_2(t) \right\} \\ & + h_i^2 \left\{ Y_{i-1}^{(n+2)} H_1(t) + Y_i^{(n+2)} H_2(t) \right\}, \quad n \geq 1 \end{aligned} \tag{13.14}$$

This has the advantage that only the form functions themselves need to be computed and not their derivatives. Furthermore, even the derivatives of higher orders remain twice continuously differentiable, and hence are very smooth. A very high accuracy can be achieved.

13.1.4 Evaluation of Radial Series Expansions

In Chapter 7, Series Expansions, we have derived a variety of radial series expansions, which are of particular interest in electron optics. These are all determined uniquely by certain axial functions, the axial harmonics. With the technique outlined above, their higher order derivatives can be computed numerically for a sequence of abscissae $z_0 \ldots z_N$ and then stored. Using the interpolation formula (13.14), it is now easy to evaluate the radial series expansions for the potential, the field strengths and even for derivatives of second order at any point (z, r) of reference within the domain of convergence. This is straightforward and is undoubtedly the fastest method of field computation.

The analytical character of the solution obtained with the BEM allows *analytical* differentiation of the axial potential, which is clearly preferable if the corresponding procedure remains reasonably simple. For the functions involved in the calculation of rotationally symmetric fields, this is certainly the case.

For a single charged ring, specified by its position (z', r') and the normalized charge 2π, the axial potential, here denoted by γ, can be easily calculated from Eq. (10.6):

$$\gamma(z - z', r') = G_0(z, 0;\ z', r') = (2R)^{-1} \tag{13.15a}$$

with

$$R = \sqrt{(z - z')^2 + r'^2} \tag{13.15b}$$

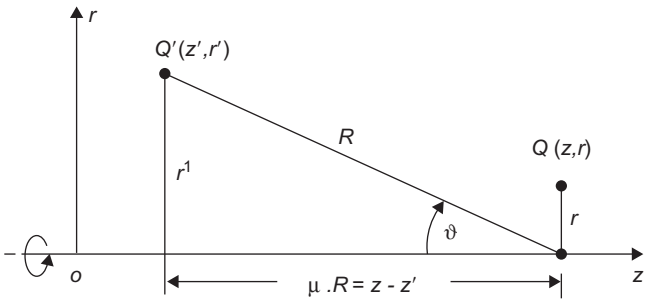


Figure 13.1
Notation used in the extended paraxial series expansion. Q is an arbitrary reference point and Q' is the trace of a ring in this meridional section.

These geometric quantities are sketched in Fig. 13.1. The derivatives of γ with respect to z can be brought into a very convenient form:

$$\gamma^{(n)}(z-z',r') = \frac{(-1)^n n! P_n(\mu)}{2R^{n+1}} \quad (n \geq 0) \tag{13.16a}$$

$P_n(\mu)$ being Legendre polynomials with argument

$$\mu = (z-z')/R = \cos\vartheta \tag{13.16b}$$

where the angle ϑ is shown in Fig. 13.1. These formulae can be evaluated efficiently.

The order in which the differentiation with respect to z and the integration over the boundary C are performed in the integral equation can be exchanged. Once the source distribution $\sigma(s)$ is known, the axial potential $\phi(z)$ and its derivatives can hence be calculated from

$$\phi^{(n)}(z_i) = \oint_C \gamma^{(n)}\big(z_i - z(s), r(s)\big)\sigma(s)r(s)ds$$
$$(i = 0, 1, 2 \ldots N, \quad n \geq 0) \tag{13.17}$$

After these values have been computed and stored, Eq. (13.14) can be employed for the calculation of $\phi^{(n)}(z)$ for arbitrary z, after which the evaluation of the radial series expansions is straightforward.

This concept can be generalized to include the superposition of aperture fields (Section 10.3) and of various multipole fields, but this will not be dealt with here.

13.2 Two-Dimensional Interpolation

Here we consider two-dimensional functions $P(u, v)$, known at the nodes of a rectangular grid. The coordinates u and v will usually be the cylindrical coordinates z and r in a

meridional section through an axisymmetric system, though this special meaning is not absolutely necessary. We now describe algorithms for calculating $P(u, \upsilon)$ and its partial derivatives at an arbitrary point Q with coordinates (u, υ).

This problem is of importance for the computation of equipotentials and Lorentz trajectories at large off-axis distances. The accurate tracing of a Lorentz trajectory through an electron optical system may require as many as 2000 calls of the field program, this number rapidly increasing with worsening smoothness of the field strength at the grid lines. When the analytic fields supplied by the boundary-element method are used, this problem does not arise, but each single call of the field program may then take so much time that it is preferable to store the values of the potential and the components of the field strength at the nodes of a suitably chosen square-shaped grid. The frequent evaluations at arbitrary points can subsequently be performed very rapidly by means of interpolation. This is particularly important when several Lorentz trajectories are to be computed, for instance in electron guns or to calculate spot patterns.

This interpolation problem has been solved in many different ways. In electron optics, different proposals have been made by Weber (1967), Lenz (1973), Kern (1978) and Kasper (1982). Two- and three-dimensional Hermite interpolation has been used, for instance by Eupper (1985). In the subsequent presentation we shall first examine simple two-dimensional Hermite interpolation, after which we consider possible improvements.

13.2.1 Hermite Interpolation

Our object is to calculate a function $P(u, \upsilon)$ at some point Q, located arbitrarily in the grid, as shown in Fig. 13.2. It is convenient to denote the partial derivatives by

$$U := P_{|u} \equiv \frac{\partial P}{\partial u}, \quad V := P_{|\upsilon} \equiv \frac{\partial P}{\partial \upsilon} \tag{13.18}$$

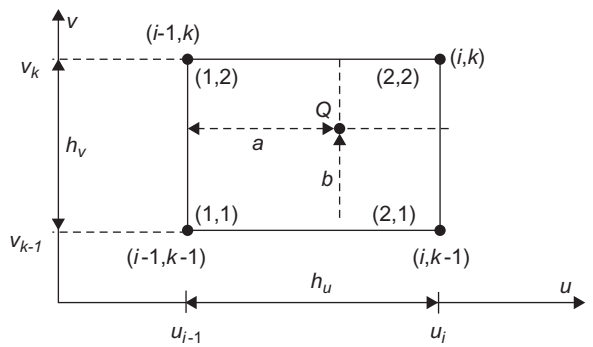


Figure 13.2
Values of the subscripts used in two-dimensional interpolation. Outside the rectangle, the subscripts J and L are those of the potential; inside the rectangle, the subscripts j and l are those of the coefficients. The lengths a and b are given by $a = h_u(1 + s)/2$ and $b = h_\upsilon(1 + t)/2$.

We assume that the nodal values of P, U and V have been computed prior to the interpolation stage and stored in two-dimensional arrays having two subscripts. The array elements $P_{i,k}$, $U_{i,k}$ and $V_{i,k}$ refer to the node with coordinates (u_i, u_k).

With this information, bivariate cubic Hermite interpolation, which is based on the form factors (13.3), can be applied. The interpolation formulae can be written explicitly in terms of these functions:

$$P(u,v) = \sum_{j=1}^{2}\sum_{l=1}^{2}\Big\{f_j(s)f_l(t)P_{J,L} + h_u g_j(s)f_l(t)U_{J,L} + h_v f_j(s)g_l(t)V_{J,L}\Big\} \tag{13.19a}$$

with the auxiliary quantities

$$\begin{aligned} h_u &= u_i - u_{i-1} \qquad & s &= (2u - u_i - u_{i-1})/h_u \\ h_v &= v_k - v_{k-1} & t &= (2v - v_k - v_{k-1})/h_v \\ J &= i + j - 2 & L &= k + l - 2 \end{aligned} \tag{13.19b}$$

The values of U and V at the point Q are obtained by the appropriate differentiations; the corresponding elementary expressions will not be given here. The derivatives U and V are still continuously differentiable on the grid lines if the arrays $[U_{i,k}]$, $[V_{i,k}]$ are calculated by applying the cubic spline technique to the potentials in the corresponding rows and columns of the grid. This method can easily be generalized to three-dimensional problems.

13.2.2 The Use of Derivatives of Higher Order

As in the one-dimensional case, the accuracy and smoothness can be improved by the use of derivatives of higher orders at the nodes of the grid. Such a proposal has been made by Kasper (1982) but this requires a particular partial differential equation to be satisfied, which is not always the case. Here we treat the most general case.

The necessary procedure is reasonably simple only if the grid is square-shaped: $u_i = ih$, $v_k = kh$ with i and k integers. We assume again that the arrays $[P_{i,k}]$, $[U_{i,k}]$ and $[V_{i,k}]$ are known beforehand. Owing to the higher accuracy required, the cubic spline technique is inadequate and a more accurate differentiation technique must be employed.

It is now of great importance that the derivatives of higher orders can be calculated in the form of *local* finite differences, so that there is no need to store them all. For conciseness, we temporarily introduce the notation presented in Fig. 13.3. The appropriate Taylor series expansions about the central node 0 yield the formulae

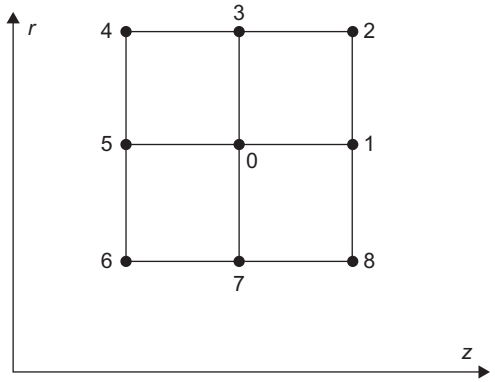


Figure 13.3
Simplified provisional numbering of the nodes in nine-point formulae for higher derivatives at the central node.

$$\begin{aligned} P_{|uu} &= 2(P_1 - 2P_0 + P_5)/h^2 - 0.5(U_1 - U_5)/h \\ P_{|vv} &= 2(P_3 - 2P_0 + P_7)/h^2 - 0.5(V_3 - V_7)/h \\ P_{|uv} &= -0.25(P_2 - P_4 + P_6 - P_8)/h^2 \\ &\quad + 0.5(V_1 - V_5 + U_3 - U_7)/h \end{aligned} \tag{13.20}$$

These derivatives refer to the central node, and their remainder is of fourth order in the mesh-length.

This method can be extended to the determination of derivatives of still higher orders. Even some of the derivatives of fifth order can be calculated in this way, but the finite differences involved then become quite numerous. For reasons of space we must confine this account to the simplest nontrivial case, which is sufficient in very many practical applications. The mixed derivatives of third order are given by fairly simple expressions:

$$\begin{aligned} P_{|uuv} &= 0.25(U_2 - U_4 + U_6 - U_8)/h^2 + O(h^2) \\ P_{|uvv} &= 0.25(V_2 - V_4 + V_6 - V_8)/h^2 + O(h^2) \end{aligned} \tag{13.21}$$

On an *axis of symmetry*, typically the optic axis ($v = 0$), some of the neighbours of the node 0 are missing. We can either extend the arrays beyond this axis and fill them up according to the symmetry or make explicit use of the symmetry. In the case of positive symmetry, $P(u,-v) = P(u, v)$, the finite differences for an axial point ($v_0 = 0$) take the simple form

$$\begin{aligned} P_{|vv} &= \frac{4(P_3 - P_0)}{h^2} - \frac{V_3}{h} \\ P_{|uv} &= P_{|uuv} = 0 \\ P_{|uvv} &= 0.5\,\frac{V_2 - V_4}{h^2} \end{aligned} \tag{13.22}$$

while $P_{|uu}$ remains unaltered.

All these finite differences are simple enough to be recalculated in every new call of the field-calculation program. We have to identify the central node 0 of Fig. 13.3 with each of the four corners in Fig. 13.2 in turn and then calculate the corresponding derivatives. This results in a 16-point configuration and the evaluation of 20 simple finite differences.

The necessary interpolations are written most concisely in the form

$$X(u,v) = \sum_{j=1}^{2}\sum_{l=1}^{2}\Big\{f_j(s)g_l(t)X_{J,L} + hg_j(s)f_l(t)X_{J,L|u} + hf_j(s)g_l(t)X_{J,L|v} + h^2 g_j(s)g_l(t)X_{J,L|uv}\Big\} \qquad (13.23)$$
$$\big(J = i + j - 2, \quad L = k + l - 2\big)$$

where the symbol X denotes P, U or V, and the subscripts J and L refer to the four corners of the mesh cell in question. Only the arrays for the potential and the first-order derivatives are stored; the rest are recalculated, but the time spent on the latter is compensated for by the saving in the computation of differentiated form functions. Eq. (13.23) implies that the same procedure is to be carried out three times, but with different coefficients. The design of this interpolation scheme is such that the field strengths — that is, the derivatives U and V — are continuously differentiable. Even the second-order derivatives, needed in a procedure to be outlined in Chapter 34, Numerical Calculation of Trajectories, Paraxial Properties and Aberrations, are fairly smooth.

In practice, many operations can be saved if the calculations are performed with a mesh-length $h = 1$, to which all stored derivatives and calculated finite differences must refer. Each computed result is finally multiplied only once by the appropriate power of the actual mesh-length. We have not presented this version here for pedagogic reasons but we recommend it for any real program. Moreover, some computation time can be saved by calculating the finite differences referring to the four corners of the *same* mesh cell only once even though these quantities are needed several times. This situation can arise if several subsequent points Q of reference in a very accurate ray-tracing program are located in the same cell. Such points can easily be identified by comparing the subscripts (i, k) with those of the previous call and by skipping the corresponding parts of the procedure when they are the same.

13.3 Interpolation and the Finite-Element Method

In the finite-element method, the results are obtained at the nodes of a mesh that is not rectangular. A method of interpolation for this situation was devised by Chmelík and Barth (1993). This is based on a set of polynomials in two variables that satisfy Laplace's equation. These polynomials are peculiar to each FEM quadrilateral; their coefficients are

established by fitting to the values at the corners of the quadrilateral and to eight neighbouring points (or four, near the axis or a boundary). Although the values and their derivatives are not continuous from one quadrilateral to the next, the resulting discontinuity still allows accurate ray tracing as the mesh can be made very dense.

For simplicity, we consider the electrostatic potential $\Phi(r, z)$ of a rotationally symmetric lens. We introduce a local coordinate system, in which distances are measured from a point (r_A, z_A) inside the quadrilateral in question:

$$\zeta = z - z_A \quad \rho = r - r_A \tag{13.24}$$

The potential is then interpolated as $P(\rho, \zeta)$,

$$P(\rho,\ \zeta) = \sum_{i=1}^{M} C_i g_i(\rho,\ \zeta) \tag{13.25}$$

The basis functions $g_i(\rho, \zeta)$ are chosen to satisfy Laplace's equation and must of course be linearly independent. A suitable set is

$$\begin{aligned}
g_1 &= 1\\
g_2 &= \zeta\\
g_3 &= -2r_A\rho + (2\zeta^2 - \rho^2)\\
g_4 &= -6r_A\zeta\rho + \zeta(2\zeta^2 - 3\rho^2)\\
g_5 &= -12r_A^2(\zeta^2 - \rho^2) + 12r_A\rho(\rho^2 - 4\zeta^2) + (3\rho^4 - 24\rho^2\zeta^2 + 8\zeta^4)\\
g_6 &= -20r_A^2\zeta(\zeta^2 - 3\rho^2) + 20r_A\rho\zeta(3\rho^2 - 4\zeta^2) + \zeta(15\rho^4 - 40\rho^2\zeta^2 + 8\zeta^4)\\
g_7 &= -40r_A^3\rho(\rho^2 - 3\zeta^2) - 20r_A^2(3\rho^4 - 21\rho^2\zeta^2 + 4\zeta^4)\\
&\quad - 30r_A\rho(\rho^4 - 12\rho^2\zeta^2 + 8\zeta^4) - (5\rho^6 - 90\rho^4\zeta^2 + 120\rho^2\zeta^4 - 16\zeta^6)
\end{aligned} \tag{13.26}$$

These must be rendered dimensionless and scaled so as to be of the order of unity at the boundaries of the quadrilateral. The values of Φ at the four corners of the quadrilateral are not sufficient to determine the multipliers C_i for $M = 7$. By adding the values at the next nearest points (Fig. 13.4), the system becomes over-determined and the multipliers can then be obtained by least-square fitting. Chmelík and Barth include a weight factor to render the transition from one quadrilateral to the next as smooth as possible; they find that the weight should be inversely proportional to the square of the distance between the point in question and the point $(r_A,\ z_A)$.

This interpolation method is used by Oral and Lencová (2009, 2013) and Oral (2010) to calculate aberration coefficients by ray tracing. By using a much denser mesh than was

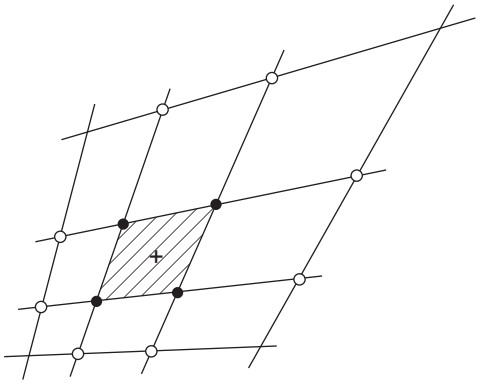


Figure 13.4
Points used to determine the multipliers in the interpolation formula
+ point of reference; • corners of the quadrilateral; ◦ next nearest neighbours.

possible for Chmelík and Barth, the accuracy was extremely high. The method is used in the EOD program (Lencová and Zlámal, 2008; Zlámal and Lencová, 2010) as well as bicubic interpolation.

In conclusion, the field interpolation can be made sufficiently accurate and fast for the purposes of ray tracing. A still more accurate but also more sophisticated method of interpolation has been worked out by Killes (1985), to which we refer for the details.

PART III

The Paraxial Approximation

CHAPTER 14

Introduction to Paraxial Equations

The general form of the trajectory equations in electromagnetic fields has been derived in Part I (3.22) but in many practical situations these equations are unnecessarily complicated. In a very large class of electron optical instruments, the electrons remain in the vicinity of a curve, frequently a straight line, which we call the optic axis. The behaviour of the various optical elements can then be characterized by simpler equations, obtained by expanding the fields and potentials about this axis and retaining only the terms of lowest order. We shall see that these equations are often second-order, linear, homogeneous differential equations and their solutions describe the linear imaging properties of lenses of various kinds. With a little care, mirrors can also be included and some aspects of electron guns and cathode lenses can even be characterized in this way.

We shall give two derivations of the paraxial equations for systems with an axis of rotational symmetry, since these are of such importance. First, we simply insert the series expansions for the components of the magnetic flux $\boldsymbol{B}$ and the electrostatic potential Φ into the general equations (3.22) and neglect all but the terms of lowest order. In the alternative derivation, we expand the characteristic function $\overline{M}$ (4.25) as a power series in the off-axis coordinates and their derivatives; the Euler equations (4.26) of the variational relation (4.33) then yield the paraxial trajectory equation if we retain only quadratic terms in the expansion. For systems of lower symmetry, we employ only one of these methods, usually the latter. The function $\overline{M}$ will almost invariably be scaled with respect to $(\mathbf{2}m_0e)^{1/2}$, as in Eq. (15.23). The momentum then scales to $\hat{\phi}^{1/2}$ and we shall often refer to this quantity as the momentum, with components $\hat{\phi}^{1/2}x'$ and $\hat{\phi}^{1/2}y'$ though it of course does not have the proper dimensions.

The presence of a magnetic field leads us to introduce a new coordinate system, twisted about the z-axis with respect to the Cartesian system in terms of which the field expansions are given in Part II. In this Part therefore, we denote the 'fixed' Cartesian system by (X, Y, z), reserving the lower-case (x, y, z) for the twisted or 'rotating' coordinate system, in terms of which all later calculations will be performed. Only static fields will be considered here. Dynamic fields are more conveniently treated separately.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00014-6
© 2018 Elsevier Ltd. All rights reserved.

CHAPTER 15

Systems with an Axis of Rotational Symmetry

Round lenses are by far the most common in electron optical instruments and we now examine their paraxial properties in detail. A typical electrostatic lens consists of two, three or more electrodes, in the form of plates in which round holes have been cut, their centres lying on a common axis (Fig. 15.1A), or of cylinders, again with the same axis (Fig. 15.1B). Although in theory the field extends indefinitely, in practice it rapidly becomes negligibly small and we speak of the field region and the field-free space outside it. Electrostatic lenses may have an overall accelerating or retarding effect, in which case the constant potential in front of the lens is not the same as that behind it (Fig. 15.1C); they are then often known as *immersion lenses* even though a real object is rarely immersed in the electrode field. An exception is the cathode lens, which is terminated by an unperforated electrode, the properties of which are to be studied, or which acts as a source. If the lens has no overall accelerating effect, in practice it very often has three electrodes (Fig. 15.1D) and is then known as an *einzel lens* or *unipotential lens*. The three-electrode design illustrated in Fig. 15.1D is typical of einzel lenses. A special case of the electrostatic lens is the electron gun (Fig. 15.1E), in which electrons are generated by a filament or cathode, in the form of a point or hairpin, and rapidly accelerated to the operating voltage of the instrument in which they are employed. Guns need special treatment, however, and are discussed in detail in Part IX.

Round magnetic lenses are devices that generate a rotationally symmetric magnetic field, effectively confined to a narrow region. The traditional design, which has changed little since its introduction by Ruska in the early 1930s (Knoll and Ruska, 1932a,b; Ruska, 1934a,b), consists of a large number of windings enclosed in an iron casing; a slot in the latter, finished with circular polepieces, concentrates the field as shown schematically in Fig. 15.2A. In some designs, the windings are in the superconducting state and carry persistent currents. In others, the entire lens is in the superconducting state and the field is confined by a diamagnetic shield (Fig. 15.2B). More radical departures from this geometry are employed for special purposes; two extreme shapes are illustrated in Figs 15.2C and D.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00015-8
© 2018 Elsevier Ltd. All rights reserved.

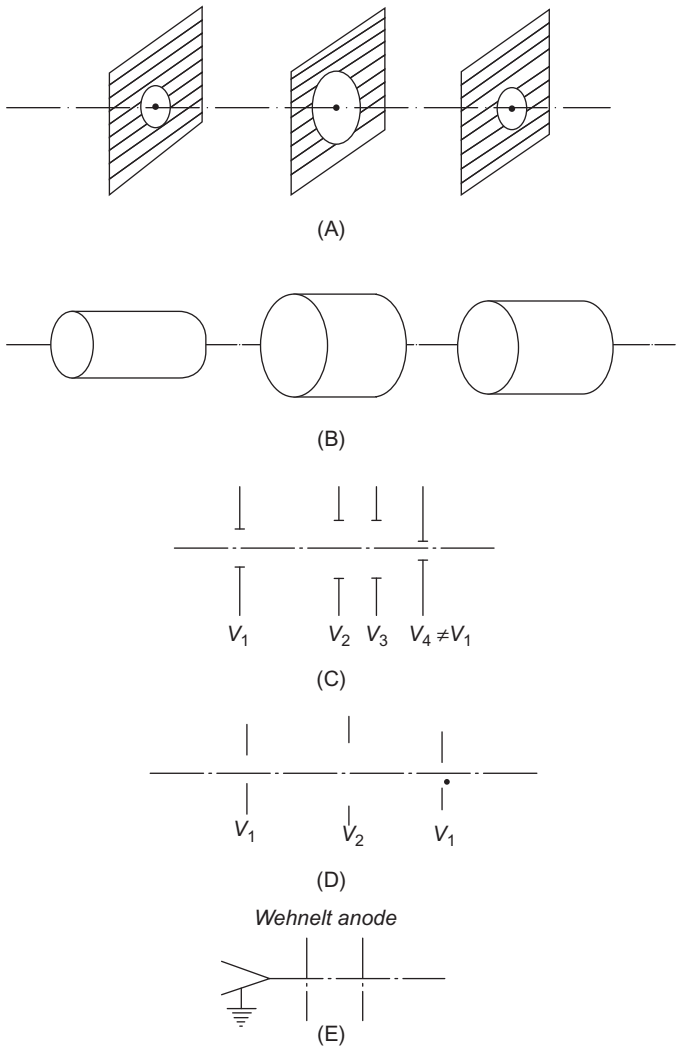


Figure 15.1
Forms of electrostatic lenses. (A) Plates with circular openings. (B) Cylinders along a common axis. (C) Lens with an overall accelerating ($V_4 < V_1$) or retarding ($V_4 > V_1$) effect. (D) Einzel or unipotential lens. (E) Gun nomenclature.

All these types of magnetic lens rely on current-carrying conductors to provide the magnetomotive force. Permanent magnets may be used instead, and have been incorporated into commercial electron microscopes, but their inflexibility is a severe handicap. Fig. 15.2E and F show permanent-magnet lenses and the axial fields within them. There has been a revival of interest in permanent-magnet lenses with the development of miniature

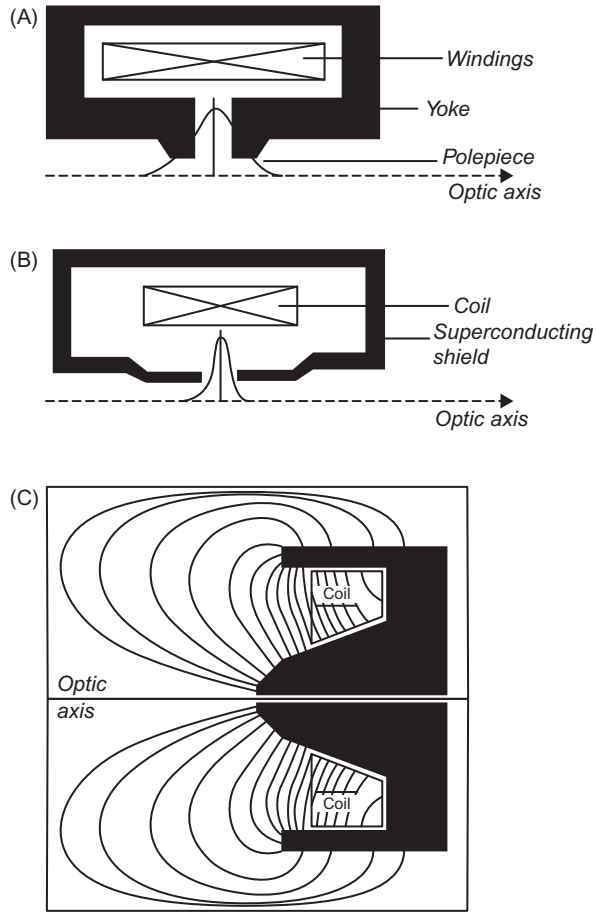


Figure 15.2
Forms of magnetic lenses. (A) Conventional lens. (B) Superconducting shielding lens. (C) Single-pole lens (with bore). (D) Laminated lens: c, polepiece; e, stigmator. (E)—(F) Permanent-magnet lenses.

scanning electron microscopes. Most of these employ electrostatic lenses but there are also some magnetic designs (Section 36.6.3 of Volume 2).

Real objects are regularly immersed deeply within the field of magnetic lenses. When discussing the properties of electron lenses, we need distinguish only two types, those in which a real object or image is situated within the field and those in which the lens transfers an intermediate image from one plane to another. Nevertheless it is usual to describe lenses in terms of the role they play. Thus we speak of condenser lenses if the 'intermediate image' being transferred is the image of the source of a microscope but of intermediates or projectors if a genuine image is in question. Geometrically, these might be

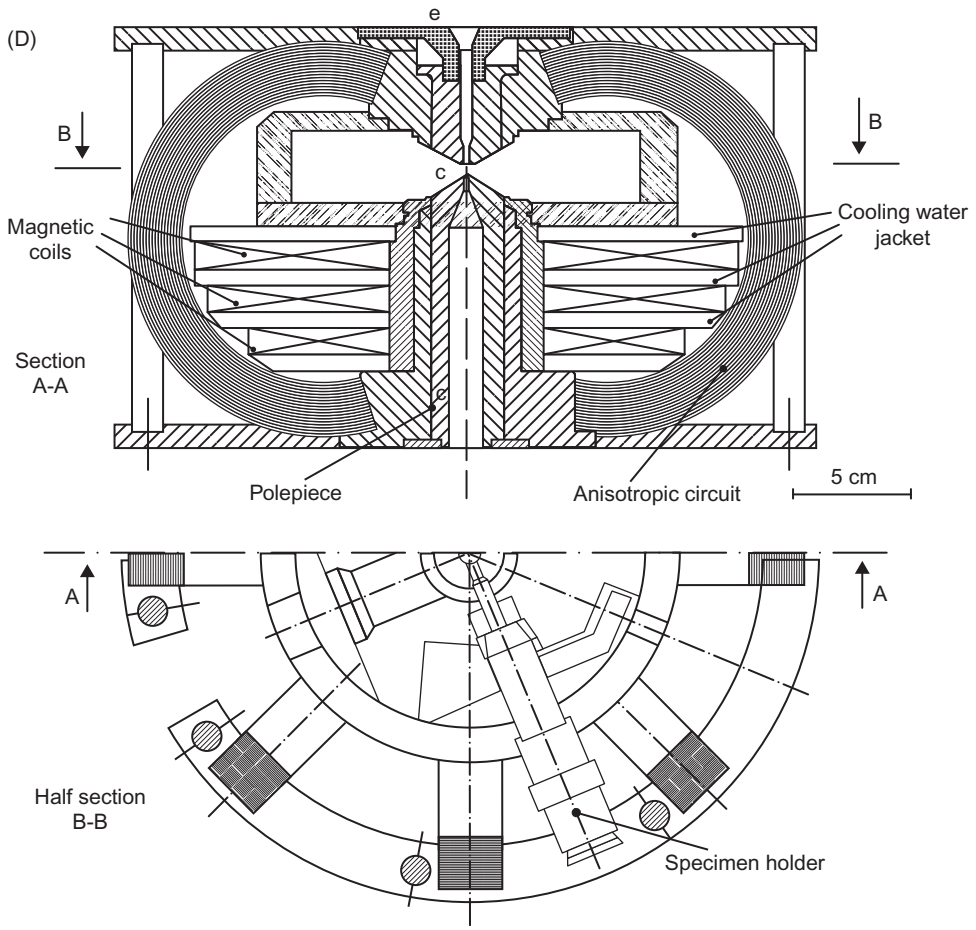


Figure 15.2
(Continued.)

very similar. Likewise, the same lens may be an objective, if the specimen is immersed in it, or a probe-forming lens, if it forms a fine probe within (or indeed outside) the field. These distinctions will become more clear in Part VII.

15.1 Derivation of the Paraxial Ray Equations from the General Ray Equations

We now derive the paraxial trajectory equations for electrostatic and magnetic lenses. Into the general trajectory equations (3.22), we substitute the expansions for $\Phi(X, Y, z)$ and

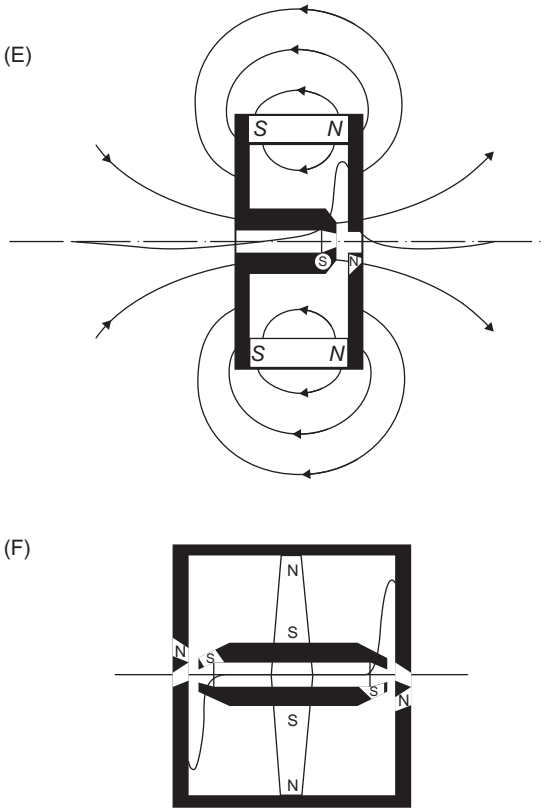


Figure 15.2
(Continued.)

$\boldsymbol{B}(X, Y, z)$ given in Part II. We now set Φ_0 equal to zero, thereby disregarding chromatic effects. From Eq. (7.14), we find

$$\begin{aligned}\frac{\partial\hat{\Phi}}{\partial X} &\approx -\frac{1}{2}(1+2\varepsilon\phi)X\phi'' = -\frac{1}{2}\gamma X\phi'' \\ \frac{\partial\hat{\Phi}}{\partial z} &\approx (1+2\varepsilon\phi)\phi' = \gamma\phi'\end{aligned} \tag{15.1}$$

where as usual $\gamma = m/m_0 = (1 - v^2/c^2)^{-1/2} = 1 + 2\varepsilon\Phi$ (2.2, 2.21). Neglecting quadratic and higher order terms in X, Y and their derivatives, the electrostatic terms on the right-hand side of Eq. (3.22a) become

$$\frac{1+X'^2+Y'^2}{2\hat{\Phi}}\left(\frac{\partial\hat{\Phi}}{\partial X} - X'\frac{\partial\hat{\Phi}}{\partial z}\right) \approx -\frac{\gamma}{4\hat{\phi}}(2X'\phi' + X\phi'') \tag{15.2}$$

with a similar expression for (3.22b).

For the magnetic term, we substitute the appropriate expansions (7.18 and 7.19); noting that B_t can be replaced by B_z since we are neglecting quadratic terms, we find

$$\frac{\eta\rho^2}{\sqrt{\hat{\Phi}}}(\rho B_Y - Y' B_t) \approx -\frac{\eta}{2\sqrt{\hat{\phi}}}(YB' + 2Y'B) \tag{15.3a}$$

and

$$\frac{\eta\rho^2}{\sqrt{\hat{\Phi}}}(-\rho B_X + X' B_t) \approx \frac{\eta}{2\sqrt{\hat{\phi}}}(XB' + 2X'B) \tag{15.3b}$$

The pair of trajectory equations (3.22) thus collapse to the following in the paraxial approximation:

$$\begin{aligned} X'' + \frac{\gamma\phi'}{2\hat{\phi}}X' + \frac{\gamma\phi''}{4\hat{\phi}}X + \frac{\eta B}{\sqrt{\hat{\phi}}}Y' + \frac{\eta B'}{2\sqrt{\hat{\phi}}}Y = 0 \\ Y'' + \frac{\gamma\phi'}{2\hat{\phi}}Y' + \frac{\gamma\phi''}{4\hat{\phi}}Y - \frac{\eta B}{\sqrt{\hat{\phi}}}X' - \frac{\eta B'}{2\sqrt{\hat{\phi}}}X = 0 \end{aligned} \tag{15.4}$$

This pair of coupled linear differential equations can be cast into a simpler form by replacing the coordinate system (X, Y, z) by a new system, rotated with respect to the former by a *variable* angle $\theta(z)$. In order to see this, we introduce the complex coordinate (7.3)

$$w = X + \mathrm{i}Y \tag{15.5}$$

so that Eq. (15.4) become

$$w'' + \frac{\gamma\phi'}{2\hat{\phi}}w' + \frac{\gamma\phi''}{4\hat{\phi}}w - \mathrm{i}\frac{\eta B}{\sqrt{\hat{\phi}}}w' - \mathrm{i}\frac{\eta B'}{2\sqrt{\hat{\phi}}}w = 0 \tag{15.6}$$

The final two terms containing i explicitly can be removed by introducing a new complex coordinate, u, such that

$$w =: u \exp \mathrm{i}\theta(z) \tag{15.7}$$

Eq. (15.6) becomes

$$u'' + u'\left(2\mathrm{i}\theta' + \frac{\gamma\phi'}{2\hat{\phi}} - \frac{\mathrm{i}\eta B}{\sqrt{\hat{\phi}}}\right) + u\left\{\mathrm{i}\theta'' - \theta'^2 + \mathrm{i}\theta'\left(\frac{\gamma\phi'}{2\hat{\phi}} - \frac{\mathrm{i}\eta B}{2\sqrt{\hat{\phi}}}\right) + \frac{\gamma\phi''}{4\hat{\phi}} - \frac{\mathrm{i}\eta B'}{2\sqrt{\hat{\phi}}}\right\} = 0 \tag{15.8}$$

and the terms explicitly involving i vanish if we choose

$$\theta' = \frac{\eta B}{2\hat{\phi}^{1/2}} \tag{15.9}$$

so that $\theta'' = \eta B'/2\hat{\phi}^{1/2} - \eta B\gamma\phi'/4\hat{\phi}^{3/2}$ and we obtain

$$u'' + \frac{\gamma\phi'}{2\hat{\phi}}u' + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}}u = 0 \tag{15.10}$$

We note that Eq. (15.9) is essentially the same as Eq. (2.39). Explicitly, writing

$$u = x + \mathrm{i}y \tag{15.11}$$

we have

$$\begin{aligned} x'' + \frac{\gamma\phi'}{2\hat{\phi}}x' + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}}x = 0 \\ y'' + \frac{\gamma\phi'}{2\hat{\phi}}y' + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}}y = 0 \end{aligned} \tag{15.12}$$

or again

$$\begin{aligned} \frac{d}{dz}(\hat{\phi}^{1/2}x') + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}}x = 0 \\ \frac{d}{dz}(\hat{\phi}^{1/2}y') + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}}y = 0 \end{aligned} \tag{15.13}$$

15.1.1 Physical Significance of the Coordinate Rotation

This transformation to the rotating coordinate system (x, y, z) is of great importance. We therefore consider it in more detail before proceeding. The complex transformation (15.7) may be written

$$\begin{aligned} X &= x\cos\theta - y\sin\theta \\ Y &= x\sin\theta + y\cos\theta \end{aligned} \tag{15.14}$$

so that in any plane $z = \text{const}$, the axes $X{-}Y$ are inclined at an angle $\theta(z)$ to $x{-}y$ (Fig. 15.3). This angle increases monotonically provided that the sign of $B(z)$ does not change and the $x{-}y$-axes therefore twist round the z-axis like the blades of a propeller or the ridge of a screw of variable pitch. Fig. 15.3A gives a perspective view of this and Fig. 15.3B shows a

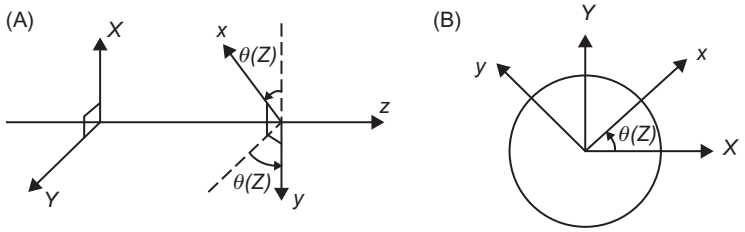


Figure 15.3
Fixed coordinates (X, Y, Z) and rotating coordinates (x, y, z). (A) Perspective view. (B) View along the optic axis.

view along the z-axis. We shall see in Part VII that $\int_{-\infty}^{\infty} B(z)\,dz = 0$ in any permanent-magnet lens and the total rotation in such a lens is hence zero.

Unlike a conventional cartesian coordinate system, the coordinate *surfaces* are not planes: the surfaces $x = 0$ and $y = 0$ are curved, though everywhere normal to each other, intersecting along the z-axis. The element of length ds in (x, y, z) is not equal to $(dx^2 + dy^2 + dz^2)^{1/2}$ but is given by

$$\begin{aligned} ds^2 &= dX^2 + dY^2 + dz^2 \\ &= dx^2 + dy^2 + dz^2\{1 + (x^2 + y^2)\theta'^2\} + 2(x\,dy - y\,dx)\theta' dz \end{aligned} \tag{15.15}$$

This rotation about the axis is closely related to the phenomenon of Larmor precession; if we express the rate of change of θ as a function of time rather than axial distance, using $d\theta/dt = \theta' dz/dt$ and $dz/dt = 2\eta\hat{\phi}/\gamma$, we find, in agreement with Eq. (2.38)

$$\frac{d\theta}{dt} = \frac{e}{2m_0}\frac{B}{\gamma} = \frac{\eta^2 B}{\gamma} \tag{15.16a}$$

or using Eq. (3.9)

$$\frac{d\theta}{d\tau} = \eta^2 B \tag{15.16b}$$

which is indeed the Larmor precession frequency. Plies (1994) notes that this is half the cyclotron frequency and explains that these are different because the centre of rotation for the Larmor rotation is the optic axis, while for the cyclotron rotation, it is the centre of curvature. Rose (2009, 2012, Section 4.1) also comments on this factor of two.

The fact that the paraxial equations separate in the rotating coordinates implies that an electron initially travelling on one of the coordinate surfaces, or on any surface $\alpha x + \beta y = 0$, remains on this surface; this leads us to ask what becomes of the angular momentum, and in particular of its axial component N (4.14). We have

$$N = (\boldsymbol{r} \times \boldsymbol{p})_z \tag{15.17}$$

in which we recall that $\boldsymbol{p}$ is the canonical momentum (4.12), $\boldsymbol{p} = \boldsymbol{g} - e\boldsymbol{A}$ and $\boldsymbol{g}$ is the kinetic momentum (2.12, 2.19). In the paraxial approximation,

$$\boldsymbol{g} \coloneqq (2m_0 e\hat{\Phi})^{1/2}\boldsymbol{t} \approx (2m_0 e\hat{\phi})^{1/2}(\boldsymbol{i}_z + r'\boldsymbol{i}_r + r\varphi'\boldsymbol{i}_\varphi)$$

for $\Phi_0 = 0$ and

$$-e\boldsymbol{A} = -eA\boldsymbol{i}_\varphi \approx -\frac{1}{2}erB\boldsymbol{i}_\varphi$$

so that

$$\boldsymbol{p} = \boldsymbol{g} - e\boldsymbol{A} = (2m_0 e\hat{\phi})^{1/2}(\boldsymbol{i}_z + r'\boldsymbol{i}_r) + \left\{(2m_0 e\hat{\phi})^{1/2}\varphi' - \frac{1}{2}eB\right\}r\boldsymbol{i}_\varphi \tag{15.18}$$

Hence

$$N = \left\{(2m_0 e\hat{\phi})^{1/2}\varphi' - \frac{1}{2}eB\right\}r^2 \tag{15.19}$$

or writing $XY' - X'Y = r^2\varphi'$

$$N = (2m_0 e\hat{\phi})^{1/2}(XY' - X'Y) - \frac{1}{2}eBr^2 \tag{15.20}$$

From Eq. (15.14) we have

$$XY' - X'Y = xy' - x'y + (x^2 + y^2)\theta' = xy' - x'y + \frac{\eta Br^2}{2\hat{\phi}^{1/2}}$$

and hence

$$N = \sqrt{2m_0 e\hat{\phi}}\,(xy' - x'y) \tag{15.21}$$

so that in the rotating coordinate system, the magnetic field does not appear explicitly in N. If N vanishes, the azimuthal angle φ remains at a constant angular distance from θ, $\varphi = \theta + \text{const}$ and the corresponding trajectories are said to be *meridional*. They lie in the curved surfaces defined by $\alpha x + \beta y = 0$, which intersect along the z-axis and are inclined at some fixed angle to the coordinate surfaces $x = 0$ and $y = 0$. Clearly any trajectory that intersects the axis at some point is a meridional trajectory, and vice versa. Rays that are not meridional are said to be *skew*.

Note: we have used Cartesian coordinates (X, Y, z) and rotating pseudo-Cartesian coordinates (x, y, z) above in preference to polar coordinates, since the calculation becomes complicated when skew rays are considered in the latter system. Extensive discussion of the correct way of handling skew rays when polars are used is to be found in most of the

earlier texts on electron optics (e.g., de Broglie, 1950; Rusterholz, 1950; Picht, 1963); the problem vanishes when Cartesians are employed (Glaser, 1952, Section 42).

15.2 Variational Derivation of the Paraxial Equations

We now take as our starting point (4.34−4.36), expanding the function $\overline{M}$ that plays the role of refractive index as a power series in x and y and retaining only quadratic terms. From Eqs (4.35) and (2.13), we have

$$\overline{M}(X, Y, X', Y', z) = \left\{2m_0 e\Phi(1+\varepsilon\Phi)(1+X'^2+Y'^2)\right\}^{1/2} - e(X'A_X + Y'A_Y + A_z) \tag{15.22}$$

Substituting for A_X, A_Y and A_z from Eqs (7.43−7.45) and for Φ from (7.36) into

$$\begin{aligned} M &:= \frac{\overline{M}}{(2m_0 e)^{1/2}} \\ &= \left\{\hat{\Phi}(1+X'^2+Y'^2)\right\}^{1/2} - \eta(X'A_X + Y'A_Y + A_z) \end{aligned} \tag{15.23}$$

we obtain a power series in X, Y and their derivatives, the quadratic terms of which, $M^{(2)}$, are given by

$$M^{(2)} = -\frac{\gamma\phi''}{8\hat{\phi}^{1/2}}(X^2+Y^2) + \frac{1}{2}\hat{\phi}^{1/2}(X'^2+Y'^2) - \frac{1}{2}\eta B(XY' - X'Y) \tag{15.24}$$

It is already clear that the Euler equations of $\delta \int M^{(2)}\, dz = 0$ will be coupled and we therefore attempt to transform the coordinates in such a way that the mixed term in $XY'-X'Y$, the source of the coupling, is eliminated. From Eq. (15.14) we obtain

$$\begin{aligned} X^2 + Y^2 &= x^2 + y^2 \\ X'^2 + Y'^2 &= x'^2 + y'^2 + 2\theta'(xy' - x'y) + \theta'^2(x^2+y^2) \\ XY' - X'Y &= xy' - x'y + \theta'(x^2+y^2) \end{aligned} \tag{15.25}$$

Substituting into $M^{(2)}$ we find

$$\begin{aligned} M^{(2)} = (x^2+y^2)&\left(-\frac{\gamma\phi''}{8\hat{\phi}^{1/2}} + \frac{\hat{\phi}^{1/2}}{2}\theta'^2 - 1/2\eta B\theta'\right) \\ &+ (x'^2+y'^2)\frac{\hat{\phi}^{1/2}}{2} + (xy'-x'y)\left(\theta'\hat{\phi}^{1/2} - 1/2\eta B\right) \end{aligned} \tag{15.26}$$

and the term in $xy'-x'y$ vanishes if we select

$$\theta' = \frac{\eta B}{2\hat{\phi}^{1/2}} \tag{15.27}$$

(as in Eq. (15.9)). The function $M^{(2)}$ becomes

$$M^{(2)} = -\frac{1}{8\hat{\phi}^{1/2}}(\gamma\phi'' + \eta^2 B^2)(x^2 + y^2) + \frac{1}{2}\hat{\phi}^{1/2}(x'^2 + y'^2) \tag{15.28}$$

Hence

$$\frac{\partial M^{(2)}}{\partial x'} = \hat{\phi}^{1/2}x', \quad \frac{\partial M^{(2)}}{\partial y'} = \hat{\phi}^{1/2}y' \tag{15.29}$$

$$\begin{aligned} \frac{\partial M^{(2)}}{\partial x} &= -\frac{1}{4\hat{\phi}^{1/2}}(\gamma\phi'' + \eta^2 B^2)x \\ \frac{\partial M^{(2)}}{\partial y} &= -\frac{1}{4\hat{\phi}^{1/2}}(\gamma\phi'' + \eta^2 B^2)y \end{aligned} \tag{15.30}$$

and the paraxial equations are thus

$$\frac{d}{dz}(\hat{\phi}^{1/2}x') + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}}x = 0 \tag{15.31}$$

with an identical equation for $y(z)$, as already found (15.13).

15.3 Forms of the Paraxial Equations and General Properties of their Solutions

15.3.1 Reduced Coordinates

In the absence of an electrostatic field, the paraxial equations take the form

$$u'' + F(z)u = 0 \tag{15.32}$$

($u = x + \mathrm{i}y$) with

$$F(z) := \frac{\eta^2 B^2}{4\hat{\phi}} \tag{15.33}$$

When $\phi(z)$ is not constant, they can again be reduced to this form by a simple transformation of the off-axis coordinates. We write

$$u(z) =: \upsilon(z)a(z) \tag{15.34}$$

in which v is a new reduced complex coordinate and $a(z)$ is a real function, chosen so that all terms involving dv/dz disappear. Substituting Eq. (15.34) into (15.10), we obtain

$$v'' + \left(2\frac{a'}{a} + \frac{\gamma\phi'}{2\hat{\phi}}\right)v' + \left(\frac{a''}{a} + \frac{a'}{a}\frac{\gamma\phi'}{2\hat{\phi}} + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}}\right)v = 0 \tag{15.35}$$

and the coefficient of v' vanishes if

$$\frac{a'}{a} = -\frac{\gamma\phi'}{4\hat{\phi}} = -\frac{\hat{\phi}'}{4\hat{\phi}} \tag{15.36}$$

or

$$a(z) = \hat{\phi}^{-1/4} \tag{15.37}$$

giving

$$v''(z) + G(z)v(z) = 0 \tag{15.38}$$

with

$$G(z) := \frac{3}{16}\left(\frac{\phi'}{\hat{\phi}}\right)^2\left(1 + \frac{4}{3}\varepsilon\hat{\phi}\right) + \frac{\eta^2 B^2}{4\hat{\phi}} \tag{15.39}$$

and

$$u(z) = v(z)/\hat{\phi}^{1/4} \tag{15.40}$$

The substitution (15.40) was introduced into electron optics by Picht (1932), and is widely known as Picht's transformation; see too Glaser (1933a–d) and Cotte (1938). This result is of interest for two reasons. First, it is simpler to perform numerical calculations with Eq. (15.38) than (15.10). Secondly, the function $G(z)$ is essentially non-negative, and we shall see that this imposes an interesting restriction on electron lenses: they always exert a converging action. We shall also find that it is better to introduce the Picht transformation before proceeding to thin-lens approximations.

15.3.2 Stigmatic Image Formation

The paraxial equations are linear, homogeneous and of second order and their most general solution is therefore of the form

$$u(z) = Au_1(z) + Bu_2(z) \tag{15.41}$$

in which $u_1(z)$ and $u_2(z)$ are any pair of linearly independent solutions of Eq. (15.10). We shall find it necessary to introduce several such pairs of solutions and we shall adopt a

consistent notation for each in subsequent chapters, but many paraxial properties are quite general and in no way depend on any particular choice. The most important result concerns the existence of stigmatic image formation. The form of Eq. (15.10) or (15.12 and 15.13) alone is sufficient to predict that pairs of planes can always be found having the properties associated with point-to-point image formation.

Consider a particular solution $h(z)$ of the paraxial equation for the complex coordinate u (15.10) that intersects the axis at $z = z_0$ and $z = z_i$: $h(z_o) = h(z_i) = 0$ (Fig. 15.4). A pencil of rays intersecting the plane $z = z_o$ at some point P_o ($u_o = x_o + \mathrm{i}y_o$) may be described by

$$u(z) = u_o g(z) + \lambda h(z) \tag{15.42}$$

in which $g(z)$ is a solution of Eq. (15.10) that is linearly independent of $h(z)$; for convenience, we have set $g(z_o) = 1$; λ is a (complex) parameter characterizing the various members of the pencil. In the plane $z = z_i$, we have

$$u(z_i) = u_o g(z_i) \tag{15.43}$$

for all λ and hence for every ray passing through P_o. Since this is true of all points in the plane $z = z_o$, the latter will be stigmatically imaged in $z = z_i$. Moreover the ratio $u(z_i)/u_o$ is constant and so the distribution of points P_i, will be identical with the distribution of P_o, apart from a change of scale: the image is a linearly magnified (or reduced) representation of the object.

If we return to the fixed Cartesian system ($w = X + \mathrm{i}Y$, z), we find

$$w(z_i) = u(z_i)\exp\left\{\mathrm{i}\theta(z_i)\right\} = g(z_i)\exp[\mathrm{i}\left\{\theta(z_i) - \theta(z_o)\right\}]w(z_o)$$

or

$$w(z_i) = M(z_i, z_o)w(z_o) \tag{15.44}$$

where

$$M(z_i, z_o) = g(z_i)\exp[\mathrm{i}\left\{\theta(z_i) - \theta(z_o)\right\}] \tag{15.45}$$

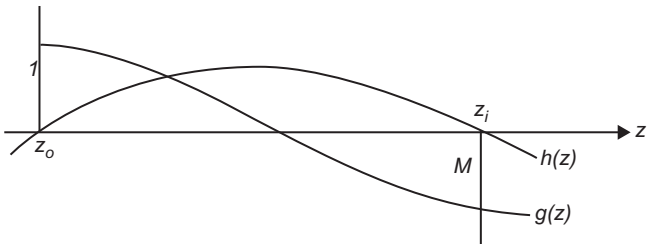


Figure 15.4
The paraxial solutions $g(z)$ and $h(z)$.

For single-stage image formation, in which a meridional ray from P_o intersects the axis only once between P_o and P_i, $g(z_i)$ is negative and we may write

$$M(z_i, z_o) = -|g(z_i)|\exp[\mathrm{i}\{\theta(z_i) - \theta(z_o)\}] = |g(z_i)|\exp[\mathrm{i}\{\theta(z_i) - \theta(z_o) + \pi\}] \tag{15.46}$$

The modulus of $g(z_i)$ and hence $|M(z_i, z_\mathbf{o})|$, is referred to as the transverse magnification and the image rotation is clearly equal to $\arg(M) - \pi$. The complex magnification $M(z_i, z_\mathbf{o})$ is rarely used and in the remainder of this book we shall reserve the symbol M for the transverse magnification, regarded as an algebraic quantity:

$$M \coloneqq g(z_i) \tag{15.47}$$

The notion of complex magnification is valuable when we need to consider the reversal of an imaging system. A pencil of rays from P_i to $P_\mathbf{o}$ will not retrace the paths of those from P_o to P_i since the direction of rotation will be opposite. This is readily seen from Eq. (15.45) which tells us that

$$M(z_o, z_i) = M^{*-1}(z_i, z_o) \tag{15.48}$$

in which the asterisk denotes the complex conjugate.

A number of useful general relations can be deduced from the form of Eq. (15.32) or (15.38) alone. Thus the fact that $F(z)$ and $G(z)$ are never negative tells us that all electron lenses have a net converging action although this need not be true of local zones of electrostatic lenses. To see this, we note that the curvature ρ of any solution of Eq. (15.32) or (15.38), given by $\rho = u''/(1+u'^2)^{3/2}$ or $\rho = \upsilon''/(1+\upsilon'^2)^{3/2}$, is always opposite in sign to u or υ respectively. Thus a solution of the appropriate paraxial equation that approaches the field parallel to the axis will be bent towards the latter. If the ray crosses the axis in the field, it will again be bent back towards it and if the field is long enough, the ray will oscillate about the axis. Thus the effect of the field is that of a converging lens. Nevertheless, care is needed here since rays can intersect the axis more than once in a strong lens and, as we shall see in the next section, the sign of the focal length will then be that associated with a divergent lens.

We have been basing our argument on the positivity of $F(z)$ or $G(z)$ and it is safe to conclude that a ray incident from field-free space parallel to the axis will intersect the axis at least once before emerging into image space. It is not, however, necessarily true that actual electron trajectories in electrostatic (or mixed electrostatic and magnetic) fields always bend towards the axis: the term in $(\phi_o/\hat{\phi})^{1/4}$ may be large enough to reverse the curvature locally. Provided the electron is not driven beyond the paraxial region, however, the convergent action will always dominate, as our reasoning based on $G(z)$ shows.

15.3.3 The Wronskian

Another property of the paraxial equations is the existence of an invariant, the Wronskian, from which a number of interesting optical relations can be derived. Let $u_1(z)$ and $u_2(z)$ be a pair of linearly independent solutions of Eq. (15.10), so that

$$\begin{aligned}\frac{d}{dz}(\hat{\phi}^{1/2}u_1') + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}}u_1 = 0\\ \frac{d}{dz}(\hat{\phi}^{1/2}u_2') + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}}u_2 = 0\end{aligned} \tag{15.49}$$

Multiplying the first equation by u_2 and the second by u_1 and subtracting, it is easy to show that

$$\frac{d}{dz}\left\{\hat{\phi}^{1/2}(u_1u_2' - u_1'u_2)\right\} = 0 \tag{15.50}$$

or

$$\hat{\phi}^{1/2}(u_1u_2' - u_1'u_2) = \text{const} \tag{15.51}$$

The same is of course true of any pair of solutions of the separate equations for $x(z)$ and $y(z)$ (15.13).

Suppose we choose $u_1(z) = h(z)$ and $u_2(z) = g(z)$, where as before $g(z_0) = 1$, $g(z_i) = M$ and $h(z_o) = h(z_i) = 0$ (15.42); we find

$$\hat{\phi}_o^{1/2}h_o' = \hat{\phi}_i^{1/2}h_i'M \tag{15.52}$$

But h_i'/h_o' is the angular magnification, M_α, and we have thus shown that

$$MM_\alpha = (\hat{\phi}_o/\hat{\phi}_i)^{1/2} \tag{15.53}$$

or, if $\phi_o = \phi_i$ as in the case of magnetic and electrostatic einzel lenses:

$$M_\alpha = \frac{1}{M} \quad (\text{when} \quad \hat{\phi}_i = \hat{\phi}_o) \tag{15.54}$$

We may rewrite Eq. (15.52) as

$$g(z_i)h'(z_i)\hat{\phi}_i^{1/2} = g(z_o)h'(z_o)\hat{\phi}_o^{1/2} \tag{15.55}$$

which is known in light optics as the Smith−Helmholtz formula (Born and Wolf, Eq. 4.4.49); it is also associated with the names of Clausius and Lagrange and was known

in a more primitive form to Cotes and Huygens (see Rayleigh, 1886 and Czapski and Eppenstein, 1924, p. 116).

A final related quantity is the longitudinal magnification: this tells us how far the image plane moves when the object plane is shifted a small distance. Consider again the pair of rays $g(z)$ and $h(z)$, which we now specify completely as the rays that satisfy the boundary conditions

$$\begin{aligned} g(z_o) &= h'(z_o) = 1 \\ g'(z_o) &= h(z_o) = 0 \end{aligned} \tag{15.56}$$

in the original object plane. In the image plane $h(z_i) = 0$ and $g(z_i) = M$. For a neighbouring object plane, distant Δz_o from $z = z_o$ (Fig. 15.5), the corresponding 'h-ray' satisfying

$$\begin{aligned} \overline{h}(z_o + \Delta z_o) &= 0 \\ \overline{h}'(z_o + \Delta z_o) &= 1 \end{aligned} \tag{15.57}$$

may be written as a linear combination of $g(z)$ and $h(z)$ since there can be only two linearly independent solutions:

$$\overline{h}(z) = Ah(z) + Bg(z) \tag{15.58}$$

Clearly

$$h(z_o + \Delta z_o) \approx \Delta z_o, \quad g(z_o + \Delta z_o) \approx 1$$

so that

$$A = 1, \quad B = -\Delta z_o$$

giving

$$\overline{h} = h - g\Delta z_o \tag{15.59}$$

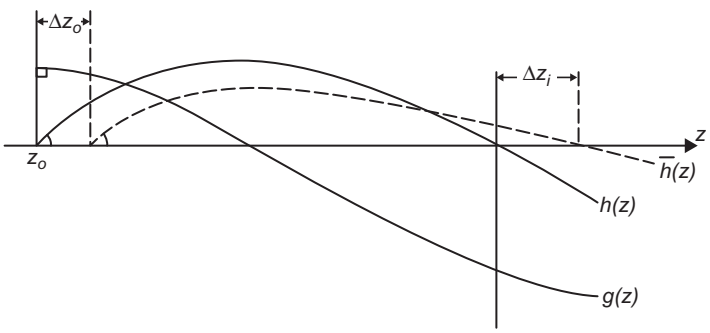


Figure 15.5
The notion of longitudinal magnification.

In the shifted image plane, $z = z_i + \Delta z_i$, $\overline{h}(z)$ vanishes and so

$$h(z_i + \Delta z_i) - g(z_i + \Delta z_i)\Delta z_o = 0$$

but $h'(z_i) = M_\alpha$ and $g(z_i) = M$ and hence

$$M_\alpha \Delta z_i - M\Delta z_o = 0$$

or

$$\frac{\Delta z_i}{\Delta z_o} =: M_l = \frac{M}{M_\alpha} = \left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{1/2} M^2 \tag{15.60}$$

The quantity M_l is known as the longitudinal magnification and we have

$$M_l M_\alpha = M \tag{15.61}$$

If $\phi_o = \phi_i$, $M_\alpha = 1/M$ (15.54) and

$$M_l = M^2 = \frac{1}{M_\alpha^2} \quad (\phi_o = \phi_i) \tag{15.62}$$

15.4 The Abbe Sine Condition and Herschel's Condition

These two conditions do not strictly belong to paraxial optics, for they are conditions under which particular sets of points are imaged stigmatically irrespective of the ray gradient. They are, however, of interest in electron optics mainly in connection with the foregoing results and we therefore make a short digression to establish them here. They are most easily derived from the invariance of the Lagrange bracket (5.34), as shown by Sturrock (1955) following the example of Herzberger (1931).

The invariance of the Lagrange bracket $\{u, v\}$ may be translated into concrete terms by considering three neighbouring rays, which we label 0, 1 and 2. The ray zero connects two points A, B as shown in Fig. 15.6; at these points, $\boldsymbol{r} = \boldsymbol{r}_a$, $\boldsymbol{p} = \boldsymbol{p}_a$ and $\boldsymbol{r} = \boldsymbol{r}_b$, $\boldsymbol{p} = \boldsymbol{p}_b$ respectively. The ray 1 is shifted by a small amount from ray 0, so that to its endpoints correspond the values $\boldsymbol{r}_a + \Delta_1\boldsymbol{r}_a$, $\boldsymbol{p}_a + \Delta_1\boldsymbol{p}_a$ and $\boldsymbol{r}_b + \Delta_1\boldsymbol{r}_b$, $\boldsymbol{p}_b + \Delta_1\boldsymbol{p}_b$; the same is true for the ray 2 except that the increments are now $\Delta_2\boldsymbol{r}_a$, $\Delta_2\boldsymbol{p}_a$, $\Delta_2\boldsymbol{r}_b$ and $\Delta_2\boldsymbol{p}_b$.

If these rays belong to a congruence, such that ray zero corresponds to the parameters (u, v), ray 1 to $(u + \Delta_1 u, v)$ and ray 2 to $(u, v + \Delta_2 v)$, the invariance of $\{u, v\}$ is equivalent to that of $\Delta_1\boldsymbol{p} \cdot \Delta_2\boldsymbol{r} - \Delta_2\boldsymbol{p} \cdot \Delta_1\boldsymbol{r}$, which is known as the *Lagrange differential invariant*. We now apply this invariance to two special cases.

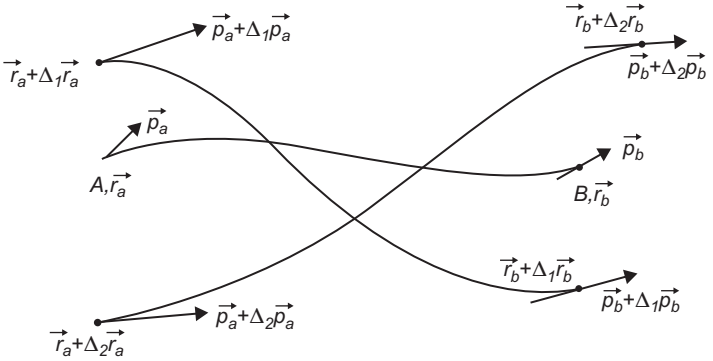


Figure 15.6
The rays employed in connection with the Lagrange invariant.

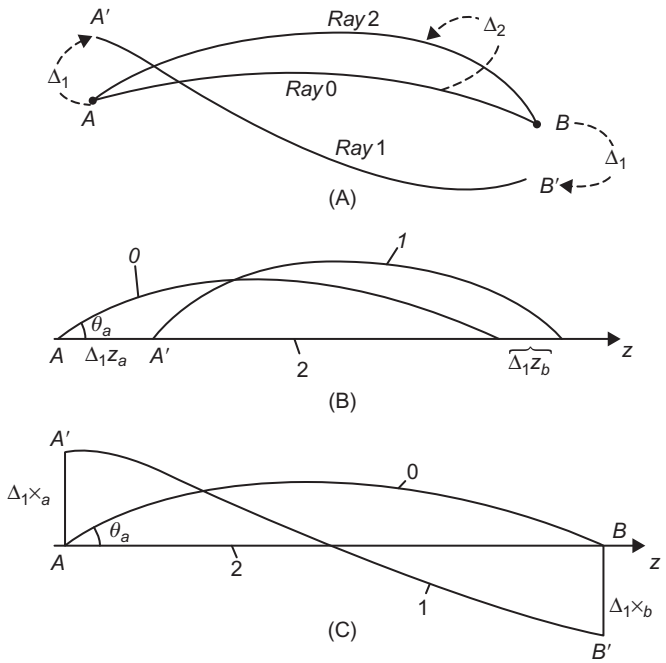


Figure 15.7
The rays employed in the derivation of (A, B) Herschel's condition and (C) the sine condition.

Suppose that the points A and B are conjugate and that to every point A' in a plane through A we can find a conjugate point B' in a plane through B. The shift from A to A' and B to B' is chosen to be the displacement Δ_1, while Δ_2 corresponds to the transition to another ray connecting A and B (Fig. 15.7A).

Thus $\Delta_2 \boldsymbol{r_a} = \Delta_2 \boldsymbol{r_b} = 0$ and $\Delta_1 x_a = \Delta_1 x_b = 0$, where the x-axes are taken perpendicular to the planes containing AA' and BB'. The invariance of $\Delta_1 \boldsymbol{p} \cdot \Delta_2 \boldsymbol{r} - \Delta_2 \boldsymbol{p} \cdot \Delta_1 \boldsymbol{r}$ shows that

$$p_a(\Delta_2 t_{ya} \cdot \Delta_1 y_a + \Delta_2 t_{za} \cdot \Delta_1 z_a) = p_b(\Delta_2 t_{yb} \cdot \Delta_1 y_b + \Delta_2 t_{zb} \cdot \Delta_1 z_b) \tag{15.63}$$

where we have written $\boldsymbol{p} = p\boldsymbol{t}$ and $\boldsymbol{t}$ is a unit vector, the components of which are the direction cosines of $\boldsymbol{p}$. The scalar p reduces to $\hat{\phi}^{1/2}$ near the axis.

We now choose the axes Oz_a, Oz_b, to coincide with the axes of a rotationally symmetric system and consider points in the planes x_a–z_a, x_b–z_b. Setting $\Delta_1 y_a = \Delta_1 y_b = 0$, $t_{za} = \cos\theta_a$, $t_{zb} = \cos\theta_b$ (Fig. 15.7B), we find

$$p_a \cdot \Delta z_a \cdot \sin\theta_a \cdot \Delta_2 \theta_a = p_b \cdot \Delta z_b \cdot \sin\theta_b \cdot \Delta_2 \theta_b \tag{15.64}$$

Writing $\Delta z_b = M_l \Delta z_a$, and integrating with respect to θ, we obtain *Herschel's condition*

$$p_a(\cos\theta_a - 1) = M_l p_b(\cos\theta_b - 1) \tag{15.65}$$

or

$$p_a \sin^2(\theta_a/2) = M_l p_b \sin^2(\theta_b/2) \tag{15.66}$$

If this condition is satisfied, an element of the axis close to A will be imaged sharply, even if the ray gradient is not small. If θ_a and θ_b are small, Eq. (15.66) becomes

$$p_a \theta_a^2 = M_l p_b \theta_b^2 \tag{15.67}$$

or

$$p_a / p_b = M_l M_\alpha^2 \tag{15.68}$$

which is equivalent to (15.61) using $p = \hat{\phi}^{1/2}$ and (15.53).

We may use the Lagrange differential invariant to derive the sine condition; the three rays are now chosen as shown in Fig. 15.7C. From Eq. (15.63), we have

$$p_a \cdot \Delta_1 y_a \cdot \cos\theta_a \cdot \Delta_2 \theta_a = p_b \cdot \Delta_1 y_b \cdot \cos\theta_b \cdot \Delta_2 \theta_b$$

note that $t_{ya} = \sin\theta_a$ and likewise for t_{yb}. Hence

$$p_a \cdot \sin\theta_a \cdot \Delta_1 y_a = p_b \cdot \sin\theta_b \cdot \Delta_1 y_b \tag{15.69}$$

which is known as the *sine condition*, the importance of which was first recognized by Ernst Abbe. When the sine condition is satisfied, a small region around the axis will be imaged sharply irrespective of the ray gradient. Looking ahead to Part IV, this implies that coma must vanish.

For small angles, Eq. (15.69) reduces to

$$p_a \theta_a \cdot \Delta_1 y_a = p_b \theta_b \cdot \Delta_1 y_b \tag{15.70}$$

which is equivalent to (15.55).

15.5 Some Other Transformations

We have seen that the paraxial equations for electrostatic lenses and mixed lenses can be cast into a more convenient form by means of Picht's transformation Eq. (15.40). We briefly mention some of the other transformations that have been proposed; we shall meet still others in connection with electron mirrors and with specific field models.

We first enquire whether the term in du/dz in (15.10) can be removed, not by a change of the transverse (dependent) variable as in the Picht transformation but by introducing a different axial (independent) variable. We write

$$\zeta = \zeta(z), \quad u = u(\zeta), \quad \phi = \phi(\zeta) \tag{15.71}$$

so that

$$\frac{du}{dz} = \frac{du}{d\zeta}\frac{d\zeta}{dz} \quad \text{and} \quad \frac{d^2u}{dz^2} = \frac{d^2u}{d\zeta^2}\left(\frac{d\zeta}{dz}\right)^2 + \frac{du}{d\zeta}\frac{d^2\zeta}{dz^2} \tag{15.72}$$

The paraxial equation (for electrostatic fields only) thus becomes

$$\ddot{u}\left(\frac{d\zeta}{dz}\right)^2 + \dot{u}\left\{\frac{\dot{\phi}\gamma}{2\hat{\phi}}\left(\frac{d\zeta}{dz}\right)^2 + \frac{d^2\zeta}{dz^2}\right\} + \gamma u\frac{\ddot{\phi}(d\zeta/dz)^2 + \dot{\phi}(d^2\zeta/dz^2)}{4\hat{\phi}} = 0 \tag{15.73}$$

in which dots denote differentiation with respect to ζ. The term in $\dot{u}$ vanishes if

$$\frac{d}{dz}\left(\hat{\phi}^{1/2}\frac{d\zeta}{dz}\right) = 0$$

or, apart from an unimportant multiplicative constant,

$$\zeta(z) = \int^z \hat{\phi}^{-1/2}(z')dz' \tag{15.74}$$

giving

$$\ddot{u} + \left\{\frac{\gamma}{4\hat{\phi}}\ddot{\phi} - \frac{1}{8}\left(\frac{\gamma\dot{\phi}}{\hat{\phi}}\right)^2\right\}u = 0 \tag{15.75}$$

An incorrect nonrelativistic form of this equation is given by Picht (1963, p. 166). Another transformation, also introduced by Picht (1932, 1963, p. 167), provides a means of designing lenses for a specific purpose, by generating the potential distribution that will create desired trajectories. We merely indicate the procedure: several examples are worked out in detail by Picht. From Eq. (15.10), in which we again set $B = 0$ and consider the nonrelativistic approximation, we see that the paraxial equation can be written

$$3\frac{u''}{u}(u\phi) + (u\phi)'' = 0 \tag{15.76}$$

Setting

$$u\phi =: T(z) \tag{15.77a}$$

and

$$3\frac{u''}{u} =: t(z) \tag{15.77b}$$

Eq. (15.76) becomes

$$T''(z) + t(z)T(z) = 0 \tag{15.78}$$

Thus, given $u(z)$ we can calculate $t(z)$, solve (15.78) for $T(z)$ and finally extract $\phi(z)$ from Eq. (15.77a). Picht gives another method of solving this problem, which we shall not describe here. For a recent attempt to solve the analogous problem in light optics, see Borghero and Demontis (2016).[1]

Hitherto, we have discussed the motion of electrons in terms of coordinates of position, deriving the ray gradients by differentiation. Position and canonical momentum are, however, conjugate variables, as explained in Part I, and we should therefore expect to be able to work in terms of either at will. Returning to the equations

$$\frac{d}{dz}\left(\frac{\partial M^{(2)}}{\partial x'}\right) = \frac{\partial M^{(2)}}{\partial x}$$

and writing $p = \partial M^{(2)}/\partial x'$, we see from Eq. (15.13) that

$$x = -\frac{4\hat{\phi}^{1/2}}{\gamma\phi'' + \eta^2 B^2}p'$$

so that substituting for x' in $p = x'\hat{\phi}^{1/2}$, we obtain

$$p = -\hat{\phi}^{1/2}\frac{d}{dz}\left(\frac{4p'\hat{\phi}^{1/2}}{\gamma\phi'' + \eta^2 B^2}\right)$$

or writing

$$G(z) := \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}} \tag{15.79}$$

[1] "In the framework of geometrical optics, [Borghero and Demontis] consider the following inverse problem: given a two-parameter family of curves (congruence) (i.e., $f(x,y,z) = c_1$, $g(x,y,z) = c_2$), construct the refractive-index distribution function $n = n(x,y,z)$ of a 3D continuous transparent inhomogeneous isotropic medium, allowing for the creation of the given congruence as a family of monochromatic light rays."

$$p'' - \frac{G'}{G}p' + \frac{G}{\hat{\phi}^{1/2}}p = 0 \tag{15.80}$$

All the rules of Gaussian optics that we shall establish in Chapter 16, Gaussian Optics of Rotationally Symmetric Systems: Asymptotic Image Formation, could equally well be derived from this equation; this duality is noted in Hawkes (1966).

CHAPTER 16

Gaussian Optics of Rotationally Symmetric Systems: Asymptotic Image Formation

16.1 Real and Asymptotic Image Formation

The fact that the paraxial trajectory equations are linear, second order and homogeneous is itself sufficient for us to anticipate that the imaging properties of the corresponding fields can be characterized by a small number of quantities. We discuss this in detail in the following paragraphs but we must first explain the notions of *real* and *asymptotic* image formation; the distinction between these is not quite the same as that between real and virtual in light optics, despite some similarities.

Since electron lenses consist of regions containing magnetic or electrostatic fields, it is possible, and in practice common, to immerse the specimen of which a magnified image is required within the field itself, particularly in the case of magnetic lenses. The lens field is thus divided into two regions playing different roles (Fig. 16.1). In a light microscope, any lenses preceding the specimen, region I in Fig. 16.1, belong to the condenser system, while the lens immediately after the specimen, region II, is the objective proper. In an electron microscope, different parts of the same lens may thus play different roles. The properties of region I will provide information about the illumination, those of region II about the image formation. In such a situation, it is clearly necessary to study the regions separated by the real object independently and the corresponding characteristics will be referred to as 'real'.

In a multi-lens system, most of the lenses will simply transfer an intermediate image from one plane to another, with the appropriate magnification, and the entire lens field contributes to this transfer. Here we must study the coordination between incoming and outgoing asymptotes, as shown in Fig. 16.2. If the intermediate image that acts as object for a lens is well outside the lens field, on the object side, the situation is exactly as in light optics. If it falls within the lens field or beyond it, then the *asymptotic object* is analogous to the familiar 'virtual object'; similar remarks apply to the image. In this context, we note that when discussing asymptotic imagery, the notions of object space and image space are used to refer not to regions of physical space – the object may lie anywhere as may the image – but to the space to which the corresponding asymptotes belong. This will become clearer when we discuss asymptotic image formation below.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00016-X
© 2018 Elsevier Ltd. All rights reserved.

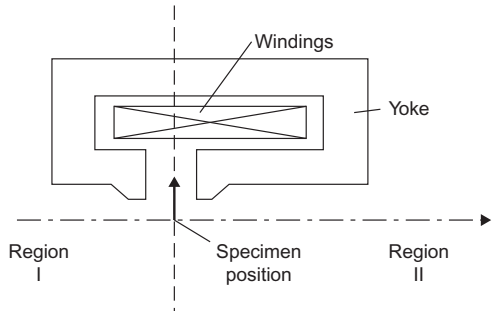


Figure 16.1
The various parts of a magnetic objective lens.

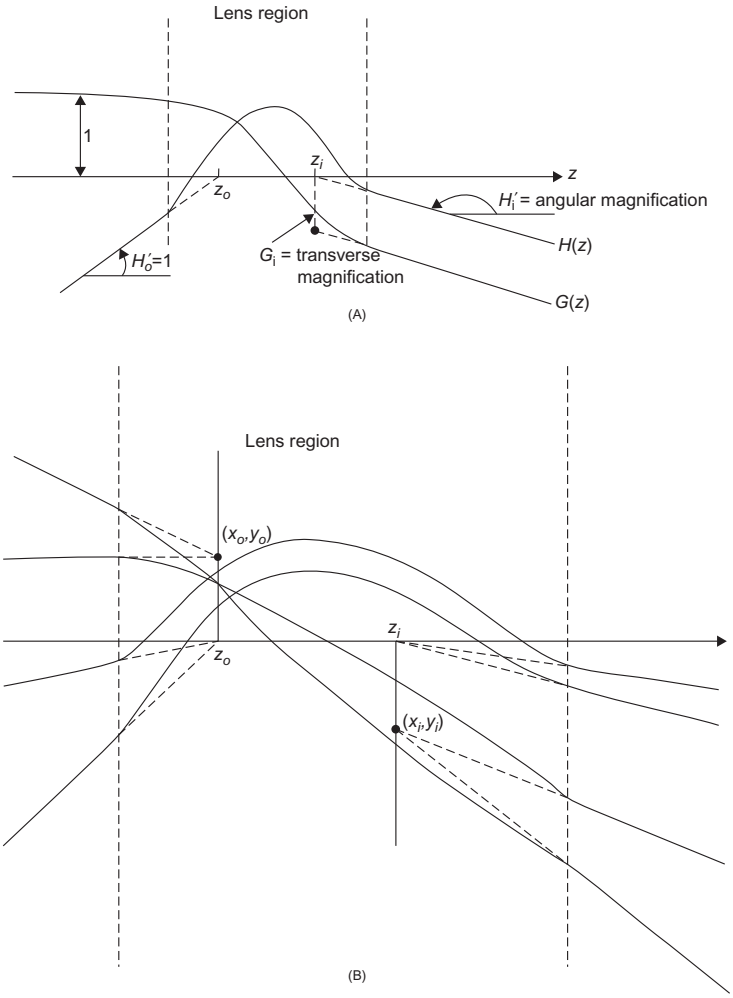


Figure 16.2
(A) The paraxial solutions $G(z)$ and $H(z)$. (B) Asymptotic image formation.

Finally, we note that both possible hybrid situations may occur: real object−asymptotic image and asymptotic object−real image (formation of a small probe within a field, for example).

16.2 Asymptotic Cardinal Elements and Transfer Matrices

We first discuss these matters in terms of specific solutions of the paraxial equations, after which we show that the same results may be obtained from the more abstract notion of bilinear transformations. In this first discussion, we use the form (15.12) of the paraxial equations and for ease of understanding we use the real x-coordinate. The reasoning for y is of course identical and we could naturally have used the complex u.

Consider a field region characterized by $\phi(z)$ and $B(z)$ (Fig. 16.3) and two solutions of the paraxial equation, $G(z)$ and $\overline{G}(z)$, satisfying the boundary conditions

$$\lim_{z\to -\infty} G(z) = 1, \qquad \lim_{z\to \infty} \overline{G}(z) = 1 \tag{16.1}$$

A general solution thus has the form

$$x(z) = AG(z) + B\overline{G}(z) \tag{16.2}$$

The rays $G(z)$ and $\overline{G}(z)$ tend to the following asymptotes:

$$\begin{aligned} \lim_{z\to \infty} G(z) &= (z - \zeta_i)G'_i \\ \lim_{z\to -\infty} \overline{G}(z) &= (z - \zeta_o)\overline{G'}_o \end{aligned} \tag{16.3}$$

Any ray incident parallel to the axis can be written as $\lambda G(z)$, where λ is a constant, and will have as its emergent asymptote $\lambda(z-\zeta_i)G'_i$. Thus rays incident parallel to the axis generate emergent asymptotes that all intersect the axis at $z = \zeta_i$ and we denote this point by $z = z_{Fi}$ and refer to it as the *asymptotic image focus*. By exactly analogous reasoning, we see that all rays that emerge parallel to the axis correspond to incident asymptotes that intersect the axis at $z = \zeta_o$; we write $\zeta_o = z_{Fo}$ and refer to this as the *asymptotic object focus*.

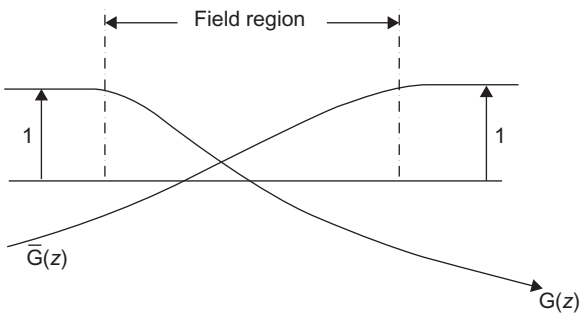


Figure 16.3
The paraxial solutions $G(z)$ and $G(z)$.

The incident asymptote to the ray $G(z)$ and its emergent asymptote intersect at a point in the plane $z = z_{Pi}$, such that

$$1 = (z_{Pi} - z_{Fi})G'_i$$

or

$$z_{Pi} = z_{Fi} + \frac{1}{G'_i} \tag{16.4}$$

Likewise the asymptotes to $\overline{G}(z)$ intersect in a plane $z = z_{Po}$,

$$1 = (z_{Po} - z_{Fo})\overline{G}'_o$$

giving

$$z_{Po} = z_{Fo} + \frac{1}{\overline{G}'_o} \tag{16.5}$$

The planes z_{Pi} and z_{Po} are known as the *asymptotic principal planes*. The distances $-1/G'_i$ and $1/\overline{G}'_o$ are known as the *asymptotic focal lengths* (Fig. 16.4). The reciprocal of a focal length is sometimes called the *convergence*. We write

$$f_i := -\frac{1}{G'_i}, \quad f_o := \frac{1}{\overline{G}'_o} \tag{16.6}$$

so that (16.4 and 16.5) become

$$z_{Pi} = z_{Fi} - f_i, \quad z_{Po} = z_{Fo} + f_o \tag{16.7}$$

This apparent inconsistency in the choice of sign is explained by considering the relation between f_o and f_i. Since (15.51)

$$\hat{\phi}^{1/2}(G\overline{G}' - G'\overline{G}) = \text{const} \tag{16.8}$$

we see that

$$\hat{\phi}_o^{1/2}\overline{G}'_o = -\hat{\phi}_i^{1/2}G'_i \tag{16.9}$$

or using (16.6),

$$f_o\hat{\phi}_o^{-1/2} = f_i\hat{\phi}_i^{-1/2} \tag{16.10}$$

In magnetic lenses, therefore, with the sign convention of (16.6), we have $f_o = f_i$ and shall frequently drop the suffix. In the class of electrostatic lenses that provide no overall acceleration, so that $\hat{\phi}_o = \hat{\phi}_i$, we again have $f_o = f_i$. Furthermore, f_o and f_i will both be positive if the rays $G(z)$ and $\overline{G}(z)$ intersect the axis only once, since electron lenses always have a convergent focusing action. As the lens is made stronger, however, there comes a point at which the emergent asymptotes are parallel to the optic axis as well as the incident

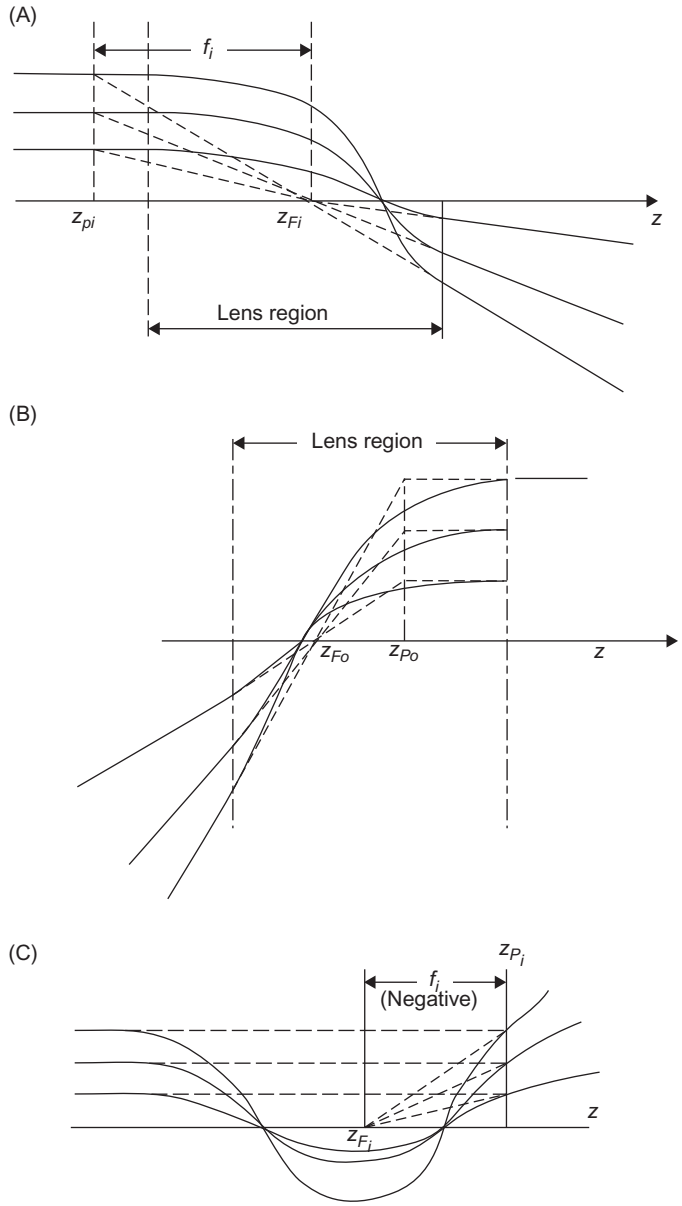


Figure 16.4

The asymptotic cardinal elements. (A) Image focus and principal plane. (B) Object focus and principal plane. (C) Image focus and principal plane for a strong lens; the image focal length has become negative.

asymptotes and the lens then behaves like a telescope (f_o and $f_i \to \infty$). Beyond this point, the rays return towards the axis but now G'_i is positive and $\overline{G}'_o$ negative (Fig. 16.4C). Formally, therefore, lenses operating in these conditions belong to the class of divergent lenses but since they have this character because they are so strongly convergent, this terminology is never used.

Returning to the general solution (16.2), we can express the incident and emergent asymptotes in the following way:

$$\lim_{z \to -\infty} x(z) = A + B\frac{z - z_{Fo}}{f_o} \tag{16.11a}$$

$$\lim_{z \to \infty} x(z) = -A\frac{z - z_{Fi}}{f_i} + B \tag{16.11b}$$

Eliminating A and B, we find, with $Q_{12} \coloneqq (z_1 - z_{Fo})(z_2 - z_{Fi})$

$$\begin{pmatrix} x_2 \\ x'_2 \end{pmatrix} = \begin{pmatrix} -(z_2 - z_{Fi})/f_i & f_o + Q_{12}/f_i \\ -1/f_i & (z_1 - z_{Fo})/f_i \end{pmatrix} \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix} \tag{16.12}$$

in which x_2 denotes $x(z)$ in some plane $z = z_2$ *on the emergent asymptote* and x'_2 the gradient of the latter ($x'_2 = -A/f_i$); x_1 denotes $x(z)$ in some plane z_1 *on the incident asymptote* and x'_1 the gradient ($x'_1 = B/f_o$). Writing

$$\boldsymbol{x} = \begin{pmatrix} x \\ x' \end{pmatrix} \tag{16.13}$$

and

$$T = \begin{pmatrix} -(z_2 - z_{Fi})/f_i & f_o + Q_{12}/f_i \\ -1/f_i & (z_1 - z_{Fo})/f_i \end{pmatrix} \tag{16.14}$$

Eq. (16.12) reduces to

$$\boldsymbol{x}_2 = T\boldsymbol{x}_1 \tag{16.15}$$

The matrix T is known as the *transfer matrix*, and we shall see that it encapsulates in a convenient way all the paraxial behaviour of the lens. From it, all the familiar imaging relations may be derived straightforwardly. Suppose that the planes P_o ($z_1 = z_o$) and P_i ($z_2 = z_i$) are conjugate, that is, that all rays from any point in P_o converge to a point in P_i. For this, the expression for x_i must be independent of x'_o and hence

$$f_o + (z_o - z_{Fo})(z_i - z_{Fi})/f_i = 0 \tag{16.16}$$

or

$$(z_o - z_{Fo})(z_i - z_{Fi}) = -f_i f_o \tag{16.17}$$

This is *Newton's lens equation*. Introducing the expression for z_{Po} and z_{Pi} (16.7), (16.17) becomes

$$(z_o - z_{Po})(z_i - z_{Pi}) - f_i(z_o - z_{Po}) + f_o(z_i - z_{Pi}) = 0$$

or

$$\frac{f_o}{z_{Po} - z_o} + \frac{f_i}{z_i - z_{Pi}} = 1 \tag{16.18}$$

Writing

$$\tilde{f} := f_o\left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{1/4} = f_i\left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/4} = (f_o f_i)^{1/2} \tag{16.19a}$$

so that $\overline{f}$ is the geometric mean of the focal lengths, and

$$\tilde{\phi} = (\hat{\phi}_o \hat{\phi}_i)^{1/2} \tag{16.19b}$$

this becomes

$$\frac{\hat{\phi}_o^{1/2}}{z_{Po} - z_o} + \frac{\hat{\phi}_i^{1/2}}{z_i - z_{Pi}} = \frac{\tilde{\phi}^{1/2}}{\tilde{f}} \tag{16.19c}$$

This is the thick-lens counterpart of the familiar thin-lens equation.

From (16.19a) we see that, irrespective of z_1 and z_2, the determinant of the transfer matrix T (16.14) has the value

$$\det T = f_o/f_i = (\hat{\phi}_o/\hat{\phi}_i)^{1/2}$$

For conjugate planes, the matrix equation simplifies to

$$\begin{pmatrix} x_i \\ x_i' \end{pmatrix} = \begin{pmatrix} -(z_i - z_{Fi})/f_i & 0 \\ -1/f_i & (z_o - z_{Fo})/f_i \end{pmatrix} \begin{pmatrix} x_o \\ x_o' \end{pmatrix} \tag{16.20}$$

and denoting the transverse magnification by M, we have

$$-(z_i - z_{Fi})/f_i = M \tag{16.21}$$

If $x_o = 0$ the ratio of x'_i to x'_o is the angular magnification M_α (15.52),

$$M_\alpha = (z_o - z_{Fo})/f_i \tag{16.22}$$

so that using (16.17),

$$M_\alpha = -f_o/(z_i - z_{Fi}) = f_o/f_i M \tag{16.23}$$

Hence

$$MM_\alpha = f_o/f_i = (\hat{\phi}_o/\hat{\phi}_i)^{1/2} \tag{16.24}$$

as already shown (15.53).

Thus

$$z_i = z_{Fi} - f_i M, \quad z_o = z_{Fo} + f_o/M \tag{16.25}$$

From (16.20), we see immediately that the principal planes are the pair of conjugate planes with unit magnification:

$$\begin{pmatrix} x_{Pi} \\ x'_i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1/f_i & f_o/f_i \end{pmatrix} \begin{pmatrix} x_{Po} \\ x'_o \end{pmatrix} \tag{16.26}$$

A third pair of axial points, the nodal points, is occasionally of interest. These are points having the property that a ray whose object asymptote intersects the axis at the object nodal point has an emergent asymptote intersecting the axis at the image nodal point, these asymptotes being parallel (Fig. 16.5). If these points are denoted by $z = z_{No}$, $z = z_{Ni}$, then

$$(z_{No} - z_{Fo})/f_i = 1 \tag{16.27}$$

so that if $\hat{\phi}_o = \hat{\phi}_i$, the nodal points and principal planes coincide. We shall see in Chapter 41 of Volume 2, that this is of interest in the design of aberration correctors. In general,

$$\begin{aligned} z_{No} &= z_{Fo} + f_i = z_{Po} - f_o + f_i \\ z_{Ni} &= z_{Fi} - f_o = z_{Pi} - f_o + f_i \end{aligned} \tag{16.28}$$

Once the foci and principal points are known, a simple construction enables us to obtain the point P_i conjugate to any object point P_o. First, a line is drawn through P_o parallel to the optic axis, intersecting the plane $z = z_{Pi}$ at Q; a line is then drawn through Q and the image focus, F_i (Fig. 16.6). Next, a line is drawn through P_o and the object focus F_o, intersecting the plane $z = z_{Po}$ at Q'; a line through Q' parallel to the optic axis intersects the line through Q and F_i at the image point P_i.

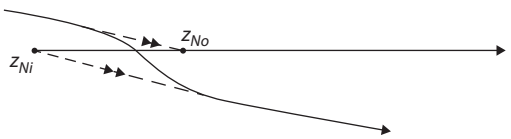


Figure 16.5
Nodal points.

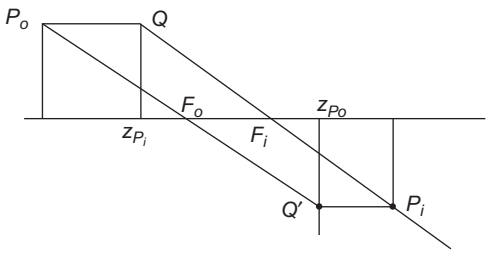


Figure 16.6

Asymptotic image construction. The image of an object P_o is obtained by first drawing a line through P_o and the object focus F_o, which intersects the object principal plane at Q'. A second line is drawn through the point Q in the image principal plane at the same height as P_o and the image focus F_i. The point of intersection of this second line and a line through Q' parallel to the optic axis is the image P_i of P_o.

The various quantities that characterize the imaging properties of a lens – the foci, principal points, nodal points and focal lengths – are known as its *cardinal elements*. In electron optics, only the focal lengths and the positions of the foci are routinely tabulated.

The matrix expression (16.12) is very convenient when we wish to calculate the cardinal elements of a doublet of two lenses, separated by a field-free region. Instead of using the general form of the matrix elements given in (16.12), however, we use the simple form that connects principal planes (16.26). We have

$$\begin{pmatrix} x(z_{Pi}^{(2)}) \\ x_i'^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1/f_i^{(2)} & f_o^{(2)}/f_i^{(2)} \end{pmatrix} \quad \begin{pmatrix} x(z_{Po}^{(2)}) \\ x_o'^{(2)} \end{pmatrix} \tag{16.29}$$

$$\begin{pmatrix} x(z_{Po}^{(2)}) \\ x_o'^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & D \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} x(z_{Pi}^{(1)}) \\ x_i'^{(1)} \end{pmatrix} \tag{16.30}$$

$$\begin{pmatrix} x(z_{Pi}^{(1)}) \\ x_i'^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1/f_i^{(1)} & f_o^{(1)}/f_i^{(1)} \end{pmatrix} \quad \begin{pmatrix} x(z_{Po}^{(1)}) \\ x_o'^{(1)} \end{pmatrix} \tag{16.31}$$

in which the superscripts (1) and (2) characterize the first and second lenses and

$$D = z_{Po}^{(2)} - z_{Pi}^{(1)} \tag{16.32}$$

is the distance between the object principal plane of the second lens and the image principal plane of the first. Multiplying the matrices, we obtain

$$\begin{pmatrix} x(z_{Pi}^{(2)}) \\ z_i'^{(2)} \end{pmatrix} = T_D \begin{pmatrix} x(z_{Po}^{(1)}) \\ z_o'^{(1)} \end{pmatrix} \tag{16.33}$$

where

$$T_D = \begin{pmatrix} -(z_{Pi}^{(2)} - \zeta_{Fi})/f_i^{II} & f_o^{II} + Q_{oi}/f_i^{II} \\ -1/f_i^{II} & (z_{Po}^{(1)} - \zeta_{Fo})/f_i^{II} \end{pmatrix} = \begin{pmatrix} 1 - D/f_i^{(1)} & Df_o^{(1)}/f_i^{(1)} \\ (D - f_i^{(1)} - f_o^{(2)})/f_i^{(1)}f_i^{(2)} & -f_o^{(1)}(D - f_o^{(2)})/f_i^{(1)}f_i^{(2)} \end{pmatrix} \tag{16.34}$$

in which ζ_{FO} and ζ_{Fi} denote the object and image foci of the doublet and f_i^{II} its image focal length; we have written $Q_{oi} \coloneqq (z_{p_o}^{(1)} - \zeta_{Fo})(z_{pi}^{(2)} - \zeta_{Fi})$ We see that

$$f_i^{II} = \frac{f_i^{(1)}f_i^{(2)}}{f_i^{(1)} + f_o^{(2)} - D} = \frac{f_i^{(1)}f_i^{(2)}}{D_F} \tag{16.35}$$

With

$$D_F \coloneqq f_i^{(1)} + f_o^{(2)} - D = z_{F_i}^{(1)} - z_{Fo}^{(2)} \tag{16.36}$$

$$\begin{aligned} \zeta_{Fi} - z_{Pi}^{(2)} &= f_i^{(2)}(f_i^{(1)} - D)/D_F \\ \zeta_{Fo} - z_{Po}^{(1)} &= f_o^{(1)}(f_o^{(2)} - D)/D_F \end{aligned} \tag{16.37}$$

We have not exhausted the forms of the transfer matrix that are occasionally useful. Thus we shall encounter the focal transfer matrix between the (nonconjugate) foci, and the Dušek matrix, in which $z_2 = z_i$. We shall discuss these as the need arises.

16.3 Gaussian Optics as a Projective Transformation (Collineation)

The reasoning in Section 16.2 has been based on the physical notion of rays and their asymptotes but the characteristic quantities of Gaussian optics, the cardinal elements, also emerge from a more abstract approach. The linearity of the equations of motion is sufficient for us to assert that the object space and image space of a field region are connected by a *projective transformation* or *collineation*. By this we mean that if (x_o, y_o, z_o) are the Cartesian coordinates of a point P_o in object space and (x_i, y_i, z_i) those of a point P_i in image space, then

$$x_i = F_1/F_4, \quad y_i = F_2/F_4, \quad z_i = F_3/F_4 \tag{16.38a}$$

where

$$F_j = a_j x_o + b_j y_o + c_j z_o + d_j \quad (j = 1, 2, 3, 4) \tag{16.38b}$$

and solving for x_o, y_o, z_o

$$\begin{aligned} x_o &= F'_1/F'_4, \quad y_o = F'_2/F'_4, \quad z_o = F'_3/F'_4 \\ F'_j &= a'_j x_i + b'_j y_i + c'_j z_i + d'_j \end{aligned} \tag{16.39}$$

From (16.38b), it is immediately obvious that the images of all points lying in the plane $F_4 = 0$ are at infinity: $F_4 = 0$ is thus the object focal plane and similarly, the plane $F'_4 = 0$ is the image focal plane.

Suppose now that the system has rotational symmetry about the z-axis and suppose too that the coordinates in image space are rotated with respect to those in object space by the appropriate angle θ (15.9 or 15.27). An object point $(0, y_o, z_o)$ will be transformed into $(0, y_i, z_i)$ but z_i will be unaltered if y_o is replaced by $-y_o$ whereas y_i will become $-y_i$. From the relations

$$y_i = \frac{b_2 y_o + c_2 z_o + d_2}{b_4 y_o + c_4 z_o + d_4}, \quad z_i = \frac{b_3 y_o + c_3 z_o + d_3}{b_4 y_o + c_4 z_o + d_4} \tag{16.40}$$

we deduce that $b_4 = b_3 = c_2 = d_2 = 0$, giving

$$y_i = \frac{b_2 y_o}{c_4 z_o + d_4}, \quad z_1 = \frac{c_3 z_o + d_3}{c_4 z_o + d_4} \tag{16.41}$$

or

$$y_o = \frac{c_4 d_3 - c_3 d_4}{b_2} \frac{y_i}{c_4 z_i - c_3}, \quad z_o = -\frac{d_4 z_i - d_3}{c_4 z_i - c_3}$$

The focal planes are thus given by the solution of

$$\begin{aligned} c_4 z_o + d_4 = 0 \quad &: \quad z_{Fo} = -d_4/c_4 \\ c_4 z_i - c_3 = 0 \quad &: \quad z_{Fi} = c_3/c_4 \end{aligned} \tag{16.42}$$

On measuring distances from these planes, by writing

$$\begin{aligned} Z_o &:= z_o + d_4/c_4 \\ Z_i &:= z_i - c_3/c_4 \end{aligned} \tag{16.43}$$

and introducing f_o, f_i thus:

$$f_o := b_2/c_4, \quad f_i := -(c_4 d_3 - c_3 d_4)/b_2 c_4 \tag{16.44}$$

we obtain

$$\frac{y_i}{y_o} = \frac{f_o}{Z_o} = -\frac{Z_i}{f_i} \tag{16.45}$$

This yields Newton's lens equation (16.17)

$$Z_o Z_i = -f_o f_i \tag{16.46}$$

Furthermore, the magnification $M = dy_i/dy_o$ for constant z_o is given by

$$M = y_1/y_o = f_o/Z_o = -Z_i/f_i \tag{16.47}$$

so that for $M = 1$ we have $Z_o = f_o$, $Z_i = -f_i$. The planes thus defined are the principal planes, situated at distances f_o and f_i from the foci as already shown (16.7).

We shall not pursue this further but clearly all the remaining results of Gaussian optics can be derived straightforwardly. For lengthy discussion of the use of projective transformations in this context, see Ollendorff (1955), Born and Wolf (1959), Czapski and Eppenstein (1924) and Carathéodory (1937).

16.4 Use of the Angle Characteristic to Establish the Gaussian Optical Quantities

Finally, we offer a third method of establishing the relations of Gaussian optics, setting out from the *angle characteristic*. This function is obtained from the point characteristic function by a Legendre transformation, which effectively changes the arguments from point coordinates to momenta. We define a function T as follows:

$$T \coloneqq S + p_a x_a + q_a y_a - p_b x_b - q_b y_b \tag{16.48}$$

in which (p, q) are the transverse components of $\boldsymbol{p}$. From (5.29) we know that

$$\Delta S \coloneqq p_b \Delta x_b + q_b \Delta y_b - (p_a \Delta x_a + q_a \Delta y_a) \tag{16.49}$$

when the integral in S is taken along a ray so that

$$\Delta T \coloneqq -(x_b \Delta p_b + y_b \Delta q_b) + x_a \Delta p_a + y_a \Delta q_a \tag{16.50}$$

and T must be a function of p_a, q_a, p_b and q_b; the function T is known as the *angle characteristic*. Provided that p_b is not proportional to p_a and q_b, to q_a, we see that

$$\begin{aligned} x_b &= -\frac{\partial T}{\partial p_b}, \quad x_a = \frac{\partial T}{\partial p_a} \\ y_b &= -\frac{\partial T}{\partial q_b}, \quad y_a = \frac{\partial T}{\partial q_a} \end{aligned} \tag{16.51}$$

Consider now pairs of points A, A' in object space and B, B' in image space. We assume that the angle characteristic between A and B, T_{AB}, is known and we calculate the new value between A', B', assuming asymptotic image formation. We write

$$T_{A'B'} = T_{A'A} + T_{AB} + T_{BB'} \tag{16.52}$$

From (16.48), we know that

$$\begin{aligned} T_{A'A} &= S_{A'A} + p_{A'} x_{A'} + q_{A'} y_{A'} - p_A x_A - q_A y_A \\ &= (z_A - z_{A'})(\hat{\phi}_A - p_A^2 - q_A^2)^{1/2} \end{aligned} \tag{16.53}$$

in which we have used $p_A = p_{A'}\ q_A = q_{A'}\ S_{AA'} = \hat{\phi}_A^{1/2}\overline{AA'}$ and

$$x_A = x_{A'} + (z_A - z_{A'})x'_A, \quad y_A = y_{A'} + (z_A - z_{A'})y'_A$$

with

$$x'_A = \frac{p_A}{(\hat{\phi}_A - p_A^2 - q_A^2)^{1/2}}, \qquad y'_A = \frac{q_A}{(\hat{\phi}_A - p_A^2 - q_A^2)^{1/2}}$$

A similar expression is obtained for $T_{BB'}$. To the paraxial approximation therefore

$$\begin{aligned} T_{A'B'} = T_{AB} &- (z_{A'} - z_A)\left(1 - \frac{p_A^2 + q_A^2}{2\hat{\phi}_A}\right)\hat{\phi}_A^{1/2} \\ &+ (z_{B'} - z_B)\left(1 - \frac{p_B^2 + q_B^2}{2\hat{\phi}_B}\right)\hat{\phi}_B^{1/2} \end{aligned} \tag{16.54}$$

Since the system has rotational symmetry, the quantities p and q can only appear in the combinations $p_A^2 + q_A^2, p_B^2 + q_B^2, p_A p_B + q_A q_B$ and we write

$$T_{AB} = \alpha(p_A^2 + q_A^2)/2 + \beta(p_B^2 + q_B^2)/2 - \overline{f}(p_A p_B + q_A q_B) \tag{16.55}$$

By applying (16.51) to (16.54), with $x_a = x_{A'}$, $p_a = p_{A'} = p_A$ etc., we see that

$$\begin{aligned} x_{A'} &= \left\{\alpha + (z_{A'} - z_A)/\hat{\phi}_A^{1/2}\right\}p_A - \overline{f}p_B \\ y_{A'} &= \left\{\alpha + (z_{A'} - z_A)/\hat{\phi}_A^{1/2}\right\}q_A - \overline{f}q_B \\ x_{B'} &= -\left\{\beta - (z_{B'} - z_B)/\hat{\phi}_B^{1/2}\right\}p_B + \overline{f}p_A \\ y_{B'} &= -\left\{\beta - (z_{B'} - z_B)/\hat{\phi}_B^{1/2}\right\}q_B + \overline{f}q_A \end{aligned} \tag{16.56}$$

If $p_B = 0$, then $x_{A'} = 0$ in the plane $z_{A'} - z_A = \alpha\hat{\phi}_A^{1/2}$ for all p_A while if $p_A = 0$, then $x_{B'}$ vanishes in $z_{B'} - z_B = \beta\hat{\phi}_B^{1/2}$. These are the foci: we write $z_{Fi} - z_B = \beta\hat{\phi}_B^{1/2}$, $z_{Fo} - z_A = \alpha\hat{\phi}_A^{1/2}$. Eliminating p_A or p_B between the equations for $x_{A'}$ and $x_{B'}$, we find

$$\overline{f}\,x_B = \left(\beta - \frac{z_{B'} - z_B}{\hat{\phi}_B^{1/2}}\right)x_A$$

or

$$\overline{f}\,x_A = \left(\alpha + \frac{z_{A'} - z_A}{\hat{\phi}_A^{1/2}}\right)x_B$$

provided that

$$(\alpha\hat{\phi}_A^{1/2} + z_{A'} - z_A)(\beta\hat{\phi}_B^{1/2} - z_{B'} + z_B) = \overline{f}^2(\hat{\phi}_A\hat{\phi}_B)^{1/2} \tag{16.57}$$

Thus $x_{B'} = x_{A'}$ if $z_{B'} - z_B = \hat{\phi}_B^{1/2}(\beta - \overline{f})$ and $z_{A'} - z_A = \hat{\phi}_A^{1/2}(\overline{f} - \alpha)$. These are therefore the principal planes,

$$\begin{aligned} z_{Pi} - z_B &= \hat{\phi}_B^{1/2}(\beta - \overline{f}) \quad \text{or} \quad z_{Pi} - z_{Fi} = -\hat{\phi}_B^{1/2}\overline{f} \\ z_{Po} - z_A &= \hat{\phi}_A^{1/2}(\overline{f} - \alpha) \quad \text{or} \quad z_{Po} - z_{Fo} = \hat{\phi}_A^{1/2}\overline{f} \end{aligned} \tag{16.58}$$

With the sign convention of (16.7), we recognize that the focal lengths are given by

$$f_o = \hat{\phi}_A^{1/2}\overline{f}, \qquad f_i = \hat{\phi}_B^{1/2}\overline{f} \tag{16.59}$$

Condition (16.57) may then be written in the form of Newton's lens equation.

16.5 The Existence of Asymptotes

In the foregoing sections, we have assumed that the curved trajectories within the field region tend to asymptotes in object and image space, 'outside' the field. In theory, however, the fields continue indefinitely, though they of course become vanishingly small, and we need to be sure that it is legitimate to use the concept of asymptotes. In particular, we need to establish conditions concerning the rate at which the field functions tend to zero for large values of $|z|$. These questions have been explored in detail by Glaser and Bergmann (1950), whom we follow closely.

If a general solution $x(z)$ of the paraxial equation tends to an asymptote, the gradient $x'(z)$ and the intercept of the tangent to $x(z)$ with an arbitrary plane perpendicular to the axis must both tend to constant values:

$$\lim_{z\to\infty} x'(z) = a, \qquad \lim_{z\to\infty} \{x(z) - z\ x'(z)\} = b \tag{16.60}$$

and similarly for $z \to -\infty$. In reduced coordinates (15.40), these conditions become

$$\lim_{z\to\infty} \hat{\phi}^{-1/4}\left\{\xi'(z) - \frac{1}{4}\frac{\gamma}{\hat{\phi}}\phi'\xi(z)\right\} = a \tag{16.61a}$$

$$\lim_{z\to\infty} \hat{\phi}^{-1/4}\left\{\xi(z)\left(1 + \frac{\gamma}{4\hat{\phi}}\phi' z\right) - z\xi'(z)\right\} = b \tag{16.61b}$$

with $\xi = x\hat{\phi}^{1/4}$. Since $\hat{\phi}$ is always finite in real fields, ξ and ξ' must be finite. The condition (16.61a) may therefore be replaced by

$$\lim_{z\to\infty} \xi'(z) = A, \qquad \lim_{z\to\infty} \xi\phi' = \overline{A} \tag{16.62}$$

ϕ' must vanish at infinity, and we assume that it falls to zero faster than $1/z$, which is easily justified. The second condition (16.61b) thus becomes

$$\lim_{z\to-\infty}(\xi-\xi' z)=B \tag{16.63}$$

which is immediately recognizable as the condition that solutions of the reduced equation (15.38) tend to asymptotes as $z\to\infty$. Multiplying (15.38) by z and integrating, we obtain (with ξ in place of v)

$$\xi-\xi' z=\int G(z)\xi z dz$$

so that for any upper bounds α or $\overline{\alpha}$

$$\begin{aligned}\lim_{z\to\infty}(\xi-\xi' z)=B&=\lim_{\alpha\to\infty}\int^{\alpha}G\xi z dz\equiv\int^{\alpha}G\xi z dz+\lim_{\alpha\to\infty}\int_{\overline{\alpha}}^{\alpha}G\xi z dz\\&=\int^{\alpha}G\xi z dz+\lim_{\alpha\to\infty}\int_{\overline{\alpha}}^{\alpha}Gz(\xi-\xi' z)dz+\lim_{\alpha\to\infty}\int_{\overline{\alpha}}^{\alpha}G\xi' z^2 dz\end{aligned} \tag{16.64}$$

Choosing $\overline{\alpha}$ very large, we find

$$B=\int^{\overline{\alpha}}G\xi z dz+B\lim_{\alpha\to\infty}\int_{\overline{\alpha}}^{\alpha}Gz dz-A\lim_{\alpha\to\infty}\int_{\overline{\alpha}}^{\alpha}Gz^2 dz \tag{16.65}$$

For the existence of asymptotes in general, therefore, the integrals

$$\int_{-\infty}^{\infty}zG(z)dz \tag{16.66}$$

and

$$\int_{-\infty}^{\infty}z^2G(z)dz \tag{16.67}$$

must converge; the existence of asymptotes parallel to the axis ($A=0$) is, however, guaranteed by the convergence of $\int_{-\infty}^{\infty}zG(z)dz$ alone.

The conditions are *necessary*; we now show that they are also *sufficient*. We can in principle solve (15.38) by an iterative procedure, taking as first approximation a linear expression of the form $\xi_0=\alpha z+\beta$ *so* that writing $\xi=\xi_0+\xi_1+\cdots+\xi_n+\cdots$,

$$\xi''_{n+1} + G(z)\xi_n = 0 \tag{16.68}$$

or

$$\xi_{n+1} = -\int_z^\infty d\zeta \int_\zeta^\infty G\xi_n dX = \int_\zeta^\infty (z+X)G\xi_n dX \tag{16.69}$$

Given the convergence of (16.66) and (16.67), we see that for $z > \overline{\alpha}$ and $\overline{\alpha}$ large,

$$\int_{\overline{\alpha}}^{z} G\,dX < \int_{\overline{\alpha}}^{z} GX\,dX < \int_{\overline{\alpha}}^{\infty} GX^2\,dX =: \varepsilon(\alpha) < 1 \tag{16.70}$$

and clearly as $\overline{\alpha} \to \infty$, $\varepsilon(\overline{\alpha}) \to 0$. From (16.69), we have

$$\xi_1 = \int (\overline{\alpha} + X)G(AX + B)dX$$

and since $X > \overline{\alpha}$,

$$\int_{\alpha}^{\infty} \alpha XG dX < \int_{\overline{\alpha}}^{\infty} X^2 G dX = \varepsilon(\alpha) \tag{16.71}$$

and so

$$|\xi_1| \le |A| \left\{ \int_{\overline{\alpha}}^{\infty} \overline{\alpha} XG dX + \int_{\overline{\alpha}}^{\infty} GX^2 dX \right\} + |B| \left\{ \int_{\overline{\alpha}}^{\infty} \overline{\alpha} G dX + \int_{\overline{\alpha}}^{\infty} \overline{\alpha} G dX + \int_{\overline{\alpha}}^{\infty} GX dX \right\}$$

From (16.70) and (16.71), we may conclude that

$$|\xi_1| \le 2(|A| + |B|)\varepsilon(\overline{\alpha})$$

Iterating, we find

$$|\xi - \xi_0| \le |\xi_1| + |\xi_2| + \cdots = (|A| + |B|)\frac{2\varepsilon(\alpha)}{1 - 2\varepsilon(\alpha)}$$

As $\overline{\alpha}$ is made larger and hence as $z \to \infty$, $\varepsilon(\overline{\alpha}) \to 0$ (16.70) and ξ tends to the linear solution ξ_0. The convergence of (16.66) and (16.67), which we have used in this derivation, is therefore not only a necessary but also a sufficient condition. In practice, the integrals (16.66) and (16.67) always do converge: since a total system must be electrically or magnetically neutral, $G(z)$ must fall off at least as fast as z^{-6} and any integral of the form $\int G(z)z^n\,dz$ will then converge for $n \le 4$.

CHAPTER 17

Gaussian Optics of Rotationally Symmetric Systems: Real Cardinal Elements

Hitherto we have considered only the coordination between asymptotes. When this is not appropriate, in microscope objective lenses for example, a different set of cardinal elements must be used. After discussing these, we enquire whether fields exist for which fixed real cardinal elements can be defined for a range of object positions. This leads us to the concept of Newtonian imaging fields, to which Glaser attached considerable importance (Glaser and Bergmann, 1950, 1951; also Glaser and Lammel, 1941, 1943); we shall follow his discussion closely. Further contributions were made by Funk (1950) and by Hutter (1945), who wrongly included several non-Newtonian distributions in the family of Newtonian fields. With the decline in the use of model fields, pedagogic purposes excepted, this class of field distributions has fallen into desuetude.

We shall use the rotating coordinate frame (x, y, z) without comment, and it must be remembered that in magnetic lenses, the surfaces $x = 0$ and $y = 0$ are not plane though we shall still speak of 'parallel' rays.

17.1 Real Cardinal Elements for High Magnification and High Demagnification

Objective lenses are conventionally operated at high magnification, and to a good approximation we may assume that the image is formed at infinity (Fig. 17.1). The family of rays that emerge parallel to the axis intersect the latter at some point F_o', which we call the *real object focus*; a family of rays emerging from the lens parallel to one another but not parallel to the axis intersect in the focal plane, the plane through F_o' perpendicular to the axis. This can be seen by introducing the solutions $G(z)$ and $\overline{G}(z)$ of the paraxial equations already used in Section 16.2 (Eq. 16.1). A family of rays parallel in image space may be written

$$x(z) = \alpha G(z) + c_k \overline{G}(z) \tag{17.1}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00017-1
© 2018 Elsevier Ltd. All rights reserved.

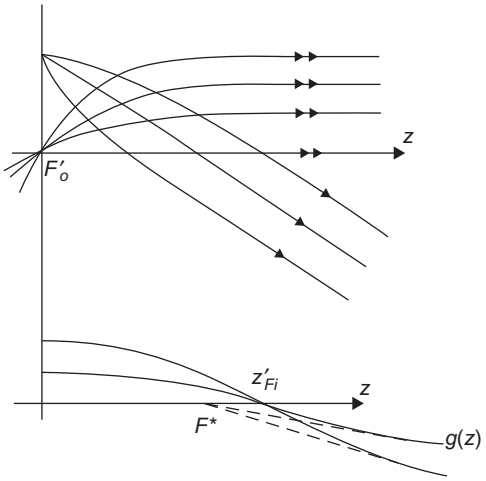


Figure 17.1
Above: The real object focus F'_o. Rays that intersect in a point in the real object focal plane emerge into image space as a parallel beam, parallel to the axis if the point lies on the axis. Below: Definition of the ray $g(z)$, which crosses the optic axis at the real image focus ($z = z'_{Fi}$). The corresponding asymptotic focus (here denoted by F^*) is in practice more useful.

in which the c_k are constants corresponding to the different rays; since

$$\lim_{z\to\infty} x(z) = \alpha G'_i(z - z_{Fi}) + c_k = -\frac{\alpha}{f_i}(z - z_{Fi}) + c_k \tag{17.2}$$

where z_{Fi} and f_i are the asymptotic focus and focal length, we see that Eq. (17.1) represents a family of rays all with image gradient $-\alpha/f_i$. In the real object focal plane, $z = z'_{Fo}$, $\overline{G}(z)$ vanishes and hence

$$x(z'_{Fo}) = \alpha G(z'_{Fo}) \qquad \text{for all} \qquad c_k \tag{17.3}$$

The real image focus is defined to be the point at which rays, parallel to the axis in the plane $z = z'_{Fo}$, intersect the axis. It is now convenient to use a different ray pair: instead of $G(z)$ and $\overline{G}(z)$, we employ $g(z)$ and $\overline{G}(z)$, where

$$g(z'_{Fo}) = 1, \quad g'(z'_{Fo}) = 0 \tag{17.4}$$

(Fig. 17.1). In the real *image focal plane* $z = z'_{Fi}$, $g(z)$ vanishes:

$$g(z'_{Fi}) = 0 \tag{17.5}$$

It is easy to show that a family of rays, parallel to one another in $z = z'_{Fo}$ but not parallel to the axis, intersect in the real image focal plane $z = z'_{Fi}$. Such a family may be written

$$x(z) = \alpha\overline{G}(z) + c_k g(z)$$

and in the plane $z = z'_{Fi}$,

$$x(z'_{Fi}) = \alpha \overline{G}(z'_{Fi}) \tag{17.6}$$

We shall see in Volume 3 that the diffraction pattern of an object placed in the real object focal plane and illuminated with a beam of electrons parallel to the axis in this plane is formed in the real image focal plane; Eq. (17.6) expresses a fundamental property of diffraction patterns, that is, that rays that are parallel in the specimen plane intersect in the 'diffraction plane', forming diffraction spots if only isolated directions occur.

In practice, it is more useful to know the location of the *asymptotic* image focus corresponding to the real object focus. This is the point at which the emergent asymptote to $g(z)$ intersects the axis (Fig. 17.1).

Focal lengths are associated with each of these foci. In the case of the real object focus, we define the real object principal plane, $z = z'_{Po}$, to be the plane perpendicular to the axis through the point of intersection of the emergent asymptote to $\overline{G}(z)$ and the tangent to $\overline{G}(z)$ at $z = z'_{Fo}$ (Fig. 17.2). The real focal length f'_o is then given by

$$f'_o = z'_{Po} - z'_{Fo} \tag{17.7}$$

and since $\lim_{z \to \infty} \overline{G} = 1$, we find

$$f'_o = \frac{1}{\overline{G}'(z'_{Fo})} \tag{17.8}$$

For the real image focal length f'_i, we have

$$f'_i = -\frac{1}{g'(z'_{Fi})} \tag{17.9}$$

The Wronskian tells us that for magnetic fields

$$\overline{G}g' - \overline{G}'g = \text{const}$$

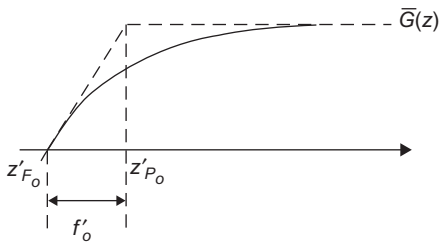


Figure 17.2
The real object principal plane and the real (objective) focal length.

and by considering the focal planes, that

$$\frac{1}{f'_o} = \frac{\overline{G}(z'_{Fi})}{f'_i} \tag{17.10}$$

For the hybrid focal length obtained by considering the emergent asymptote to $g(z)$, which we denote f''_i, we have

$$f''_i = -\frac{1}{g'(\infty)} \tag{17.11}$$

and the Wronskian tells us that

$$f''_i = f'_o \tag{17.12}$$

Finally, we consider the practical situation in which a specimen is placed close to but not exactly at $z = z'_{Fo}$, so that the magnification is high but not infinite. From Fig. 17.3A, we see that the magnification is given by

$$|M| = \frac{\alpha_o}{\alpha_i} = \frac{Z_i}{f'_o + \zeta} \approx \frac{Z_i}{f'_o} \tag{17.13}$$

in which $\zeta \ll f'_o$ and Z_i is the distance to the image from z'_{Po}. From Fig. 17.3B, however, we can express the magnification as

$$|M| = \left|\frac{x_i}{x_o}\right| \approx \frac{f'_o}{\zeta} \tag{17.14}$$

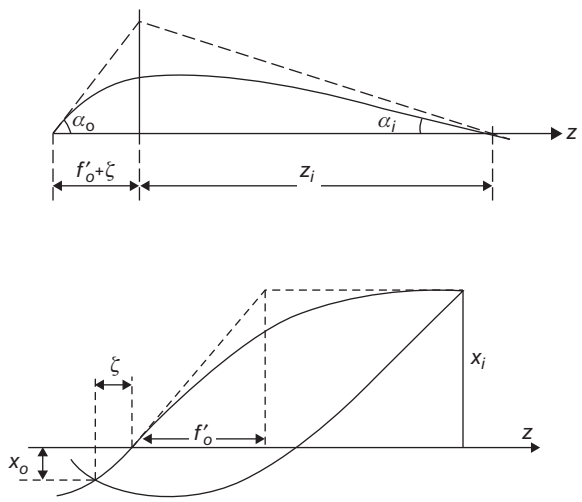


Figure 17.3
High but not infinite magnification in an objective lens.

(by similar triangles). Thus $\zeta Z_i = f_o'^2$, in which we once again recognize Newton's lens equation (now valid only for small values of ζ).

17.2 Osculating Cardinal Elements

The concept of osculating cardinal elements was introduced by Glaser (Glaser and Lammel, 1941, 1943; Glaser and Bergmann, 1950, 1951) in an attempt to establish whether generalized cardinal elements can be defined that collapse to the asymptotic cardinal elements when both object and image lie outside the field and which also allow us to use the lens equation in Newton's form,

$$M = \frac{\overline{f}_o}{z_o - \overline{z}_{Fo}} = -\frac{z_i - \overline{z}_{Fi}}{\overline{f}_i} \tag{17.15}$$

even when the object or the image is located in the field. These new cardinal elements, $\overline{z}_{Fo}, \overline{z}_{Fi}, \overline{f}_o$ and $\overline{f}_i$ will in general vary with object position z_o and hence with magnification M. From Eq. (17.15), we have

$$\begin{gathered} \frac{1}{\overline{f}_o} = \frac{d}{dz_o}\left(\frac{1}{M}\right), \quad \frac{1}{\overline{f}_i} = -\frac{dM}{dz_o}\frac{dz_o}{dz_i} \\ \overline{z}_{Fo} = z_o - \overline{f}_o/M, \quad \overline{z}_{Fi} = z_i + M\overline{f}_i \end{gathered} \tag{17.16}$$

and we regard these as *definitions* of new *osculating* cardinal elements, so called from their geometrical interpretation. From the relation between longitudinal and lateral magnification (Eq. 15.60), we note that

$$\frac{\overline{f}_i}{\hat{\phi}_i^{1/2}} = \frac{\overline{f}_o}{\hat{\phi}_o^{1/2}} \tag{17.17}$$

The osculating focal lengths and the positions of the osculating foci may be calculated by considering an arbitrary pair of solutions of the paraxial equation, $s(z)$ and $t(z)$, such that the general solution is

$$x(z) = as(z) + bt(z) = \frac{t(z)}{t(z_o)}x_o + a\left\{s(z) - \frac{s(z_o)}{t(z_o)}t(z)\right\} \tag{17.18}$$

with $t(z_o) \neq 0$, $x_o \coloneqq x(z_o)$ and $b = \{x_o - as(z_o)\}/t(z_o)$. The image plane conjugate to an object plane $z = z_o$ is the plane in which the term multiplied by a vanishes:

$$s(z_i) - \frac{s(z_o)}{t(z_o)}t(z_i) = 0 \tag{17.19}$$

whereupon

$$x(z_i) = \frac{t(z_i)}{t(z_o)} x_o \tag{17.20}$$

The magnification is thus

$$M = \frac{t(z_i)}{t(z_o)} = \frac{s(z_i)}{s(z_o)} \tag{17.21}$$

Differentiating, we find

$$\frac{dM}{dz_i} = \frac{1}{t(z_o)} \frac{dt(z_i)}{dz_i} - \frac{t(z_i)}{t^2(z_o)} \frac{dt(z_o)}{dz_o} \frac{dz_o}{dz_i} \tag{17.22}$$

Eq. (17.19) gives

$$\begin{aligned} \frac{dz_o}{dz_i} &= -\frac{t(z_o)s'(z_i) - s(z_o)t'(z_i)}{t'(z_o)s(z_i) - s'(z_o)t(z_i)} \\ &= -\frac{t(z_o)}{t(z_i)} \cdot \frac{t(z_o)s'(z_i) - s(z_o)t'(z_i)}{t'(z_o)s(z_o) - s'(z_o)t(z_o)} \end{aligned} \tag{17.23}$$

where primes denote differentiation with respect to the argument z_o or z_i. From Eqs (17.16) and (17.22–17.23), we see that

$$\begin{aligned} \frac{1}{\overline{f}_i} &= -\frac{t'(z_o)s'(z_i) - t'(z_i)s'(z_o)}{t(z_o)s'(z_o) - t'(z_o)s(z_o)} \\ \frac{1}{\overline{f}_o} &= -\frac{t'(z_o)s'(z_i) - t'(z_i)s'(z_o)}{t(z_i)s'(z_i) - t'(z_i)s(z_i)} \end{aligned} \tag{17.24}$$

Likewise from Eqs (17.16) and (17.24), we obtain (17.25)

$$\begin{aligned} z_i - \overline{z}_{Fi} &= -\frac{t(z_i)s'(z_o) - t'(z_o)s(z_i)}{t'(z_i)s'(z_o) - t'(z_o)s'(z_i)} \\ z_o - \overline{z}_{Fo} &= -\frac{t(z_o)s'(z_i) - t'(z_i)s(z_o)}{t'(z_o)s'(z_i) - t'(z_i)s'(z_i)} \end{aligned} \tag{17.25}$$

The geometrical meanings of these expressions can be extracted by considering the particular ray

$$\xi(z) = t(z)s'(z_i) - s(z)t'(z_i) \tag{17.26}$$

In the plane $z = z_i$, $\xi'(z_i) = 0$ and so the tangent to $\xi(z)$ is parallel to the axis in this plane. The osculating focal length $\bar{f}_o$ is simply given by

$$\bar{f}_o = \frac{\xi(z_i)}{\xi'(z_o)} \tag{17.27}$$

and the position of the focus by

$$z_o - \bar{z}_{Fo} = \frac{\xi(z_o)}{\xi'(z_o)} \tag{17.28}$$

An osculating object principal plane may be defined to be the point of intersection of the tangent to $\xi(z)$ in the object plane and the tangent in the image plane. It is easy to show that

$$\bar{z}_{Po} - \bar{z}_{Fo} = \bar{f}_o \tag{17.29}$$

Similar relations may be established straightforwardly for the image osculating cardinal elements. These geometrical relations are illustrated in Fig. 17.4A.

These osculating cardinal elements can be used to study image formation in the neighbourhood of a given pair of conjugate points. Consider an object placed in a plane distant Δz_o from z_o; the magnification will change from M to $M + \Delta M$. We see that

$$M + \Delta M = M - \frac{\Delta z_i}{\bar{f}_i} = -\frac{z_i + \Delta z_i - \bar{z}_{Fi}}{\bar{f}_i} \tag{17.30}$$

Also

$$\frac{1}{M} + \Delta\left(\frac{1}{M}\right) = \frac{1}{M} + \frac{\Delta z_o}{\bar{f}_o} \tag{17.31}$$

Hence

$$M + \Delta M = -\frac{z_i + \Delta z_i - \bar{z}_{Fi}}{\bar{f}_i} = \frac{\bar{f}_o}{z_o + \Delta z_o - \bar{z}_{Fo}} \tag{17.32}$$

and so Newton's lens equation is satisfied for the new object position, for small displacements Δz_o.

An example of the variation of the osculating focal length as a function of magnification for the magnetic lens field $B(z) = B_0 \exp(-z^2/a^2)$ is shown in Fig. 17.4B for various values of the lens strength.

From Fig. 17.1, it is immediately clear that the osculating object focus corresponding to an image at infinity is identical with the real object focus discussed in Section 17.1 and the osculating object principal plane likewise coincides with the real object principal plane.

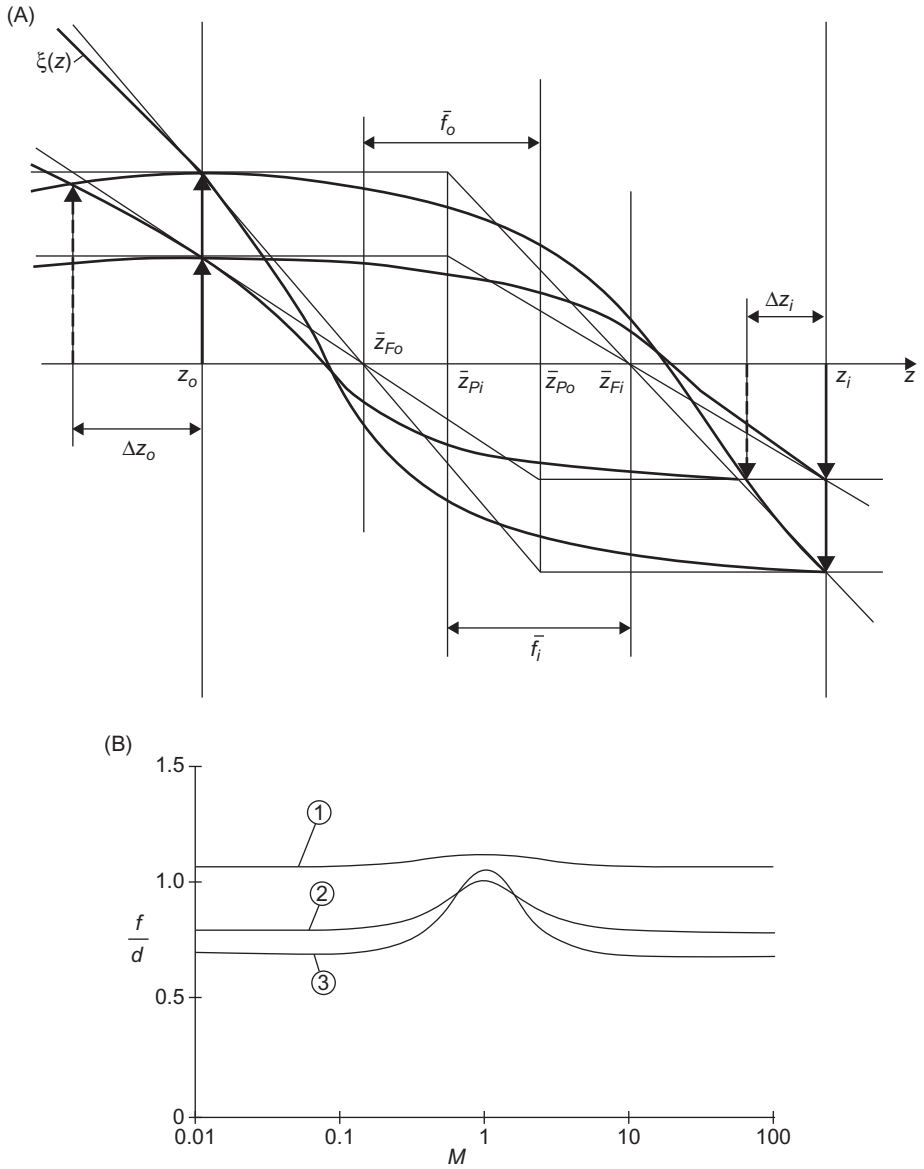


Figure 17.4

(A) The osculating cardinal elements. (B) The osculating focal length as a function of magnification and lens strength for $B(z) = B_0 \exp(-z^2/a^2)$. 1: $J^2/\hat{\phi} = 142$ A^2/V; 2: 285 A^2/V; 3: 427 A^2/V. *After Glaser (1952), Courtesy Springer Verlag.*

Finally, we consider the question of whether field distributions exist for which the osculating cardinal elements are stationary and thus do not vary with object position (or magnification). This question was very thoroughly studied by Glaser (Glaser and Lammel, 1941), and contributions to the topic have also been made by Hutter (1945) and

Funk (1950). The list of fields given by Hutter is misleading, however, as it contains several fields that do not possess fixed osculating elements. The fullest account is to be found in Glaser (1952, Section 68), which we follow closely.

We set out from Eq. (17.15) and render the axial coordinate dimensionless by writing

$$\zeta := \sigma z/a \tag{17.33}$$

where a is some convenient dimension of the lens and σ is defined below. We no longer add bars to ζ where these were necessary for z. Solving Eq. (17.15) for ζ_i we obtain the projective relation

$$\zeta_i = \frac{\zeta_{Fi}\zeta_o - \overline{f}_o\overline{f}_i - \zeta_{Fi}\zeta_{Fo}}{\zeta_o - \zeta_{Fo}} \tag{17.34}$$

and shift the origin to the point midway between F_o and F_i: $\zeta_{Fo} = -\zeta_{Fi}$ Anticipating a result proved in Section 17.3, we set $\zeta_{Fi}^2 - \overline{f}_o\overline{f}_i$ equal to $-\sigma^2$ and Eq. (17.34) becomes

$$\zeta_i = \frac{\zeta_{Fi}\zeta_o - \sigma^2}{\zeta_o + \zeta_{Fi}} \tag{17.35}$$

We now map the whole z- or ζ-axis onto a finite domain by writing

$$\zeta =: \sigma \cot \psi \quad 0 \leq \psi \leq \pi \tag{17.36}$$

and choose ω such that

$$\zeta_{Fi} =: \sigma \cot \frac{\pi}{\omega} \tag{17.37}$$

The projective relation (17.35) then becomes

$$\cot \psi_i = \frac{\cot(\pi/\omega)\ \cot \psi_o - 1}{\cot \psi_o + \cot(\pi/\omega)} = \cot(\psi_o + \pi/\omega) \tag{17.38}$$

or

$$\psi_i = \psi_o + \pi/\omega \tag{17.39}$$

We now substitute these transformations into the paraxial equations of motion; it will be sufficient to consider the magnetic case ($\phi = \text{const}$). Replacing z by ψ (Eqs 17.33 and 17.36) and introducing a new variable ξ,

$$x =: \xi \operatorname{cosec}\ \psi \tag{17.40}$$

the paraxial equation $x'' + (\eta^2 B^2/4\hat{\phi})x = 0$ becomes

$$\ddot{\xi} + \left\{1 + k^2 b^2 (a\ \cot \psi)\ \operatorname{cosec}^4 \psi\right\}\ \xi = 0 \tag{17.41}$$

in which

$$k^2 := \frac{\eta^2 B_0^2 a^2}{4\hat{\phi}} \tag{17.42}$$

and

$$B(z) =: B_0 b(z) \tag{17.43}$$

where B_0 is the greatest value of $B(z)$ so that $b(z) \leq 1$.

Let $s(\psi)$ and $t(\psi)$ be two independent solutions of Eq. (17.41), corresponding to the solutions $x_s(\sigma \cot \psi) = s(\psi) \operatorname{cosec} \psi$, $x_t(\sigma \cot \psi) = t(\psi)\operatorname{cosec} \psi$; then Eq. (17.15) may be written

$$M = \frac{s(\psi_i)}{s(\psi_o)} \frac{\sin \psi_o}{\sin \psi_i} = \frac{t(\psi_i)}{t(\psi_o)} \frac{\sin \psi_o}{\sin \psi_i} \tag{17.44}$$

Setting

$$r(\psi_o) := \frac{s}{t} = \frac{x_s(\sigma \cot \psi)}{x_t(\sigma \cot \psi)} \tag{17.45}$$

Eq. (17.19) becomes

$$r(\psi_o) = r(\psi_i) \tag{17.46}$$

Since $\psi_i = \psi_o + \pi/\omega$ Eq. (17.39), the function $r(\psi)$ must be such that

$$r(\psi_o + \pi/\omega) = r(\psi_o) \tag{17.47}$$

that is, it must be *periodic* with period π/ω. Differentiating Eq. (17.45) gives

$$\dot{r} = \frac{\dot{s}t - s\dot{t}}{t^2} \tag{17.48}$$

and using the Wronskian ($\dot{s}t - s\dot{t} = \text{const} = C$), we see that

$$\dot{r} = C/t^2 \tag{17.49}$$

From this we can deduce that $t(\psi)$ must be semi-periodic,

$$t(\psi + \pi/\omega) = -t(\psi) \tag{17.50}$$

and the same is true of $s(\psi)$. The derivatives of these functions are also semi-periodic and returning to Eq. (17.41), we see that the field function $b^2(a \cot \psi) \operatorname{cosec}^4 \psi$ must be periodic with period π/ω. Writing

$$b^2(a \cot \psi) \operatorname{cosec}^4 \psi = F^2(\psi) \tag{17.51}$$

we can conclude that all 'Newtonian' fields, that is, fields for which the osculating cardinal elements are stationary (do not vary with object position), must be of the form

$$b(a\ \cot\psi) = F(\psi)\sin^2\psi, \quad F(\psi+\pi/\omega) = F(\psi) \tag{17.52}$$

Not all such fields are Newtonian, however: this condition is necessary but not sufficient. Such a condition is provided by Eq. (17.47), which we may restate as follows: the solutions $s(\psi)$ and $t(\psi)$ of the Hill differential equation

$$\ddot{\xi} + \left\{1 + k^2F^2(\psi)\right\}\xi = 0, \quad F(\psi+\pi/\omega) = F(\psi) \tag{17.53}$$

must be semi-periodic with half-period π/ω. (For further details of Hill equations, see Kamke, 1977, Section C.2, Eq. 2.30; Whittaker and Watson, 1927, Chapter XIX). The simplest function $F(\psi)$ satisfying Eq. (17.52) is clearly

$$F(\psi) = 1 \tag{17.54}$$

giving

$$b(a\cot\psi) = \sin^2\psi$$

or

$$B(z) = \frac{B_0}{1+(z/a)^2} \tag{17.55}$$

This field distribution, introduced by Glaser (1940) and very extensively studied, is known as *Glaser's bell-shaped model*.

For more complicated forms of $F(\psi)$, an example of which is

$$\begin{aligned} B(a\cot\psi) &= B_0\, R^{1/2}\cos^2\psi \\ R &= 1 + \frac{1+k^2}{k^2}C_1\frac{1-C_2\sin^2\omega(\pi-\psi)}{\left\{1-C_3\sin^2\omega(\pi-\psi)\right\}^2}\cdot\sin^2\omega(\pi-\psi) \end{aligned} \tag{17.56}$$

where C_1, C_2 and C_3 are constants (some examples are shown in Fig. 17.5), see Glaser (1952, Section 68). We shall say no more about these fields here, however, because the field shape is a function of the parameter $k^2 \propto B_0^2/\hat{\phi}$. Thus any given member of this family of fields will be Newtonian for only one value $B_0^2/\hat{\phi}$: they are of academic interest only. For some values of the parameters, they appear to represent grossly saturated superconducting lenses quite well (see the field distributions in Bonjour (1974, 1975)) but there is little or no incentive to use models of such restricted utility. The attractive feature of model fields is that they enable us to study the general behaviour of a lens as all the

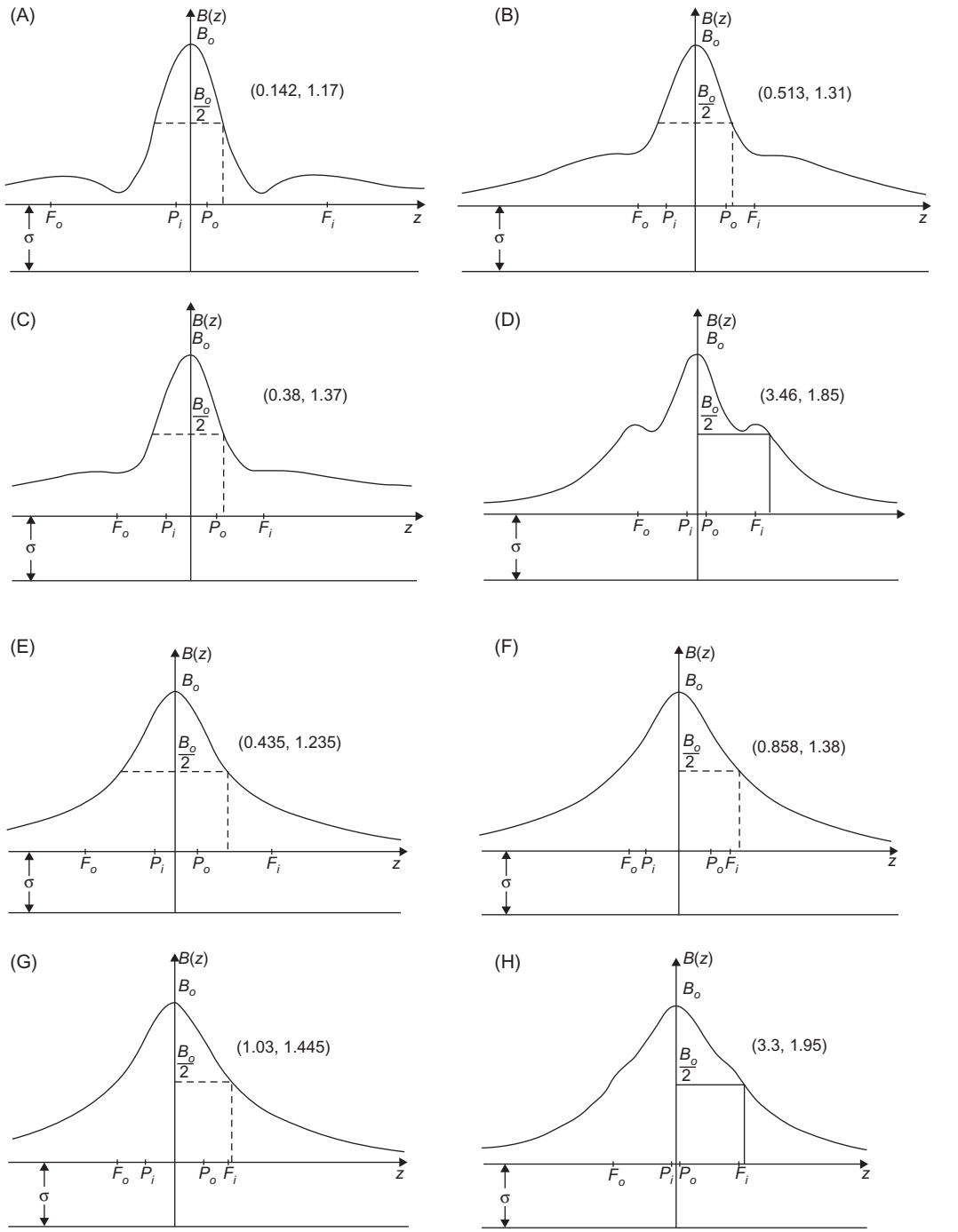


Figure 17.5

Various Newtonian fields of the form (17.56). (A)–(D), $C_1 = -0.724$, $C_2 = 1.049$, $C_3 = 0.559$; (E)–(H), $C_1 = -0.298$, $C_2 = 1.035$, $C_3 = 0.331$. The parameters (k^2, ω) are as indicated. *After Glaser (1952, Courtesy Springer Verlag.*

parameters are varied, even if the numerical values are not exact; if a model is not capable of this, there is even less reason to eschew numerical trajectory tracing (Chapter 34: Numerical Calculation of Trajectories, Paraxial Properties and Aberrations).

Another aspect of Newtonian fields, which we shall not consider here, concerns the existence of osculating cardinal elements of higher order, that is, cardinal elements for which Eq. (17.15) is valid to better than a linear approximation. The quadratic case is mentioned by Glaser and Bergmann (1951) and the general case has been studied by Putz (1951), whose unpublished work is recapitulated in detail by Glaser (1952, Section 68).

17.3 Inversion of the Principal Planes

As we have seen, the principal planes may be defined in several ways. For certain definitions, which will emerge from the discussion, the object and image principal planes are crossed in the sense that the object principal plane is on the image-space side of the image principal plane: $z_{Pi} < z_{po}$. Consider a pair of conjugate planes z_o, z_i; from Eq. (16.25),

$$z_i = \frac{z_{Fi}z_o - f_o f_i - z_{Fo}z_{Fi}}{z_o - z_{Fo}} \tag{17.57}$$

as in Eq. (17.15). Setting the origin of coordinates midway between z_{Fo} and z_{Fi}, so that

$$-z_{Fo} = z_{Fi} =: z_F > 0 \tag{17.58}$$

we find

$$z_i = \frac{z_F z_o - (f_o f_i - z_F^2)}{z_o + z_F} \tag{17.59}$$

This projective transformation must have real or imaginary self-corresponding points, that is, points for which $z_i = z_o =: \zeta$, say. For this

$$\zeta^2 = -f_o f_i + z_F^2 \tag{17.60}$$

If both object and image are real, there can be no self-corresponding points: a ray setting out from the axial point in $z = z_o$ cannot intersect the axis again in this plane. The self-corresponding points are therefore imaginary and

$$f_o f_i - z_F^2 > 0 \tag{17.61}$$

If $f_o = f_i =: f$ we may conclude that

$$z_F < f \tag{17.62}$$

since $z_F > 0$ by hypothesis and the fact that electron lenses are convergent implies that $f > 0$ for single stage imaging. But

$$z_{Pi} = -z_{Po} = z_F - f < 0$$

and so

$$z_{Pi} < z_{Po} \tag{17.63}$$

If $f_o \neq f_i$, we have

$$z_F^2 < f_o f_i \tag{17.64}$$

and since

$$\frac{1}{4}(f_o - f_i)^2 > 0$$

we may write

$$z_F^2 < \frac{1}{4}\left(f_0^2 + 2f_o f_i + f_i^2\right)$$

or

$$0 < z_F < \frac{1}{2}(f_o + f_i) \tag{17.65}$$

or

$$z_{Pi} = z_F - f_i < \frac{1}{2}(f_o - f_i)$$

$$z_{Po} = -z_F + f_o > \frac{1}{2}(f_o - f_{ii})$$

so that finally $z_{Po} > z_{Pi}$ as before Eq. (17.63).

The above proof is valid only if the object and image are real and if the image is formed in a single stage: there must be no intermediate image, for otherwise we could not take the square roots (Eqs 17.62–17.65). This inversion of the principal planes was first found experimentally, by Ruska (1934b). A situation that appears to conflict with this has been found by Sturrock (1951, 1955); we return to this in Part VII, where we examine field models.

17.4 Approximate Formulae for the Cardinal Elements: The Thin-Lens Approximation and the Weak-Lens Approximation

Simple expressions for the focal lengths and positions of the foci can be obtained if the lens field is *weak*. We shall find that all weak lenses can be treated as thin lenses, that is, as

lenses in which the principal planes coincide when $\phi_i = \phi_o$, but the converse is not necessarily true; when $\phi_i \neq \phi_o$, we are obliged to define thin lenses slightly differently. We set out from the reduced equation (15.38) in the form $v'' = -Gv$ and integrate twice to give

$$v'(z) = a_0' - \int_{z_0}^{z} Gv \; d\zeta := a_0' - G_1(z)$$

$$v(z) = a_0 - a_0'(z - z_0) - \int_{z_0}^{z} G_1(\zeta)\, d\zeta$$

which may be integrated by parts to yield

$$v(z) = a_0 + a_0'(z - z_0) - \int_{z_0}^{z} (z - \zeta) G(\zeta) v(\zeta)\, d\zeta \tag{17.66}$$

which is a Volterra-type integral equation. A formal solution is given by the Picard–Lindelöff iterative procedure

$$v_{n+1}(z) = a_0 + a_0'(z - z_0) - \int_{z_0}^{z} (z - \zeta) G(\zeta) v_n(\zeta)\, d\zeta \qquad \text{for} \quad n = 0, \; 1, \ldots \tag{17.67}$$

Suppose now that

$$\frac{1}{\tilde{f}} := \int_{-\infty}^{\infty} G(z)\, dz \ll \frac{1}{L} \tag{17.68}$$

where L is some length characteristic of the lens producing the field $\phi(z)$ or $B(z)$, typically a gap or bore or total length. For convenience, we set the origin at the 'centre of gravity' of $G(z)$, so that its first moment vanishes:

$$\int_{-\infty}^{\infty} zG(z)\, dz = 0 \tag{17.69}$$

We denote the second moment or 'moment of inertia' of $G(z)$ by D:

$$D := \int_{-\infty}^{\infty} z^2 G(z)\, dz \tag{17.70}$$

Clearly

$$D \ll L \ll \tilde{f} \tag{17.71}$$

Let us now evaluate $v_1(z)$ from Eq. (17.67), taking as zero-order approximation $v_0(z) = a_0 + a_0'(z - z_0)$. In field-free object space, we have

$$v_1(z) = a_0 + a_0'(z - z_0) \tag{17.72}$$

and in field-free image space,

$$v_1(z) = a_0 + a_0'(z - z_0) + \frac{(a_0' z_0 - a_0)z}{\tilde{f}} + a_0' D \tag{17.73}$$

Neglecting the final term, we see that incident and emergent rays intersect in the centre-of-gravity plane, $z = 0$. The focal lengths and foci are easily found by expressing the above relation between asymptotes in matrix form and comparing with Eq. (16.12). In the plane $z_1 = z_2 = 0$ (Dušek matrix), we have

$$\begin{aligned} \begin{pmatrix} v_1^{(i)}(0) \\ v_1^{(i)'} \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ -1/\tilde{f} & 1 \end{pmatrix} \begin{pmatrix} v_1^{(0)}(0) \\ v_1^{(i)'} \end{pmatrix} \\ &\equiv \begin{pmatrix} pz_F/f_i & p\left(f_o + z_{Fo}z_{Fi}/f_i\right) \\ -p/f_i & -pz_{Fo}/f_i \end{pmatrix} \begin{pmatrix} v_1^{(0)}(0) \\ v_1^{(i)'} \end{pmatrix} \end{aligned} \tag{17.74}$$

where $p = (\hat{\phi}_i/\hat{\phi}_o)^{1/4}$, so that

$$\begin{aligned} f_i &= p\tilde{f}, \quad f_o = -z_{Fo}/p = \tilde{f}/p \\ z_{Fi} &= f_i/p = \tilde{f}, \quad z_{Fo} = -f_i/p = -\tilde{f} \end{aligned} \tag{17.75}$$

If $\phi_i = \phi_o$ ($p = 1$), the lens behaves like a thin lens situated at the centre of gravity of $G(z)$, the principal planes Eq. (16.7) coinciding in this plane as expected. If, however, $\phi_i \neq \phi_o$, the object and image focal lengths are related to $\tilde{f}$ explicitly as follows

$$f_o = \left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/4} \tilde{f}, \quad f_i = \left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{1/4} \tilde{f} \tag{17.76}$$

and so $f_o f_i = \tilde{f}^2$: the intermediate quantity $\tilde{f}$ (Eq. 16.19a) is the geometric mean of the object and image focal lengths. The foci are equidistant from the centre-of-gravity plane: $z_{Fi} = -z_{Fo} = \tilde{f}$. The principal planes do not, however, coincide in this plane but are separated by a distance

$$\begin{aligned} z_{Pi} - z_{Po} &= z_{Fi} - f_i - z_{Fo} - f_o = 2\tilde{f} - f_i - f_o \\ &= \tilde{f}\left\{2 - \left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{1/4} - \left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/4}\right\} \end{aligned} \tag{17.77}$$

Despite this, the lens is still regarded as thin and we have therefore proved that every weak electron lens can be regarded as a thin lens, where the criterion for weakness is Eq. (17.68).

In the cases of magnetic lenses and electrostatic lenses, we obtain the following expressions:

Magnetic Lenses

We set $\phi = \text{const}$ in $G(z)$ and find

$$\frac{1}{f} = \frac{1}{\tilde{f}} = \frac{\eta^2}{4\hat{\phi}} \int_{-\infty}^{\infty} B^2(z)\ dz \tag{17.78}$$

This expression, first derived by Busch (1927), is known as *Busch's formula* (cf. Ollendorff and Wendt, 1932).

Electrostatic Lenses

We now set $B = 0$ in $G(z)$ and obtain

$$\begin{aligned} \frac{1}{f_o} &= \frac{3}{16}\left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{1/4} \int_{-\infty}^{\infty} \left(\frac{\phi'}{\hat{\phi}}\right)^2 \left(1 + \frac{4}{3}\varepsilon\hat{\phi}\right) dz \\ \frac{1}{f_i} &= \frac{3}{16}\left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/4} \int_{-\infty}^{\infty} \left(\frac{\phi'}{\hat{\phi}}\right)^2 \left(1 + \frac{4}{3}\varepsilon\hat{\phi}\right) dz \end{aligned} \tag{17.79}$$

or nonrelativistically ($\varepsilon\phi << 1,\ \hat{\phi} \to \phi$):

$$\begin{aligned} \frac{1}{f_o} &= \frac{3}{16}\left(\frac{\phi_i}{\phi_o}\right)^{1/4} \int_{-\infty}^{\infty} \frac{\phi'^2}{\phi^2} dz \\ \frac{1}{f_i} &= \frac{3}{16}\left(\frac{\phi_o}{\phi_i}\right)^{1/4} \int_{-\infty}^{\infty} \frac{\phi'^2}{\phi^2} dz \end{aligned} \tag{17.80}$$

We note that it is important to derive Eq. (17.79) or (17.80) from the reduced equation. If ordinary coordinates (x, y, z) are employed, it is easy to obtain the wrong result; this point is discussed by Sturrock (1955, p. 15ff.), who shows how the confusion arises.

CHAPTER 18

Electron Mirrors

18.1 Introduction

In the discussion of the paraxial properties of rotationally symmetric systems, we have assumed that not only do the electron trajectories remain in the vicinity of the axis but that their gradients also remain small. If, however, the potential barrier in a retarding electrostatic lens is sufficiently high to prevent electrons from passing, the latter will be returned towards object space and the lens will have a mirror action. We shall see that to a first approximation, which it is convenient to call the paraxial approximation, the familiar cardinal elements can again be used for such *contracurrent* or *catoptric* systems, but special precautions must clearly be taken in the vicinity of the turning point, where the gradient is locally very large, passing through infinity.

Several ways of circumventing this difficulty have been proposed, two of which have been studied in detail. In each case, the distance along the optic axis, z, is replaced by some other independent variable, preferably one which, unlike z, increases monotonically as the electron proceeds and hence does not lead to infinite gradients when the electron turns round. The most obvious choice is the time, or a quantity very closely related to it (Recknagel, 1936, 1937; Hottenroth, 1936, 1937; Nicholl, 1938; Zworykin et al., 1945; Regenstreif, 1951; Glaser, 1952; Septier, 1953; Schiske, 1957). Another possibility is to use a transformation introduced by Hahn (1965) as a means of unifying lens studies, in which the entire z-axis is mapped onto a finite region; its suitability for the study of mirrors is pointed out in Hahn (1971). Another suggestion comes from Bernard (1952), who writes $s^2 = -z$, with the origin of z at the point of reflection of an electron on the axis. The incident part of the trajectory then corresponds to $s < 0$ and the reflected part to $s > 0$; dx/ds and dy/ds do not become infinite.

A transformation that has been extensively studied employs *Cartesian* coordinates but in such a way that the 'paraxial' equations remain valid. It is still required that the charged particles remain close to the axis but not that their gradients remain small (Kel'man et al., 1971a–c, 1972a,b, 1973a,b; Daumenov et al., 1978; Ximen et al., 1983). We present this theory, which has been developed in considerable detail, in Section 18.3.

We assume throughout this chapter that although the z-component of the electron velocity is reduced to zero by the potential barrier in the mirror, the same is not true of the retarding

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00018-3
© 2018 Elsevier Ltd. All rights reserved.

field: $E_z \neq 0$ in the plane in which the electron is instantaneously stationary. In the exceptional case in which this did occur, the electrons would be trapped; this remote contingency is envisaged by Recknagel (1937).

Electron mirrors are not the only systems in which the gradients of electron trajectories become too large for the conditions of paraxial imagery to be satisfied. In electron emission structures – guns, cathode lenses and image converters in particular – the electron velocity is locally small and the gradients steep. Guns are dealt with separately in Part IX. The methods of Kel'man et al. and Hahn have been used to study cathode lenses in some detail and any technique that enables mirrors to be analysed is in principle equally suitable for cathode lenses, in which the rays resemble those returning from a mirror except that the equipotentials are determined by the cathode surface. We now say a few words about the instrumental aspects of these various devices.

Electron mirrors are used for reflecting the particle beam in certain types of energy analyser and mass spectrometer. A typical example of the former is the analyser built by Castaing and Henry (1962), which is shown in Fig. 18.1. Here, the direction of the beam is reversed by a repulsive electrostatic field, thus causing the particles to pass twice through

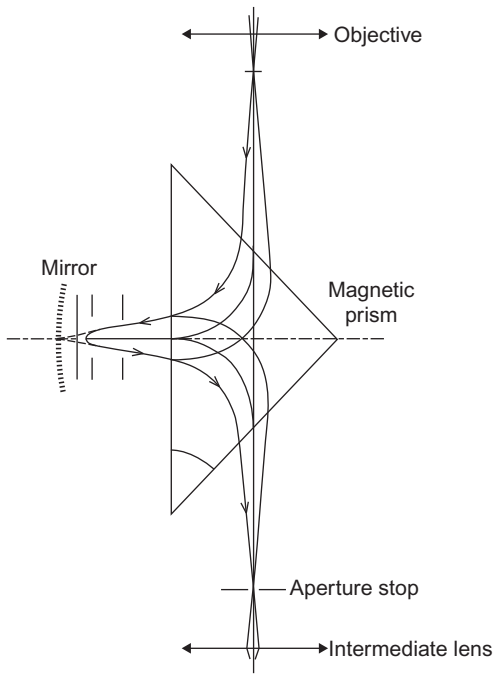


Figure 18.1
Combination of electrostatic mirror and magnetic prism that permits energy-filtered images or energy-loss spectra to be formed.

the magnetic deflection field. The points of reflection are located somewhere in the field, away from any material surface.

Electron mirrors are also the essential elements of mirror or reflection electron microscopes. Here the beam direction is reversed just in front of the surface of the specimen, and contrast is produced in the image by local variations of some property of the specimen surface: variation in height or magnetic field distribution or electrostatic potential. For a detailed review, see Bok et al. (1971). An extensive list of publications in which mirror optics is considered is included in (Hawkes, 2012). Mirrors are also essential components in low-energy electron microscopes and photoemission electron microscopes, which are described at length in Bauer (2008, 2012, 2014, 2018) and Feng and Scholl (2008, 2018).

In another device (Lichte, 1983), the electron mirror is used as an interferometer (Fig. 18.2). Here, the electrons come so close to the material surface that the reflected beam is modulated by any roughness, as in the mirror microscope. The height distribution over

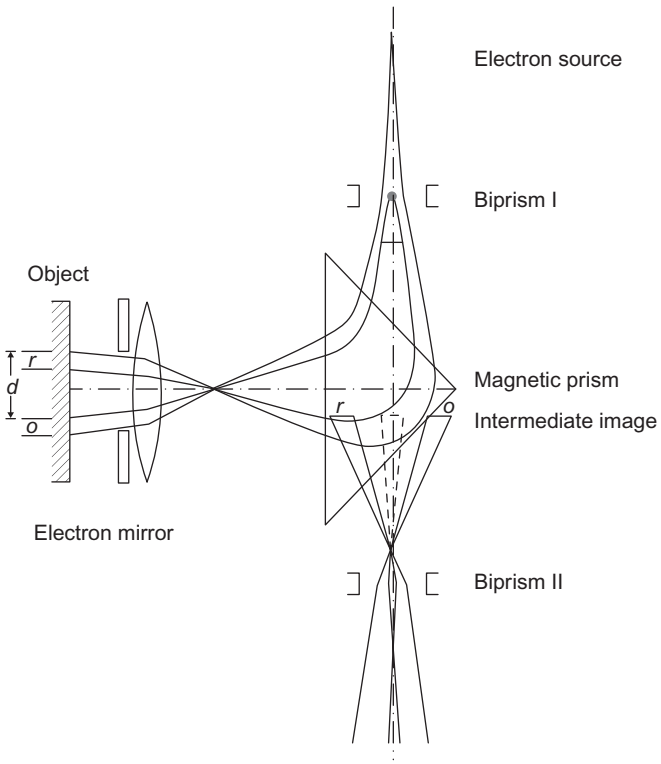


Figure 18.2
The arrangement of prism, mirror and biprism in an electron mirror interference microscope. The presence of the two biprisms enables phase differences between the partial waves *r* and *o* to be measured interferometrically. *After Lichte (1983) Courtesy Physical Society of Japan.*

the surface is now determined by analysing the interference pattern obtained by superposing a uniform beam on the reflected beam. Since this effect cannot be fully understood without going beyond geometrical electron optics, we defer further discussion to Volume 3.

Cathode lenses are employed in *electron emission microscopes* and in *image-converters*. Emission microscopes are mainly employed in metallurgical investigations. A typical cathode lens is shown schematically in Fig. 18.3. Here, the specimen acts as a (cold) cathode and electrons are ejected from it by lateral bombardment with electrons (Möllenstedt and Düker, 1953), photons (Koch, 1967a,b) or ions. The local intensity of the ensuing electron emission depends to some extent on the intensity of the irradiation, which is assumed to be practically uniform over the area illuminated, and depends principally on material properties such as the work function. The electrons emitted are then used to form an image of the cathode surface on a recording medium, where the distribution of the properties that determined the image is now observable. Here, therefore, the cathode plays two roles: first, it serves as an electron emitter, and in addition, it forms the first electrode of the electrostatic lens in front of it. Such a lens is often known as an 'immersion objective'. For surveys, see Möllenstedt and Lenz ((1963), Müllerová and Frank (2003), the series of articles by Tromp (2011, 2015a, b), Tromp et al. (2010, 2012, 2013), Geelen et al. (2015) and the book by Bauer (2014). Another use of cathode lenses is in the rapidly growing area of ultrafast electron microscopy (UEM), in which extremely short electron

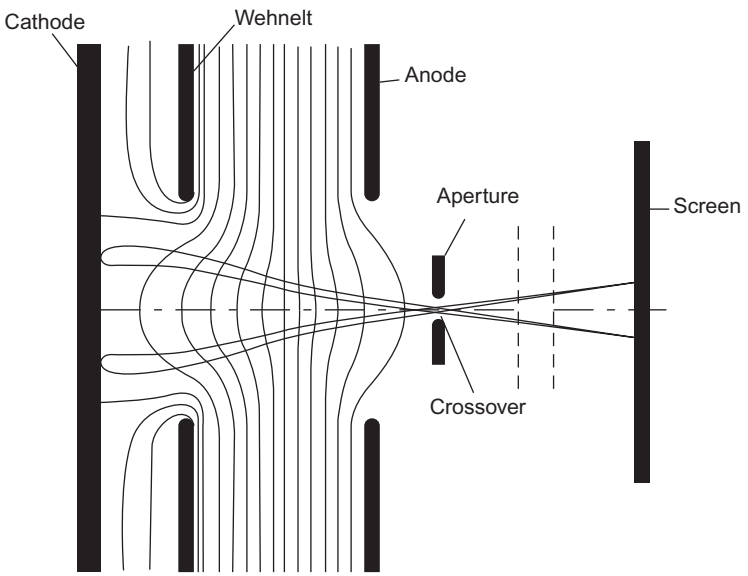


Figure 18.3
Electrostatic equipotentials and electron trajectories in an electron microscope with a cathode lens.

pulses are generated by bombarding the cathode with pulses of photons or electrons (or ions). This makes it possible to track rapidly changing features of the specimen either by imaging or diffraction. Fragile specimens could perhaps be imaged before being damaged by the beam. We give some account of this in Chapter 37 of Volume 2.

In *image converters* (Fig. 18.4), the function of the cathode is similar. Here, the cathode is a thin spherical layer, transparent to infrared or visible light. The inner surface of the cathode is coated with a layer of conducting material with a low work function so that when light falls on the cathode, photoelectrons are emitted. By imaging the cathode surface onto a viewing screen with electrons of sufficiently high energy, the feeble light image projected onto the cathode surface can be intensified. These introductory remarks are merely intended to give the reader a general idea of these devices; a few more technical details will be found in Chapter 37 of Volume 2.

The two methods of analysing the properties of mirrors that have been examined in most detail are the modified temporal representation (Section 18.2) and the cartesian representation (Section 18.3). In the first, the independent varriable of the equations of motion is a time-like quantity while in the other, time is eliminated from the outset. This work was briefly described in the first edition but with the growing use of electron mirrors in low-energy-electron microscopes (LEEM) and photoemission electron microscopes (PEEM), we now include a more complete description (see Chapter 37 of Volume 2). Aberration coefficients have been derived by means of both forms of the theory; when the temporal representation is chosen, an additional step is needed to extract aberration coefficients corresponding to position and angle from those involving time. More details are given in Chapter 28, The Aberrations of Mirrors and Cathode Lenses.

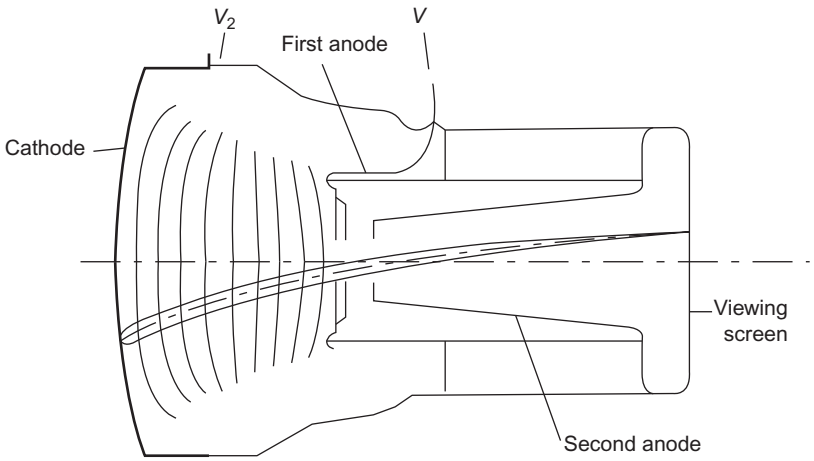


Figure 18.4
An electrostatic three-electrode image converter for infrared light.

In both cases, the position of an arbitrary electron is measured as usual by its distance from the optic axis and by its distance, parallel to the optic axis, from a *reference electron* at a given time. The latter travels along the optic axis in the negative z direction towards the mirror, is reversed and returns along the optic axis in the positive z direction. The point of reflection, at which $dz/dt = 0$, is denoted by ζ_T, and we choose this point as the origin of the electrostatic potential:

$$\Phi(0, 0, \zeta_T) = 0 \tag{18.1}$$

Thus the coordinates of the reference electron at time t are

$$x(t) = 0, \quad y(t) = 0, \quad z(t) =: \zeta(t) \tag{18.2a}$$

and for an arbitrary electron,

$$x(t), \quad y(t), \quad z(t) = \zeta(t) + h(t) \tag{18.2b}$$

The ζ- component of the Lorentz equation (2.1) for the reference electron is

$$\ddot{\zeta} = \frac{\phi'}{2\phi_0}$$

and on multiplying this by $\dot{\zeta}$ and integrating with respect to time, we find

$$\dot{\zeta}^2 = \frac{\phi}{\phi_0} \tag{18.3}$$

where ϕ_0 is the accelerating voltage and dots indicate differentiation with respect to time. Outside the mirror field, $\dot{\zeta} = \pm 1$. It will prove convenient to use the time τ (3.9, 3.10) instead of t; we recall that $d\tau = 2\eta\phi_0^{1/2}dt$.

Note that in this chapter, we neglect relativistic effects, which are rarely of importance in mirror systems. If required, a relativistically correct treatment of the temporal representation can be found in the work of Preikszas (1995; Preikszas and Rose, 1997), reproduced by Rose (2012, Chapter 10). See also Preikszas and Rose (1994, 1996a,b).

18.2 The Modified Temporal Representation

After a short calculation (given in full by Preikszas and Rose, 1997), the following equation for $u = x + iy$ in the rotating coordinate system is obtained:

$$\ddot{u} + \frac{\phi'' + \eta^2 B^2}{4\phi_0} u = F_u \tag{18.4a}$$

The longitudinal coordinate h satisfies

$$\ddot{h} - \frac{\phi''}{2\phi_0} h = F_h \tag{18.4b}$$

$$F_u = F_u^{(e)} + F_u^{(m)} \tag{18.5a}$$

in which

$$\begin{aligned} F_u^{(e)} &= \frac{1}{32}\frac{\phi^{(4)}}{\phi_0} u(uu^*) - \frac{1}{4}\frac{\phi'''}{\phi_0} uh - \frac{1}{8}\frac{\phi^{(4)}}{\phi_0} uh^2 \\ F_u^{(m)} &= \mathrm{i}\hat{\eta}\left\{ -\frac{1}{4}B''\dot{u}(uu^*) + B'\dot{u}h + \frac{1}{2}B''\dot{u}h^2 + \frac{1}{2}B'u\dot{h} + \frac{1}{2}B''uh\dot{h} \right\} \\ &\quad + \frac{1}{8}\hat{\eta}^2 u\left(uu^*BB'' - 4hBB' - 2h^2BB''\right) \\ &\quad + \mathrm{i}\dot{\zeta}\left\{ -\frac{1}{16}B'''u(uu^*) + \frac{1}{2}B''uh + \frac{1}{4}B'''uh^2 \right\} \end{aligned} \tag{18.5b}$$

and

$$\begin{aligned} F_h &= F_h^{(e)} + F_h^{(m)} \\ F_h^{(e)} &= -\frac{1}{8}\frac{\phi'''}{\phi_0} uu^* + \frac{1}{128}\frac{\phi^{(5)}}{\phi_0}(uu^*)^2 - \frac{1}{8}\frac{\phi^{(4)}}{\phi_0}(uu^*)h + \frac{1}{4}\frac{\phi'''}{\phi_0} h^2 \\ F_h^{(m)} &= \frac{1}{4}\hat{\eta}(B' + hB'')\left\{\mathrm{i}(\dot{u}u^* - u\dot{u}^*) - \hat{\eta}uu^*B\right\}. \end{aligned} \tag{18.6}$$

As usual, primes denote derivatives with respect to z or ζ and $\hat{\eta}$ denotes $\eta/\phi_0^{1/2}$.

The Gaussian approximation is obtained by setting $F_u = F_h = 0$. We seek a symmetric and an antisymmetric solution of (18.4a), $p(\tau)$ and $q(\tau)$ respectively. These must satisfy the following conditions at the point of reversal, τ_T:

$$\begin{aligned} p(\tau_T) &= -1, \quad \dot{p}(\tau_T) = 0 \\ q(\tau_T) &= 0, \quad \dot{q}(\tau_T) = 1 \end{aligned} \tag{18.7}$$

and the Wronskian is thus

$$\dot{p}q - p\dot{q} = 1$$

The general form of these solutions is sketched in Figs. 18.5A and B; in this parametric representation, there is no singularity or ambiguity.

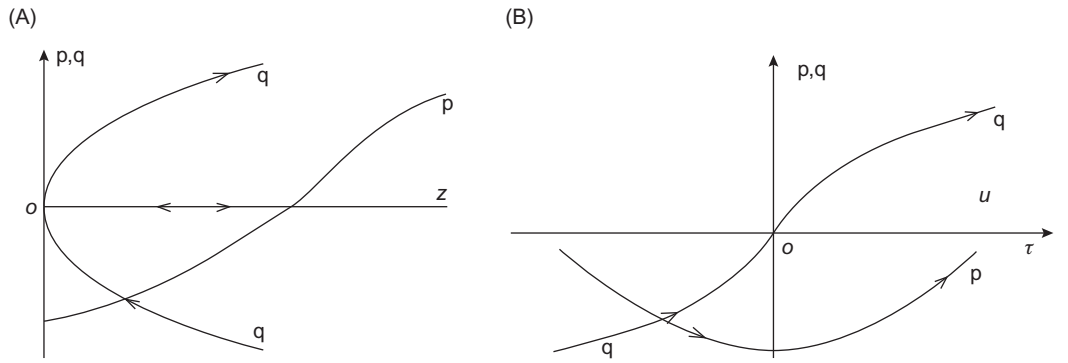


Figure 18.5
Appearance of the fundamental solutions u and v (A) as functions of the axial coordinate z and (B) in parametric form together with $z(\sigma)$.

We denote the symmetric and antisymmetric solutions of (18.4b) by h_σ and h_ν. For h_ν, we have

$$h_\nu = \dot{\zeta} = \pm\left(\frac{\phi}{\phi_0}\right)^{1/2} \tag{18.8}$$

in which the plus sign corresponds to $\tau \geq \tau_T$ and the minus sign to $\tau < \tau_T$. With

$$\dot{h}_\sigma h_\nu - h_\sigma \dot{h}_\nu = 1$$

we see that

$$h_\sigma = h_\nu \int_{\overline{\tau}}^{\tau} \frac{d\tau}{h_\nu^2} = \dot{\zeta} \int_{\overline{\tau}}^{\tau} \frac{\phi_0}{\phi} d\tau \tag{18.9}$$

The lower limit $\overline{\tau}$ is chosen in such a way that $h_\sigma(\tau)$ is indeed symmetric: $\dot{h}_\sigma(\tau_T) = 0$. Outside the region of the mirror field, we have

$$\dot{h}_\sigma = h_\nu = \pm 1, \quad \dot{h}_\nu = 0 \tag{18.10}$$

We now investigate image formation by an electron mirror. Consider an object plane, situated at $z = z_o \geq 0$. First, we have to determine the corresponding parameter $\tau_o \leq 0$ from the inverse function $\tau = \tau(z)$; there is always exactly one such solution. We next introduce new fundamental solutions $G(\tau)$, $H(\tau)$ adapted to the particular object plane. These must satisfy equations of the form of (18.4a with $F_u = 0$) with the initial conditions

$$G(\tau_o) = \dot{H}(\tau_o) = 1, \quad \dot{G}(\tau_o) = H(\tau_o) = 0 \tag{18.11}$$

instead of Eq. (18.7). The linear relations between G, H and p, q are given by

$$G(\tau) = \dot{p}(\tau_o)q(\tau) - \dot{q}(\tau_o)p(\tau) \tag{18.12a}$$

$$H(\tau) = -p(\tau_o)q(\tau) + q(\tau_o)p(\tau) \tag{18.12b}$$

The Wronskians of both pairs of solutions are equal to unity, as can easily be verified.

The particular trajectory with the initial velocity components

$$\dot{x}(\tau_o) \coloneqq \dot{x}_o, \quad \dot{y}(\tau_o) \coloneqq \dot{y}_o, \quad \dot{z}(\tau_o) \coloneqq -\sqrt{\phi(z_o)/\phi_0} \tag{18.13a}$$

at the starting point

$$x(\tau_o) \coloneqq x_o, \quad y(\tau_o) \coloneqq y_o, \quad z(\tau_o) \coloneqq z_o \tag{18.13b}$$

is given by $z = z(\tau)$ and

$$x(\tau) = x_o G(\tau) + \dot{x}_o H(\tau) \tag{18.14a}$$

$$y(\tau) = y_o G(\tau) + \dot{y}_o H(\tau) \tag{18.14b}$$

Just as in the case of ordinary round lenses the position of the conjugate image plane is determined by the next zero of the function $H(\tau)$: $H(\tau_i) = 0$ with $\tau_i > \tau_o$. Usually the focusing fields are so weak that $\tau_i > 0$, which means that the image is formed after the reflection. The position of the image plane is then given by $z_i = z(\tau_i)$ and the lateral image coordinates are

$$x_i = x_o G(\tau_i), \quad y_i = y_o G(\tau_i) \tag{18.15}$$

The image is rotated relative to the object; the corresponding angle is

$$\theta_i = \frac{\eta}{2\phi_0^{1/2}} \int_{\tau_o}^{\tau_i} B(z(\tau))d\tau \tag{18.16}$$

From this formula, which is quite unambiguous, it is obvious that the sense of rotation is maintained after the reflection.

Image formation by a cathode lens is a special case of the foregoing situation. In this case the object is located in the cathode plane $z_o = 0$; hence $\tau_o = 0$ and consequently $G(\tau) \equiv q(\tau)$, $H(\tau) \equiv p(\tau)$, as is obvious on comparing Eq. (18.7) with (18.11). Other than this simplification, there is nothing new about image formation by a cathode lens. A narrow aperture is often brought into the beam in order to filter out electrons with very large lateral velocity components. In this way the chromatic aberration of the cathode lens can be reduced, but this can only be fully understood after studying the aberrations. Here we merely point out that the appropriate plane for this aperture is given by

$$z_a \coloneqq z(\tau_a) \text{ with } q(\tau_a) = 0 \tag{18.17a}$$

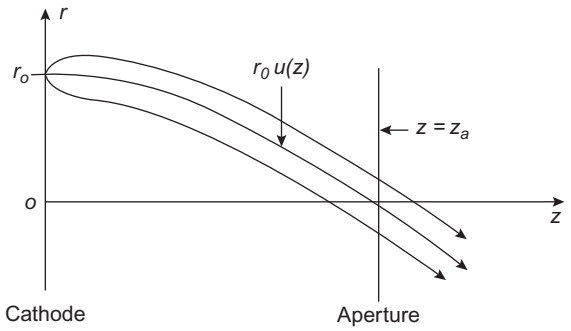


Figure 18.6
Appropriate location of an aperture in a cathode lens.

as sketched in Fig. 18.6. The lateral coordinates

$$x_a = \dot{x}_o p(\tau_a), \quad y_a = \dot{y}_o p(\tau_a) \tag{18.17b}$$

are then independent of x_o and y_o, so that there is no vignetting.

We could go on to define asymptotic cardinal elements – foci, principal planes and focal lengths – exactly as for round lenses. If we adopt the same sign conventions in the definitions, namely $z_{Fi} - z_{Pi} =: f_i$ and $z_{Po} - z_{Fo} =: f_o$ (Eq. 16.7), we find that now $f_i = -f_o =: f$, so that with (Eq. 16.25)

$$\begin{aligned} z_i - z_{Fi} &= -f_i M \\ z_o - z_{Fo} &= f_o / M \end{aligned} \tag{18.18}$$

we have

$$(z_i - z_{Fi})(z_o - z_{Fo}) = -f_o f_i = f^2 \tag{18.19a}$$

or

$$\frac{1}{z_o - z_{Po}} + \frac{1}{z_i - z_{Pi}} = \frac{1}{f} \tag{18.19b}$$

If we compare this with the treatment of mirrors in light optics (e.g., Born and Wolf, 1959 Section 4.3, where the sign convention for f_i is the opposite of that adopted here), we see that convergent mirrors correspond to positive values of f_o and hence negative values of f; for divergent mirrors, $f_o < 0$ and hence $f > 0$. We return to mirror optics in Section 18.3.

18.3 The Cartesian Representation

Although the trajectories reverse their direction and have large gradients in the vicinity of the turning point, a 'paraxial' ray equation can nevertheless be derived in the conventional

Cartesian form. This approach has been investigated by Kel'man et al., who first used it to facilitate the study of cylindrical mirrors (Kel'man et al., 1971a–c, 1972a,b) and subsequently employed it to analyse rotationally symmetric mirrors including magnetic fields (Kel'man et al., 1973a,b). Their theory can be derived from a variational principle, as shown in detail by Daumenov et al. (1978); we shall return to this in Chapter 28, The Aberrations of Mirrors and Cathode Lenses, in connection with the aberrations. This theory was redeveloped by Sekunova and Yakushev (1975), Sekunova (1977), Daumenov et al., (1981), Kolesnikov and Monastyrskii (1988), Monastyrskii (1989), Bimurzaev et al. (1999, 2003, 2004a–c) and Bimurzaev and Yakushev (2004); connected accounts are to be found in Yakushev and Sekunova (1986) and Yakushev (2013), which we follow closely.

We set out from the same equations as for the temporal representation but instead of the time-like quantity τ, we use the parameter ζ (Eq. 18.3) as independent variable. With $z = \zeta + h$, and noting that

$$\frac{d}{d\tau} = \dot{\zeta}\frac{d}{d\zeta}, \qquad \frac{d^2}{d\tau^2} = \ddot{\zeta}\frac{d}{d\zeta} + \dot{\zeta}^2\frac{d^2}{d\zeta^2} \tag{18.20}$$

we obtain

$$\phi(\zeta)u''(\zeta) + \frac{1}{2}\phi'(\zeta)u'(\zeta) + \frac{1}{4}\left\{\phi''(\zeta) + \eta^2 B^2(\zeta)\right\}u(\zeta) = S^{(3)} \tag{18.21a}$$

$$2\phi h' - \phi' h = s^{(3)} \tag{18.21b}$$

in which primes denote differentiation with respect to ζ. The non-linear terms have the form

$$\begin{aligned} S^{(3)} &= \frac{\phi^{iv} + 4\eta^2 BB''}{32}(uu^*)u - \frac{(\phi'' + \eta^2 B^2)'}{4}uh \pm \mathrm{i}\frac{\eta\phi^{1/2}B''}{16}(uu^{*\prime} - u^*u')u \\ s^{(3)} &= -\phi u'u^{*\prime} - \frac{\phi'' + \eta^2 B^2}{4}uu^* \pm \mathrm{i}\frac{\eta\phi^{1/2}B}{2}(uu^{*\prime} - u^*u') + \hat{\varepsilon} \end{aligned} \tag{18.22}$$

Where the reference particle has zero energy (at its point of reflection), the arbitrary electron has energy $\hat{\varepsilon}$.

The paraxial behaviour is obtained by setting $S^{(3)} = s^{(3)} = 0$. These paraxial ray equations have the same formal structure as those for ordinary round lenses. The novel fact here is that very small values are required only for $x^2 + y^2$, whereas the gradients may be arbitrarily large. The range of validity of the paraxial approximation has clearly been extended. Nevertheless, Eq. (18.21) also have disadvantages: great care must be taken in numerical work in the vicinity of the singularity; it is also necessary to be vigilant concerning the signs of the variables before and after reflection.

Since Eq. (18.21a) has a singularity (a first-order pole) at the point at which $\phi(\zeta_T) = 0$, the linearly independent solutions are of the form $p(\zeta)$ and $q(\zeta) = \phi^{1/2} g(\zeta)$, in which it is $p(\zeta)$ and $g(\zeta)$ that are analytic functions; the latter satisfies the equation

$$\phi g'' + \frac{3}{2}\phi' g' + \frac{3\phi'' + \eta^2 B^2}{4} g = 0 \tag{18.23}$$

and

$$\phi^{1/2}(pq' - p'q) = \frac{1}{2}\phi'_T \tag{18.24}$$

The functions $p(\zeta)$ and $g(\zeta)$ are chosen so that

$$p(\zeta_T) = g(\zeta_T) = 1 \tag{18.25}$$

and in the neighbourhood of ζ_T, therefore, we have

$$p'_T = -\frac{\phi''_T + \eta^2 B_T^2}{2\phi'_T}, \quad g'_T = -\frac{3\phi''_T + \eta^2 B_T^2}{6\phi'_T} \tag{18.26}$$

The general solution thus has the form

$$u(\zeta) = \alpha p(\zeta) + \beta q(\zeta) \tag{18.27}$$

When the object plane ($\zeta = \zeta_o$) lies in field-free space, a ray passing through the point u_o in $\zeta = \zeta_o$ with gradient u'_o may be written

$$u(\zeta) = \frac{p(\zeta)}{R p'_o}\left(u_o - u'_o \frac{q_o}{q'_o}\right) \pm \frac{q(\zeta)}{R g'_o}\left(u_o - u'_o \frac{p_o}{p'_o}\right) \tag{18.28}$$

The double sign again indicates incident (minus) and reflected (plus). We now consider some specific rays.

When $u'_o = p'_o u_o / p_o$, the second term vanishes and the incident ray returns along the same path. It will thus intersect the optic axis at the centre of curvature of the mirror, $\zeta = \zeta_c$. Outside the field of the mirror, the ray $p(\zeta)$ tends to the asymptote

$$p(\zeta) \to \frac{\zeta - \zeta_c}{p'_o} \tag{18.29}$$

When $u'_o = q'_o u_o / q_o$, the first term in Eq. (18.28) vanishes and the resulting ray will be reflected symmetrically about the optic axis. The incident and reflected asymptotes will therefore intersect at the (virtual) vertex of the mirror, ζ_v:

$$q(\zeta) \to (\zeta - \zeta_v) q'_o \tag{18.30}$$

We see that the quantity R, hitherto undefined, may be identified with the radius of curvature of the mirror:

$$R = \frac{p_o}{p'_o} - \frac{q_o}{q'_o} = \zeta_v - \zeta_c \tag{18.31}$$

and Eq. (18.24) shows that

$$R = \frac{\phi'_T}{2p'_o q'_o \phi_o^{1/2}} \tag{18.32}$$

The focal length (f) and position of the focus (ζ_F) can be deduced from the expression for a ray incident parallel to the optic axis:

$$u = \frac{u_o}{R}\left(\frac{p(\zeta)}{p'_o} \pm \frac{q(\zeta)}{q'_o}\right) \tag{18.33}$$

from which we see that

$$f = R/2 \tag{18.34}$$

and

$$\zeta_F = \frac{\zeta_v + \zeta_c}{2} \tag{18.35}$$

The principal plane coincides with the vertex. If the image plane conjugate to ζ_o is denoted by ζ_i, we have

$$p_o q_i + p_i q_o = 0$$

or

$$\frac{1}{\zeta_o - \zeta_c} + \frac{1}{\zeta_i - \zeta_c} = \frac{2}{R} = \frac{1}{f} \tag{18.36}$$

as in light optics and in agreement with Eq. (18.19b).

For numerical evaluation, this form of the theory is less convenient than the parametric form dealt with in Section 18.2. In the case of a mirror, one has first to solve Eq. (18.21a) for ζ decreasing from ζ_0 to a small value L. A system of linear equations, arising from the requirement that as many derivatives as possible should be continuous at $\zeta = L$, must then be solved. The trajectory, in the form of a series expansions, has then to be traced until it again reaches the plane $\zeta = L$. The numerical tracing is then resumed (Fig. 18.7). This is a fairly high price to pay for any conceptual advantages

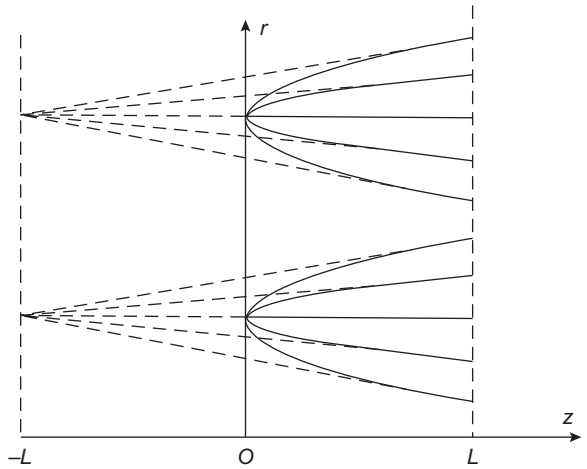


Figure 18.7
Trajectories approximated by parabolae in the zone $0 \leq z < L$. The tangents at $z = L$ all intersect in $z = -L$.

gained by avoiding the parametric representation – the parametric form clearly has many attractions.

18.4 A Quadratic Transformation

We mention briefly one other transformation designed to avoid the problems arising from the large ray gradients at the turning point. This was introduced by Bernard (1952), who wrote

$$z_v - z =: s^2 \tag{18.37}$$

for an electron initially travelling in the negative z direction and reflected at $z = z_v$. The negative sign is to be taken before the square root for the incident part of the trajectory and the positive sign for the reflected part, which already shows that this transformation is by no means ideal. If we consider only the electrostatic case ($B = 0$), Eq. (18.4a) becomes, after some calculation,

$$\frac{d^2x}{ds^2} + \left(\frac{1}{2\phi}\frac{d\phi}{ds} - \frac{1}{s}\right)\frac{dx}{ds} + \frac{1}{4\phi}\left(\frac{d^2\phi}{ds^2} - \frac{1}{s}\frac{d\phi}{ds}\right)x = 0 \tag{18.38}$$

with a similar equation for $y(s)$. The term in dx/ds is eliminated by writing

$$\overline{x} = \phi^{1/4} s^{-1/2} x \tag{18.39}$$

(this is similar to the Picht transformation) whereupon Eq. (18.38) reduces to

$$\frac{d^2\overline{x}}{ds^2} + G\overline{x} = 0 \tag{18.40}$$

with

$$G(s) = \frac{3}{16}\left(\frac{d\phi/ds}{\phi}\right)^2 - \frac{3}{4s^2} \tag{18.41}$$

Eq. (18.40) may then be analysed in the usual way and cardinal elements defined.

CHAPTER 19

Quadrupole Lenses

Hitherto, we have been studying systems with an axis of rotational symmetry, which are by far the most common in electron optical devices in which the electron accelerating voltage does not exceed one or a few hundred kilovolts or, exceptionally, a few megavolts (in practice, 3 or 4 MV maximum). At higher voltages, focusing elements with lower symmetry are more commonly used, and, in particular, elements with planes of electrical symmetry forming quadrupoles. These possess the property of 'strong focusing', by which we mean that their fields exert a force directly on the electrons, towards or away from the axis, whereas in round magnetic lenses, the focusing force is more indirect, arising from the coupling between B_z and the azimuthal component of the electron velocity. Quadrupoles have also been very thoroughly studied as elements for use at conventional accelerating voltages for a quite different reason. We shall see in Section 24.3 that one of the most undesirable aberrations of round electron lenses cannot be eliminated in any straightforward manner but can in principle be cancelled by introducing quadrupoles and octopoles into the system. This has provided the incentive for exhaustive studies of quadrupole lens properties, comparable in thoroughness with those on round lenses. In this chapter, we derive the paraxial equations of quadrupole systems, and introduce the notion of an orthogonal system.

A brief survey of the history of quadrupole studies is to be found at the beginning of Chapter 39 of Volume 2. Here we merely remark that although strong focusing at high energy and aberration correction have been the principal stimuli for research on quadrupoles, their properties were very fully explored long before either of these applications was known: the first study, which was thorough and meticulous, appeared as a Berlin dissertation in 1943 (Melkich, 1947).

In the discussion of Section 7.2.3, the terms in 2φ in the potential expansion (7.37 or 7.59) and their magnetic counterparts (7.43−7.45) were described as quadrupole terms. Here we use the term in a slightly less restrictive sense: a magnetic or electrostatic quadrupole is characterized by the presence of two planes of symmetry in the potential or field (Fig. 19.1), and rotationally symmetric fields as well as octopoles (and higher order $2n$-poles) are not excluded. From (7.36), we see that the term in $p_2(z)$ describes a potential for which the planes $x = 0$ and $y = 0$ are symmetry planes, $\Phi(x, y, z) = \Phi(\pm x, \pm y, z)$, and a typical electrostatic quadrupole thus has the form shown in Fig. 7.2 and Fig. 19.1A.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00019-5
© 2018 Elsevier Ltd. All rights reserved.

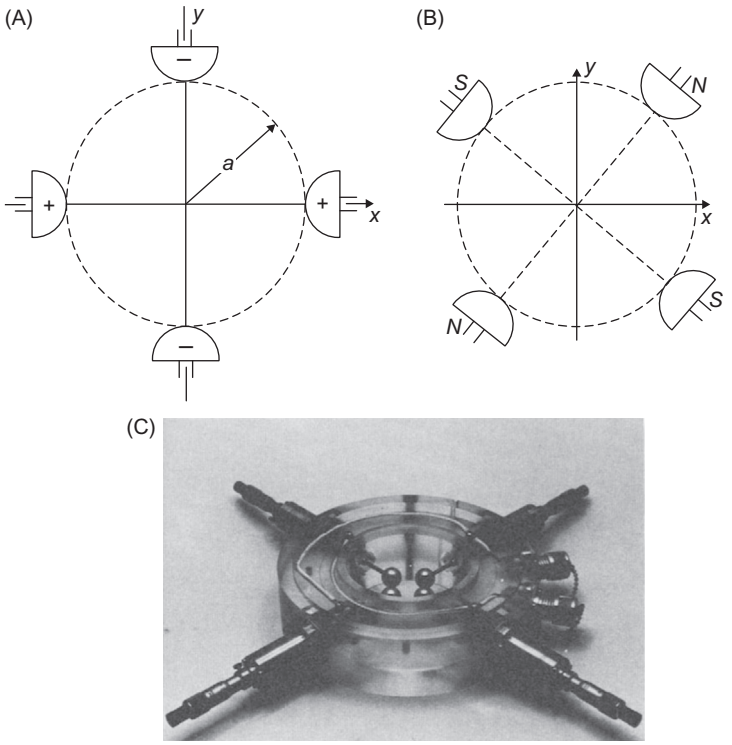


Figure 19.1
Quadrupoles. (A) Electrostatic quadrupole characterized by $p_2(z)$, with equal and opposite voltages on the electrodes. (B) Magnetic quadrupole characterized by $Q_2(z)$. (C) Photograph of an electrostatic quadrupole.

The term in $q_2(z)$ simply corresponds to a similar quadrupole inclined at 45° to that of Fig. 19.1A. In the magnetic case, the planes $x = 0$ and $y = 0$ are planes of symmetry for $P_2(z)$ and planes of antisymmetry for $Q_2(z)$. For a field described by Q_2 only, $B_x(x, y, z) = B_x(-x, y, z) = -B_z(x, -y, z) = -B_z(-x, -y, z)$, with analogous relations for B_y (Fig. 19.1B). We shall learn that uncoupled equations can be obtained if the quadrupoles are orientated with respect to the coordinate axes in such a way that $q_2 = 0$ and $P_2 = 0$, that is, as shown in Fig. 19.1.

19.1 Paraxial Equations for Quadrupoles

We set out from the general case in which rotationally symmetric magnetic and electrostatic fields may be present as well as the quadrupole fields themselves. The fields are thus characterized by six axial functions, $\phi(z)$ and $B(z)$ for the round lens components, $p_2(z)$ and $q_2(z)$ for the electrostatic quadrupoles and $Q_2(z)$ and $P_2(z)$ for the magnetic quadrupoles.

We substitute the field expansions (7.36) and (7.43−7.45) into the function $M, M = \overline{M}/(2m_0e)^{1/2}$ (15.23) as in Section 15.2 and expand M as a power series in X, Y and their derivatives. We find

$$\begin{aligned} M^{(0)} &= \hat{\phi}^{1/2} \\ M^{(2)} &= -\frac{\gamma\phi''}{8\hat{\phi}^{1/2}}\left(X^2+Y^2\right) + \left(\frac{\gamma p_2}{4\hat{\phi}^{1/2}} - \frac{1}{2}\eta Q_2\right)\left(X^2-Y^2\right) \\ &\quad + \left(\frac{\gamma q_2}{2\hat{\phi}^{1/2}} + \eta P_2\right)XY + \frac{1}{2}\hat{\phi}^{1/2}\left(X'^2+Y'^2\right) \\ &\quad - \frac{1}{2}\eta B(XY'-X'Y) \end{aligned} \tag{19.1}$$

The term in $XY'-X'Y$ can be removed by introducing the rotating coordinate system employed in connection with round magnetic lenses (15.7, 15.9) whereupon $M^{(2)}$ becomes

$$\begin{aligned} M^{(2)} &= -\left(\frac{\gamma\phi''}{8\hat{\phi}^{1/2}} + \frac{\eta^2B^2}{8\hat{\phi}^{1/2}}\right)\left(x^2+y^2\right) \\ &\quad + \left\{\left(\frac{\gamma p_2}{2\hat{\phi}^{1/2}} - \eta Q_2\right)\cos 2\theta + \left(\frac{\gamma q_2}{2\hat{\phi}^{1/2}} + \eta P_2\right)\sin 2\theta\right\}\frac{x^2-y^2}{2} \\ &\quad + \left\{-\left(\frac{\gamma p_2}{2\hat{\phi}^{1/2}} - \eta Q_2\right)\sin 2\theta + \left(\frac{\gamma q_2}{2\hat{\phi}^{1/2}} + \eta P_2\right)\cos 2\theta\right\}xy \\ &\quad + \frac{1}{2}\hat{\phi}^{1/2}\left(x'^2+y'^2\right) \end{aligned} \tag{19.2}$$

with $\theta' = \eta B/2\hat{\phi}^{1/2}$. The paraxial equations, $\partial M^{(2)}/\partial x_i = d(\partial M^{(2)}/dx'_i)/dz$ ($x_1 = x$, $x_2 = y$), still do not separate, however, unless

$$\tan\theta(z) = \frac{\gamma q_2/\hat{\phi}^{1/2} + 2\eta P_2}{\gamma p_2/\hat{\phi}^{1/2} + 2\eta Q_2} \tag{19.3}$$

This is known as the *orthogonality condition* and has been known in various forms for many years (Melkich, 1947; Glaser, 1956, Section 37; Dušek, 1959). This condition can be satisfied in several ways but only one is a practical possibility. Most generally, $\theta(z)$ may vary with z, in which case electrodes and polepieces must be devised and constructed of

such shapes that condition (19.3) is everywhere satisfied. More reasonably, $\theta(z)$ may be a constant, not necessarily zero; this requires $B(z) = 0$ and $\gamma q_2 + 2\eta P_2 \hat{\phi}^{1/2} \propto \gamma p_2 - 2\eta Q_2 \hat{\phi}^{1/2}$. Finally we may set $\theta(z)$ equal to zero, so that $q_2(z) = P_2(z) \equiv 0$, and at least one of $p_2(z)$ and $Q_2(z)$ is not zero everywhere. We retain only this final case.

From now on, then, we consider only orthogonal systems, and we assume that the electrodes and polepieces are disposed so that only $\phi(z)$, $p_2(z)$ and $Q_2(z)$ are allowed to be nonzero. The paraxial equations then take the form

$$\frac{d}{d_z}\left(\hat{\phi}^{1/2}x'\right) + \frac{\gamma\phi'' - 2\gamma p_2 + 4\eta Q_2 \hat{\phi}^{1/2}}{4\hat{\phi}^{1/2}}x = 0 \tag{19.4a}$$

$$\frac{d}{d_z}\left(\hat{\phi}^{1/2}y'\right) + \frac{\gamma\phi'' + 2\gamma p_2 - 4\eta Q_2 \hat{\phi}^{1/2}}{4\hat{\phi}^{1/2}}y = 0 \tag{19.4b}$$

or nonrelativistically

$$x'' + \frac{\phi'}{2\phi}x' + \frac{\phi'' - 2p_2 + 4\eta Q_2 \phi^{1/2}}{4\phi}x = 0 \tag{19.5a}$$

$$y'' + \frac{\phi'}{2\phi}y' + \frac{\phi'' + 2p_2 - 4\eta Q_2 \phi^{1/2}}{4\phi}y = 0 \tag{19.5b}$$

or in reduced form, $\xi_x \coloneqq x\hat{\phi}^{1/4}, \xi_y \coloneqq y\hat{\phi}^{1/4}$,

$$\xi''_x + \left\{\frac{3}{16}\left(\frac{\phi'}{\hat{\phi}}\right)^2\left(1 + \frac{4}{3}\varepsilon\hat{\phi}\right) - \frac{p_2 - 2\eta Q_2\hat{\phi}^{1/2}}{2\hat{\phi}}\right\}\xi_x = 0 \tag{19.6a}$$

$$\xi''_y + \left\{\frac{3}{16}\left(\frac{\phi'}{\hat{\phi}}\right)^2\left(1 + \frac{4}{3}\varepsilon\hat{\phi}\right) + \frac{p_2 - 2\eta Q_2\hat{\phi}^{1/2}}{2\hat{\phi}}\right\}\xi_y = 0 \tag{19.6b}$$

Each paraxial equation is a linear, homogeneous, second-order differential equation and by any of the lines of reasoning set out in Chapter 16, Gaussian Optics of Rotationally Symmetric Systems: Asymptotic Image Formation, we may again establish the existence of cardinal elements. For quadrupoles, it is usually sufficient to list the asymptotic cardinal elements, though real (and osculating) elements can of course be defined if needed. Unlike round lenses, however, two sets of cardinal elements are needed, one to characterize the $x{-}z$ plane, the other the $y{-}z$ plane. In the absence of any rotationally symmetric lens field ($\phi = \text{const}$), the lens action in one of these planes will be convergent while in the other it will be divergent. This is readily seen from (19.4−19.6).

The action of a quadrupole lens on electron rays is most conveniently characterized by a pair of transfer matrices, similar to (16.12) except that the matrix describing the coordination between object and image space is different in the two planes. Before writing down these matrices, we must first introduce the notion of *astigmatic objects and images.* Quadrupoles are commonly employed as multiplets – the quadruplet can have most attractive features – and if the cardinal elements are different in the $x{-}z$ and $y{-}z$ planes, the image plane corresponding to a given object plane may clearly be different as well: the system will not produce a stigmatic image of a point object. If a further lens follows, this *astigmatic image* will be the object for the subsequent stage, and we must thus expect to have to deal with *astigmatic objects.*

The most general transfer matrix relates position and slope in some plane in object space, $z = z_1$, to the same quantities in a plane in image space, $z = z_2$ (cf. 16.15):

$$\begin{pmatrix} x_2 \\ x_2' \end{pmatrix} = T_x \begin{pmatrix} x_1 \\ x_1' \end{pmatrix} \quad , \quad \begin{pmatrix} y_2 \\ y_2' \end{pmatrix} = T_y \begin{pmatrix} y_1 \\ y_1' \end{pmatrix} \tag{19.7}$$

and as in (16.12), we write

$$T_x = \begin{pmatrix} -\dfrac{z_2 - z_{F_i}^{(x)}}{f_{xi}} & \dfrac{\left(z_2 - z_{F_i}^{(x)}\right)\left(z_1 - z_{Fo}^{(x)}\right)}{f_{xi}} + f_{xo} \\ -\dfrac{1}{f_{xi}} & \dfrac{z_1 - z_{F_o}^{(x)}}{f_{xi}} \end{pmatrix} \tag{19.8a}$$

$$T_y = \begin{pmatrix} -\dfrac{z_2 - z_{F_i}^{(y)}}{f_{yi}} & \dfrac{\left(z_2 - z_{F_i}^{(y)}\right)\left(z_1 - z_{Fo}^{(y)}\right)}{f_{yi}} + f_{yo} \\ -\dfrac{1}{f_{yi}} & \dfrac{z_1 - z_{F_o}^{(y)}}{f_{yi}} \end{pmatrix} \tag{19.8b}$$

The cardinal elements are most conveniently defined with the aid of the rays $G_x(z)$, $G_y(z)$, $\overline{G}_x(z)$ and $\overline{G}_y(z)$, which satisfy conditions analogous to (16.1):

$$\lim_{z \to -\infty} G_x(z) = \lim_{z \to -\infty} G_y(z) = 1 \qquad \lim_{z \to \infty} \overline{G}_x(z) = \lim_{z \to \infty} \overline{G}_y(z) = 1 \tag{19.9}$$

The rays $G_x(z)$ and $\overline{G}_x(z)$ satisfy (19.4a) while $G_y(z)$ and $\overline{G}_y(z)$ satisfy (19.4b). The image foci are then the points of intersection of the image asymptotes to G_x and G_y with the axis; the rays $\overline{G}_x$ and $\overline{G}_y$ similarly define the object foci. The focal lengths are given by

$$\begin{aligned} f_{xi} &= -1/G_{xi}', \quad f_{yi} = -1/G_{yi}' \\ f_{xo} &= 1/\overline{G}_{xo}', \quad f_{yo} = 1/\overline{G}_{yo}' \end{aligned} \tag{19.10}$$

Let us suppose that the planes $z = z_{xo}$ and $z = z_{xi}$ are conjugate, in the sense that $(T_x)_{12} = 0$, or

$$\left(z_{xi} - z_{F_i}^{(x)}\right)\left(z_{xo} - z_{Fo}^{(x)}\right) = -f_{xi}f_{xo} \tag{19.11}$$

so that

$$T_x = \begin{pmatrix} M_x & 0 \\ -1/f_{xi} & \left(\hat{\phi}_0/\hat{\phi}_i\right)^{1/2}/M_x \end{pmatrix} \tag{19.12}$$

in which we have used the Wronskian of (19.4a) to show that $f_{xo}/f_{xi} = \left(\hat{\phi}_0/\hat{\phi}_i\right)^{1/2}$ a similar relation is of course true for f_{yo}/f_{yi}. The magnitude M_x is the height of the image asymptote to $G_x(z)$ in the plane $z = z_{xi}$. In general, however, $(T_y)_{12} \neq 0$ when (19.11) is satisfied and so although rays from a point $P(x_o, y_o)$ all have the same x-coordinate $x_i = M_x x_o$ in $z = z_{xi}$, their y-coordinate is a function of both y_o and the gradient y'_o:

$$T_y = \begin{pmatrix} -\dfrac{z_{xi} - z_{F_i}^{(y)}}{f_{yi}} & \dfrac{\left(z_{xi} - z_{F_i}^{(y)}\right)\left(z_{xo} - z_{Fo}^{(y)}\right)}{f_{yi}} + f_{yo} \\ -\dfrac{1}{f_{yi}} & \dfrac{z_{xo} - z_{F_o}^{(y)}}{f_{yi}} \end{pmatrix} \tag{19.13}$$

A point $P(x_0, y_0)$ is therefore imaged as a line in the plane $z = z_{xi}$ parallel to the y-axis. Likewise, if we consider a pair of planes $z = z_{yo}$ and $z = z_{yi}$ for which $(T_y)_{12} = 0$, we find that in general $(T_x)_{12}$ does not vanish and again a line image is formed, now parallel to the x-axis. These line images are thus at right-angles to one another and separated by the *astigmatic difference*. One or both may be virtual (Fig. 19.2). If we move an axial point object along the axis, the line foci will move and there will always be real or virtual object positions for which the line foci coincide and the image is stigmatic. In general, however, the magnifications in the two planes will not be equal.

The astigmatic difference can be expressed in terms of the cardinal elements and magnification. Using the quadrupole analogue of (16.25),

$$\begin{aligned} z_{xi} - z_{Fi}^{(x)} &= -f_{xi}M_x \\ z_{yi} - z_{Fi}^{(y)} &= -f_{yi}M_y \end{aligned} \tag{19.14}$$

and

$$\begin{aligned} z_{xo} - z_{Fo}^{(x)} &= f_{xo}/M_x \\ z_{yo} - z_{Fo}^{(y)} &= f_{yo}/M_y \end{aligned} \tag{19.15}$$

we see that

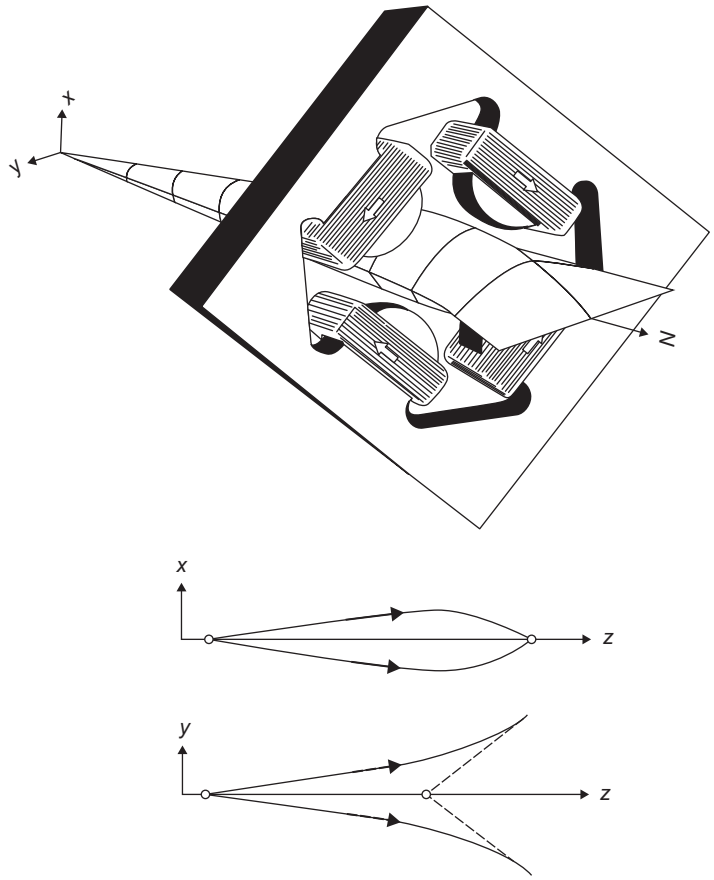


Figure 19.2
Formation of a line image in a magnetic quadrupole. The arrows show the directions of the currents in the windings. Courtesy D.F. Hardy

$$\begin{aligned} \Lambda_i &\coloneqq z_{xi} - z_{yi} = \Lambda_{Fi} - f_{xi}M_x + f_{yi}M_y \\ \Lambda_o &\coloneqq z_{xo} - z_{yo} = \Lambda_{Fo} + f_{xo}M_x - f_{yo}M_y \end{aligned} \tag{19.16}$$

where

$$\begin{aligned} \Lambda_{Fi} &\coloneqq z_{Fi}^{(x)} - z_{Fi}^{(y)} = \Lambda_i\left(M_x = M_y = 0\right) \\ \Lambda_{Fo} &\coloneqq z_{Fo}^{(x)} - z_{Fo}^{(y)} = \Lambda_o\left(M_x = M_y \to \infty\right) \end{aligned} \tag{19.17}$$

From (19.13), it is readily seen that quadratic equations are obtained for M_x and M_y if we attempt to satisfy the stigmatic imaging condition, $\Lambda_i = \Lambda_o = 0$. The discriminant δ is the same for M_x and M_y and we find

$$M_x = \frac{\Lambda_{Fo}\Lambda_{Fi} - f_{xo}f_{xi} + f_{yo}f_{yi} \pm \delta}{2\Lambda_{Fo}f_{xi}}$$
$$M_y = \frac{-\Lambda_{Fo}\Lambda_{Fi} - f_{xo}f_{xi} + f_{yo}f_{yi} \pm \delta}{2\Lambda_{Fo}f_{yi}} \tag{19.18a}$$

with

$$\begin{aligned}\delta^2 &= \left(f_{xo}f_{xi} - f_{yo}f_{yi}\right)^2 + \Lambda_{Fo}^2\Lambda_{Fi}^2 + 2\left(f_{xo}f_{xi} + f_{yo}f_{yi}\right)\Lambda_{Fo}\Lambda_{Fi} \\ &= \left(f_{xo}f_{xi} + f_{yo}f_{yi} + \Lambda_{Fo}\Lambda_{Fi}\right)^2 - 4f_{xo}f_{xi}f_{yo}f_{yi}\end{aligned} \tag{19.18b}$$

In the usual case in which the signs of f_x and f_y are different, δ^2 is positive and there are two real roots and hence two pairs of stigmatic conjugates.

The cardinal elements of multiplets are established most easily by multiplying the transfer matrices of the individual lenses; these must be separated by transfer matrices corresponding to the *drift spaces*, the spaces between the planes $z = z_2$ for one lens and $z = z_1$ for the next. We recall that these planes may be chosen in various ways; $z_1 = z_2 = 0$, in which case incident position and gradient are related to emergent position and gradient *in the same plane*, conventionally the mid-plane of the lens, is one good choice, thoroughly explored by Dušek (1959). Here we have

$$\begin{pmatrix} x_2 \\ x_2' \end{pmatrix} = \begin{pmatrix} z_{Fi}^{(x)}/f_{xi} & z_{Fi}^{(x)}z_{Fo}^{(x)}/f_{xi} + f_{xo} \\ -1/f_{xi} & -z_{Fo}^{(x)}/f_{xi} \end{pmatrix} \begin{pmatrix} x_1 \\ x_1' \end{pmatrix}$$
$$\begin{pmatrix} y_2 \\ y_2' \end{pmatrix} = \begin{pmatrix} z_{Fi}^{(y)}/f_{yi} & z_{Fi}^{(y)}z_{Fo}^{(y)}/f_{yi} + f_{yo} \\ -1/f_{yi} & -z_{Fo}^{(y)}/f_{yi} \end{pmatrix} \begin{pmatrix} y_1 \\ y_1' \end{pmatrix} \tag{19.19}$$

(We note that Dušek's matrices are trivially different since he used the vectors $(x'\ x)^T$ and $(y'\ y)^T$.)

Another convenient choice involves using different pairs of planes for T_x and T_y, namely the focal planes, since the diagonal matrix elements then vanish:

$$T_x = \begin{pmatrix} 0 & f_{xo} \\ -1/f_{xi} & 0 \end{pmatrix}, \quad T_y = \begin{pmatrix} 0 & f_{yo} \\ -1/f_{yi} & 0 \end{pmatrix} \tag{19.20}$$

This choice has been studied in great detail by Regenstreif (1966, 1967), who has established straightforward rules for writing down the transfer matrices of an arbitrary number of quadrupoles. His procedure can be applied to Dušek matrices (Hawkes, 1970), which we temporarily write as follows:

$$T_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \tag{19.21}$$

where T_i denotes either T_x or T_y for the i-th quadrupole of a sequence and

$$a_i \coloneqq \frac{z_{Fi}}{f_i}, \quad b_i \coloneqq \frac{z_{Fo}z_{Fi}}{f_i} + f_o$$
$$c_i \coloneqq -\frac{1}{f_i}, \quad d_i \coloneqq -\frac{z_{Fo}}{f_i} \tag{19.22}$$

The separation between the i-th and $(i+1)$-th quadrupoles is denoted by $L_{i,i+1}$ and we write

$$T\left(L_{i,i+1}\right) \coloneqq \begin{pmatrix} 1 & L_{i,i+1} \\ 0 & 1 \end{pmatrix} \tag{19.23}$$

Introducing the distances $X_{i,i+1}$ between the image focus of the i-th quadrupole and the object focus of the $(i+1)$-th quadrupole,

$$X_{i,i+1} \coloneqq \frac{a_i}{c_i} + L_{i,i+1} + \frac{d_{i+1}}{c_{i+1}} \tag{19.24}$$

we can show (Regenstreif, 1966, 1967; Hawkes, 1970) that the transfer matrix of n quadrupoles separated by $n-1$ drift spaces is given by

$$T^{(n)} = \begin{pmatrix} a^{(n)} & b^{(n)} \\ c^{(n)} & d^{(n)} \end{pmatrix}$$
$$a^{(n)} = a_n \prod_{i=1}^{n-1} c_i \alpha_n \quad , \quad b^{(n)} = d_1 a_n \prod_{i=2}^{n-1} c_i \beta_n \tag{19.25}$$
$$c^{(n)} = \prod_{i=1}^{n} c_i \gamma_n, \quad d^{(n)} = d_1 \prod_{i=2}^{n} c_i \delta_n$$

in which

$$\alpha_n = \left(X_{n-1,n} - \frac{1}{a_n c_n}\right)\gamma_{n-1} - \frac{\gamma_{n-2}}{c_{n-1}^2}$$
$$\beta_n = \left(X_{n-1,n} - \frac{1}{a_n c_n}\right)\delta_{n-1} - \frac{\delta_{n-2}}{c_{n-1}^2} \tag{19.26}$$
$$\gamma_n = X_{n-1,n}\gamma_{n-1} - \frac{\gamma_{n-2}}{c_{n-1}^2}$$
$$\delta_n = X_{n-1,n}\delta_{n-1} - \frac{\delta_{n-2}}{c_{n-1}^2}$$

and

$$\gamma_0 = \delta_0 = 0, \quad \gamma_1 = \delta_1 = 1$$
$$\gamma_2 = X_{1,2}, \quad \delta_2 = X_{1,2} - \frac{1}{c_1 d_1}$$

Another expression for the elements of the transfer matrix between an arbitrary pair of planes, $z = z_n$ and $z = z_0$, may be derived by using the transfer matrix between the principal planes. Writing

$$\begin{aligned} D^{(i)} &:= z_{Po}^{(i)} - z_{Pi}^{(i-1)} \quad 2 \le i \le n-1 \\ D^{(1)} &:= z_{Po}^{(1)} - z_0, \quad D^{(n)} := z_n - z_{Pi}^{(n-1)} \end{aligned} \tag{19.27}$$

we form the matrix

$$T(z_0, z_n) = D_n T_{n-1} D_{n-1} \ldots T_2 D_2 T_1 D_1 \tag{19.28}$$

The elements can be written as Gaussian brackets (Herzberger, 1943, 1958; Hawkes, 1967; Dymnikov, 1968), which are defined as follows:

$$\begin{aligned} [x_1] &= x \\ [x_1 x_2] &= x_1 x_2 + 1 \\ [x_1 x_2 x_3] &= x_1 x_2 x_3 + x_1 + x_3 \\ [x_1 x_2 x_3 \ldots x_n] &= [x_1 x_2 x_3 \ldots x_{n-2}] + x_n [x_1 x_2 x_3 \ldots x_{n-1}] \end{aligned} \tag{19.29}$$

We find

$$T(z_0, z_n) = \begin{pmatrix} [D_n c_{n-1} \ldots D_2 c_1] & [D_n c_{n-1} \ldots D_2 c_1 D_1] \\ [c_{n-1} D_{n-1} \ldots c_1] & [c_{n-1} D_{n-1} \ldots c_1 D_1] \end{pmatrix} \tag{19.30}$$

For a proof see Hawkes (1967). Gaussian brackets are also employed by Chechulin and Yavor (1969) in connection with prisms; renewed interest has been shown in them in light optics too (e.g., Tanaka, 1981, 1982, 1983, 1986).

One common requirement for quadrupole multiplets is that their overall behaviour should be the same as that of a round lens; for this, the cardinal elements in the $x{-}z$ and $y{-}z$ planes must coincide and the focal lengths f_x and f_y must be equal. If we consider quadrupoles that may have a round electrostatic lens component, provided that the latter has no overall accelerating or retarding effect ($\phi_i = \phi_o$), we can see on symmetry grounds that one at least of these conditions is satisfied by imposing a certain symmetry on the system. In particular, we perceive that the focal lengths in the $x{-}z$ and $y{-}z$ planes are automatically equal in *antisymmetric multiplets.* The latter are defined as follows. If a multiplet consists of $2N$ quadrupoles ($N \ge 1$) such that the central plane is a plane of geometrical symmetry and electrical antisymmetry, we say that the multiplet is *antisymmetric.* The case of $N = 2$ was extensively studied by a group in Leningrad (Yavor, 1962; Dymnikov and Yavor, 1963; Dymnikov et al., 1963a,b, 1964a,b, 1965; Shpak and Yavor, 1964) and has come to be known as the *Russian quadruplet* (Fig. 19.3). Consider two rays, $G_x(z)$ and $\overline{G}_y(z)$. Because of the electrical antisymmetry, the sequence of functions $p_2(z)$ or $Q_2(z)$ encountered by $G_x(z)$ as it proceeds in the positive z-direction will be exactly the same as that traversed by $\overline{G}_y(z)$ if we imagine it travelling in the negative z-direction. The gradients of

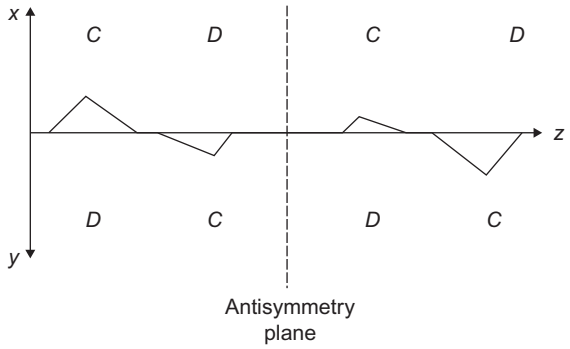


Figure 19.3
The antisymmetric or 'Russian' quadruplet. The centre plane is a plane of geometrical symmetry and electrical antisymmetry. *C*, *D* denote convergent and divergent action respectively.

the emergent asymptotes will hence be equal but opposite in sign, and the focal lengths f_x and f_y will hence be equal (since $\phi_i = \phi_o$, we already have $f_{xi} \coloneqq f_{xo} =: f_x$ and $f_{yi} \coloneqq f_{yo} =: f_y$). If we wish to arrange that an antisymmetric multiplet has the same image-forming properties as a round lens, therefore, we have only to ensure that the foci (or principal points) in the $x-z$ and $y-z$ planes coincide. For a given geometry, we need only vary the relative excitation until the condition is satisfied. We therefore obtain a set of pairs of excitations, which is known as the *load characteristic* of the quadruplet. Some examples are given in Chapter 39 of Volume 2.

19.2 Transaxial Lenses

The foregoing discussion has been confined to the optics of quadrupoles in general and we have said little about the shapes of the electrodes and polepieces and hence about the effects of any additional symmetries. Two particular additional symmetry properties are of interest, however; one leads to cylindrical lenses, at least in the electrostatic case, as we see in Chapter 20, Cylindrical Lenses. A different symmetry characterizes *transaxial lenses*, in which the field or potential is rotationally symmetric but the optic axis is no longer the same as the symmetry axis but is perpendicular to it. The electron beam now passes between rotationally symmetric, typically plane electrodes and is focused by the fields in any gaps. Fig. 19.4 shows such a system. The electrodes are ideally circular or annular but are in practice reduced to sectors since the electron beam occupies so little of the space available.

It can be seen by comparison with Chapter 20, Cylindrical Lenses, that such structures bear some resemblance to cylindrical lenses but differ from the latter in that the potential is not independent of one of the transverse cartesian coordinates but is the same for all azimuthal

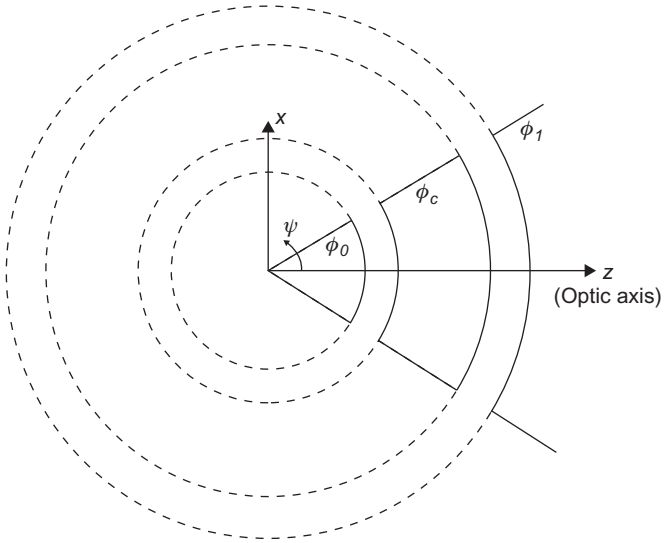


Figure 19.4
A transaxial lens. The electrodes shown lie in some plane $y = \text{const}$ and an identical set lies in the plane $y = -\text{const}$.

angles φ at a given radial distance from the axis of rotational symmetry. The optical behaviour of such systems was first investigated formally by Strashkevich (1962), to whom we owe the name 'transaxial lenses'; the theory was set out in some detail in his book of 1966. In the early 1970s, it was realized (by V.M. Kel'man and colleagues in Alma-Ata) that certain features of these structures rendered them attractive for use in the collimator of a prism spectrometer, and their properties were investigated in some detail (Glikman et al., 1971; Brodskii and Yavor, 1970, 1971; Karetskaya et al., 1970, 1971a,b). This work is presented in full in Kel'man et al. (1979), one of the three chapters of which is devoted wholly to these lenses.

The symmetry conditions are now such that

$$\begin{aligned} \Phi(x, y, z) &= \Phi(x, -y, z) \\ \Phi(0, y, z) &= \Phi\left(x, y, \left(z^2 - x^2\right)^{1/2}\right) \end{aligned} \tag{19.31}$$

For small values of x and y, therefore,

$$\begin{aligned} \Phi(x, y, z) = \phi(z) + \frac{\phi'}{2z}x^2 + \frac{1}{z^2}\left(\phi'' - \frac{\phi'}{z}\right)x^4 \\ + \phi_2 y^2 + \frac{\phi_2'}{2z}x^2y^2 + \phi_4 y^4 \end{aligned} \tag{19.32}$$

with

$$\begin{aligned}\phi_2 &\coloneqq -\frac{1}{2}\left(\phi'' + \frac{\phi'}{z}\right)\\ \phi_4 &\coloneqq -\frac{1}{12}\left(\phi_2'' + \frac{\phi_2'}{z}\right)\end{aligned} \tag{19.33}$$

Comparing Eq. (19.32) with (7.36), we see that

$$p_2(z) = \phi'(z)/z + \phi''/2 \tag{19.34}$$

The paraxial equations have the form

$$x'' + \frac{\phi'}{2\phi}x' - \frac{\phi'}{2z\phi}x = 0 \tag{19.35a}$$

$$y'' + \frac{\phi'}{2\phi}y' - \frac{\phi_2}{\phi}y = 0 \tag{19.35b}$$

All the theory for quadrupoles can hence be employed without further discussion. The form of (19.35a) is, however, such that simple expressions can be obtained for the focal length and foci. Thus on writing

$$\xi \coloneqq x/z \tag{19.36}$$

Eq. (19.35a) becomes

$$\xi'' + \frac{2 + \phi' z/2\phi}{z}\xi' = 0 \tag{19.37}$$

and after some trivial calculation, we find

$$\begin{aligned}x &= \frac{x_o z}{z_o} + \left(x_o' z_o - x_o\right)\phi_o^{1/2} z \int_{z_o}^{z} \frac{d\zeta}{\phi^{1/2}\zeta^2}\\ &= x_o' z\left(x_o - x_o' z_o\right)\phi_o^{1/2}\left(\frac{1}{\phi^{1/2}} + \frac{z}{2}\int_{z_o}^{z}\frac{\phi' d\zeta}{\phi^{3/2}\zeta}\right)\end{aligned} \tag{19.38}$$

giving the transfer matrix

$$\begin{pmatrix} x \\ x' \end{pmatrix} = \begin{pmatrix} \left(\phi_o/\phi\right)^{1/2} - z/f_i & z - z_o\left(\phi_o/\phi\right)^{1/2} + zz_o/f_i \\ -1/f_i & 1 + z_o/f_i \end{pmatrix}\begin{pmatrix} x_o \\ x_o' \end{pmatrix} \tag{19.39}$$

with

$$\frac{1}{f_i} = -\frac{\phi_o^{1/2}}{2}\int_{-\infty}^{\infty}\frac{\phi' dz}{\phi^{3/2} z} \tag{19.40}$$

in which we have extended the limits of integration to infinity in f_i since it is asymptotic imagery that will be of interest. The planes z_o and z_i will be conjugate if

$$z_i - z_o\left(\frac{\phi_o}{\phi_i}\right)^{1/2} + \frac{z_o z_i}{f_i} = 0 \tag{19.41}$$

or

$$\frac{1}{\phi_o^{1/2} z_o} - \frac{1}{\phi_i^{1/2} z_i} = -\frac{1}{f_i \phi_o^{1/2}} \tag{19.42}$$

and the transfer matrix becomes

$$\begin{pmatrix} z_i/z_o & 0 \\ -1/f_i & 1 + z_o/f_i \end{pmatrix} = \begin{pmatrix} z_i/z_o & 0 \\ -1/f_i & z_o\phi_o^{1/2}/z_i\phi_i^{1/2} \end{pmatrix} \tag{19.43}$$

In the converging or $y-z$ plane, there is no such simple solution of the paraxial equation.

CHAPTER 20

Cylindrical Lenses

Cylindrical lenses are electrostatic or magnetic devices in which the potential or field is constant in some direction perpendicular to the optic axis. They are the electron optical analogues of glass lenses, the surfaces of which are not spheres, as in rotationally symmetric or 'round' lenses, but cylinders, whence their name. This nomenclature has not always been adhered to in electron optics, and rotationally symmetric lenses are not infrequently referred to as cylindrical, since their central opening is indeed a circular hole and some electrostatic lenses consist of a sequence of metal cylinders.

Cylindrical lenses have a long history, going back to one of the earliest publications on lens properties, the note by Davisson and Calbick (1931) on the lens-like behaviour of round openings and slits. The paraxial properties of such lenses were first discussed by Picht (1939b) and in very much more detail by Gratsiatos (1940). In the same year Strashkevich (1940a,b) gave the ray equations for electrostatic cylindrical lenses, including the reduced form (15.40), and these equations are again to be found in Leitner (1942). Many of the properties were elucidated and rediscovered over the years; the relevant papers are listed in the bibliography to this chapter. We single out the work of Kel'man et al. (1954), Yavor (1955) and Kel'man and Yavor (1955) on magnetic cylindrical lenses, recapitulated in detail in Kel'man and Yavor (1968), and the studies of Hawkes (1966/1967) and Rose (1966/1967, 1972).

The potential distributions in cylindrical lenses have been calculated or measured, in particular in connection with mass spectrometer design; see Wallington (1970, 1971), Harting and Read (1976), Mulvey and Wallington (1973) and Boerboom (1959, 1960a,b).

We now assume that the potentials Φ (7.36) and W (7.41) are functions of X and z only, so that in the general expansion of Eq. (7.36), we have

$$\begin{aligned} &p_2(z) = -\frac{1}{2}\phi''(z), \quad p_4(z) = \frac{1}{8}\phi^{(4)}(z) \\ &q_2(z) = q_4(z) = 0 \end{aligned} \tag{20.1}$$

and in Eqs (7.46−7.48)

$$\begin{aligned} &P_2(z) = \frac{1}{2}B'(z), \quad P_4(z) = -\frac{1}{8}B'''(z) \\ &Q_2(z) = Q_4(z) = 0 \end{aligned} \tag{20.2}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00020-1
© 2018 Elsevier Ltd. All rights reserved.

It is immediately clear from Eq. (19.3) that an electrostatic cylindrical lens forms an orthogonal system whereas a magnetic one does not. We have

$$\begin{aligned}
\Phi(x,y,z) &= \phi(z) - \frac{1}{2}\phi'' x^2 + \frac{1}{24}\phi^{(4)} x^4 \\
A_x(x,y,z) &= -\frac{1}{2} yB + \frac{1}{48} B'' y(9x^2 + y^2) \\
A_y(x,y,z) &= \frac{1}{2} xB - \frac{1}{48} B'' y(5x^2 - 3y^2) \\
A_z(x,y,z) &= -\frac{1}{2} xyB' + \frac{1}{48} B''' xy(3x^2 + y^2)
\end{aligned} \tag{20.3}$$

Expanding the function M (Eq. 15.23), we obtain

$$\begin{aligned}
M^{(0)} &= \hat{\phi}^{1/2} \\
M^{(2)} &= -\frac{\gamma\phi''}{4\hat{\phi}^{1/2}} x^2 + \frac{1}{2}\hat{\phi}^{1/2}\left(x'^2 + y'^2\right) + \frac{1}{2}\eta xyB' \\
&\quad - \frac{1}{2}\eta B(xy' - x'y)
\end{aligned} \tag{20.4}$$

giving the paraxial equations

$$\frac{d}{dz}(\hat{\phi}^{1/2} x') + \frac{\gamma\phi''}{2\hat{\phi}^{1/2}} x + \eta B y' = 0 \tag{20.5a}$$

$$\frac{d}{dz}(\hat{\phi}^{1/2} y' - \eta B x) = 0 \tag{20.5b}$$

(We note in passing that expression (20.4) for $M^{(2)}$ could have been simplified by the use of Sturrock's partial-integration rule (1955), which tells us that when a term of the form $g(x,y,z)\dfrac{df(x,y,z)}{dz}$ occurs in a variational function such as $M^{(2)}$ it may be replaced by $-fdg/dz$; here we could have reduced the terms in B to $-\eta Bxy'$.) Integrating Eq. (20.5b), we find

$$y' = \frac{C + \eta Bx}{\hat{\phi}^{1/2}} \tag{20.6}$$

in which C is a constant and Eq. (20.5a) may then be written

$$x'' + \frac{\gamma\phi'}{2\hat{\phi}} x' + \frac{\gamma\phi'' + 2\eta^2 B^2}{2\hat{\phi}} x = -\frac{\eta CB}{\hat{\phi}} \tag{20.7}$$

This can be recast into reduced form by writing $\upsilon = x\hat{\phi}^{1/4}$ (15.40).

Eq. (20.7) is a linear, second order, differential equation but is no longer homogeneous when a magnetic field is present. In the absence of such fields ($B = 0$), the paraxial equation for $x(z)$ is almost the same as that for round electrostatic lenses, the only difference being the factor 2 in the denominator of the term in x. The equation for $y(z)$ can be solved immediately, in this case:

$$y(z) = C\int^{z} \hat{\phi}^{-1/2}(\zeta)d\zeta \tag{20.8}$$

The lens action in the $x-z$ plane is thus described by a transfer matrix of the type Eq. (16.12). It is interesting to note that the best thin-lens approximation to the focal length is now obtained by writing

$$v := x\hat{\phi}^{1/2}$$

(and not $v = x\hat{\phi}^{1/4}$ as in 15.40 and 17.79). In the nonrelativistic approximation, Eq. (20.7) becomes

$$\frac{d}{dz}\left(\frac{v'}{\phi^{1/2}}\right) = -\frac{v\phi'^2}{2\phi^{5/2}}$$

and the focal length is then

$$\frac{1}{f_i} = \frac{\phi_o^{1/2}}{2}\int \frac{\phi'^2}{\phi^{5/2}}dz$$

This expression was obtained by Brodskii and Yavor (1971), who find that it is substantially more accurate than that given by writing $v = x\hat{\phi}^{1/4}$. Glikman et al. (1967a) have demonstrated that electrostatic cylindrical lenses always have a converging action in the $x-z$ plane, using the exact trajectory equation rather than the paraxial equation.

In the $y-z$ plane, we have

$$\begin{pmatrix} y_2 \\ y_2' \end{pmatrix} = \begin{pmatrix} 1 & \int_{z_1}^{z_2} (\hat{\phi}_1/\hat{\phi})^{1/2}d\zeta \\ 0 & (\hat{\phi}_1/\hat{\phi}_2)^{1/2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_1' \end{pmatrix} \tag{20.9}$$

This is the type of transfer matrix that we meet in light optics for a parallel-plane glass plate separating media of different refractive index.

In the general case when $B \neq 0$, Eq. (20.7) is solved using the method of variation of parameters. If $g(z)$ and $h(z)$ are two linearly independent solutions of the homogeneous equation obtained from Eq. (20.7), satisfying the initial conditions

$$\begin{aligned} g(z_o) &= h'(z_o) = 1 \\ g'(z_o) &= h(z_o) = 0 \end{aligned} \tag{20.10}$$

in some object plane, $z = z_o$, the general solution is

$$x(z) = x_o g(z) + x'_o h(z) \tag{20.11}$$

The solution of the inhomogeneous equation is thus

$$\begin{aligned} x(z) &= x_o g(z) + x'_o h(z) \\ &\quad - \frac{\eta C h(z)}{\hat{\phi}_o^{1/2}} \int_{z_o}^{z} \frac{g(\zeta) B(\zeta)}{\hat{\phi}^{1/2}(\zeta)} d\zeta + \frac{\eta C g(z)}{\hat{\phi}_o^{1/2}} \int_{z_o}^{z} \frac{h(\zeta) B(\zeta)}{\hat{\phi}^{1/2}(\zeta)} d\zeta \\ &=: x_o g(z) + x'_o h(z) + \eta C j(z) \end{aligned} \tag{20.12}$$

with

$$j(z) := -\frac{1}{\phi_0^{1/2}} \left(h(z) \int \frac{gB}{\phi^{1/2}} d\zeta - g(z) \int \frac{hB}{\phi^{1/2}} d\zeta \right)$$

From Eq. (20.6), we have

$$C = \hat{\phi}_o^{1/2} y'_o - \eta B_o x_o \tag{20.13}$$

and so

$$x(z) = x_o \left\{ g(z) - \eta^2 B_o j(z) \right\} + x'_o h(z) + y'_o \eta \hat{\phi}_o^{1/2} j(z) \tag{20.14}$$

Substituting into Eq. (20.6) and integrating, we obtain

$$\begin{aligned} y(z) &= y_o + y'_o \hat{\phi}_o^{1/2} \left(\int_{z_o}^{z} \frac{d\zeta}{\hat{\phi}^{1/2}} + \eta^2 \int_{z_o}^{z} \frac{Bj}{\hat{\phi}^{1/2}} d\zeta \right) \\ &\quad + x_o \eta \left(-B_o \int_{z_o}^{z} \frac{d\zeta}{\hat{\phi}^{1/2}} + \int_{z_o}^{z} \frac{Bg}{\hat{\phi}^{1/2}} d\zeta - \eta^2 B_o \int_{z_o}^{z} \frac{Bj}{\hat{\phi}^{1/2}} d\zeta \right) \\ &\quad + x'_o \eta \int_{z_o}^{z} \frac{Bh}{\hat{\phi}^{1/2}} d\zeta \end{aligned} \tag{20.15}$$

Systems with optical properties as complicated as Eqs (20.14) and (20.15) suggest are unattractive in practice. Their general behaviour has been studied in the context of ophthalmological optics by Gullstrand (1900, 1906, 1908, 1915, 1924) and a more accessible account is given by Carathéodory (1937). The electron optical situation has been explored in detail by Kel'man et al. (1954), Yavor (1955), Kel'man and Yavor (1954, 1955) and Rose (1966/1967 and especially 1972).

PART IV

Aberrations

CHAPTER 21

Introduction to Aberration Theory

The paraxial approximation, characterizing the linear coupling between two spaces, object space and image space, describes the dominant behaviour of the various electron optical components but small departures from this can rarely be neglected. In order to assess the magnitude of such nonlinear effects, we must proceed to the next higher order approximation, which involves retaining quartic terms in the expression for the refractive index M (Eq. 4.35) for systems with a straight optic axis. We shall find that the results of this calculation can be compactly expressed in terms of coefficients, the geometrical aberration coefficients, which vary in number with the symmetry of the system. The object of most aberration calculations is to obtain expressions for these coefficients in terms of the axial potential or field functions and of particular solutions of the paraxial ray equations.

The calculation may be performed in two very different ways. The most straightforward consists in writing down the ray equations as in Section 15.1 but now retaining higher order terms in the various field and potential expansions. For the most common components, round lenses, quadrupoles and mirrors, this generates linear, second-order, ordinary differential equations, which are now inhomogeneous; the corresponding homogeneous equations are identical with the paraxial equations. These inhomogeneous equations are solved by variation of parameters or by writing down the appropriate Green's function; the difference is purely formal and almost trivial. This procedure is commonly known as the *trajectory method.* The method is best understood by an example, and the reader is referred to Section 24.2, where round lens aberrations are studied in this way. The trajectory method has been used to study systems with arbitrary curved axes in considerable detail. We examine such systems in Part X.

Alternatively, a general perturbation theory may be developed, which enables us to answer the following question: given the solutions of ray equations derived from the paraxial refractive index, $M = M^{(2)}$, what will be the solutions for a slightly different refractive index, $M = M^{(2)} + M^{(p)}$? (Here $M \propto \overline{M}$, see Eqs (4.35) and (22.2).) By choosing $M^{(p)} = M^{(4)}$ for example, we obtain the primary geometrical aberrations of systems with a straight axis. Although this second approach may seem more abstract, it has one distinct advantage, which will become apparent below: any inter-relations between the various aberration coefficients emerge automatically, whereas they may be by no means obvious when the trajectory method is adopted. This second procedure, which we shall mostly use, is known

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00021-3
© 2018 Elsevier Ltd. All rights reserved.

as the *eikonal method* (from the term introduced by Bruns (1895) in his *Das Eikonal*) or *method of characteristic functions* (following Hamilton, 1828−1837, see Conway and Synge (1931)).

These *geometrical aberrations*, arising from higher order terms in the field expansions, or more evocatively, from allowing the electrons to stray a little way beyond the truly paraxial region, are not the only type that can arise. The next most important are the *chromatic aberrations.* These are the analogue in electron optics of effects caused by variations in refractive index with wavelength in light optics; they can arise in several ways. We have been assuming that the electron energy and any magnetic lens fields and electrostatic lens potentials are static and fixed. In reality, however, the lens excitations will inevitably fluctuate somewhat (except in superconducting lenses operating in the persistent-current mode) and the electrons of the incident beam will never be perfectly monoenergetic: their energies will span a small range, narrow but not negligibly so, for reasons that are examined in Part IX. We shall see in Volume 3 that a wavelength (λ) proportional to $\hat{\Phi}^{-1/2}$ can be associated with electrons accelerated through a voltage drop of Φ, and effects caused by variations in electron energy relative to the lens excitation may thus be regarded as due to a change in the focusing properties with λ. They are hence known generically as *chromatic aberrations.* In studying these, we establish *chromatic aberration coefficients*, which characterize the optical consequences of small changes in electron energy, electrostatic lens potential or magnetic lens field strength. In addition, in electron microscopes, some electrons commonly lose energy within the specimen and this too contributes to the energy spread in the beam and hence to the chromatic blur.

Various other perturbations have been characterized by aberration coefficients and calculated by the methods set out in this part: effects due to the inclusion of relativity, those caused by moderate space charge forces and those provoked by small departures from the assumed symmetry of the system. Most formulae are now available in relativistically correct form and the incentive to explore the difference between nonrelativistic and relativistic behaviour by perturbation theory has vanished. Conversely, it can still be useful to enquire what effect space charge will have when it has been neglected in a preliminary calculation and here perturbation theory is convenient. The final group of aberrations, associated with small mechanical or electrical imperfections in the lenses, is of extreme practical importance; these *mechanical* or *parasitic aberrations* must always be kept small if lens performance is not to be degraded.

As we saw in Part III, it is necessary to distinguish between real and asymptotic coupling between object and image space and this remains true of aberration calculations. To illustrate this, consider a real object immersed in a magnetic field; the image will be affected by the aberrations introduced by the part of the field downstream from the object. If, however, the same magnetic field is used to magnify an intermediate image, the entire

field will contribute to the aberrations and we then characterize the field by the coupling between incident and emergent asymptotes. In the first case, we speak of real aberrations, in the second, of asymptotic aberrations.

The study of electron lens aberrations is almost as old as the electron microscope itself, for the first calculations of aberration coefficients were made in the early 1930s by Scherzer, who preferred the trajectory method, and by Glaser, who introduced the eikonal method into electron optics (Glaser, 1933a–c, 1935, 1936a,b, 1937, 1938, 1949; Scherzer, 1933, 1936b, 1937; see also Funk, 1936, 1937; Gratsiatos, 1936; Rogowski, 1937; Ramberg, 1939; Busch and Brüche, 1937). During that decade, the real aberrations of round lenses were thoroughly studied and several equivalent formulae for the various coefficients were derived. In particular, Scherzer (1936b) demonstrated that the spherical aberration, which is of particular concern since it limits the resolution of microscopes, always has the same sign, whatever the lens design. This finding is so important that it is known as *Scherzer's theorem*, the only named theorem in the subject.

During the 1940s, the aberration coefficients of quadrupoles were obtained by the trajectory method (Melkich, 1947) and in a celebrated paper, again by Scherzer (1947), a number of possible ways of cancelling spherical aberration were adumbrated. Also during this period, Grinberg (1942, 1943a,b, republished 1948) derived highly general ray equations, assuming no particular symmetry, and these, supplemented by the work of Vandakurov (1955a,b, 1956a,b, 1957) and Kas'yankov (1956a,b, 1957, 1958a), have been widely used by Russian authors, notably Strashkevich and Pilat (1951, 1952) and Strashkevich and Gluzman (1954), who made the first attempts to follow Grinberg's approach in the study of aberrations.

The late 1940s and early 1950s saw several substantial contributions. Tretner returned to Scherzer's proof that the spherical aberration coefficient cannot be made to vanish by ingenious lens design, without infringing any of the conditions required by the proof, and established minimum values for this coefficient and for that of chromatic aberration (for which a similar rule holds true), subject to reasonable practical constraints on lens dimensions and excitation (Tretner, 1950, 1954, 1955, 1956 and especially 1959). Sturrock (1951a,b, 1952, 1955) re-examined the eikonal theory of electron lens aberrations and succeeded in placing it on a firm theoretical foundation; he showed how higher order aberrations can be calculated and how the theory can be used to establish asymptotic aberration coefficients. The credit for recognizing the importance of distinguishing between real and asymptotic aberrations goes to Lenz (1956, 1957), however, closely followed by Seman (1958a). Sturrock also examined in detail the effect of small imperfections in magnetic lens construction – small departures from circularity of the bore, for example, – expressing his results in terms of *parasitic* aberration coefficients. These last studies were inspired by the work of Bertein on electrostatic lenses, and simpler analysis had already led Rang (1949a) and Hillier and Ramberg (1947) to introduce the stigmator, a weak

quadrupole designed to cancel the principal parasitic aberration, the axial astigmatism (Sturrock, 1949, 1951b; Bertein, 1947c,e, 1948a).

We have mentioned that the formulae for the aberration coefficients can be cast into many equivalent forms. The technique originally employed for this required partial integration and substitution from the paraxial equations to eliminate second derivatives of $x(z)$ and $y(z)$, using (15.12) for round lenses, for example. This is not only laborious but has the severe disadvantage that the form of the result cannot be predicted: we cannot know, until all the possible equivalent forms of a given coefficient have been established, which terms can be eliminated simultaneously and whether there are any that can never be removed. A procedure that permits us to write down a general expression containing all possible forms of every coefficient was introduced by Seman (1951, 1954, 1955a–c, 1958b) for round lenses and has been extensively used by Hawkes (1966/1967b, 1967b) for quadrupoles and round lenses.

The methods of Sturrock, whose familiarity with the work of Hamilton and of Herzberger (1931) had enabled him to consolidate the fundamental studies of Glaser, were later applied to quadrupole lenses by Hawkes (1965a–c), who derived formulae for all their geometrical aberration coefficients, more compact than those of Melkich (1947). The chromatic aberrations of such lenses were analysed by Kel'man and Yavor (1961), who derived the condition that must be satisfied if a mixed electrostatic–magnetic quadrupole is to be achromatic, subsequently rediscovered by Septier (1963) and generalized by Hawkes (1965c).

In 1963, Verster noticed in the course of his work on 'gauze lenses' (see Chapter 41 of Volume 2) that some aberration coefficients can be written as polynomials in reciprocal magnification, the polynomial coefficients being determined by lens geometry and excitation. Such a representation had been known to the members of the van Heel school (see van Heel, 1949, 1964) but had not hitherto been recognized in electron optics. This observation led Hawkes (1968, 1970b–f) to examine the structure of the asymptotic aberration coefficients and of the real coefficients of Newtonian fields in some detail; formulae were established for the coefficients occurring in the various polynomials and, by expressing the results in matrix form (cf. Brouwer, 1957, 1964), expressions for the aberrations of multiplets could be obtained explicitly. The aberration polynomials were also obtained by Ade (1973, 1982), by a rather different route.

During the 1960s, the basic aberration theory was again reformulated by Rose (1968, 1968/69; Rose and Petri, 1971), who devised a systematic way of handling higher order aberrations, suitable for studying the complicated combinations of electron optical elements needed for practical aberration correction in an electron microscope.

The most far-reaching innovation of the 1960s was, however, a consequence of the increasing availability of powerful computers. With these, it became possible to calculate

optical properties easily and accurately and to apply sophisticated minimization techniques to the task of finding lens combinations possessing some particular property (Moses, 1970, 1971a–c, 1972, 1973, 1974; Rose and Moses, 1973). The 1970s saw the introduction of the finite-element and boundary-element methods for field calculation (Munro, 1970, 1971, 1972, 1973; Adams and Read, 1972a,b; Harrington, 1968; Read et al., 1971; Rauh, 1971), and the extremely difficult problem of calculating electron gun behaviour accurately was gradually solved (see Lauer, 1982; Kasper, 1982; and Part IX). Systems of great complexity can now be analysed in detail and the subject has reached maturity.

Computers have been used not only for numerical solution of field and trajectory equations but also to derive expressions for aberration coefficients. Here, we require them to perform not arithmetic but algebra and a number of special languages have been devised for this purpose. In the early years, REDUCE was widely used (Hearn, 1985; Fitch, 1985) and CAMAL had the great attraction that it was easy to use (Barton and Fitch, 1972; Fitch, 1979, 2009; see Ng, 1979 for an overall view of the subject and Davenport et al. (1988), Grabmeier et al. (2003) or von zur Gathen and Gerhard (2013) for more recent, authoritative texts). CAMAL was employed to calculate aberration coefficients by Hawkes (1977a,b, 1980a, 1983a) and REDUCE by Goto and Soma (1977) and Soma (1977), see Chapter 34, Numerical Calculation of Trajectories, Paraxial Properties and Aberrations, where other specialized programs are mentioned. Today, several such programs are in everyday use, MAPLE and MATHEMATICA for example.

Computer algebra has also been found useful in the latest addition to the family of methods used to study aberrations, in which the properties of Lie algebra and the associated groups are exploited. These techniques were introduced into particle and later light optics by Dragt (Dragt, 1979, 1982, 1987, 1990; Dragt and Forest, 1983, 1986; Douglas and Dragt, 1983); full accounts are to be found in Dragt and Forest (1986), Dragt et al. (1986) and Mondragón and Wolf (1986). The aberration coefficients of electron mirrors could hardly have been established by the modified temporal approach (Sections 18.2 and 28.1) without the use of the specially designed computer algebra program MOPS (Preikszas, 1995).

The foregoing survey is restricted to some of the highlights in the history of electron lens aberration theory. Many contributions of arguably comparable importance have not been mentioned and this account should therefore be regarded as no more than a preliminary glance at the field, charted more fully in the remainder of this part and in the bibliography, where more detailed attributions are to be found. In particular, we have not mentioned the use of lens models and the calculation of explicit forms of the coefficients for the most important of these; references to these are to be found in Part VII on instrumental optics (especially Chapters 35, 36 and 39 of Volume 2), but the rise of fast and accurate numerical methods is accompanied by the fall, not to say the demise, of many of these models, which were introduced precisely because accurate calculation was unthinkably laborious.

CHAPTER 22

Perturbation Theory: General Formalism

We set out from the variational principle obtained from (4.32) and (4.34):

$$\int_{P_1}^{P_2} M\, dz \rightarrow \text{extr.} \tag{22.1}$$

in which we have written

$$\begin{aligned} M &\coloneqq \overline{M}(2m_o e)^{-1/2} \\ &= \hat{\Phi}^{1/2}(1+x'^2_1+x'^2_2)^{1/2} - \eta(A_1 x'_1 + A_2 x'_2 + A_z) \end{aligned} \tag{22.2}$$

for electrons, using (2.19). We adopt the suffix notation here because the coordinates (x_1, x_2, z) may be curvilinear, the optic axis coinciding with the z-axis. At this stage, this simply means that the z-axis is the path of a possible ray. Although much of the text is devoted to systems with a straight axis, whereupon (x_1, x_2, z) become the familiar cartesian coordinates, we shall meet a very important family of curved-axis systems in Part X. We recall that (22.1) asserts that the first-order variation of the integral vanishes when the integration is taken along a ray, provided that the endpoints remain fixed. We now consider the value of this integral for an arbitrary path of integration, and to prevent any possible confusion, we write

$$\tilde{S} \coloneqq \int_{P_1}^{P_2} M\, dz \tag{22.3}$$

reserving S for the value of the integral when the integration is taken over a ray. We find that if the path of integration is altered from $x_j(z)$ to $x_j(z) + \Delta x_j(z)$, then $\tilde{S}$ changes by an amount $\Delta\tilde{S}$, where

$$\Delta\tilde{S} = \int_{z_1}^{z_2} \left(\Delta x_j \frac{\partial M}{\partial x_j} + \Delta x'_j \frac{\partial M}{\partial x'_j} \right) dz \tag{22.4a}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00022-5
© 2018 Elsevier Ltd. All rights reserved.

or integrating by parts,

$$\Delta\tilde{S} = -\int_{z_1}^{z_2} \Delta x_j \left(\frac{d}{dz}\frac{\partial M}{\partial x_j'} - \frac{\partial M}{\partial x_j} \right) dz + \left[\frac{\partial M}{\partial x_j'} \right]_{z_2} \Delta x_j(z_2) - \left[\frac{\partial M}{\partial x_j'} \right]_{z_1} \Delta x_j(z_1) \tag{22.4b}$$

Summation over $j = 1, 2$ is implicit here. Furthermore, the variation is restricted to the surfaces $z = z_1$ and $z = z_2$, so that the z-coordinates of the endpoints of the path of integration are held constant (Fig. 22.1). We know from (22.1) that $\Delta\tilde{S}$ must vanish when $\Delta x_{j1} = \Delta x_{j2} = 0$ ($j = 1, 2$) if the path of integration follows a ray; we have written $\Delta x_j(z_k) =: \Delta x_{jk}$. The ray equations

$$\frac{d}{dz}\frac{\partial M}{\partial x_j'} = \frac{\partial M}{\partial x_j} \tag{22.5}$$

are then satisfied. For this situation, we denote the value of $\tilde{S}$ by S and from (22.4), we have

$$\Delta S(z_1, z_2) = \left(\frac{\partial M}{\partial x_j'} \right)_2 \Delta x_{j2} - \left(\frac{\partial M}{\partial x_j'} \right)_1 \Delta x_{j1} \tag{22.6}$$

Introducing the ray vector

$$p_j = \frac{\partial M}{\partial x_j'} \tag{22.7}$$

(22.6) becomes

$$\Delta S(z_1, z_2) = p_{j2}\Delta x_{j2} - p_{j1}\Delta x_{j1} \tag{22.8}$$

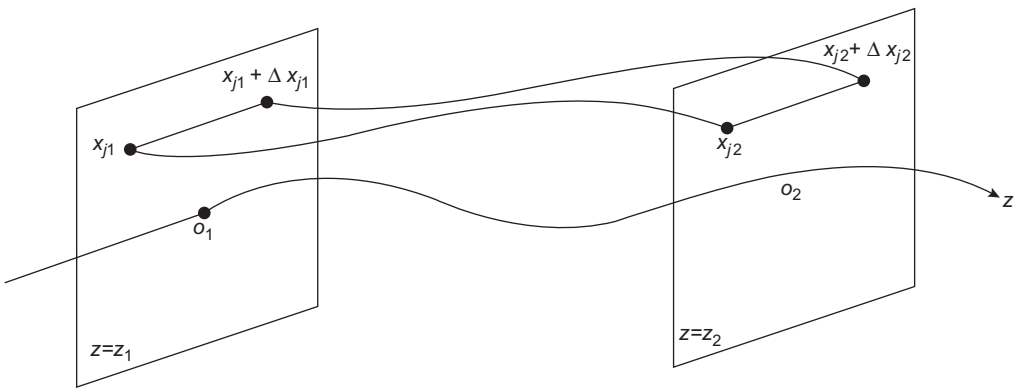


Figure 22.1
Perturbation from $x_j(z)$ to $x_j(z) + \Delta x_j(z)$.

It is convenient to write ΔS_{12} instead of $\Delta S(z_1, z_2)$. We recall that, apart from the trivial change of scale relating S and $\overline{S}$, the function S is Hamilton's point characteristic function and its arguments are x_{j1} and x_{j2}, the word 'point' underlining the fact that these are the coordinates of points P_1 and P_2. Characteristic functions with other arguments are sometimes convenient, such as the angle characteristic already encountered in Section 15.5.

Suppose now that the system is perturbed in some way, and that a parameter λ characterizes this perturbation. We need to introduce two perturbation operators, one of which enables us to represent the perturbed form of a function while the other expresses the effect of perturbing its arguments. For the former, we write the perturbed form of a function f as $P_f f$,

$$P_f = 1 + \lambda P_f^I + \lambda^2 P_f^{II} + \ldots \tag{22.9}$$

For M in particular, we write

$$P_f M = M + \lambda M^I + \lambda^2 M^{II} + \ldots \tag{22.10}$$

For the argumental perturbation operator P_a, we use the Taylor expansion of a function in the form $f(x + \varepsilon) = \exp(\varepsilon\partial/\partial x) f(x)$ so that

$$P_a = \exp\left(\lambda D^I + \lambda^2 D^{II} + \ldots\right) \tag{22.11}$$

where

$$\begin{aligned} D^I &= x_j^I \frac{\partial}{\partial x_j} + x_j' I \frac{\partial}{\partial x_j'} \\ D^{II} &= x_j^{II} \frac{\partial}{\partial x_j} + x_j' II \frac{\partial}{\partial x_j'} \end{aligned} \tag{22.12}$$

The total perturbation is then

$$P = P_a P_f \tag{22.13}$$

We now apply these general remarks to the function S_{12}. Since $S_{12} = \int_{z_1}^{z_2} M\, dz$, we may write

$$PS_{12} = \int_{z_1}^{z_2} PM\, dz \tag{22.14}$$

It is permissible to take the operator P under the integral since the argument z is not varied by P_a. From (22.8), we have

$$\Delta PS_{12} = Pp_{j2} \cdot \Delta Px_{j2} - Pp_{j1} \cdot \Delta Px_{j1} \tag{22.15}$$

Combining (22.10, 22.11 and 22.13), we find

$$\begin{aligned} PS_{12} &= \int_{z_1}^{z_2} \left\{ 1 + \lambda D^I + \lambda^2 D^{II} + \frac{1}{2}\lambda^2 (D^I)^2 + \ldots \right\} \times (M + \lambda M^I + \lambda^2 M^{II} + \ldots)\, dz \\ &= \int_{z_1}^{z_2} \left[M + \lambda (D^I M + M^I) + \lambda^2 \left\{ M^{II} + D^I M^I + D^{II} M + \frac{1}{2}(D^I)^2 M \right\} + \ldots \right] dz \end{aligned} \tag{22.16}$$

In order to keep the notation reasonably compact, we write

$$PS_{12} = S_{12}^{(0)} + \lambda S_{12}^{(1)} + \lambda^2 S_{12}^{(2)} + \ldots \tag{22.17}$$

so that arabic indices (1), (2), . . . indicate the order of *total* perturbation whereas roman numerals I, II, . . . signify the order of *functional* perturbation. Then, comparing (22.16) and (22.17), it is immediately clear that

$$S_{12}^{(1)} = \int_{z_1}^{z_1} (M^I + D^I M)\, dz \tag{22.18}$$

From (22.8), we see that for any pair ξ_j,

$$\int_{z_1}^{z_2} \left(\xi_j \frac{\partial M}{\partial x_j} + \xi_j' \frac{\partial M}{\partial x_j'} \right) dz = p_{j2}\xi_{j2} - p_{j1}\xi_{j1} \tag{22.19}$$

so that for $\xi_j = x_j^{(1)}$ we have

$$\int_{z_1}^{z_2} D^I M\, dz = p_{j2} x_{j2}^{(1)} - p_{j1} x_{j1}^{(1)}$$

and hence

$$S_{12}^{(1)} = p_{j2} x_{j2}^{(1)} - p_{j1} x_{j1}^{(1)} + S_{12}^I \tag{22.20}$$

where

$$S_{12}^{I} \coloneqq \int_{z_1}^{z_2} M^{I}\, dz \tag{22.21}$$

The final function, S_{12}^{I} is known as the *first-order perturbation characteristic function* and we shall use it frequently. It is in fact the functional contribution to S_{12}^{1}, the argumental part coming from $\int_{z_1}^{z_2} D^{I} M dz$.

Returning to (22.15), we have

$$\begin{aligned}\Delta S_{12}^{(1)} = {} & p_{j2}^{(1)} \cdot \Delta x_{j2} + p_{j2} \cdot \Delta x_{j2}^{(1)} \\ & - p_{j1}^{(1)} \cdot \Delta x_{j1} - p_{j1} \cdot \Delta x_{j1}^{(1)}\end{aligned} \tag{22.22}$$

so that from (22.20) and (22.22),

$$\begin{aligned}\Delta S_{12}^{I} = {} & \left(p_{j2}^{(1)} \cdot \Delta x_{j2} - x_{j2}^{(1)} \cdot \Delta p_{j2}\right) \\ & - \left(p_{j1}^{(1)} \cdot \Delta x_{j1} - x_{j1}^{(1)} \cdot \Delta p_{j1}\right)\end{aligned} \tag{22.23}$$

This is a *first-order perturbation relation* but it is much more general in form than we commonly need. We may perturb the rays subject to various constraints, which correspond to particular physical situations. Two choices of constraints are of especial relevance. Consider first the implications of setting

$$x_{j1}^{(1)} = p_{j1}^{(1)} = 0 \tag{22.24}$$

and, for simplicity, suppose that the plane $z = z_1$ lies in field-free space. Then

$$p_j = \hat{\phi}^{1/2} \frac{x_j'}{(1 + x'^2 + y'^2)^{1/2}} \approx \hat{\phi}^{1/2} x_j' \left\{1 - \frac{1}{2}(x'^2 + y'^2)\right\}$$

and writing

$$\begin{aligned} p_j &= p_j^{(p)} + p_j^{(1)} \\ x_j' &= x_j^{(p)'} + x_j^{(1)'} \coloneqq \alpha_j + \alpha_j^{(1)} \end{aligned}$$

we see that

$$\begin{aligned} p_j^{(p)} &= \hat{\phi}^{1/2} \alpha_j \\ p_j^{(1)} &= \alpha_j^{(1)} - \frac{1}{2} \alpha_j \sum_{1}^{2} \alpha_j^2 \end{aligned}$$

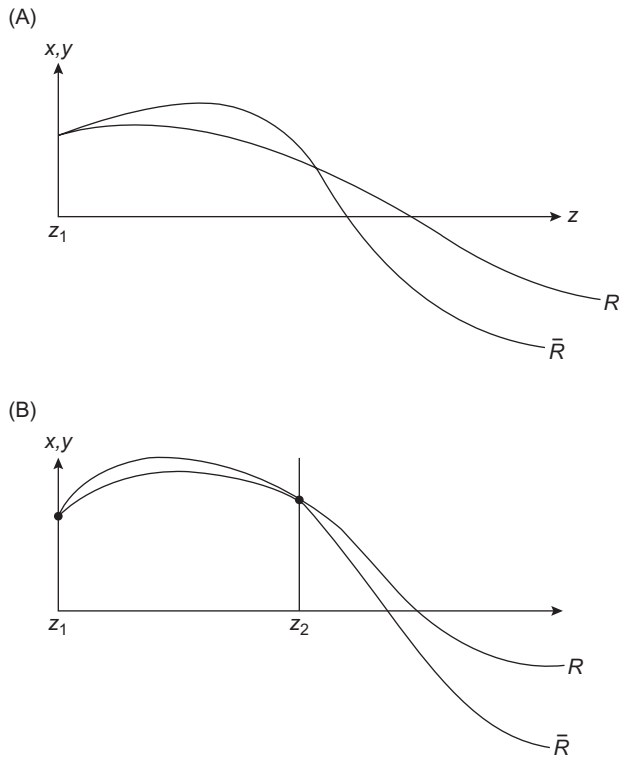


Figure 22.2
The constraints described by (A) (22.24) and (B) by (22.25).

The constraints (22.24) imply that the position of the ray in $z = z_1$ does not change whereas the gradient is altered. Thus in Fig. 22.2A, if R is the unperturbed ray, the perturbed ray $\overline{R}$ sets out from the same point with a slightly different gradient. By using this set of constraints, aberration coefficients expressed in terms of position and gradient in the object plane are obtained, but the aberrations of gradient do not emerge straightforwardly. We shall deal with this small complication in connection with asymptotic aberrations in Chapter 25 (25.1−25.7).

The other physically significant choice of constraints is

$$x_{j1}^{(1)} = x_{j2}^{(1)} = 0 \tag{22.25}$$

so that the unperturbed and perturbed rays must pass through the same point in two different planes, z_1 and z_2. This situation is shown in Fig. 22.2B; the ray $\overline{R}$ is clearly not the same as that in Fig. 22.2A.

Some caution is needed when selecting the constraints, since not all combinations are permissible on every occasion. Thus (22.25) cannot be used when the planes $z = z_1$ and $z = z_2$ are conjugate, since all the rays through a point P_2' in the neighbourhood of P_2 will intersect in a point close to P_1; apart from this point, the neighbourhood of P_1 will not be connected by rays to P_2'.

We shall therefore study (22.23) in two forms. For constraints (22.24), we have

$$\Delta S_{12}^I = p_{j2}^{(1)} \cdot \Delta x_{j2} - x_{j2}^{(1)} \cdot \Delta p_{j2} \tag{22.26}$$

and for (22.25),

$$\Delta S_{12}^I = p_{j2}^{(1)} \Delta x_{j2} - p_{j1}^{(1)} \Delta x_{j1} \tag{22.27}$$

From (22.26), we obtain the very important pair of equations

$$\begin{aligned} \frac{\partial S_{12}^I}{\partial x_{k1}} &= p_{j2}^{(1)} \frac{\partial x_{j2}}{\partial x_{k1}} - x_{j2}^{(1)} \frac{\partial p_{j2}}{\partial x_{k1}} \\ \frac{\partial S_{12}^I}{\partial p_{k1}} &= p_{j2}^{(1)} \frac{\partial x_{j2}}{\partial p_{k1}} - x_{j2}^{(1)} \frac{\partial p_{j2}}{\partial p_{k1}} \end{aligned} \tag{22.28}$$

Equation (22.27) we shall handle slightly differently. We shall use it to express aberrations in some arbitrary plane, frequently the image plane, in terms of position coordinates in the object plane (suffix *o*) and aperture plane (suffix *a)* in a system. Returning to (22.23), we first set $z_1 = z_o$, $x_{ja}^{(1)} = 0$, giving

$$\Delta S_{o2}^I = p_{j2}^{(1)} \cdot \Delta x_{j2} - x_{j2}^{(1)} \cdot \Delta p_{j2} - p_{jo}^{(1)} \cdot \Delta x_{jo} \tag{22.29a}$$

and then set $z_1 = z_a$, $x_{jo}^{(1)} = 0$, giving

$$\Delta S_{a2}^I = p_{j2}^{(1)} \cdot \Delta x_{j2} - x_{j2}^{(1)} \cdot \Delta p_{j2} - p_{ja}^{(1)} \cdot \Delta x_{ja} \tag{22.29b}$$

Hence

$$\begin{aligned} \frac{\partial S_{o2}^I}{\partial x_{ka}} &= p_{j2}^{(1)} \frac{\partial x_{j2}}{\partial x_{ka}} - x_{j2}^{(1)} \frac{\partial p_{j2}}{\partial x_{ka}} \\ \frac{\partial S_{a2}^I}{\partial x_{ko}} &= p_{j2}^{(1)} \frac{\partial x_{j2}}{\partial x_{ko}} - x_{j2}^{(1)} \frac{\partial p_{j2}}{\partial x_{ko}} \end{aligned} \tag{22.30}$$

These may be solved for the unknown quantities $p_{j2}^{(1)}$ and $x_{j2}^{(1)}$. The determinant of (22.30) is given by

$$\Delta = -\frac{\partial(p_{j2}, x_{j2})}{\partial(x_{jo}, x_{ja})} \tag{22.31}$$

and may be expanded as a product of Lagrange brackets,

$$\Delta = -[x_{1o}, x_{2o}][x_{1a}, x_{2a}] + [x_{1o}, x_{1a}][x_{2o}, x_{2a}] - [x_{1o}, x_{2a}][x_{2o}, x_{1a}] \tag{22.32}$$

where $[\lambda, \mu]$ denotes

$$\frac{\partial(p_{12}, x_{12})}{\partial(\mu, \lambda)} + \frac{\partial(p_{22}, x_{22})}{\partial(\mu, \lambda)} \tag{22.33}$$

We have seen that the Lagrange brackets are invariant (5.34) so that Δ is likewise independent of z. Hence

$$\Delta = \frac{\partial(p_{1o}, p_{2o})}{\partial(x_{1a}, x_{2a})} = \frac{\partial(p_{1a}, p_{2a})}{\partial(x_{1o}, x_{2o})} \tag{22.34}$$

If there exists an image plane $z = z_i$ in which the object plane $z = z_o$ is imaged stigmatically, all rays from a point in z_o will be reunited in z_i irrespective of their point of intersection with the plane $z = z_a$:

$$\frac{\partial x_{ki}}{\partial x_{ja}} = 0 \tag{22.35}$$

and, solving (22.30) when (22.35) is satisfied, we find

$$\begin{aligned} x_{1i}^{(1)} &= \frac{1}{\Delta}\left(\frac{\partial p_{2i}}{\partial x_{1a}}\frac{\partial S_{oi}^{I}}{\partial x_{2a}} - \frac{\partial p_{2i}}{\partial x_{2a}}\frac{\partial S_{oi}^{I}}{\partial x_{1a}}\right) \\ x_{2i}^{(1)} &= \frac{1}{\Delta}\left(\frac{\partial p_{1i}}{\partial x_{2a}}\frac{\partial S_{oi}^{I}}{\partial x_{1a}} - \frac{\partial p_{1i}}{\partial x_{1a}}\frac{\partial S_{oi}^{I}}{\partial x_{2a}}\right) \end{aligned} \tag{22.36}$$

with

$$\Delta = \frac{\partial(p_{1i}, p_{2i})}{\partial(x_{1a}, x_{2a})} \tag{22.37}$$

We note that only the perturbation characteristic S_{oi}^{I} appears in (22.36).

Second-order perturbations

Returning to (22.16) and (22.17), we see that

$$S_{12}^{(2)} = \int_{z_1}^{z_2} \left\{ M^{II} + D^I M^I + D^{II} M + \frac{1}{2}(D^I)^2 M \right\} dz \tag{22.38}$$

and using (22.19) with $\xi_j = x_j^{(2)}$, which tells us that

$$\int_{z_1}^{z_2} D^{II} M \, dz = p_{j2} x_{j2}^{(2)} - p_{j1} x_{j1}^{(2)} \tag{22.39}$$

we may write

$$S_{12}^{(2)} = \int_{z_1}^{z_2} \left\{ M^{II} + D^I M^I + \frac{1}{2}(D^I)^2 M \right\} dz + p_{j2} \cdot x_{j2}^{(2)} - p_{j1} \cdot x_{j1}^{(2)} \tag{22.40}$$

The perturbed form of (22.19) enables us to simplify this further. Applying the operator P to (22.19), we find

$$\int_{z_1}^{z_2} \left(\xi_j \frac{\partial (PM)}{\partial x_j} + \xi_j' \frac{\partial (PM)}{\partial x_j'} \right) dz = \xi_{j2} P p_{j2} - \xi_{ji} P p_{j1} \tag{22.41}$$

so that, for $\xi_j = x_j^{(1)}$,

$$\int_{z_1}^{z_2} D^I (M^I + D^I M) \, dz = p_{j2}^{(1)} x_{j2}^{(1)} - p_{j1}^{(1)} x_{j1}^{(1)} \tag{22.42}$$

This enables us to eliminate either $D^I M^I$ or $\frac{1}{2}(D^I)^2 M$ from (22.40); we obtain

$$S_{12}^{(2)} = p_{j2} x_{j2}^{(2)} + p_{j2}^{(1)} x_{j2}^{(1)} - \left(p_{j1} x_{j1}^{(2)} + p_{j1}^{(1)} x_{j1}^{(1)} \right) + \int \left\{ M^{II} - \frac{1}{2}(D^I)^2 \right\} dz \tag{22.43a}$$

or

$$S_{12}^{(2)} = p_{j2} x_{j2}^{(2)} + \frac{1}{2} p_{j2}^{(1)} x_{j2}^{(1)} - \left(p_{j1} x_{j1}^{(2)} + \frac{1}{2} p_{j1}^{(1)} x_{j1}^{(1)} \right) + \int \left\{ M^{II} + \frac{1}{2} D^I M^I \right\} dz \tag{22.43b}$$

and writing

$$\begin{aligned} S_{12}^{II} &= \int (M^{II} - \frac{1}{2}(D^I)^2 M) \, dz \\ \tilde{S}_{12}^{II} &= \int (M^{II} + \frac{1}{2} D^I M^I) \, dz \end{aligned} \tag{22.44}$$

it is easily shown that

$$\begin{aligned}
\Delta S_{12}^{II} &= \left(p_{j2}^{(2)} \cdot \Delta x_{j2} - x_{j2}^{(2)} \cdot \Delta p_{j2}\right) \\
&\quad - \left(p_{j1}^{(2)} \cdot \Delta x_{j1} - x_{j1}^{(2)} \cdot \Delta p_{j1}\right) \\
&\quad - \left(x_{j2}^{(1)} \cdot \Delta p_{j2}^{(1)} - x_{j1}^{(1)} \cdot \Delta p_{j1}^{(1)}\right) \\
\Delta \tilde{S}_{12}^{II} &= \left(p_{j2}^{(2)} \cdot \Delta x_{j2} - x_{j2}^{(2)} \cdot \Delta p_{j2}\right) \\
&\quad - \left(p_{j1}^{(2)} \cdot \Delta x_{j1} - x_{j1}^{(2)} \cdot \Delta p_{j1}\right) \\
&\quad - \frac{1}{2}\left(x_{j2}^{(1)} \cdot \Delta p_{j2}^{(1)} - p_{j2}^{(1)} \cdot \Delta p_{j2}\right) \\
&\quad + \frac{1}{2}\left(x_{j1}^{(1)} \cdot \Delta p_{j1}^{(1)} - p_{j1}^{(1)} \cdot \Delta x_{J1}^{(1)}\right)
\end{aligned} \tag{22.45}$$

Hence

$$\begin{aligned}
p_{j2}^{(2)} \frac{\partial x_{j2}}{\partial x_{ka}} - x_{j2}^{(2)} \frac{\partial p_{j2}}{\partial x_{ka}} &= \frac{\partial S_{o2}^{II}}{\partial x_{ka}} + x_{j2}^{(1)} \frac{\partial p_{j2}^{(1)}}{\partial x_{ka}} \\
p_{j2}^{(2)} \frac{\partial x_{j2}}{\partial x_{ko}} - x_{j2}^{(2)} \frac{\partial p_{j2}}{\partial x_{ko}} &= \frac{\partial S_{o2}^{II}}{\partial x_{ko}} + x_{j2}^{(1)} \frac{\partial p_{j2}^{(1)}}{\partial x_{ko}}
\end{aligned} \tag{22.46}$$

which may be solved for $x_j^{(2)}$ and $p_j^{(2)}$. In an *image* plane, $z = z_i$, the solution collapses to the simpler form

$$\begin{aligned}
x_{1i}^{(2)} &= \frac{1}{\Delta}\left\{\frac{\partial p_{2i}}{\partial x_{1a}}\left(\frac{\partial S_{oi}^{II}}{\partial x_{2a}} + x_{ji}^{(1)} \frac{\partial p_{ji}^{(1)}}{\partial x_{2a}}\right) - \frac{\partial p_{2i}}{\partial x_{2a}}\left(\frac{\partial S_{oi}^{II}}{\partial x_{1a}} + x_{ji}^{(1)} \frac{\partial p_{2a}^{(1)}}{\partial x_{1a}}\right)\right\} \\
x_{2i}^{(2)} &= \frac{1}{\Delta}\left\{\frac{\partial p_{1i}}{\partial x_{2a}}\left(\frac{\partial S_{oi}^{II}}{\partial x_{1a}} + x_{ji}^{(1)} \frac{\partial p_{ji}^{(1)}}{\partial x_{1a}}\right) - \frac{\partial p_{1i}}{\partial x_{1a}}\left(\frac{\partial S_{oi}^{II}}{\partial x_{2a}} + x_{ji}^{(1)} \frac{\partial p_{ji}^{(1)}}{\partial x_{2a}}\right)\right\}
\end{aligned} \tag{22.47}$$

where Δ is the corresponding determinant. (Note that the summation is implicit over j, $j = 1, 2$, whereas i is simply a label signifying "image".)

Equations (22.47) are complicated in appearance but a physical meaning may be associated with the various groups of terms. When we analyse a particular system, we first examine its paraxial properties, then its primary aberrations, characterized essentially by M^I and hence S^I. If we then proceed to the secondary aberrations, a contribution will come from M^{II} and hence S^{II}. A further contribution arises because it is the paraxial solutions that we substitute in M^{II} and these are not the best approximation available since we already know the primary aberrations. This further contribution is the secondary aberration term arising from the primary aberrations.

Before applying these general formulae to particular types of system, we must mention one further degree of complication that can arise. In the case of primary aberrations, those obtained by first-order perturbation theory, the results are additive, in the sense that if M^I consists of two parts, $M^I = M_1^I + M_2^I$, the calculations may be performed separately for ${M_1}^I$ and ${M_2}^I$ and the results added. This is no longer true for second-order perturbations, and we may need to introduce a new parameter, μ, in addition to λ. This typically arises when geometrical and chromatic aberrations are being considered; for primary aberrations, we have merely to calculate ${M_1}^I$ for the geometrical aberrations and ${M_2}^I$ for the chromatic aberrations and add the resulting perturbation terms $x_j^{(1)}$. For the secondary aberrations, there will be mixed terms, arising from the interplay between the two contributions. These are incorporated in the theory in the following way. We denote functional perturbations with respect to μ by the superscript J and $P_f M$ now becomes

$$P_f M = M + \lambda M^I + \mu M^J + \lambda^2 M^{II} + \lambda\mu M^{IJ} + \mu^2 M^{JJ} + \ldots \tag{22.48}$$

The order in which the perturbations are considered is of no importance, as we can readily see on physical grounds. If the contribution to the secondary geometrical aberrations is gradually increased from a negligibly small value, the chromatic contribution remaining fixed, the resulting effects must be the same as those observed if the geometrical aberrations are fixed at their final value and the chromatic aberration gradually increased from a small value.

The formulae for a single perturbation parameter λ given above enable us to manipulate all the terms of an expansion of the form (22.48) except the mixed term $\lambda\mu M^{IJ}$. We merely state the results. The single-parameter reasoning may be followed provided that the following replacements are made:

$$\begin{aligned} M^{II} &\to M^I M^J \\ (D^I)^2 M &\to 2 D^I D^J \\ D^I M^I &\to D^I M^J + D^J M^I \end{aligned} \tag{22.49}$$

We define

$$\begin{aligned} S_{12}^{IJ} &:= \int_{z_1}^{z_2} \left(M^{IJ} - D^I D^J M \right) dz \\ \tilde{S}_{12}^{IJ} &:= \int_{z_1}^{z_2} \left\{ M^I M^J + \frac{1}{2} (D^I M^J + D^J M^I) \right\} dz \end{aligned} \tag{22.50}$$

for which

$$\begin{aligned}
\Delta S_{12}^{IJ} &= -\left(x_{j2}^{I}\cdot\Delta p_{j2}^{J}+x_{j2}^{J}\cdot\Delta p_{j2}^{I}\right)\\
&\quad+\left(x_{j1}^{I}\cdot\Delta p_{j1}^{J}+x_{j1}^{J}\cdot\Delta p_{j1}^{I}\right)\\
&\quad+\left(p_{j1}^{IJ}\cdot\Delta x_{j2}-x_{j2}^{IJ}\cdot\Delta p_{j2}\right)\\
&\quad-\left(p_{j1}^{IJ}\cdot\Delta x_{j1}-x_{j1}^{IJ}\cdot\Delta p_{j1}\right)\\
\Delta\tilde{S}_{12}^{IJ} &= -\frac{1}{2}\left(x_{j2}^{I}\cdot\Delta p_{j2}^{J}+x_{j2}^{J}\cdot\Delta p_{j2}^{I}\right)\\
&\quad+\frac{1}{2}\left(x_{j1}^{I}\cdot\Delta p_{j1}^{J}+x_{j1}^{J}\cdot\Delta p_{j1}^{I}\right)\\
&\quad+\frac{1}{2}\left(p_{j2}^{I}\cdot\Delta x_{j2}^{J}+p_{j2}^{J}\cdot x_{j2}^{I}\right)\\
&\quad-\frac{1}{2}\left(p_{j1}^{I}\cdot\Delta x_{j1}^{J}+p_{j1}^{J}\cdot\Delta x_{j1}^{I}\right)\\
&\quad+\left(p_{j2}^{IJ}\cdot\Delta x_{j2}-x_{j2}^{IJ}\cdot\Delta p_{j2}\right)\\
&\quad-\left(p_{j1}^{IJ}\cdot\Delta x_{j1}-x_{j1}^{IJ}\cdot\Delta p_{j1}\right)
\end{aligned}\tag{22.51}$$

in which we have been obliged to label x_j and p_j with I and J to distinguish the different contributions. Thus x_j^I might represent the primary geometrical aberration, for example, and x_j^J the chromatic aberration.

The mixed forms of (22.46) are as follows:

$$\begin{aligned}
p_{j2}^{(2)}\frac{\partial x_{j2}}{\partial x_{ka}}-x_{j2}^{(2)}\frac{\partial p_{j2}}{\partial x_{ka}} &= \frac{\partial S_{o2}^{IJ}}{\partial x_{ka}}+x_{j2}^{I}\frac{\partial p_{j2}^{J}}{\partial x_{ka}}+x_{j2}^{J}\frac{\partial p_{j2}^{I}}{\partial x_{ka}}\\
p_{j2}^{(2)}\frac{\partial x_{j2}}{\partial x_{ko}}-x_{j2}^{(2)}\frac{\partial p_{j2}}{\partial x_{ko}} &= \frac{\partial S_{o2}^{IJ}}{\partial x_{ko}}+x_{j2}^{I}\frac{\partial p_{j2}^{J}}{\partial x_{ko}}+x_{j2}^{J}\frac{\partial p_{j2}^{I}}{\partial x_{ko}}
\end{aligned}\tag{22.52}$$

Note on nomenclature: We follow Rose in speaking of purely geometric aberrations of different order, of purely chromatic aberrations of different degree and of mixed aberrations of different rank (rank = degree + order).

CHAPTER 23

The Relation Between Permitted Types of Aberration and System Symmetry

23.1 Introduction

The nature of the aberrations of any given system is entirely determined by its symmetry. A classification of the various system types was proposed by Sturrock (1951a) and a detailed exploration of the relation between aberrations and symmetry is to be found in Hawkes (1965a).

Consider a ray passing through an entirely general system, intersecting three planes (or even surfaces) normal to the optic axis, itself in general curved. The point of intersection of the ray with an arbitrary 'current' plane, $z = z_c$, is a function of the point of intersection with the other two planes, which we refer to as the object plane $z = z_o$ and the aperture plane $z = z_a$ (Fig. 23.1). Thus

$$w_c = X_c + \mathrm{i}Y_c = f(w_o, \overline{w}_o, w_a, \overline{w}_a) \tag{23.1}$$

In this chapter, a bar over a symbol denotes its complex conjugate (thus $\overline{w} = X - \mathrm{i}Y$ and likewise for the four-digit notation for aberration coefficients introduced below). Since we are interested in the *optical* behaviour of the system, and hence in rays that remain in the vicinity of the optic axis, we expand w_c as a power series:

$$w_c = \sum_{\alpha,\beta,\gamma,\delta \geq 0} (\alpha\beta\gamma\delta) w_o^{\alpha} \overline{w}_o^{\beta} w_a^{\gamma} \overline{w}_a^{\delta} \tag{23.2}$$

in which the complex coefficient $(\alpha\beta\gamma\delta)$ is a function of z. Suppose now that the system is such that, on simultaneously moving the object point w_o to $w_o \exp(2\pi\mathrm{i}/N)$ and the aperture point w_a to $w_a \exp(2\pi\mathrm{i}/N)$, the point w_c moves to $w_c \exp(2\pi\mathrm{i}/N)$. We then say that the system is N-fold symmetrical. Clearly, not all values of N and shapes of axis are compatible, the higher symmetries only being found with a straight optic axis, for example. From (23.2), it is readily seen that in an N-fold symmetrical system, the indices α, β, γ and δ can take only those nonnegative integral values for which

$$\alpha - \beta + \gamma - \delta = 1 + kN \tag{23.3}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00023-7
© 2018 Elsevier Ltd. All rights reserved.

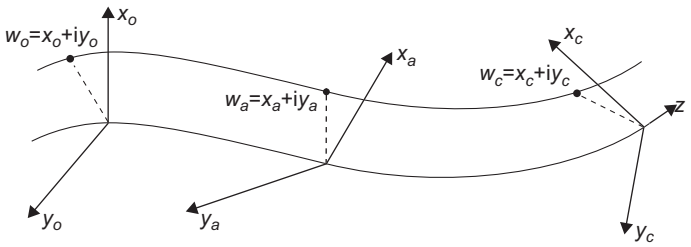


Figure 23.1
General ray intersecting the object, aperture and current surfaces at the points w_o, w_a and w_c respectively.

where k is a positive or negative integer or zero. For rotationally symmetric systems, only the values that satisfy

$$\alpha - \beta + \gamma - \delta = 1 \tag{23.4}$$

are acceptable. The permitted values of α, β, γ and δ for values of N from 1 to 6 are shown in Table 23.1 (Hawkes and Cosslett, 1962; Hawkes, 1965a).

The condition that restricts the values of α, β, γ and δ contains only the combinations $(\alpha + \gamma)$ and $(\beta + \delta)$, and this suggests a simpler method of obtaining the permissible members in the expansion for w_c. We need consider only one of two particular pencils of rays, either the pencil through the axial object point (0, 0, z_o) or the pencil through the axial aperture point (0, 0, z_a). The point w_c can now be written

$$w_c = \sum_{\xi,\eta \geq 0} (\xi\eta) w_o^{\xi} \overline{w}_o^{\eta} \quad \text{or} \quad w_c = \sum_{\xi,\eta \geq 0} (\xi\eta) w_a^{\xi} \overline{w}_a^{\eta} \tag{23.5}$$

and hence

$$\xi - \eta = 1 + kN$$

We now have only to replace w_o or w_a by $w_o + w_a$ for the condition $\alpha - \beta + \gamma - \delta = 1 + kN$ to be automatically satisfied, since the binomial theorem ensures that $\alpha + \gamma = \xi$ and $\beta + \delta = \eta$. In optical terms, we have considered first two particular pencils, of which one could be affected only by distortions, and could not produce an indistinct image, while the other could be perturbed only by wholly aperture-dependent aberrations. To obtain the remaining aberrations we replaced these pencils by off-axial pencils, passing through neither the object nor the aperture origin.

As an example of this, we might consider the secondary aberrations of $N = 2$ systems, which are fifth order. From Table 23.1, we see that there are 56 of these. The condition that $\xi - \eta = 1 + 2k$ subject to $\xi + \eta = 5$ allows $(\xi\eta)$ to take the values (50), (41), (32), (23), (14)

Table 23.1: The values of α, β, γ and δ for $N = 1$ to $N = 6$ and rotational symmetry

Rotational Symmetry: $(\alpha + \gamma) - (\beta + \delta) = 1$

	D	−2	−1	0	1	2	3
I	1000				0010		
III	2100		2001	1011	0021	0120	
					1110		
V	3200	3002	3101	2111	0032	1220	0230
			2012	1022	2210	1031	
					1121		

In all the subsequent lists, the rotationally symmetric terms just tabulated all recur; they are therefore not repeated in every list.

$N = 6$. $(\alpha + \gamma) - (\beta + \delta) = -5, 1$

	D	−5	−4	−3	−2	−1
V	0500	0005	0104	0203	0302	0401

$N = 5$. $(\alpha + \gamma) - (\beta + \delta) = -4, 1$

	D	−4		−3	−2	−1
IV	0400	0004		0103	0202	0301

$N = 4$. $(\alpha + \gamma) - (\beta + \delta) = -3, 1, 5$

	D	−4		−3	−2	−1
III	0300			0003	0102	0201
V	1400	1004		1103	1202	1301
	5000			0014	0113	0212
	0	5	4	3	2	1
V	0311	0050	1040	2030	3020	0410
						4010

$N = 3$. $(\alpha + \gamma) - (\beta + \delta) = -5, -2, 1, 4$

	D	−5	−4	−3	−2	−1
II	0200				0002	0101
IV	1300			1003	1102	1201
	4000				0013	0112
V	0500	0005	0104	0203	0302	0401
	0	4		3	2	1
IV	0211	0040		1030	2020	0310
						3010

$N = 2$. $(\alpha + \gamma) - (\beta + \delta) = \pm 1, \pm 3, \pm 5$

	D	−5	−4	−3	−2	−1		0	5	4	3	2	1
I	0100					0001							
III	3000			0003	1002	0012	III	0111			0030	1020	0210
	0300				0102	0201							2010
	1200					1101							
V	5000	0005	1004	2003	2102	2201	V	1211	0050	1040	1130	0320	1310
	4100		0104	0203	1202	1301		0311		0140	0041	3020	3110
	2300			1103	1013	4001		3011			2030	2120	0221
	1400			0014	0113	1112		0122				1031	2021
	0500				0302	0401							4010
						0023							0410
						0212							

(*Continued*)

Table 23.1: (Continued)

$N = 1.\ (\alpha+\gamma)-(\beta+\delta) = 0,\ \pm1,\ \pm2,\ \pm3,\ \pm4,\ \pm5$												
+ 1 is included in this list												
	D	−5	−4	−3	−2	−1		0	5	4	3	2	1
0	0000												
I	1000					0001	I						0010
	0100												
II	2000				0002	0101	II	0011				0020	1010
	1100					1001							0110
	0200												
III	3000			0003	0102	2001	III	1011			0030	1020	2010
	2100				1002	0201		0111				0120	0210
	1200					1101							1110
	0030					0012							0021
IV	4000		0004	0103	2002	3001	IV	2011		0040	1030	2020	3010
	3100			1003	0202	2101		1111			0130	1120	2110
	2200				1102	1201		0211				0220	1210
	1300				0013	0301		0022				0031	0310
	0400					1012							1021
						0112							0121
V	5000	0005	0104	2003	3002	4001	V	3011	0050	1040	2030	3020	4010
	4100		1004	0203	2102	3101		2111		0140	1130	2120	3110
	3200			1103	1202	2201		1211			0230	1220	2210
	2300			0014	0302	1301		0311			0041	0320	1310
	1400				1013	0401		1022				1031	0410
	0500				0113	2012		0122				0131	2021
						0212							1121
						1112							0221
						0023							0032

Each row contains terms for which $\alpha+\beta+\gamma+\delta$ is the same (roman numeral), and each column all the terms with some particular value of $\gamma-\delta$ (arabic numeral). The column headed '*D*' contains distortions, for which $\gamma=\delta=0$.

and (05). Of these, $(\xi\eta) = (50)$ generates six aberrations and $(\xi\eta) = (14)$ generates ten aberrations; these together are characteristic of $N = 4$. The term $(\xi\eta) = (32)$ generates the 12 aberrations characteristic of rotationally symmetrical systems. The six aberrations characteristic of $N = 6$ (which also afflict $N = 3$ systems, therefore) are generated by $(\xi\eta) = (05)$. Finally, $(\xi\eta) = (41)$ and $(\xi\eta) = (23)$ generate the 22 aberrations peculiar to $N = 2$ systems. The full results are set out in Table 23.2.

A particularly simple way of discussing the effects of each aberration term, $(\alpha\beta\gamma\delta)w_o^{\alpha}\overline{w}_o^{\beta}w_a^{\gamma}\overline{w}_a^{\delta}$, or group of aberration terms, has been explored in detail by Chako (1957), who employed only Cartesian coordinates and considered only rotationally symmetric

Table 23.2: The values of ($\xi\eta$) corresponding to each type of symmetry

	$\xi+\eta=1$	$\xi+\eta=2$	$\xi+\eta=3$	$\xi+\eta=4$	$\xi+\eta=5$
$N=1$	10(R)	20	30(2)	40(3)	50(4)
	01(2)	11	21(R)	31	41(2)
		02(3)	12(2)	22	32(R)
			03(4)	13(3)	23(2)
				04(5)	14(4)
					05(6)
$N=2$	10(R)		30		50(4)
	01		21(R)		41
			12		32(R)
			03(4)		23
					14(4)
					05(6)
$N=3$	10(R)	02	21(R)	40	32(R)
				13	05(6)
$N=4$	10(R)		21(R)		50
			03		32(R)
					14
$N=5$	10(R)			04	32(R)
$N=6$	10(R)				05
rotational symmetry	10		21		32

The bracketed symbols indicate the symmetry class of which the corresponding aberrations are typical. *R* signifies rotationally symmetric.

systems. Amboss (1959) used the same method with complex coordinates to examine slightly imperfect systems. If we wish to interpret geometrically an expression of the form

$$w_c = \sum(\alpha\beta\gamma\delta)w_o^{\alpha}\overline{w}_o^{\beta}w_a^{\gamma}\overline{w}_a^{\delta} \tag{23.6}$$

we can obtain a curve that is typical of each type of aberration on a general plane in image space by considering a pencil of rays that emerges from an object point w_o = constant and is restricted by an annular stop in the aperture plane $|w_a|$ = constant = r_a. These are the curves that Chako calls 'characteristic curves'. We can then write

$$w_c = \sum r_o^{\alpha+\beta}r_a^{\gamma+\delta}|\alpha\beta\gamma\delta|\exp\ \mathrm{i}\left\{\widehat{\alpha\beta\gamma\delta} + (\alpha-\beta)\theta_o + (\gamma-\delta)\theta_a\right\} \tag{23.7}$$

in which $(\alpha\beta\gamma\delta)\ = :|\alpha\beta\gamma\delta|\ \mathrm{e}^{\mathrm{i}\widehat{\alpha\beta\gamma\delta}}$. For each term, therefore, we have

$$w_c = (AB\Gamma\Delta)\mathrm{e}^{\mathrm{i}(\gamma-\delta)\Theta_a} \quad \begin{cases} (AB\Gamma\Delta) = |\alpha\beta\gamma\delta|r_o^{\alpha+\beta}r_a^{\gamma+\delta} \\ \Theta_a = \theta_a + \dfrac{\widehat{\alpha\beta\gamma\delta} + (\alpha-\beta)\theta_o}{\gamma-\delta}. \end{cases} \tag{23.8}$$

If $\gamma = \delta = 0$, we can consider the pencil that passes through a fixed point in the aperture plane and intersects the object plane in an annulus, $|w_o| = r_o = \text{constant}$.

Not all the coefficients $(\alpha\beta\gamma\delta)$ are independent, as we can see by considering the point characteristic function S, which we expand as a power series in X, X', Y and Y':

$$S = S^{(0)} + S^{(1)} + S^{(2)} + \cdots + S^{(r)} + \cdots \tag{23.9}$$

in which each group of terms, $S^{(r)}$, contains only products of degree r in the off-axial coordinates. Of these groups, $S^{(0)}$ is a constant and will not concern us further; $S^{(1)}$ is zero since the axis must also be a possible ray; the Gaussian or first-order (primordial) properties are given by $S^{(2)}$ provided that $N \leq 2$ and the primary, secondary and further aberrations are given by successive nonzero terms of S.

The function S is a definite integral of a physical quantity along a real path between two points; it must therefore be real and must also satisfy the same symmetry conditions as the whole system. This suggests that we might write each of the even components of S as a quadratic or Hermitian form

$$S^{(r)} = Q_r^T R Q_r \quad \text{or} \quad S^{(r)} = H_r^* r H_r \tag{23.10}$$

Q is a column matrix and Q^T its transpose; H is also a column matrix and H^* is the conjugate complex matrix of H^T; R and r are square matrices, which we shall refer to as the coefficient matrices.

We have already used a four-symbol notation $(\alpha\beta\gamma\delta)$ to enable the coefficients of the terms in the series expansion of w_c to be written down directly. We shall employ a similar notation for the elements of the coefficient matrices, R and r. Each element will be denoted by a four-figure symbol $(pqrs)$ that is automatically associated with the term $w_o^p \overline{w}_o^q w_a^r \overline{w}_a^s$. Whereas $\alpha + \beta + \gamma + \delta$ is equal to the order of the corresponding aberration, $p + q + r + s$ is equal to the order of the terms in the characteristic function that correspond to the same aberration, namely $\alpha + \beta + \gamma + \delta + 1$. Since S must be real, we can immediately conclude that

$$(pqrs) = (\overline{qpsr}) \tag{23.11}$$

All the relations that connect the aberration coefficients stem from this property. In the simplest case, that of rotational symmetry, the relation which we encounter below between the two parts of the coma coefficient is a consequence of

$$(1012) = \left(\overline{0121}\right) \tag{23.12}$$

Similarly, the fact that under certain conditions, quadrupole lens systems are fully characterized by three, and not four, aperture aberration coefficients can be deduced from

$$(0013) = \left(\overline{0031}\right) \tag{23.13}$$

The condition used by de Broglie (1950, pp.131−2) to obtain the interrelations for axially symmetric systems can be straightforwardly derived from Eq. (23.11). This relation between the elements of the characteristic function does, however, bring out the fact that de Broglie's condition is a consequence of more fundamental relations which remain true, even though the actual aberration coefficients may no longer be so simply connected.

If the N-fold symmetry of a system is reinforced by a symmetry plane, the coefficients $(pqrs)$ are either all real or all purely imaginary. For, if the system possesses a plane of symmetry, S will be invariant under a change from right-handed to left-handed axes, and hence $(-1)^{p+q+r+s}(pqrs) = (qpsr)$. Since we have already shown that $(pqrs) = \overline{(qpsr)}$, $(pqrs)$ must be real if $p + q + r + s$ is even, and imaginary if $p + q + r + s$ is odd.

For simplicity, we study the aberrations in field-free image space, taking $z = z_a$ as the exit pupil of the system. Then in image space, we have

$$w = w_a + (z - z_a) \left(\frac{\partial w}{\partial z}\right)_{z=z_a} \tag{23.14}$$

or

$$w(z) = \left\{1 + \frac{\mathrm{i}B(z - z_a)}{A}\right\} w_a + 2\frac{z - z_a}{A}\frac{\partial S_{oa}}{\partial \overline{w}_a}$$

in which we have used the fact that the terms in $S^{(2)}$ containing X' and Y' can always be cast into the form $\frac{1}{2}A\left(X'^2 + Y'^2\right) + B(X'Y - XY')$. Thus $w(z)$ has the form

$$w(z) = Cw_a + D\frac{\partial S_{oa}}{\partial \overline{w}_a} \tag{23.15}$$

in which C may be complex and D is real.

It is to be noted that the analysis set out here enables us to identify all aberrations permitted by symmetry but some of these may be forbidden for other reasons. Shao (1987a,b) and Shao and Crewe (1987) pointed out that a $2N$-pole exhibits an axial aberration similar to that of a round lens of order $2N-3$. Thus a sextupole is useful to correct the third-order spherical aberration of a round lens (for $N = 3$, $2N - 3 = 3$) but for a decapole ($N = 5$), $2N - 3 = 7$ and for a 14-pole ($N = 7$), $2N - 3 = 11$. This is confirmed by Rose, who introduces the useful notion of *parity* (Rose, 2012, Section 8.2.1). Thus quadrupoles, octopoles, dodecapoles (N even) have even parity: each electrode (or pole) of a multipole has the same polarity as the electrode opposite it. Sextupoles, decapoles and 14-poles (N odd) have odd parity: the polarity of the opposite electrode is now different.

We now consider the image formation and aberrations for each type of symmetry, as characterized by N.

23.2 N = 1

This is the most general case: the system is wholly arbitrary, the axis may be straight or curved, and few generalizations can be made about the image formation. The function $S^{(2)}_{oa}$ is of the form[1]

$$\begin{pmatrix} w_o \\ \overline{w}_o \\ w_a \\ \overline{w}_a \end{pmatrix}^T \begin{pmatrix} 2000 & \overline{1100}(r) & \overline{1010} & \overline{1001} \\ 0 & \overline{2000} & \overline{1001} & \overline{1010} \\ 0 & 0 & 0020 & \overline{0011}(r) \\ 0 & 0 & 0 & \overline{0020} \end{pmatrix} \begin{pmatrix} w_o \\ \overline{w}_o \\ w_a \\ \overline{w}_a \end{pmatrix} \tag{23.16}$$

and using Eq. (23.15), we find that these coefficients are related to the $(\alpha\beta\gamma\delta)$ of Eq. (23.6) thus:

$$(1000) = D(1001); \quad (0100) = D\left(\overline{1010}\right); \quad (0001) = 2D\left(\overline{0020}\right)$$
$$\Re(0010) = D(0011) + \Re(C); \quad \Im(0010) = \Im(C)$$

To understand what imagery such a system would produce, we consider a pencil of rays emerging from a fixed object point and restricted by an annular stop, $|w_a| = r_a = \text{constant}$. We denote $(1000)w_o + (0100)\overline{w}_o$ by U, and we write $(\alpha\beta\gamma\delta) = |\alpha\beta\gamma\delta| \exp \mathrm{i}\widehat{\alpha\beta\gamma\delta}$ and $w = re^{\mathrm{i}\theta}$. We find

$$w_c = U + r_a|0010|e^{\mathrm{i}\left(\theta_a + \widehat{0010}\right)} + r_a|0001|e^{-\mathrm{i}\left(\theta_a - \widehat{0001}\right)}$$

or

$$(w_c - U)e^{-\frac{1}{2}\mathrm{i}\left(\widehat{0010} + \widehat{0001}\right)} = r_a|0001|e^{\mathrm{i}\Theta_a} + r_a|0010|e^{-\mathrm{i}\Theta_a} \tag{23.17}$$

in which

$$\Theta_a = \theta_a + \frac{1}{2}\left(\widehat{0010} - \widehat{0001}\right)$$

In general, therefore, the pencil will emerge from the system as an astigmatic bundle of rays, which collapses to two mutually perpendicular focal lines in two particular current planes where $|0010| + |0001| = 0$ and $|0010| - |0001| = 0$; these focal lines are inclined to the axes w_c at an angle that depends upon the position of the object point. If a point object approaches the system from infinity, the focal lines may execute any one of three possible manœuvres. They may swivel round the axis of the system, turning always in the same

[1] (r) indicates that the adjoining element is real.

sense; in this case ($\overline{0010} + \overline{0001}$) changes monotonically. Systems in which the focal lines behave in this way were called '*gedrehte'* or *'tordierte Systeme'* by Gullstrand, who first discussed them (1915, p.20). Alternatively, the lines may swivel round the axis until the object point reaches one of the two 'orthogonal points', beyond which they turn in the opposite sense; when the object point reaches the second orthogonal point, the focal lines revert to their original sense of rotation. These are Gullstrand's '*zurückgedrehte'* or *'retordierte Systeme'*. Finally, these two orthogonal points may coincide. Between the two orthogonal points of the preceding type of system, the focal lines unwind through a quarter-turn; in these '*halbgedrehte*' or '*semitordierte Systeme*', the focal lines turn through the same angle instantaneously, with the result that as the object point crosses the coincident orthogonal point, the focal lines appear abruptly to change places, but continue to rotate in the same sense. This behaviour is also discussed by Herzberger (1931, p. 91).

In a stigmatic system, the image of a pair of straight lines $x_o = \text{const}$ and $y_o = \text{const}$ will be

$$y_i = -\frac{\Re(1000-0100)}{\Im(1000-0100)}x_i + \frac{\Re(1000+0100)\Re(1000-0100)+\Im(1000+0100)\Im(1000-0100)}{\Im(1000-0100)}x_o \tag{23.18a}$$

and

$$y_i = \frac{\Im(1000+0100)}{\Re(1000+0100)}x_i + \frac{\Re(1000+0100)\Re(1000-0100)+\Im(1000+0100)\Im(1000-0100)}{\Im(1000+0100)}y_o \tag{23.18b}$$

which represents a pair of straight lines, no longer at right angles; a rectangle is therefore imaged as a parallelogram.

The primary aberrations are second order, and correspond to $S^{(3)}$; using $(pqrs) = (\overline{qpsr})$, we can write the latter in the form

$$S^{(3)} = \begin{pmatrix} w_o \\ \overline{w}_o \\ w_a \\ \overline{w}_a \end{pmatrix}^T \begin{pmatrix} 3000 & 1200 & 0 & 1020 & 1002 & 1011 \\ \overline{1200} & \overline{3000} & 0 & \overline{1002} & \overline{1020} & \overline{1011} \\ 2010 & \overline{2001} & \overline{1110} & 0030 & \overline{0021} & 0 \\ 2001 & \overline{2010} & \overline{1110} & 0021 & \overline{0030} & 0 \end{pmatrix} \begin{pmatrix} w_o^2 \\ \overline{w}_o^2 \\ w_o\overline{w}_o \\ w_a^2 \\ \overline{w}_a^2 \\ w_a\overline{w}_a \end{pmatrix} \tag{23.19}$$

No element is necessarily real.

In an arbitrary image plane, the primary aberrations w_c^I are given by Eq. (22.30); we substitute

$$w_c = (1000)w_o + (0100)\overline{w}_o + (0010)w_a + (0001)\overline{w}_a$$

$$p_1 + \mathrm{i}p_2 = Aw_c' - \mathrm{i}Bw_c$$

and obtain

$$Dw_c^I = 2A \begin{vmatrix} 0010 & 0001 & 0001 & \partial S_{oc}^{(3)}/\partial \overline{w}_a \\ 1000 & 0100 & 0100 & \partial S_{ac}^{(3)}/\partial \overline{w}_o \\ 0001 & 0010 & 0010 & \partial S_{oc}^{(3)}/\partial w_a \\ 0100 & 1000 & 1000 & \partial S_{ac}^{(3)}/\partial w_o \end{vmatrix} \tag{23.20}$$

Each of the elements in the fourth column is composed of 10 members. For further details, see Hawkes (1965a).

23.2.1 N = 1. Systems with a Plane of Symmetry

The axis of the system is now a curve lying in a plane, and we shall suppose this to be the plane which also contains the y-axis. Certain deflection systems and β-spectrometers fall within this class. The primary aberrations are still second order, but the function $S^{(3)}$ must be independent of the signs of x_o and x_a; it is therefore of the form

$$\begin{pmatrix} x_o^2 \\ x_o x_a \\ x_a^2 \\ y_o^2 \\ y_a^2 \end{pmatrix}^T \begin{pmatrix} 2100 & 2001 \\ 1110 & 1011 \\ 0120 & 0021 \\ 0300 & 0201 \\ 0102 & 0003 \end{pmatrix} \begin{pmatrix} y_o \\ y_a \end{pmatrix} \tag{23.21}$$

and so

$$\begin{aligned} k_x x_c^I &= h_{xc}\{2[2100]x_o y_o + 2[2001]x_o y_a + [1110]y_o x_a + [1011]x_a y_a\} \\ &\quad - g_{xc}\{(1110)x_o y_o + (1011)x_o y_a + 2(0120)y_o x_a + 2(0021)x_a y_a\} \\ k_y y_c^I &= h_{yc}\{[2100]x_o^2 + [1110]x_o x_a + [0120]x_a^2 \\ &\quad + 3[0300]y_o^2 + 2[0201]y_o y_a + [0120]y_a^2\} \\ &\quad - g_{yc}\{(2001)x_o^2 + (1011)x_o x_a + (0021)x_a^2 \\ &\quad + (0201)y_o^2 + 2(0102)y_o y_a + 3(0003)y_a^2\} \end{aligned} \tag{23.22}$$

In a general plane, therefore, ten coefficients suffice to characterize the imagery, and in the image plane of a stigmatic system, eight.

The aperture aberrations are of the form

$$x_c^I = (0011)x_a y_a, \quad y_c^I = (0020)x_a^2 + (0002)y_a^2 \tag{23.23}$$

which represents an ellipse, centred on the point $(0, \frac{1}{2}\{(0020) + (0002)\}r_a^2)$ with axes $\frac{1}{2}r_a^2(0011)$, $\frac{1}{2}r_a^2\{(0020) - (0002)\}$. For a circular aperture, the envelope of this family of ellipses is a pair of straight lines, inclined to the y-axis at an angle $\operatorname{arccot} 2\sqrt{(0020)(0002)}/(0011)$. In a stigmatic orthomorphic system, $(0011) = 2(0020)$, and the angle becomes $\operatorname{arccot} \sqrt{(0002/(0020)}$ or $\operatorname{arccot} \sqrt{3(0003/(0021)}$.

The distortions are similar in nature; the remaining aberrations are types of astigmatism, and the aberration curve for a fixed object point and an annular aperture is a tilted centred ellipse.

The secondary aberrations are third order; the contribution from $S^{(4)}$ is analysed in Hawkes (1965a).

23.3 N = 2

Systems for which $N = 2$ and there is no symmetry plane may consist of any combination of round lenses and quadrupole lenses in any orientation; the lenses may be either electrostatic or magnetic and the electrostatic lenses need not be excited symmetrically nor need they be geometrically symmetric. The aberrations of quadrupole systems in which the azimuthal alignment of the individual members is imperfect thus fall within this class. The Gaussian imagery is no simpler than that of general systems, but the primary aberrations are now third order. Both the primary and secondary (fifth-order) aberrations are analysed in detail in Hawkes (1965a).

Electron optical systems with straight axes may be orthogonal without possessing a plane of symmetry. These are the systems which comprise Dušek's '*erster Hauptfall*' (1959) and since they involve a complicated and delicate balance of electric and magnetic forces, they have as yet found no practical employment. The surfaces in which electrons experience no expulsive force are not planes but curved surfaces, which twist about the axis within the lens fields.

Since such systems are orthogonal, it is simpler to use Cartesian coordinates, and $S^{(4)}$ can therefore be written

$$S^{(4)} = \begin{pmatrix} x_o^2 \\ y_o^2 \\ x_a^2 \\ y_a^2 \\ x_o x_a \end{pmatrix}^T A \begin{pmatrix} x_o^2 \\ y_o^2 \\ x_a^2 \\ y_a^2 \\ x_o x_a \\ y_o y_a \\ x_o y_a \\ x_a y_o \\ x_o y_o \\ x_a y_a \end{pmatrix} \tag{23.24}$$

where

$$A = \begin{pmatrix} 4000 & 2200 & 2020 & 2002 & 3010 & 2101 & 3001 & 2110 & 3100 & 2011 \\ 0 & 0400 & 0220 & 0202 & 1210 & 0301 & 1201 & 0310 & 1300 & 0211 \\ 0 & 0 & 0040 & 0022 & 1030 & 0121 & 1021 & 0130 & 1120 & 0031 \\ 0 & 0 & 0 & 0004 & 1012 & 0103 & 1003 & 0112 & 1102 & 0013 \\ 0 & 0 & 0 & 0 & 0 & 1111 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The aberrations associated with perturbation characteristic functions of this form resemble the secondary aberrations of $N = 1$ systems.

The aperture aberrations are of the form

$$\begin{aligned} x_c^I &= \frac{h_{xc}}{k_x}\left\{[1030]x_a^3 + [1012]x_a^2 y_a + [1012]x_a y_a^2 + [1003]y_a^3\right\} \\ &\quad - \frac{g_{xc}}{k_x}\left\{4(0040)x_a^3 + 3(0031)x_a^2 y_a + 2(0022)x_a y_a^2 + (0013)y_a^2\right\}, \\ y_c^I &= \frac{h_{yc}}{k_y}\left\{[0130]x_a^3 + [0121]x_a^2 y_a + [0112]x_a y_a^2 + [0103]y_a^3\right\} \\ &\quad - \frac{g_{yc}}{k_y}\left\{(0031)x_a^3 + 2(0022)x_a^2 y_a + 3(0013)x_a y_a^2 + 4(0004)y_a^4\right\} \end{aligned} \tag{23.25}$$

When the imagery is stigmatic, the aberration curve in the stigmatic image plane simplifies to the form

$$\begin{aligned} -Ah'_{xi}x_i^I &= \alpha_1 x_a^3 + 3\beta x_a^2 y_a + \gamma x_a y_a^2 + \delta y_a^3 \\ -Ah'_{yi}y_i^I &= \beta x_a^3 + \gamma x_a^2 y_a + 3\delta x_a y_a^2 + \alpha_2 y_a^3 \end{aligned}$$

The secondary aberrations involve $S^{(6)}$.

23.3.1 N = 2. Systems Possessing a Plane of Symmetry

Systems belonging to this class are automatically orthogonal; they represent the special case of the preceding class for which the curved orthogonal surfaces collapse into a pair of mutually perpendicular planes (Dušek's '*verdrehungsfreie Orthogonalsysteme*'). The elements of such systems may be electrostatic and magnetic quadrupole lenses, and round electrostatic lenses; all the electrodes of the electrostatic quadrupoles must lie in the same pair of (mutually perpendicular) azimuthal planes, however, and all the polepieces of the magnetic quadrupoles must lie in the pair of azimuthal planes that are inclined to the electrode planes at 45°. The electrodes need be symmetrical in neither excitation nor geometry, provided of course that any asymmetry is compatible with the symmetry plane.

In complex notation, the presence of the symmetry plane implies that all the elements of the coefficient matrices are real; in the Cartesian notation, $S^{(4)}$ and $S^{(6)}$ contain only even powers of x and even powers of y. Examining Eq. (23.24), we can see that $S^{(4)}$ is obtained by selecting the first six columns of the coefficient matrix, and retaining the first six elements of the ten-element column matrix, thus:

$$S^{(4)} = \begin{pmatrix} x_o^2 \\ y_o^2 \\ x_a^2 \\ y_a^2 \\ x_o x_a \end{pmatrix}^T \begin{pmatrix} 4000 & 2200 & 2020 & 2002 & 3010 & 2101 \\ 0 & 0400 & 0220 & 0202 & 1210 & 0301 \\ 0 & 0 & 0040 & 0022 & 1030 & 0121 \\ 0 & 0 & 0 & 0004 & 1012 & 0103 \\ 0 & 0 & 0 & 0 & 0 & 1111 \end{pmatrix} \begin{pmatrix} x_o^2 \\ y_o^2 \\ x_a^2 \\ y_a^2 \\ x_o x_a \\ y_o y_a \end{pmatrix} \tag{23.26}$$

The primary aberrations in a general plane are therefore given by

$$\begin{aligned} k_x x_c^I = {} & h_{xc}\{4[4000]x_0^3\} + 2[2200]x_o y_o^2 + [1030]x_a^3 + [1012]x_a y_a^2 \\ & + x_a\left(3[3010]x_0^2 + [1210]y_0^2\right) + 2[2101]x_o y_o y_a \\ & + 2[2020]x_o x_a^2 + 2[2002]x_o y_a^2 + [1111]y_o x_a y_a\} \\ & - g_{xc}\{(3010)x_o^3 + (1210)x_o y_o^2 + 4(0040)x_a^3 + 2(0022)x_a y_a^2 \\ & + x_a(2(2020)x_o^2 + 2(0220)y_o^2) + (1111)x_o y_o y_a \\ & + 3(1030)x_o x_a^3 + (1012)x_o y_a^2 + 2(0121)y_o x_a y_a\} \\ k_y y_c^I = {} & h_{yc}\{2[2200]x_o^2 y_o + 4[0400]y_o^3 + [0121]x_a^2 y_a + [0103]y_a^3 \\ & + y_a([2101]x_o^2 + 3[0301]y_o^2) + 2[1210]x_o y_o x_a \\ & + 2[0220]y_o x_a^2 + 2[0202]y_o y_a^2 + [1111]x_o x_a y_a\} \\ & - g_{yc}\{(2101)x_o^2 y_o + (0301)y_o^3 + 2(0022)x_a^2 y_a + 4(0004)y_a^3 \\ & + y_a(2(2002)x_o^2 + 2(0202)y_o^2) + (1111)x_o y_o x_a \\ & + (0121)y_o x_a^2 + 3(0103)y_o y_a^2 + 2(1012)x_o x_a y_a\} \end{aligned} \tag{23.27}$$

In a stigmatic system, there exists an image plane in which $h_x(z_i) = h_y(z_i) = 0$; as we have remarked earlier, the aperture aberrations are then described by only three coefficients:

$$\begin{aligned} A h'_{xi} x_i^I &= -4(0040)x_a^3 - 2(0022)x_a y_a^2 \\ A h'_{yi} y_i^I &= -2(0022)x_a^2 y_a - 4(0004)y_a^3 \end{aligned}$$

provided the slopes of $h_x(z)$ and $h_y(z)$ are equal at the image plane. $S^{(6)}$ is given by

$$\begin{aligned} S^{(6)} = {} & (x_o^3 \;\; y_o^3 \;\; x_a^3 \;\; y_a^3 \;\; x_o y_o x_a \;\; x_o y_o y_a \;\; x_o x_a y_a \;\; y_o x_a y_a) \\ & VI\left(x_o^3 \;\; y_o^3 \;\; x_a^3 \;\; y_a^3 \;\; x_o^2 y_o \;\; x_o^2 x_a \;\; x_o^2 y_a \;\; y_o^2 x_o \;\; y_o^2 x_a \;\; y_o^2 y_a \right. \\ & \left. x_a^2 x_o \;\; x_a^2 y_o \;\; x_a^2 y_a \;\; y_a^2 x_o \;\; y_a^2 y_o \;\; y_a^2 x_a \;\; x_o y_o x_a \;\; x_o y_o y_a \;\; x_o x_a y_a \;\; y_o x_a y_a\right)^T \end{aligned} \tag{23.28}$$

and

$$VI^T = \begin{pmatrix} 6000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0600 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3030 & 0 & 0060 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0303 & 0 & 0006 & 0 & 0 & 0 & 0 \\ 0 & 2400 & 0 & 2103 & 0 & 0 & 0 & 0 \\ 5010 & 0 & 2040 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2301 & 0 & 2004 & 0 & 0 & 0 & 0 \\ 4200 & 0 & 1230 & 0 & 0 & 0 & 0 & 0 \\ 3210 & 0 & 0240 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0501 & 0 & 0204 & 0 & 0 & 0 & 0 \\ 4020 & 0 & 1050 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0420 & 0 & 0123 & 0 & 0 & 0 & 0 \\ 0 & 0321 & 0 & 0024 & 0 & 0 & 0 & 0 \\ 4002 & 0 & 1032 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0402 & 0 & 0105 & 0 & 0 & 0 & 0 \\ 3012 & 0 & 0042 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1410 & 0 & 1113 & 2220 & 0 & 0 & 0 \\ 4101 & 0 & 1131 & 0 & 0 & 2202 & 0 & 0 \\ 0 & 1311 & 0 & 1014 & 2121 & 0 & 2022 & 0 \\ 3111 & 0 & 0141 & 0 & 0 & 1212 & 0 & 0222 \end{pmatrix} \tag{23.29}$$

23.4 N = 3

Systems for which $N = 3$ may contain any number of round lenses, together with elements of a new kind; these latter may consist of a diaphragm with a triangular opening, for example, or of a symmetric sextupole. Such a device has been employed by Amboss (1959) in an attempt to combat 'anticoma'. If the system does contain round lenses, then the first-order properties will be those of an ordinary rotationally symmetric system and the primary aberrations will be due to the three-fold symmetric element alone. If the system consists only of sextupolar elements, the properties of lowest order arise from $S^{(3)}$ and the primary aberrations from $S^{(4)}$.

When both sextupole elements and round lenses are present, the image-forming properties of the system are described by

$$w_c = (1000)w_o + (0010)w_a \tag{23.30}$$

in which (1000) is real, and equal to $g(z)$, and (0010), also real, is equal to $h(z)$.

The component $S^{(3)}$ of S is of the form

$$S^{(3)} = \begin{pmatrix} w_o^2 \\ \overline{w}_o^2 \\ w_a^2 \\ \overline{w}_a^2 \end{pmatrix}^T \begin{pmatrix} 3000 & \overline{0} & 2010 & \overline{0} \\ 0 & \overline{3000} & 0 & \overline{2010} \\ 1020 & \overline{0} & 0030 & \overline{0} \\ 0 & \overline{1020} & 0 & \overline{0030} \end{pmatrix} \begin{pmatrix} w_o \\ \overline{w}_o \\ w_a \\ \overline{w}_a \end{pmatrix} \tag{23.31}$$

so that

$$\begin{aligned} Dw_c^I &= h\frac{\partial S_{ac}^{(3)}}{\partial \overline{w}_o} - g\frac{\partial S_{oc}^{(3)}}{\partial \overline{w}_a} \\ &= h\left\{3[\overline{3000}]\overline{w}_o^2 + 2[\overline{2010}]\overline{w}_o w_a + [\overline{1020}]\overline{w}_a^2\right\} \\ &\quad - g\left\{\left(\overline{2010}\right)\overline{w}_o^2 + 2\left(\overline{1020}\right)\overline{w}_o w_a + \left(\overline{0030}\right)\overline{w}_a^2\right\} \end{aligned} \tag{23.32}$$

In the Gaussian image plane, $h(z)$ vanishes. In any plane, however, there are three primary aberrations: a distortion, $(0200)\overline{w}_o^2$, where

$$(0200) = \frac{h}{D}3[\overline{3000}] - \frac{g}{D}\left(\overline{2010}\right) \tag{23.33a}$$

an aperture aberration, $(0002)\ \overline{w}_a^2$,

$$(0002) = \frac{h}{D}[\overline{1020}] - \frac{g}{D}3\left(\overline{0030}\right) \tag{23.33b}$$

and an astigmatism, $(0101)\overline{w}_o w_a$,

$$(1010) = \frac{h}{D}2[\overline{2010}] - \frac{g}{D}2\left(\overline{1020}\right) \tag{23.33c}$$

We have so far made no assumptions about the alignment of the triangular elements, and the aberration coefficients can therefore all be complex. If there is only one such element, however, or if corresponding points of different elements all lie in the same meridian plane, then all the coefficients will be imaginary, provided the y-axes (say) lie in this plane.

The secondary aberrations will be the same as the primary aberrations of the round lenses, in nature at least; their values will be modified, however, by the presence of the second-order (primary) aberrations. This property, to which Hawkes drew attention in 1965, is exploited to correct the spherical aberration of round lenses (see Chapter 41).

If no round lenses are present, we obtain a system from which all the familiar characteristics of a lens system have vanished. The effect of lowest order is no longer a

combination of anisotropic magnification, defocusing and astigmatism, as in the most general cases of $N = 1$ and $N = 2$ systems. Instead we find

$$w_c = (0200)\overline{w}_o^2 + (0002)\overline{w}_a^2 + (0101)\overline{w}_o\overline{w}_a \tag{23.34}$$

so that even if (0002) and (0101) can be reduced simultaneously to zero, the magnification is not linear.

The primary aberrations will be third order and the secondary aberrations fourth.

23.5 N = 4

Apart from rotationally symmetric lenses, the system may contain any number of electrostatic or magnetic octopoles, in any orientation. The Gaussian imagery is identical with that of ordinary round systems, and the primary aberrations are now third order. The corresponding component of S, namely $S^{(4)}$, is most compactly written as the Hermitian form

$$S^{(4)} = \begin{pmatrix} w_o^2 \\ \overline{w}_o^2 \\ w_a^2 \\ \overline{w}_a^2 \\ w_o w_a \end{pmatrix}^T \begin{pmatrix} \overline{2200} & 4000 & 2002 & 2020 & \overline{2101} \\ \overline{4000} & 0 & 0 & 0 & \overline{3010} \\ \overline{2002} & 0 & \overline{0022} & 0040 & \overline{1012} \\ \overline{2020} & 0 & \overline{0040} & 0 & \overline{1030} \\ \overline{2101} & 3010 & 1012 & 1030 & 1111 \end{pmatrix} \begin{pmatrix} \overline{w}_o^2 \\ w_o^2 \\ \overline{w}_a^2 \\ w_a^2 \\ \overline{w}_o\overline{w}_a \end{pmatrix} \tag{23.35}$$

The aberrations in a general plane are therefore given by

$$\begin{aligned} Dw_c^I &= h_c \frac{\partial S_{ac}^{(4)}}{\partial \overline{w}_o} - g_c \frac{\partial S_{oc}^{(4)}}{\partial \overline{w}_a} \\ &= h \begin{pmatrix} w_o^2 \\ \overline{w}_o^2 \\ w_a^2 \\ \overline{w}_a^2 \\ w_o w_a \end{pmatrix}^T \begin{pmatrix} 2[2200] & [2101] \\ 4[\overline{4000}] & 3[\overline{3010}] \\ 2[\overline{2002}] & [\overline{1012}] \\ 2[\overline{2020}] & [\overline{1030}] \\ 2[\overline{2101}] & [1111] \end{pmatrix} \begin{pmatrix} \overline{w}_o \\ \overline{w}_a \end{pmatrix} \\ &\quad - g \begin{pmatrix} w_o^2 \\ \overline{w}_o^2 \\ w_a^2 \\ \overline{w}_a^2 \\ w_o w_a \end{pmatrix}^T \begin{pmatrix} (2101) & 2(2002) \\ (\overline{3010}) & 2(\overline{2020}) \\ (\overline{1012}) & 2(0022) \\ 3(\overline{1030}) & 4(\overline{0040}) \\ (1111) & 2(\overline{1012}) \end{pmatrix} \begin{pmatrix} \overline{w}_o \\ \overline{w}_a \end{pmatrix} \end{aligned} \tag{23.36}$$

There are therefore two aperture aberrations: $(0003)\overline{w}_a^3$ and the ordinary spherical aberration of round lenses, $(0021)w_a^2\overline{w}_a$; two distortions: $(0300)\overline{w}_o^3$ and the round lens distortion,

$(2100)w_o^2\overline{w}_o$; two astigmatisms: $(0201)\overline{w}_o^2\overline{w}_a$ and the round lens astigmatism, $(2001)\ w_o^2\overline{w}_a$; the round lens coma terms $(1011)w_o w_a\overline{w}_a$ and $(0120)\overline{w}_o w_a^2$ together with an 'anticoma' term $(0102)\overline{w}_o\overline{w}_a^2$; and finally, the round lens field curvature, $(1110)w_o\overline{w}_o w_a$.

If all the elements are aligned in such a way that the system possesses a plane of symmetry—this implies that one electrode of each electrostatic element lies in a single azimuthal plane, and a polepiece of each magnetic element in a plane which is inclined to the electrode plane at $22\frac{1}{2}^\circ$—all the elements of the coefficient matrix will be real.

The secondary aberrations are fifth order. $S^{(6)}$ can be written as a Hermitian form:

$$S^{(6)} = \begin{pmatrix} w_o^3 \\ w_a^3 \\ w_o\overline{w}_o^2 \\ w_o\overline{w}_a^2 \\ \overline{w}_o^2 w_a \\ w_a\overline{w}_a^2 \\ w_o\overline{w}_o\overline{w}_a \\ \overline{w}_o w_a\overline{w}_a \end{pmatrix}^T A \begin{pmatrix} \overline{w}_o^3 \\ \overline{w}_a^3 \\ w_o^2\overline{w}_o \\ \overline{w}_o w_a^2 \\ w_o^2\overline{w}_a \\ w_a^2\overline{w}_a \\ w_o\overline{w}_o w_a \\ w_o w_a\overline{w}_a \end{pmatrix} \tag{23.37}$$

where

$$A = \begin{pmatrix} 3300(r) & 3003 & 5100 & 3120 & 5001 & 3021 & 4110 & 4001 \\ \overline{3003} & 0033(r) & 2130 & \overline{1005} & 2031 & 0051 & 1140 & 1041 \\ \overline{5100} & \overline{2130} & 0 & \overline{3102} & 3201 & \overline{2112} & 0 & 0 \\ \overline{3120} & 1005 & 3102 & 1122(r) & 0 & 1023 & 0 & 2013 \\ \overline{5001} & \overline{2031} & \overline{3201} & 0 & 2211(r) & 0 & 0 & 0 \\ \overline{3021} & \overline{0051} & 2112 & \overline{1023} & 0 & 0 & 0 & 0 \\ \overline{4110} & \overline{1140} & 0 & 0 & 0 & 0 & 0 & 0 \\ \overline{4011} & \overline{1041} & 0 & \overline{2013} & 0 & 0 & 0 & 0 \end{pmatrix}$$

The derivatives of $S^{(6)}$ with respect to $\overline{w}_o$ and $\overline{w}_a$ can each be written in the form

$$\begin{pmatrix} w_o^3 \\ w_a^3 \\ w_o\overline{w}_o^2 \\ w_o\overline{w}_a^2 \\ \overline{w}_o^2 w_a \\ w_a\overline{w}_a^2 \\ w_o\overline{w}_o\overline{w}_a \\ \overline{w}_o w_a\overline{w}_a \end{pmatrix}^T \Upsilon \begin{pmatrix} w_o^2 \\ \overline{w}_o^2 \\ w_a^2 \\ \overline{w}_a^2 \\ w_o w_a \end{pmatrix} \tag{23.38}$$

For $\partial S_{ac}^{(6)}/\partial \overline{w}_o$,

$$\Upsilon = \begin{pmatrix} [5100] & 3[3300] & [3120] & [3102] & [4110] \\ [2130] & 3[\overline{3003}] & [\overline{1005}] & [\overline{1023}] & [1140] \\ 0 & 5[\overline{5100}] & 3[\overline{3102}] & 3[\overline{3120}] & 3[\overline{3201}] \\ 0 & 0 & [1122] & [\overline{1140}] & [2112] \\ 0 & 5[\overline{5001}] & 0 & 3[\overline{3021}] & 0 \\ 0 & 0 & 0 & [\overline{1041}] & 0 \\ 2[3201] & 4[4110] & 2[\overline{2112}] & 2[\overline{2130}] & 2[2211] \\ 0 & 4[4011] & 2[\overline{2013}] & 2[\overline{2031}] & 0 \end{pmatrix}$$

and for $\partial S_{oc}^{(6)}/\partial \overline{w}_a$,

$$\Upsilon = \begin{pmatrix} (5001) & (3201) & (3021) & 3(3003) & (4011) \\ (2031) & (\overline{2013}) & (0051) & 3(0033) & (1041) \\ 0 & (\overline{4110}) & (\overline{2112}) & 3(\overline{2130}) & (2211) \\ 0 & 0 & 3(1023) & 5(1005) & 3(2013) \\ 0 & (\overline{4011}) & 0 & 3(\overline{2031}) & 0 \\ 0 & 0 & 0 & 5(\overline{0051}) & 0 \\ 2(3102) & 2(\overline{3120}) & 2(1122) & 4(\overline{1140}) & 2(2112) \\ 0 & 2(\overline{3021}) & 2(\overline{1023}) & 4(\overline{1041}) & 0 \end{pmatrix} \tag{23.39}$$

These aberrations are all members of the class $N = 2$, and we shall therefore discuss them no further.

23.6 N = 5 and 6

(a) $N = 5$. A system possessing this symmetry would consist of 'decapole' elements: these could be produced in the form of ordinary lenses with pentagonal openings instead of round ones, or as elements with ten poles (or electrodes) symmetrically disposed about the axis. Round lenses could of course be present also, and since the function of the decapole element would probably be to correct or diminish aberrations, the primordial properties and primary aberrations of any practical $N = 5$ system would be most likely to be those of a round lens system. The secondary aberrations arising from the decapole unit would then be fourth order and, just as the quadrupole action of a stigmator is used to annul the parasitic astigmatism of round lens systems, these fourth-order aberrations are potentially capable of combating either the seventh-order axial aberrations of an axially symmetric system, or certain mechanical aberrations due to constructional shortcomings of the system. Note that for $N = 5$, $2N - 3 = 7$.

We shall not discuss this case in any detail; we simply state that the component $S^{(5)}$ which leads to these fourth-order aberrations is of the form

$$S^{(5)} = \begin{pmatrix} w_o^4 \\ w_o^2\overline{w}_a^2 \\ w_a^4 \\ \overline{w}_o^4 \\ \overline{w}_o\overline{w}_a^2 \\ \overline{w}_a^4 \end{pmatrix}^T \begin{pmatrix} 5000 & 4010 & 0 & 0 \\ 3020 & 2030 & 0 & 0 \\ 1040 & 0050 & 0 & 0 \\ 0 & 0 & \overline{5000} & \overline{4010} \\ 0 & 0 & \overline{3020} & \overline{2030} \\ 0 & 0 & \overline{1040} & \overline{0050} \end{pmatrix} \begin{pmatrix} w_o \\ w_a \\ \overline{w}_o \\ \overline{w}_a \end{pmatrix} \tag{23.40}$$

so that

$$\partial S_{ac}^{(5)}/\partial\overline{w}_o = 5[\overline{5000}]\overline{w}_o^4 + 4[\overline{4010}]\overline{w}_o^3\overline{w}_a + 3[\overline{3020}]\overline{w}_o^2\overline{w}_a^2 + 2[\overline{2030}]\overline{w}_o\overline{w}_a^3 + [\overline{1040}]\overline{w}_a^4$$

and

$$\partial S_{oc}^{(5)}/\partial\overline{w}_a = (\overline{4010})\overline{w}_o^4 + 2(\overline{3020})\overline{w}_o^3\overline{w}_a + 3(\overline{2030})\overline{w}_o^2\overline{w}_a^2 + 4(\overline{1040})\overline{w}_o\overline{w}_a^3 + 5(\overline{0050})\overline{w}_a^4 \tag{23.41}$$

(b) $N = 6$. The same general remarks also apply to these systems; the aberrations are the same as certain of those which afflict systems for which $N = 3$.

23.7 Systems with an Axis of Rotational Symmetry

The primordial properties and primary aberrations of these systems have been very thoroughly studied. The primary aberrations, which are third order, are most simply derived from the Hermitian form

$$S^{(4)} = \begin{pmatrix} \overline{w}_o^2 \\ \overline{w}_o\overline{w}_a \\ \overline{w}_a^2 \end{pmatrix}^T \begin{pmatrix} 2200(r) & \overline{2101} & \overline{2002} \\ 2101 & 1111(r) & \overline{1012} \\ 2002 & 1012 & 0022(r) \end{pmatrix} \begin{pmatrix} w_o^2 \\ w_o w_a \\ w_a^2 \end{pmatrix} \tag{23.42}$$

The secondary aberrations are fifth order; the function $S^{(6)}$ can be written

$$S^{(6)} = \begin{pmatrix} \overline{w}_o^3 \\ \overline{w}_o^2\overline{w}_a \\ \overline{w}_o\overline{w}_a^2 \\ \overline{w}_a^3 \end{pmatrix}^T \begin{pmatrix} 3300(r) & \overline{3201} & \overline{3102} & \overline{3003} \\ 3201 & 2211(r) & \overline{2112} & \overline{2013} \\ 3102 & 2112 & 1122(r) & \overline{1023} \\ 3003 & 2013 & 1023 & 0033(r) \end{pmatrix} \begin{pmatrix} w_o^3 \\ w_o^2 w_a \\ w_o w_a^2 \\ w_a^3 \end{pmatrix} \tag{23.43}$$

so that

$$\frac{\partial S_{ac}^{(6)}}{\partial\overline{w}_o} = \begin{pmatrix} \overline{w}_o^2 \\ \overline{w}_o\overline{w}_a \\ \overline{w}_a^2 \end{pmatrix}^T \begin{pmatrix} 3[3300] & 3[\overline{3201}] & 3[\overline{3102}] & 3[\overline{3003}] \\ 2[3201] & 2[2211] & 2[\overline{2112}] & 2[\overline{2013}] \\ [3102] & [2112] & [1122] & [\overline{1023}] \end{pmatrix} \begin{pmatrix} w_o^3 \\ w_o^2 w_a \\ w_o w_a^2 \\ w_a^3 \end{pmatrix} \tag{23.44a}$$

and

$$\frac{\partial S_{oc}^{(6)}}{\partial \overline{w}_a} = \begin{pmatrix} \overline{w}_o^2 \\ \overline{w}_o \overline{w}_a \\ \overline{w}_a^2 \end{pmatrix}^T \begin{pmatrix} (3201) & (2211) & \left(\overline{2112}\right) & \left(\overline{2013}\right) \\ 2(3102) & 2(2112) & 2(1122) & 2\left(\overline{1023}\right) \\ 3(3003) & 3(2013) & 3(1023) & 3(0033) \end{pmatrix} \begin{pmatrix} w_o^3 \\ w_o^2 w_a \\ w_o w_a^2 \\ w_a^3 \end{pmatrix} \tag{23.44b}$$

In the stigmatic image plane, the contribution to the secondary aberrations that arises from $S^{(6)}$ contains the following terms:

$$\begin{aligned}
\text{distortion: } (3200) &= -\frac{(3201)}{Ah_i'} \\
\text{aperture aberration: } (0032) &= -3\frac{(0033)}{Ah_i'}, \quad \text{real} \\
\text{astigmatism and field curvature: } (2210) &= -\frac{(2211)}{Ah_i'}, \quad \text{real} \\
(3101) &= -2\frac{(3102)}{Ah_i'} \\
\text{comas: } (1220) &= -\frac{\left(\overline{2112}\right)}{Ah_i'} \\
(2111) &= -2\frac{(2112)}{Ah_i'} \\
(3002) &= -3\frac{(3003)}{Ah_i'} \\
\text{terms in } r_a^3\text{: } (0230) &= -\frac{\left(\overline{2013}\right)}{Ah_i'} \\
(1121) &= -2\frac{(1122)}{Ah_i'}, \quad \text{real} \\
(2012) &= -3\frac{(2013)}{Ah_i'} \\
\text{terms in } r_a^4\text{: } (0131) &= -2\frac{\left(\overline{1023}\right)}{Ah_i'} \\
(1022) &= -3\frac{(1023)}{Ah_i'}
\end{aligned} \tag{23.45}$$

In an electrostatic system, all the coefficients are real, and the aberrations resemble those of glass lenses. If, however, the system is magnetic, 'anisotropic aberrations' also appear as we should expect, by which we mean that certain of the aberration coefficients are complex numbers.

Consider, for example, the terms in the first power of the aperture coordinates, (2210) and (3101), the astigmatisms:

$$w_c^{II} = (2210)w_o^2\overline{w}_o^2 w_a + (3101)w_o^3\overline{w}_o\overline{w}_a$$

or

$$w_c^{II}\mathrm{e}^{\mathrm{i}\varphi} = r_o^4 r_a\left\{(2210)\mathrm{e}^{\mathrm{i}\Theta_a} + |3101|\mathrm{e}^{-\mathrm{i}\Theta_a}\right\}$$

in which

$$\varphi = -\left(\frac{1}{2}\widehat{3101} + \theta_o\right) \text{ and } \Theta_a = \theta_a - \left(\frac{1}{2}\widehat{3101} + \theta_o\right)$$

which represents a tilted ellipse, with semiaxes $r_o^4 r_a$ $\{(2210) \pm |3101|\}$.

Likewise, the comas produce an overall effect

$$w_c^{II} = Aw_o\overline{w}_o^2 w_a^2 + 2\overline{A}w_o^2\overline{w}_o w_a\overline{w}_a + Bw_o^3\overline{w}_a^2$$

or

$$w_c^{II}\mathrm{e}^{\mathrm{i}\varphi}/r_o^3 r_a^2 = 2\overline{A}\mathrm{e}^{\mathrm{i}(\theta_o+\varphi)} + |A|\mathrm{e}^{2\mathrm{i}\Theta_a} + |B|\mathrm{e}^{-\mathrm{i}\Theta_a}$$

where now

$$\varphi = -\theta_o - \frac{1}{2}\left(\tilde{A} + \tilde{B}\right) \text{ and } \Theta_a = \theta_a - \theta_o + \frac{1}{4}\left(\tilde{A} - \tilde{B}\right)$$

which represents a family of ellipses for different values of r_a, centred on the line

$$w_c^{II} = 2r_o^3 r_a^2\overline{A}\mathrm{e}^{\mathrm{i}\theta_o}$$

with semi-axes $r_o^3 r_a^2(|A| \pm |B|)$. $\tilde{A}$ and $\tilde{B}$ denote the arguments of A and B respectively.

The terms in r_a^4 give

$$w_c^{II} = 2A\overline{w}_o w_a^3\overline{w}_a + 3\overline{A}w_o w_a^2\overline{w}_a^2$$

or

$$w_c^{II}/r_o r_a^4 = 2|A|\mathrm{e}^{2\mathrm{i}\Theta_a} + 3\overline{A}\mathrm{e}^{\mathrm{i}\theta_o}$$

in which $\Theta_\alpha = \theta_a + \frac{1}{2}\left(\tilde{A} - \theta_o\right)$, which represents a family of circles, centred on the line $w_c^{II} = 3r_o r_a^4\overline{A}\mathrm{e}^{\mathrm{i}\theta_o}$, radii $2r_o r_a^4|A|$.

Finally, the terms in r_a^3 lead to

$$w_c^{II} = 3Aw_o^2 w_a\overline{w}_a^2 + \overline{A}\overline{w}_o^2 w_a^3 + Bw_o\overline{w}_o w_a^2\overline{w}_a$$

or

$$w_c^{II}/r_o^2 r_a^3 = 3|A|e^{i(\overline{A}+2\theta_o-2\theta_a)} + |A|e^{i(-\overline{A}-2\theta_o+3\theta_a)} + Be^{i\theta_a}$$

The terms in $|A|$ can be written as

$$e^{-i\varphi}\{3|A|e^{-i\Theta_a} + |A|e^{3i\Theta_a}\}$$

in which $\varphi = -\frac{1}{2}\tilde{A} - \theta_o$ and $\Theta_a = \theta_a - \theta_o - \frac{1}{2}\tilde{A}$ and if we write $\upsilon = w_c^{II}e^{i\varphi}/r_o^2 r_a^3|A| = x' + iy'$, we have

$$x' = 4\cos^3\theta, \quad y' = -4\sin^3\theta$$

which represents an astroid.

23.8 Note on the Classification of Aberrations

We see from the earlier parts of this chapter that although the number of aberration coefficients needed to characterize complex systems may be large, the aberrations form families, the importance of which is often very different in different components. In electron microscopes, for example, the objective lens suffers from the same aberrations as the final projector but for the former only the aperture aberrations and coma are important, whereas for the latter all but the distortions are negligible. We therefore introduce a convenient nomenclature, which brings together members of the same family of aberrations for any optical element. Each aberration is associated with a term of the form $x_o^p y_o^q x_a^r y_a^s$ and in the case of stigmatic imagery and primary aberrations, these can be related back to particular terms of S^I.

With the inclusion of aberration correctors in electron microscopes, the literature of parasitic aberrations has grown rapidly and several notations for the numerous coefficients have been introduced, These are tabulated in Section 31.6.

23.8.1 Terms Independent of x_o, y_o ($p = q = 0$): Aperture Aberrations

For these aberrations, the aberration figure is independent of the choice of object point and, in particular, their effects do not vanish or become small if the latter lies on the optic axis. They are known generically as *aperture defects* or *aperture aberrations* and include as important special cases the spherical aberration of round lenses and the aperture aberrations of quadrupoles. The defocus also belongs to this group, since an image formed on a plane that is not conjugate to the object plane is blurred in an aperture-dependent way. The axial astigmatism associated, for example, with imperfect roundness of a lens intended to be rotationally symmetric and the primary (axial) chromatic aberration are also aperture defects.

23.8.2 Terms Independent of x_a, y_a ($r = s = 0$): Distortions

These are at the other extreme from aperture aberrations, for they depend only on the position of the object point. Hence all rays from a given object point will be displaced by the same amount from their paraxial points of arrival and, if the latter lie in a stigmatic image plane, such terms will shift the paraxial image point but not blur it. These aberrations are known as distortions and include the isotropic and anisotropic distortions of round lenses and the chromatic aberrations other than the axial term.

23.8.3 Intermediate Terms

The intermediate terms may be classified into those linear in x_o or y_o; those linear in x_a and y_a; and any others. The former may be regarded as comas, by analogy with the primary coma of round lenses. Terms linear in x_a and y_a form the group of field curvatures and astigmatism. The others, which arise when we consider fourth and higher order aberrations, have acquired a variety of names (see Chapter 31, Parasitic Aberrations).

23.8.4 Phase Shifts

If we are considering a pair of conjugate planes in an imaging system, any terms in S^I that are independent of x_a and y_a will not contribute to the aberrations, which are determined by $\partial S^I/\partial x_a$ and $\partial S^I/\partial y_a$. Such terms become important when we consider the dependence of the aberration coefficients on object and aperture position in Chapter 25 and will reappear in Volume 3 in connection with the wave theory of aberrations.

23.8.5 Parasitic Aberrations

With the incorporation of aberration correctors in electron microscopes, the literature of parasitic aberrations has exploded and names have been chosen for many of the axial aberrations: star aberration, three-lobe aberration, fourfold, fivefold and sixfold astigmatism, pentacle and chaplet aberration for example. These are defined in Section 31.6.

CHAPTER 24

The Geometrical Aberrations of Round Lenses

24.1 Introduction

We now obtain explicit formulae for the aberration coefficients of systems of round magnetic and electrostatic lenses in which the fields may overlap. In order to avoid repetition, we deal with the real aberrations expressed in terms of object and aperture coordinates in detail, then describe briefly the changes necessary to convert these coefficients when position and gradient in the object plane are used. We then show (in Chapter 25, Asymptotic Aberration Coefficients) how a similar calculation yields the asymptotic aberrations. These calculations will be based on the eikonal method but the main stages in the reasoning when the trajectory method is employed are also given.

Within this overall plan, certain decisions remain to be taken: the most important concerns the use of reduced coordinates (Eq. 15.40). If these are employed, as they are by Sturrock (1955), the expressions for the various coefficients are very different in appearance from those obtained in conventional coordinates, so different indeed that Glaser (1952) believed them to be wrong. (In fact, they can be shown to be equivalent but the demonstration is undeniably laborious.) We have chosen to use conventional coordinates but we list one form of the coefficients in reduced coordinates for reference purposes (Section 24.8).

The other decision concerns the use of complex coordinates. There is no doubt that these give the analysis a more compact appearance, but expressions involving the third and fourth powers of complex coordinates are much less easy to picture than those involving first and second powers, which we have met in the paraxial domain. We therefore retain a certain flexibility, giving some important formulae in both forms; we avoid repeating very similar equations for x and y by the use of suffix notation, as in Chapter 22, Perturbation Theory: General Formalism.

24.2 Derivation of the Real Aberration Coefficients

The paraxial ray equations for round lenses were obtained in Section 15.2 by writing down the Euler equations corresponding to the variational relation $\delta \int M^{(2)} dz = 0$. If we retain the next higher order terms, $M^{(4)}$ in the series expansion for M and again write down the Euler

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00024-9

© 2018 Elsevier Ltd. All rights reserved.

equations, we obtain the inhomogeneous second-order equations from which the aberrations are obtained in the trajectory method; alternatively, we may set $M^{(p)} = M^{(4)}$ and use the perturbation theory of Chapter 22, Perturbation Theory: General Formalism. In both cases, we need $M^{(4)}$. Substituting terms of higher order for Φ and A_x, A_y into (Eq. 15.23), we find

$$\begin{aligned} M^{(4)} = {} & \frac{1}{128\hat{\phi}^{1/2}}\left(\gamma\phi^{(4)} - \frac{\phi''^2}{\hat{\phi}}\right)(X^2+Y^2)^2 \\ & - \frac{\gamma\phi''}{16\hat{\phi}^{1/2}}(X^2+Y^2)(X'^2+Y'^2) \\ & - \frac{\hat{\phi}^{1/2}}{8}(X'^2+Y'^2)^2 \\ & + \frac{\eta B''}{16}(X^2+Y^2)(XY'-X'Y) \end{aligned} \tag{24.1}$$

and introducing the rotating coordinate system (x, y, z) with the aid of Eq. (15.26)

$$\begin{aligned} M^{(4)} = {} & -\frac{1}{4}L_1(x^2+y^2)^2 - \frac{1}{2}L_2(x^2+y^2)(x'^2+y'^2) \\ & - \frac{1}{4}L_3(x'^2+y'^2)^2 - R(xy'-x'y)^2 \\ & - P\hat{\phi}^{1/2}(x^2+y^2)(xy'-x'y) \\ & - Q\hat{\phi}^{1/2}(x'^2+y'^2)(xy'-x'y) \end{aligned} \tag{24.2}$$

in which

$$\begin{aligned} L_1 &= \frac{1}{32\hat{\phi}^{1/2}}\left(\frac{\phi''^2}{\hat{\phi}} - \gamma\phi^{(4)} + \frac{2\gamma\phi''\eta^2B^2}{\hat{\phi}} + \frac{\eta^4B^4}{\hat{\phi}} - 4\eta^2BB''\right) \\ L_2 &= \frac{1}{8\hat{\phi}^{1/2}}(\gamma\phi'' + \eta^2B^2) \\ L_3 &= \frac{1}{2}\hat{\phi}^{1/2} \\ P &= \frac{\eta}{16\hat{\phi}^{1/2}}\left(\frac{\gamma\phi''B}{\hat{\phi}} - B'' + \frac{\eta^2B^3}{\hat{\phi}}\right) \\ Q &= \frac{\eta B}{4\hat{\phi}^{1/2}} \\ R &= \frac{\eta^2B^2}{8\hat{\phi}^{1/2}} \end{aligned} \tag{24.3}$$

Note that in purely electrostatic lenses, P, Q and R vanish, and that in purely magnetic lenses, $L_2 = R$; Q is half the rate of rotation (Eq. 15.9) of the rotating frame (x, y, z).

24.2.1 The Trajectory Method

We now write

$$\delta \int (M^{(2)} + M^{(4)}) dz = 0 \tag{24.4}$$

or

$$\frac{d}{dz}\left(\frac{\partial M^{(2)}}{\partial x_j'}\right) - \frac{\partial M^{(2)}}{\partial x_j} = -\frac{d}{dz}\left(\frac{\partial M^{(4)}}{\partial x_j'}\right) + \frac{\partial M^{(4)}}{\partial x_j} \tag{24.5}$$

Here and elsewhere, $j = 1, 2$; $x_1 = x$ and $x_2 = y$. Substituting for $M^{(2)}$ and $M^{(4)}$ we find

$$\frac{d}{dz}(\hat{\phi}^{1/2} x_j') + \frac{\gamma\phi'' + \eta^2 B^2}{4\hat{\phi}^{1/2}} x_j = \Lambda_j \tag{24.6}$$

where

$$\begin{aligned}
\Lambda_1 \coloneqq\ & (x^2 + y^2)\{ - L_1 x + L_2' x' + L_2 x'' - (P\phi^{1/2})' y - 2P\phi^{1/2} y'\} \\
& + (x'^2 + y'^2)\{ - L_2 x + L_3' x' + L_3 x'' - (Q\phi^{1/2})' y - 2Q\phi^{1/2} y'\} \\
& + 2(xy' - x'y)\{ - P\phi^{1/2} x + (Q\phi^{1/2})' x' + Q\phi^{1/2} x'' - R' y - 2Ry'\} \\
& + 2(xx' + yy')(L_2 x' - P\phi^{1/2} y) \\
& + 2(x'x'' + y'y'')(L_3 x' - Q\phi^{1/2} y) \\
& + 2(xy'' - x''y)(Q\phi^{1/2} x' - 2Ry)
\end{aligned} \tag{24.7}$$

and Λ_2 is obtained by writing $x \to y$, $y \to -x$. Eq. (24.6) is solved by replacing x and y by their paraxial expressions on the right-hand side, which thereby becomes a known function of z, and employing the method of variation of parameters. Suppose that $a(z)$ and $b(z)$ are two linearly independent solutions of the homogeneous equation, which is also the paraxial equation. We seek a solution of the inhomogeneous equation (24.6) of the form

$$x_j(z) = A_j(z) a(z) + B_j(z) b(z) \tag{24.8}$$

Selecting $A_j(z)$ and $B_j(z)$ in such a way that

$$A_j' a + B_j' b = 0 \tag{24.9}$$

and substituting Eq. (24.8) into (24.6), we obtain

$$\hat{\phi}^{1/2}(A_j' a' + B_j' b') = \Lambda_j \tag{24.10}$$

Solving Eqs. (24.9, 24.10) for A_j' and B_j' yields

$$\begin{aligned} A_j' &= -\frac{\Lambda_j b}{\hat{\phi}^{1/2}(ab' - a'b)} = -\frac{\Lambda_j b}{W} \\ B_j' &= \frac{\Lambda_j a}{\hat{\phi}^{1/2}(ab' - a'b)} = \frac{\Lambda_j a}{W} \end{aligned} \tag{24.11}$$

where W denotes the constant Wronskian (15.51)

$$W = \hat{\phi}^{1/2}(ab' - a'b) \tag{24.12}$$

Hence

$$\begin{aligned} A_j(z) &= A_j(z_1) - \frac{1}{W}\int_{z_1}^{z} \Lambda_j b \, d\zeta \\ B_j(z) &= B_j(z_2) + \frac{1}{W}\int_{z_2}^{z} \Lambda_j a \, d\zeta \end{aligned} \tag{24.13}$$

We have retained the possibility of using different lower limits in the integrals appearing in Eq. (24.13) because we need to study two sets of boundary conditions; when these govern ray position in two different planes, this extra flexibility is indispensable.

If the aberrations are expressed in terms of position coordinates in two planes, the object and aperture planes $z = z_o$ and $z = z_a$, then the aberration terms vanish there. Introducing the paraxial solutions $s(z)$, $t(z)$,

$$\begin{aligned} a(z) &\Rightarrow s(z), \qquad s(z_o) = t(z_a) = 1 \\ b(z) &\Rightarrow t(z), \qquad s(z_a) = t(z_o) = 0 \end{aligned} \tag{24.14}$$

we find

$$x_j(z) = A_j(z_1)s(z) + B_j(z_2)t(z) - \frac{s(z)}{W_s}\int_{z_1}^{z} \Lambda_j t \, d\zeta + \frac{t(z)}{W_s}\int_{z_2}^{z} \Lambda_j \, s d\zeta \tag{24.15}$$

where $W_s := \hat{\phi}^{1/2}(st' - s't) = \hat{\phi}_o^{1/2} t_o' = -\hat{\phi}_a^{1/2} s_a'$ is the Wronskian for the choice (24.14). In $z = z_o$, $x_j = x_{jo}$ so that $A_j(z_1) = x_{jo}$, $z_1 = z_o$. In $z = z_a$, $x_j = x_{ja}$ and hence $B_j(z_2) = x_{ja}$, $z_2 = z_a$. Finally,

$$\begin{aligned} x_j(z) = {} & x_{jo}s(z) + x_{ja}t(z) \\ & + \frac{1}{W_s}\left\{ t(z)\int_{z_a}^{z} \Lambda_j s \, d\zeta - s(z)\int_{z_o}^{z} \Lambda_j t \, d\zeta \right\} \end{aligned} \tag{24.16}$$

In the image plane, $z = z_i$, $t(z)$ is again zero and $s(z_i) = M$. It is usual to refer the aberrations back to the object plane by considering the quantity $\{x_j(z_i) - Mx_{jo}\}/M$:

$$\frac{x_j(z_i) - Mx_{jo}}{M} =: \Delta x_{ji} = -\frac{1}{W_s}\int_{z_o}^{z_i} \Lambda_j t \, dz \tag{24.17}$$

The individual aberration coefficients are then extracted by substituting $x_j(z) = x_{jo}s(z) + x_{ja}\,t(z)$ in the Λ_j and collecting terms of various degree in x_o, y_o, x_a and y_a. We shall perform this step in detail below when using the eikonal method.

If the aberrations are expressed in terms of position and gradient in the object plane, we use the pair of paraxial solutions denoted by $g(z)$ and $h(z)$ (Eq. 15.56):

$$\begin{aligned} a(z) &\Rightarrow g(z), \quad & g(z_o) &= h'(z_o) = 1 \\ b(z) &\Rightarrow h(z), \quad & g'(z_o) &= h(z_o) = 0 \end{aligned} \tag{24.18}$$

and now

$$\begin{aligned} x_j(z) = {} & A_j(z_1)g(z) + B_j(z_2)h(z) \\ & + \frac{1}{W_g}\left\{ -g(z)\int_{z_1}^{z} \Lambda_j h \, d\zeta + h(z)\int_{z_2}^{z} \Lambda_j g \, d\zeta \right\} \end{aligned} \tag{24.19}$$

where $W_g := \hat{\phi}^{1/2}(gh' - g'h) = \hat{\phi}_o^{1/2}$ is the Wronskian for Eq. (24.18). In $z = z_o$, $x_j(z_o) = x_{jo}$ and $x_j'(z_o) = x_o'$ so that as before $A_j(z_1) = x_{jo}$, $z_1 = z_o$; now, however, z_2 is also equal to z_o and $B_j(z_2) = x_{jo}'$:

$$\begin{aligned} x_j(z) = {} & x_{jo}g(z) + x_{jo}'h(z) \\ & + \frac{1}{W_g}\left\{ h(z)\int_{z_o}^{z} \Lambda_j g \, d\zeta - g(z)\int_{z_o}^{z} \Lambda_j h \, d\zeta \right\} \end{aligned} \tag{24.20}$$

or in the image plane, $h(z_i) = 0$,

$$\frac{x_{ji} - Mx_{jo}}{M} = \Delta x_{ji} = -\frac{1}{W_g}\int_{z_o}^{z_i} \Lambda_j h \, dz \tag{24.21}$$

Despite their formal resemblance, there is an important difference between Eqs. (24.17) and (24.21) or (24.16) and (24.19). Since both $t(z)$ and $h(z)$ vanish in the object and image planes, these paraxial solutions are proportional, $t(z) \propto h(z)$. The rays $s(z)$ and $g(z)$, though of course linearly related, are not proportional:

$$\begin{aligned} g(z) &= s(z) + g(z_a)t(z) = s(z) - \frac{s'(z_o)}{t'(z_o)}t(z) \\ s(z) &= g(z) + s'(z_o)h(z) = g(z) - \frac{g(z_a)}{h(z_a)}h(z) \end{aligned} \tag{24.22}$$

With the aid of Eq. (24.17), therefore, the effect of aperture position on the aberration coefficients can be explored.

24.2.2 The Eikonal Method

Instead of using $M^{(4)}$ to derive Eq. (24.6), we can equally well use perturbation theory, regarding $M^{(2)}$ as the unperturbed refractive index and setting $M^{(4)} = M^{(p)}$ For round lenses, the formulae of Chapter 22, Perturbation Theory: General Formalism, reduce to a much simpler form since $M^{(2)}$ (15.29) contains terms in $(x^2 + y^2)$ and $(x'^2 + y'^2)$ only so that paraxially,

$$p_j = \hat{\phi}^{1/2} x'_j \tag{24.23}$$

For the boundary conditions (22.24), we have $x_j = x_{jo} g + x'_{jo} h$ and hence

$$\begin{aligned} \frac{\partial x_j(z)}{\partial x_{ko}} &= g(z)\delta_{jk} & \frac{\partial x_j(z)}{\partial p_{ko}} &= \frac{h(z)}{\hat{\phi}_o^{1/2}}\delta_{jk} \\ \frac{\partial p_j(z)}{\partial x_{ko}} &= \hat{\phi}^{1/2}(z) g'(z)\delta_{jk} & \frac{\partial p_j(z)}{\partial p_{ko}} &= h'(z)\delta_{jk} \end{aligned} \tag{24.24}$$

and (22.28) become

$$\begin{aligned} \frac{\partial S^I_{12}}{\partial x_{jo}} &= p^{(1)}_{j2} g(z_2) - x^{(1)}_{j2} \hat{\phi}^{1/2}(z_2) g'(z_2) \\ \hat{\phi}_o^{1/2} \frac{\partial S^I_{12}}{\partial p_{jo}} &= p^{(1)}_{j2} h(z_2) - x^{(1)}_{j2} \hat{\phi}^{1/2}(z_2) h'(z_2) \end{aligned} \tag{24.25}$$

Solving for $x^{(1)}_j(z_2)$ and $p^{(1)}_j(z_2)$, we find

$$\begin{aligned} W_g x^{(1)}_j(z_2) &= h(z_2) \frac{\partial S^I_{12}}{\partial x_{jo}} - \hat{\phi}_o^{1/2} g(z_2) \frac{\partial S^I_{12}}{\partial p_{jo}} \\ W_g p^{(1)}_j(z_2) &= \hat{\phi}^{1/2}(z_2) \left\{ h'(z_2) \frac{\partial S^I_{12}}{\partial x_{jo}} - \hat{\phi}_o^{1/2} g'(z_2) \frac{\partial S^I_{12}}{\partial p_{jo}} \right\} \end{aligned} \tag{24.26}$$

or using Eq. (24.23),

$$W_g x_j{}^{(1)}(z_2) = h(z_2) \frac{\partial S^I_{12}}{\partial x_{jo}} - g(z_2) \frac{\partial S^I_{12}}{\partial x'_{jo}} \tag{24.27a}$$

$$\frac{W_g p_j{}^{(1)}(z_2)}{\hat{\phi}^{1/2}(z_2)} = h'(z_2) \frac{\partial S^I_{12}}{\partial x_{jo}} - g'(z_2) \frac{\partial S^I_{12}}{\partial x'_{jo}} \tag{24.27b}$$

Eq. (24.27a) gives the aberrations of position in $z = z_2$. We have already mentioned that the boundary condition $p_{jo}^{(1)} = 0$ is not equivalent to $x_{j0}^{\prime(1)} = 0$; in order to obtain the aberrations of gradient, additional manipulations are required. We pursue these in full in Chapter 25, Asymptotic Aberration Coefficients.

If the aberrations are to be expressed in terms of position in the object and aperture planes, we use Eq. (22.30) with $x_j(z) = x_{jo}s(z) + x_{ja}t(z)$ so that

$$\begin{aligned}\frac{\partial S_{o2}^I}{\partial x_{ja}} &= p_{j2}^{(1)} t(z_2) - x_{j2}^{(1)} \hat{\phi}_2^{\,1/2} t'(z_2)\\ \frac{\partial S_{a2}^I}{\partial x_{jo}} &= p_{j2}^{(1)} s(z_2) - x_{j2}^{(1)} \hat{\phi}_2^{\,1/2} s'(z_2)\end{aligned} \tag{24.28}$$

with solution

$$\begin{aligned}W_s x_j^{(1)}(z_2) &= t(z_2)\frac{\partial S_{a2}^I}{\partial x_{jo}} - s(z_2)\frac{\partial S_{o2}^I}{\partial x_{ja}}\\ W_s p_j^{(1)}(z_2) &= \hat{\phi}_2^{\,1/2}\left\{ t'(z_2)\frac{\partial S_{a2}^I}{\partial x_{jo}} - s'(z_2)\frac{\partial S_{o2}^I}{\partial x_{ja}} \right\}\end{aligned} \tag{24.29}$$

The final stage of the calculation involves substituting the paraxial solutions for $x_j(z)$ into $M^{(4)}$ and reorganizing the result into convenient groups of terms. We reintroduce x and y, setting

$$\begin{aligned}x_1(z) &= x_o s(z) + x_a t(z)\\ x_2(z) &= y_o s(z) + y_a t(z)\end{aligned} \tag{24.30}$$

whereupon $M^{(4)}$ Eq. (24.2) takes the form

$$\begin{aligned}M^{(4)} = &-\frac{1}{4}(L_1 s^4 + 2L_2 s^2 s'^2 + L_3 s'^4)(x_o^2 + y_o^2)^2\\ &-\frac{1}{4}(L_1 t^4 + 2L_2 t^2 t'^2 + L_3 t'^4)(x_a^2 + y_a^2)^2\\ &-\left\{ L_1 s^2 t^2 + 2L_2 s s' t t' + L_3 s'^2 t'^2 - R(st' - s't)^2 \right\}(x_o x_a + y_o y_a)^2\\ &-\frac{1}{2}\left\{ L_1 s^2 t^2 + L_2(s^2 t'^2 + s'^2 t^2) + L_3 s'^2 t'^2 \right.\\ &\qquad \left. + 2R(st' - s't)^2 \right\}(x_o^2 + y_o^2)(x_a^2 + y_a^2)\\ &-\left\{ L_1 s^3 t + L_2 s s'(st)' + L_3 s'^3 t' \right\}(x_o^2 + y_o^2)(x_o y_a + y_o y_a)\end{aligned}$$

$$
\begin{aligned}
&- \left\{L_1 st^3 + L_2 (st)' tt' + L_3 s' t'^3\right\}(x_a^2 + y_a^2)(x_o y_a + y_o y_a) \\
&- W_s (Ps^2 + Qs'^2)(x_o^2 + y_o^2)(x_o y_a - x_a y_o) \\
&- 2W_s (Pst + Qs't')(x_o x_a + y_o y_a)(x_o y_a - x_a y_o) \\
&- W_s (Pt^2 + Qt'^2)(x_a^2 + y_a^2)(x_o y_a - x_a y_o)
\end{aligned} \tag{24.31}
$$

We write[1]

$$
\begin{aligned}
A &= \frac{1}{W}\int \left\{L_1 s^2 t^2 + 2L_2 ss' tt' + L_3 s'^2 t'^2 - R(st' - s't)^2\right\} dz \\
C &= \frac{1}{W}\int \left(L_1 t^4 + 2L_2 t^2 t'^2 + L_3 t'^4\right) dz \\
D &= \frac{1}{W}\int \left(L_1 s^3 t + L_1 ss'(st)' + L_3 s'^3 t'\right) dz \\
F &= \frac{1}{W}\int \left\{2L_1 s^2 t^2 + L_2 (st' + s't)^2 + 2L_3 s'^2 t'^2 + R(st' - s't)^2\right\} dz \\
E &= \frac{1}{W}\int \left(L_1 s^4 + 2L_2 s^2 s'^2 + L_3 s'^4\right) dz \\
K &= \frac{1}{W}\int \left\{L_1 st^3 + L_2 (st)' tt' + L_3 s' t'^3\right\} dz \\
a &= 2\int (Pst + Qs't') dz \\
d &= \int \left(Ps^2 + Qs'^2\right) dz \\
k &= \int \left(Pt^2 + Qt'^2\right) dz
\end{aligned} \tag{24.32}
$$

It is clear both on general symmetry grounds and from Eq. (24.31) that the quantities x_o, y_o, x_a and y_a can occur in S^I only as the rotationally invariant groups r_o, r_a, V and υ, where

$$
\begin{aligned}
r_o^2 &\coloneqq x_o^2 + y_o^2 = u_o u_o^* \\
r_a^2 &\coloneqq x_a^2 + y_a^2 = u_a u_a^* \\
V &\coloneqq x_o x_a + y_o y_a = r_o r_a \cos\left(\varphi_a - \varphi_o\right) = \frac{1}{2}\left(u_o u_a^* + u_o^* u_a\right) \\
\upsilon &\coloneqq x_o y_a - x_a y_o = r_o r_a \sin\left(\varphi_a - \varphi_o\right) = \frac{\mathrm{i}}{2}\left(u_o u_a^* - u_o^* u_a\right)
\end{aligned} \tag{24.33a}
$$

[1] The notation adopted here is not that widely encountered in the literature, notably in Glaser (GdE, HdP), but is mnemonically superior: we shall see that A and a are associated with astigmatism, C with spherical aberration, D and d with distortion, F with field curvature and K and k with coma.

with $x_o =: r_o \cos \varphi_o$, $y_o =: r_o \sin \varphi_o$ and similarly for x_a, y_a; the complex numbers $u_o = x_o + \mathrm{i}y_o$ and $u_a = x_a + \mathrm{i}y_a$ are those defined in Eq. (15.11). It will also be convenient to write

$$\varphi := \varphi_a - \varphi_o \tag{24.33b}$$

The quantities S^{I}_{a2} and S^{I}_{o2} have the following generic form, only the lower limit of integration changing:

$$\begin{aligned}
-\frac{S^I}{W} &= \frac{E}{4} r_0^4 + \frac{C}{4} r_a^4 + \frac{A}{2}\left(V^2 - \upsilon^2\right) \\
&\quad + \frac{F}{2} r_o^2 r_a^2 + D r_o^2 V + K r_a^2 V \\
&\quad + \upsilon\left(d r_o^2 + k r_a^2 + aV\right) \\
&= \frac{1}{4} C r_a^4 + r_a^3 r_o (K \cos\varphi + k \sin\varphi) \\
&\quad + \frac{1}{2} r_a^2 r_o^2 (A \cos 2\varphi + a \sin 2\varphi) + \frac{1}{2} F r_a^2 r_o^2 \\
&\quad + r_a r_o^3 (D \cos\varphi + d \sin\varphi) + \frac{1}{4} E r_o^2 \\
&= \begin{pmatrix} u_o^{*2} \\ u_o^* u_a^* \\ u_a^{*2} \end{pmatrix}^T \begin{pmatrix} E/4 & (D - id)/2 & (A - ia)/4 \\ (D + id)/2 & F/2 & (K - ik)/2 \\ (A + ia)/4 & (K + ik)/2 & C/4 \end{pmatrix} \begin{pmatrix} u_o^2 \\ u_o u_a \\ u_a^2 \end{pmatrix}
\end{aligned} \tag{24.34}$$

We add the suffix o to the coefficients $A, C, \ldots, d, k$ when the integration runs from z_o to an arbitrary plane z_2; we add a when the lower limit is z_a; when no suffix is added, the integrals run from the object plane z_o to the image plane $z = z_i$.

In a general plane

$$\begin{aligned}
x^{(1)}(z) &= r_o^2 \Big\{ x_o (D_o s - E_a t) + x_a (F_o s - A_o s - D_a t) - y_o d_o s \Big\} \\
&\quad + r_a^2 \Big\{ x_o (K_o s - F_a t + A_a t) + x_a (C_o s - K_a t) \\
&\quad - y_o (k_o s + a_a t) - y_a k_a t \Big\} \\
&\quad + V \big\{ 2x_o (A_o s - D_a t) + 2x_a (K_o s - A_a t)
\end{aligned}$$

$$
\begin{aligned}
&\quad - y_o(a_o s + d_a t)\Big\} \\
&\quad + \upsilon\big\{x_0(a_a s - 3d_a t) + 2x_a(k_0 s - a_a t)\Big\}
\end{aligned}
\tag{24.35a}
$$

$$
\begin{aligned}
y^{(1)}(z) = {} & r_o^2\Big\{y_o(D_o s - E_a t) + y_a(F_o s - A_o s - D_a t) + x_o d_o s\Big\} \\
& + r_a^2\Big\{y_o(K_o s - F_a t + A_a t) + y_a(C_o s - K_a t) \\
& + x_o(k_o s + a_a t) + x_a k_a t\Big\} \\
& + V\Big\{2y_o(A_o s - D_a t) + 2y_a(K_o s - A_a t) \\
& + x_o(a_o s + d_a t)\Big\} \\
& + \upsilon\big\{y_o(a_a s - 3d_a t) + 2y_a(k_o s - a_a t)\Big\}
\end{aligned}
\tag{24.35b}
$$

or

$$
u^{(1)}(z) = \begin{pmatrix} r_o^2 \\ r_a^2 \\ V \\ \upsilon \end{pmatrix}^T \begin{pmatrix} (D_o + \mathrm{i}d_o)s - E_a t & F_o s - A_o s - D_a t \\ (K_o + \mathrm{i}k_o)s - (F_a - A_a - \mathrm{i}a_a)t & C_o s - (K_a - \mathrm{i}k_a)t \\ (2A_o + \mathrm{i}a_o)s - (2D_a - \mathrm{i}d_a)t & 2(K_o s - A_a t) \\ a_o s - 3d_a t & 2(k_o s - a_a t) \end{pmatrix} \begin{pmatrix} u_o \\ u_a \end{pmatrix} \tag{24.35c}
$$

In the image plane, where $t(z)$ vanishes and $s(z_i) = M$,

$$
\begin{aligned}
\frac{x^{(1)}(z_i)}{M} =: \Delta x_i = {} & x_a\big(Cr_a^2 + 2KV + 2k\upsilon + (F - A)r_o^2\big) \\
& + x_o\big(Kr_a^2 + 2AV + a\upsilon + Dr_o^2\big) \\
& - y_o\big(kr_a^2 + aV + dr_o^2\big) \\
\frac{y^{(1)}(z_i)}{M} =: \Delta y_i = {} & y_a\big(Cr_a^2 + 2KV + 2k\upsilon + (F - A)r_o^2\big) \\
& + y_o\big(Kr_a^2 + 2AV + a\upsilon + Dr_o^2\big) \\
& + x_o\big(kr_a^2 + aV + dr_o^2\big)
\end{aligned}
\tag{24.36}
$$

Setting

$$
\Delta u_i \coloneqq \frac{\Delta x_i + \mathrm{i}\Delta y_i}{M} \tag{24.37}
$$

the pair of equations (24.36) can be combined into the compact expression

$$\Delta u_i = \begin{pmatrix} u_o^* \\ 2u_a^* \end{pmatrix} \begin{pmatrix} D + \mathrm{i}d & F & K - \mathrm{i}k \\ \frac{1}{2}(A + \mathrm{i}a) & K + \mathrm{i}k & \frac{1}{2}C \end{pmatrix} \begin{pmatrix} u_o^2 \\ u_o u_a \\ u_a^2 \end{pmatrix} \tag{24.38a}$$

which could have been derived directly from Eq. (24.34) since $\partial/\partial x + \mathrm{i}\partial/\partial y = 2\partial/\partial u^*$. Eq. (24.38a) may also be written in the form

$$\begin{aligned} \Delta u_i = {} & C r_a^2 u_a \quad \text{(spherical aberration)} \\ & + 2(K + \mathrm{i}k) r_a^2 u_o + (K - \mathrm{i}k) u_a^2 u_o^* \quad \text{(coma)} \\ & + (A + \mathrm{i}a) u_o^2 u_a^* \quad \text{(astigmatism)} \\ & + F r_o^2 u_a \quad \text{(field curvature)} \\ & + (D + \mathrm{i}d) r_o^2 u_o \quad \text{(distortion)} \end{aligned} \tag{24.38b}$$

$$\begin{aligned} = {} & C r_a^2 u_a \\ & + 2(K + \mathrm{i}k) r_a^2 r_o \mathrm{e}^{\mathrm{i}\varphi_o} + (K - \mathrm{i}k) r_a^2 r_o \mathrm{e}^{\mathrm{i}(2\varphi_a - \varphi_o)} \\ & + (A + \mathrm{i}a) r_o^2 r_a \mathrm{e}^{\mathrm{i}(2\varphi_o - \varphi_a)} + F r_o^2 r_a \mathrm{e}^{\mathrm{i}\varphi_a} \\ & + (D + \mathrm{i}d) r_o^3 \mathrm{e}^{\mathrm{i}\varphi_o} \end{aligned} \tag{24.38c}$$

The numerous terms of which Δx_i and Δy_i are composed depend in different ways on the object and aperture coordinates and are in practice of unequal importance. We have indicated in Eq. (24.38b) the names by which they are commonly known and in Sections 24.3–24.6, we consider each in turn, pointing out general properties and other features of interest. Formulae for particular field or potential models and information about the dependence of the various coefficients on lens geometry and excitation will be found in Part VII.

The perturbation characteristic function S^I has hitherto appeared as an accessory that enables us to obtain the aberrations Δx_i and Δy_i by differentiation. We saw in Section 5.3, however, that pencils of rays propagate in such a way that they are always orthogonal to the surfaces $S = \text{const}$, $\boldsymbol{p}(\boldsymbol{r}) = \text{grad}\, \overline{S}(\boldsymbol{r})$ (5.12) or in the absence of magnetic field, $\boldsymbol{g}(\boldsymbol{r}) = \text{grad}\, \overline{S}(\boldsymbol{r})$. This relationship enables us to interpret Eq. (24.34) in physical terms and sheds additional light on the origin of the various types of aberration.

For simplicity, we assume that object and image are in field-free space (or virtual) and we introduce the entrance and exit pupils at $z = z_{ao}$ and $z = z_{ai}$. These are the images of the true aperture by the parts of the lens that precede and follow it, respectively. Fig. 24.1 shows the two pairs of conjugate planes, object and image and the two pupils; in the paraxial approximation, the point P_o is conjugate to P_i and the latter is shifted by aberrations to $\overline{P}_i$.

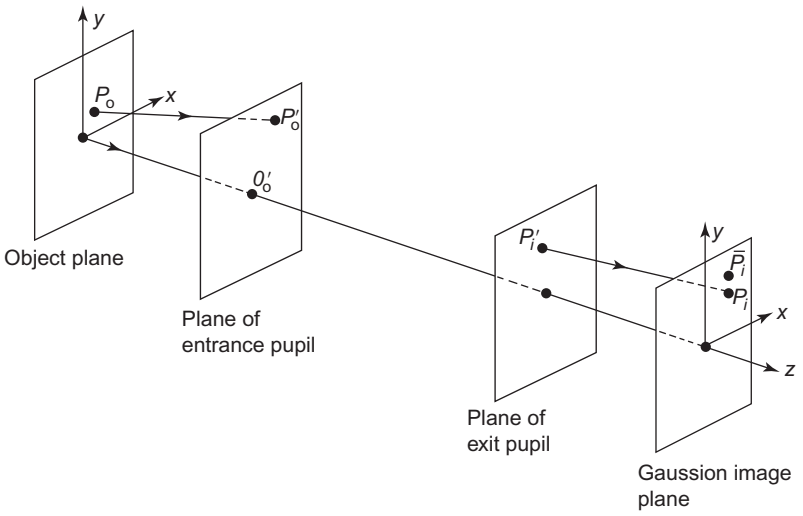


Figure 24.1
The object and image planes are conjugate as are the pupil planes. In the paraxial approximation, P_i is the image of P_o; aberrations shift P_i to $\overline{P}_i$.

Consider now a ray from P_o that intersects the entrance pupil at P'_o and the image pupil at P'_i. In the absence of aberrations, all rays from P_o are orthogonal to a set of spheres centred on P_o; the action of the lens converts the incident sphere into a sphere centred on P_i, and the rays in image space are all orthogonal to a set of spheres likewise centred on P_i. In the presence of aberrations, these spheres in image space are replaced by aspheric surfaces, and the distance between the paraxial spheres and the aberration surfaces is measured by S^I. Fig. 24.2 shows the spherical surface S_o centred on P_i that intersects the axis at the exit pupil and the aspheric surface S to which the electron trajectories are orthogonal in the presence of aberrations; we have again chosen the member of the set of such surfaces that intersects the axis at the exit pupil. The ray connecting P'_i to $\overline{P}_i$ cuts the reference sphere S_o at Q_o and the aspheric surface S at $\overline{Q}$. The distance $S(\overline{Q}, Q_o)$ is given by S^I (24.34) and we therefore associate each term of the latter with a particular distortion of the surface S. We return to this interpretation of S briefly in each of the following sections[2].

[2] Readers familiar with similar treatments of the aberrations in light optics, Sections 5.1−5.3 of Born and Wolf (1959, 2002) for example, may note that the role of the radius of the Gaussian reference sphere (R in Born and Wolf), which occurs in the aberration formulae, is absorbed into the Wronskian in electron optics. Thus (5.1.10) of Born and Wolf is exactly equivalent to (24.29a) for $t(z_2) = 0$.

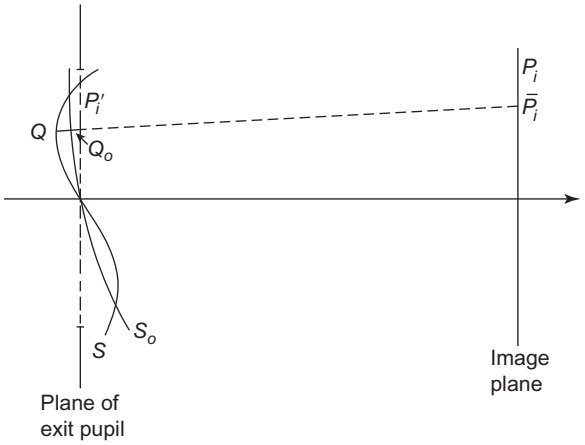


Figure 24.2
The reference sphere S_o is centred on P_i. Aberrations distort S_o into the aspheric surface S.

24.3 Spherical Aberration (Terms in x_a and y_a only)

We first examine the *spherical* or *aperture aberration,* of the greatest importance in the first (objective) lens of magnifying systems and in the final lens of probe-forming systems. From Eq. (24.36), we see that a pencil of rays from some object point (x_o, y_o) intersects the image plane not at the Gaussian image point (Mx_o, My_o) but at points (x_i, y_i) such that

$$\begin{aligned} x_i &= M\left\{x_o + Cx_a\left(x_a^2 + y_a^2\right)\right\} \\ y_i &= M\left\{y_o + Cy_a\left(x_a^2 + y_a^2\right)\right\} \end{aligned} \tag{24.39}$$

Hence $\Delta x_{ji} = Cx_{ja}r_a^2$ $(j = 1, 2)$. Each Gaussian image point is replaced by an aberration figure, as explained in Chapter 23, The Relation Between Permitted Types of Aberration and System Symmetry. If the beam is confined within a circular aperture, radius r_A, so that $x_a^2 + y_a^2 \leq r_A^2$, the aberration figure is circular

$$\left\{(\Delta x_i)^2 + (\Delta y_i)^2\right\}^{1/2} \leq Cr_A^3 \tag{24.40}$$

Thus all rays from a fixed object point that intersect the aperture around a circle of radius r_a intersect the image plane around a circle, centred on the Gaussian image point, of radius MCr_a^3. These rays intersect in a plane distant ζ from the Gaussian image plane. For simplicity, we consider an axial object point (Fig. 24.3); a ray in the neighbourhood of the image plane may be written

$$\begin{aligned} x(z_i + \zeta) &= x_a t(z_i + \zeta) + MCx_a r_a^2 \\ &\approx x_a t'(z_i)\zeta + MCx_a r_a^2 \end{aligned} \tag{24.41}$$

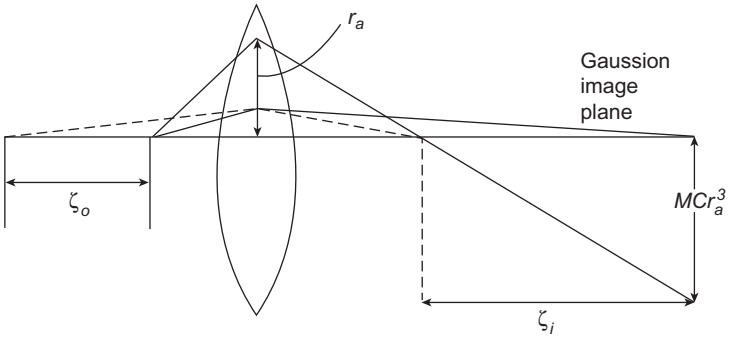


Figure 24.3
Spherical aberration. In the Gaussian image plane, the image of a point object is a circular disc of radius MCr_a^3. The outermost ray intersects the axis at a distance ζ_i from the Gaussian image plane. The axial object point of which this point of intersection is the Gaussian image is distant ζ_o from the original object plane.

and likewise for $y(z_i + \zeta)$. This ray crosses the axis in the plane for which

$$\zeta = -\frac{MC}{t_i'} r_a^2 \tag{24.42}$$

and clearly all rays for which r_a is constant intersect the axis at this point. The coefficient MC/t_i' may be written $W_s C/\hat{\phi}_i^{1/2} t_i'^2$; it therefore has the same sign as CW_s and hence as the integral in the expression for C (24.32b). We shall see that this integral is never negative, and so ζ is always negative (24.42). *Axial rays affected by spherical aberration intersect the axis before reaching the image plane* (Fig. 24.3). The value of ζ corresponding to r_A is known as the longitudinal spherical aberration:

$$\zeta_i := -\frac{MC}{t_i'} r_A^2 = -\frac{M^2 C}{t_o'} \left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{1/2} r_A^2 \tag{24.43a}$$

or referred back to object space,

$$\zeta_o := -\frac{\zeta_i}{M^2} = \frac{C}{t_o'} \left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{1/2} r_A^2 \tag{24.43b}$$

Another way of picturing the fact that ζ_i is always negative is to say that rays far from the axis are focused more strongly than those closer to it.

It is frequently more convenient to measure the size of the spherical aberration disc in terms of the range of ray gradients at the object instead of the radius of the aperture. Rather than return to the formula for aberration coefficients expressed in terms of position and gradient at the object, we can simply transform (24.39) by recalling that the rays of the pencil from an object point (x_o, y_o) are described by $x = x_o s + x_a t$, $y = y_o s + y_a t$ so that $x'_o = x_o s'_o + x_a t'_o$, $y'_o = y_o s'_o + y_a t'_o$. Thus

$$\begin{aligned}\Delta x_i &= C x_a (x_a^2 + y_a^2) = \frac{C}{t'^3_o} x'_o (x'^2_o + y'^2_o) + \text{mixed terms} \\ \Delta y_i &= C y_a (x_a^2 + y_a^2) = \frac{C}{t'^3_o} y'_o (x'^2_o + y'^2_o) + \text{mixed terms}\end{aligned} \tag{24.44}$$

in which the mixed terms contribute to other aberrations but none of the latter contributes to spherical aberration. The coefficient C/t'^3_o is usually denoted by C_s, (the suffix recalling the 'spherical' origin of this defect in the case of glass lenses):

$$\begin{aligned}\Delta x_i &=: C_s x'_o (x'^2_o + y'^2_o) \\ \Delta y_i &=: C_s y'_o (x'^2_o + y'^2_o)\end{aligned} \tag{24.45}$$

and from Eq. (24.32), we see that

$$C_s = \frac{1}{\hat{\phi}_o^{1/2}} \int_{z_o}^{z_i} (L_1 h^4 + 2L_2 h^2 h'^2 + L_3 h'^4) dz \tag{24.46}$$

in which we have used the fact that $h(z) = t(z)/t'_o$. In works on aberration correction, C_s is often denoted by C_3 (the next higher order axial aberration being denoted by C_5).

The radius of the spherical aberration disc in the Gaussian image plane is MCr_A^3. Is it smaller in any neighbouring plane? We return to Eq. (24.41), which we now write

$$x(z_i - \zeta) = -x_a t'_i \zeta + MC x_a r_a^2 \tag{24.47}$$

The ray that is most distant from the axis in the image plane intersects the aperture plane around the circle $x_A^2 + y_A^2 = r_A^2$; for a given magnification, this ray and a general ray are equidistant from the axis in some plane $z = z_i - \zeta$ if $x(z_i - \zeta) = -x_A t'_i \zeta + MC x_A r_A^2$ or

$$\zeta t'_i (x_A - x_a) = MC(x_A^3 - x_a^3)$$

or again

$$\zeta = MC(x_A^2 + x_A x_a + x_a^2)/t'_i \tag{24.48}$$

(in which we have set $y_a = 0$ to simplify the calculation—no generality is lost). In this plane

$$x(z_i - \zeta) = -MC(x_a + x_A) x_a x_A \tag{24.49}$$

and this is smallest in the plane for which $dx(z_i - \zeta)/dx_a = 0$, namely, that in which $x_a = -x_A/2$. For this value, Eq. (24.48) tells us that

$$\zeta = \frac{3}{4}\frac{MC}{t_i'}x_A^2 = \frac{3}{4}|\zeta_i| \tag{24.50}$$

and the beam radius is

$$x(z_i - \zeta) = \frac{1}{4}MCx_A^3 \tag{24.51}$$

Thus the radius of this *disc of least confusion* is only one quarter of that of the spherical aberration disc in the Gaussian image plane; the disc is formed in a plane three-quarters of the way from the Gaussian image plane to the plane of the marginal focus.

In practice, the spherical aberration coefficient is tabulated for the case of greatest interest, objective lenses, used either as the first lens beyond the specimen in a magnifying system or as the final lens just before the target in a demagnifying (probe-forming) system. How are these values related? We shall discuss such relations in detail in the case of asymptotic aberrations in Chapter 25, Asymptotic Aberration Coefficients, but a simple argument enables us to relate the values of the real spherical aberration coefficient in the two situations. When the aberration is expressed in terms of aperture coordinates, the question is almost trivial since the only ray occurring in the integral in C is $t(z)$, which is unaffected by interchanging z_o and z_i. The size of the aberration disc, referred back to the particular object plane, is governed by W and hence by t_o' and $\hat{\phi}_o^{1/2}$ only. When the aberration is expressed in terms of gradient (x_o', y_o'), however, confusion may arise. Consider the situation illustrated in Fig. 24.4A, which shows three rays; ray 1 is a paraxial ray connecting the axial object and image points P_o and P_i. Ray 2 sets out from P_o but strikes the image plane at $\overline{P}_i(\overline{x}_i, \overline{y}_i)$, owing to the spherical aberration; ray 3, also affected by spherical aberration, passes through P_i and must hence have set out from some point $\overline{P}_o$ off the axis. Rays 1 and 2 have the same gradient at P_o, and rays 1 and 3 at P_i. We have $P_i\overline{P}_i = MC_s^{(oi)}\theta_o^3$ and $P_o\overline{P}_o = C_s^{(io)}\theta_i^3/M$, where $C_s^{(oi)}$ and $C_3^{(io)}$ are the values of C_s for the cases in which the electron travel from P_o towards P_i and the reverse, respectively. From Eqs. (24.39, 24.45), however, we see that

$$0 = MP_o\overline{P}_o + MC_s^{(oi)}\theta_o^3$$

or

$$P_o\overline{P}_o = -C_s^{(oi)}\theta_0^3 \tag{24.52}$$

Hence

$$\frac{1}{M}C_s^{(io)}\theta_i^3 = -C_s^{(oi)}\theta_0^3$$

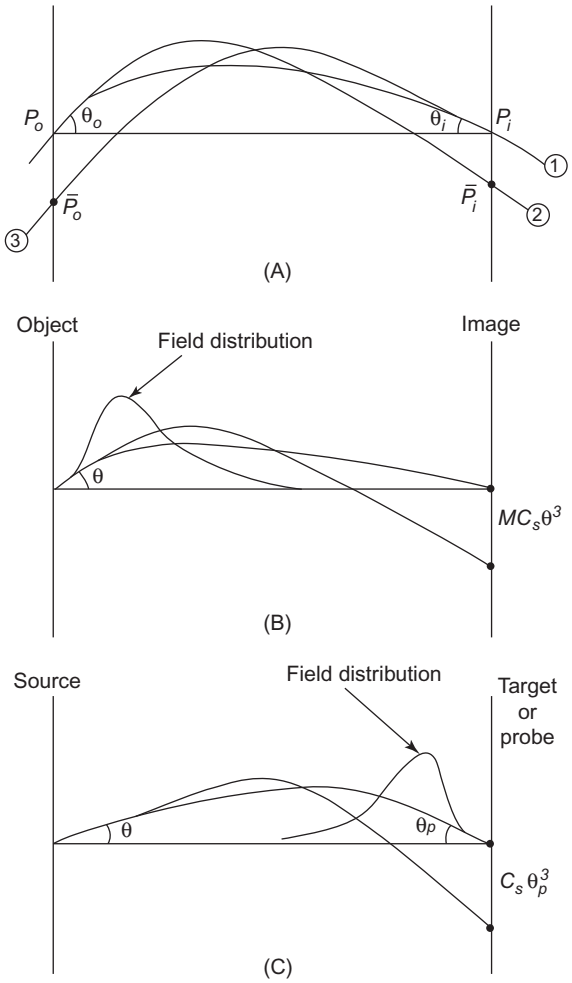


Figure 24.4
Forward and backward aberration coefficients. (A) Calculation of the relation between the coefficients. (B) Aberration disc ($|M| \gg 1$). (C) Probe size ($|M| \ll 1$).

or

$$C_s^{(io)} = -M\left(\frac{\theta_o}{\theta_i}\right)^3 C_s^{(oi)} = M^4\left(\frac{\hat{\phi}_i}{\hat{\phi}_o}\right)^{3/2} C_s^{(oi)} \tag{24.53}$$

or for magnetic lenses

$$C_s^{(io)} = M^4 C_s^{(oi)} \tag{24.54}$$

The forward and backward spherical aberration coefficients are thus very different. Eq. (24.54) is an obvious consequence of the relation between C_s and C (24.44). Figs 24.4B

and C depict a field distribution used as a magnifying and demagnifying system. Eq. (24.52) shows that if the spherical aberration coefficient for the magnifying situation is C_s, the radius of the corresponding probe will be given by $C_s\theta_p^3$, where θ_p denotes the angular aperture at the probe. That this is consistent with Eq. (24.53) is easily seen from the fact that the probe radius r_p is given by

$$\begin{aligned} r_p &= M^{(io)} C_s^{(io)} \theta^3 \\ &= \frac{1}{M} M^4 \left(\frac{\hat{\phi}_i}{\hat{\phi}_o} \right)^{3/2} C_s^{(oi)} \frac{1}{M^3} \left(\frac{\hat{\phi}_o}{\hat{\phi}_i} \right)^{3/2} \theta_p^3 \\ &= C_s^{(oi)} \theta_p^3 \end{aligned} \tag{24.55}$$

In practice, these relations are often required for magnetic lenses alone, and the important result is then Eq. (24.54). In the study of electron guns, however, Eq. (24.53) is indispensable.

We mentioned at the end of Section 24.2 that each aberration can be associated with a characteristic distortion of the surfaces $S = \text{const}$. For the spherical aberration, we have $S'/W = -\frac{1}{4} C r_a^4$ and so, as shown in Fig. 24.5, the true surface S (including aberrations) is everywhere closer to the image plane than the reference sphere S_o, except at the axis where the two touch. Since the curvature of S is greater than that of S_o, it is immediately obvious that outer rays will be more strongly focused than those close to the axis and hence the marginal focus retreats towards the lens as the aperture is opened more widely.

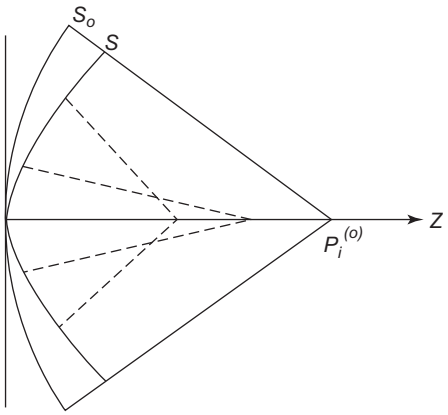


Figure 24.5
Spherical wave surface S_o centred on the Gaussian image point and aspherical wave surface S distorted by spherical aberration. As the aperture is widened, the image point corresponding to the outermost rays, or wave-normals, retreats towards the lens.

We now turn to the coefficient itself, C or C_s. The formulae for these can be rewritten in numerous ways, very different in appearance but otherwise equivalent. These expressions can be obtained either by partial integration using the paraxial equations to replace second derivatives of the paraxial rays s and t or g and h whenever they occur, or by an ingenious differential technique introduced by Seman (1951, 1954, 1955a–c, 1958b) and exploited extensively by Hawkes (1966/7b, 1967b). Seman in fact applied his method not to individual aberration coefficients but to the characteristic function S^I, as we explain in Section 24.9. We introduce it now, however, as it provides an extremely convenient way of analysing individual coefficients for which a formula is already available, without returning to the characteristic function. We set out from Eq. (24.46),

$$C_s = \frac{1}{\hat{\phi}_o^{1/2}} \int (L_1 h^4 + 2L_2 h^2 h'^2 + L_3 h'^4) dz$$

in which L_1, L_2 and L_3 are given by Eq. (24.3). Retention of the relativistic terms in the electrostatic case always renders the calculation extremely heavy, however, and we shall therefore apply the method to the nonrelativistic form of C_s, obtained by setting $\varepsilon = 0$, $\hat{\phi} = \phi$ in Eq. (24.3). The relativistic results are merely listed.

We begin by noticing that the quantities ϕ, B and h and the operator (d/dz) occur in only a few combinations in the integrand of C_s, namely $(d/dz)^2\ (B^2/\phi^{1/2})h^4$ and $(d/dz)^4\ \phi^{1/2}h^4$. A set of terms with comparable structure is obtained by differentiating once all the terms generated by $(d/dz)(B^2/\phi^{1/2})h^4$ and $(d/dz)^3\phi^{1/2}h^4$, eliminating h'' as necessary with the aid of the paraxial equation, $h'' = -(\phi'/2\phi)h' - (\phi'' + \eta^2 B^2)h/4\phi$. The first term generates

$$\frac{BB'}{\phi^{1/2}}h^4; \quad \frac{B^2\phi'}{\phi^{3/2}}h^4; \quad \text{and} \quad \frac{B^2}{\phi^{1/2}}h^3h'$$

and the other,

$$\begin{aligned} &\frac{\phi'''}{\phi^{1/2}}h^4; \quad \frac{\phi'\phi''}{\phi^{3/2}}h^4; \quad \frac{\phi'^3}{\phi^{5/2}}h^4; \\ &\frac{\phi''}{\phi^{1/2}}h^3h'; \quad \frac{\phi'^2}{\phi^{3/2}}h^3h'; \\ &\frac{\phi'}{\phi^{1/2}}h^2h'^2; \quad \text{and} \quad \phi^{1/2}hh'^3 \end{aligned} \tag{24.56}$$

Each of these is differentiated and then formally integrated; the term in $\phi^{1/2}hh'^3$, for example, yields

$$\frac{d}{dz}(\phi^{1/2}hh'^3) = \frac{\phi'}{2\phi^{1/2}}hh'^3 + \phi^{1/2}h'^4 - 3\phi^{1/2}hh'^2\left(\frac{\phi'}{2\phi}h' + \frac{\phi'' + \eta^2B^2}{4\phi}h\right) \tag{24.57a}$$

so that

$$\left[\phi^{1/2}hh'^3\right]_{z_o}^{z_i} + \int_{z_o}^{z_i}\left(\frac{\phi'}{\phi^{1/2}}hh'^3 + \frac{3}{4}\frac{\phi'' + \eta^2B^2}{\phi^{1/2}}h^2h'^2 - \phi^{1/2}h'^4\right)dz \equiv 0 \tag{24.57b}$$

In all we have ten such identities. After first multiplying each by an arbitrary constant p_j, $j = 1-10$, these are all added to the integral in C_s, giving

$$\begin{aligned}
32\phi_o^{1/2}C_s = \int_{z_o}^{z_i}\phi^{1/2}\Bigg[&h'^4(16 + p_1) + h'^3h\frac{\phi'}{\phi}(-p_1 + 2p_2) \\
&+ h'^2h^2\left\{\frac{\phi''}{\phi}\left(8 - \frac{3}{4}p_1 + p_2 + 3p_3\right) + \left(\frac{\phi'}{\phi}\right)^2\left(-\frac{3}{2}p_2 + 3p_4\right)\right. \\
&\left.+ \frac{\eta^2B^2}{\phi}\left(8 - \frac{3}{4}p_1 + 3p_5\right)\right\} \\
&+ h'h^3\left\{\frac{\phi'''}{\phi}(p_3 + 4p_6) + \frac{\phi''\phi'}{\phi^2}\left(-\frac{1}{2}p_2 - p_3 + 2p_4 + 4p_7\right)\right. \\
&+ \left(\frac{\phi'}{\phi}\right)^3(-2p_4 + 4p_8) + \frac{\eta^2B^2\phi'}{\phi^2}\left(-\frac{1}{2}p_2 - p_5 + 4p_9\right) \\
&\left.+ \frac{\eta^2BB'}{\phi}(2p_5 + 4p_{10})\right\} \\
&+ h^4\left\{\frac{\phi^{(4)}}{\phi}(p_6 - 1) + \frac{\phi'''\phi'}{\phi^2}\left(-\frac{1}{2}p_6 + p_7\right)\right. \\
&+ \left(\frac{\phi''}{\phi}\right)^2\left(1 - \frac{1}{4}p_3 + p_7\right) - \frac{5}{2}\left(\frac{\phi'}{\phi}\right)^4p_8 \\
&+ \frac{\phi''\phi'^2}{\phi^3}\left(-\frac{1}{4}p_4 - \frac{3}{2}p_7 + 3p_8\right) + \frac{\eta^2B^2\phi''}{\phi^2}\left(2 - \frac{p_3}{4} - \frac{p_5}{4} + p_9\right) \\
&\left.+ \frac{\eta^2B^2\phi'^2}{\phi^3}\left(-\frac{1}{4}p_4 - \frac{3}{2}p_9\right) + \frac{\eta^2BB'\phi'}{\phi^2}\left(2p_9 - \frac{1}{2}p_{10}\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\eta^4B^4}{\phi^2}\left(1-\frac{1}{4}p_5\right)+\frac{\eta^2BB''}{\phi}\left(-4+p_{10}\right)+\frac{\eta^2B'^2}{\phi}p_{10}\Bigg\}\Bigg]dz\\
&-\Bigg[\phi^{1/2}\Big(p_1hh'^3+p_2\frac{\phi'}{\phi}h^2h'^2+p_3\frac{\phi''}{\phi}h^3h'\\
&+p_4\frac{\phi'^2}{\phi^2}h^3h'+p_5\frac{\eta^2B^2}{\phi}h^3h'+p_6\frac{\phi'''}{\phi}h^4\\
&+P_7\frac{\phi''\phi'}{\phi^2}h^4+p_8\frac{\phi'^3}{\phi^3}h^4+p_9\frac{\eta^2B^2\phi'}{\phi^2}h^4\\
&+p_{10}\frac{\eta^2BB'}{\phi}h^4\Bigg]_{z_o}^{z_i}
\end{aligned}
\tag{24.58}
$$

Since the multipliers p_i are arbitrary, we may choose them in any self-consistent way to eliminate terms that are for some reason undesirable. For numerical work, for example, we usually eliminate high derivatives of ϕ and B; if a field model that permits the integrals to be evaluated in closed form is being studied, we may well prefer to have as few terms to integrate as possible. Again, we might wish to establish whether C_s can change sign, in which case we enquire whether or not the integrand can be written as a sum of squared terms. The integrated terms all vanish since $h(z)$ appears undifferentiated in each of them. They must not, however, be completely forgotten, for they can affect the thin-lens approximation, examined below.

In practice, the simultaneous presence of electrostatic and magnetic fields is rare and we therefore list a number of forms of C_s for the two separate cases, $B=0$ and $\phi=\text{const}$. The mixed forms may be obtained by manipulating Eq. (24.58).

24.3.1 Electrostatic case ($\mathbf{B=0}$, $\phi\neq$ const)

General Relativistic Expression

$$
\begin{aligned}
32\hat{\phi}_0^{1/2}C_s=\int&(A_0h^4+2A_1h^3h'+2A_2h^2h'^2\\
&+2A_3hh'^3+A_4h'^4)dz
\end{aligned}
\tag{24.59}
$$

in which

$$
\begin{aligned}
A_0=&\frac{\gamma\phi^{(4)}}{\hat{\phi}^{1/2}}(p_6-1)+\frac{\phi''^2}{\hat{\phi}^{3/2}}\left\{\left(-\frac{p_3}{4}+1\right)\gamma^2+p_7\right\}\\
&-4\varepsilon\frac{\phi''^2}{\hat{\phi}^{1/2}}(1+p_7')+\frac{\phi'\phi'''}{\hat{\phi}^{3/2}}\left(-\frac{1}{2}\gamma^2p_6+p_7\right)\\
&-4\varepsilon\frac{\phi'\phi'''}{\hat{\phi}^{1/2}}\left(p_7'-\frac{1}{2}p_6\right)-\gamma\frac{\phi'^2\phi''}{\hat{\phi}^{5/2}}\left(-3p_8+\frac{3}{2}p_7+\frac{1}{4}p_4\right)
\end{aligned}
$$

$$
\begin{aligned}
&- 2\varepsilon \frac{\phi'^2\phi''}{\hat{\phi}^{3/2}}\left\{p'_{10} + \varepsilon\phi\, p'_{11} - \gamma(p'_7 + p'_5)\right\} \\
&- \frac{5}{2}\frac{\gamma^2\phi'^4}{\hat{\phi}^{7/2}}p_8 + \frac{\varepsilon\phi'^4}{\hat{\phi}^{5/2}}(\gamma p'_{10} + 2p_8 + \gamma\varepsilon\phi\, p'_{11}) \\
&- \frac{2}{3}\frac{\varepsilon^2\phi'^4}{\hat{\phi}^{3/2}}p'_{11}
\end{aligned}
\tag{24.60a}
$$

$$
\begin{aligned}
A_1 = {} & \frac{\gamma\phi'''}{\hat{\phi}^{1/2}}\left(\frac{1}{2}p_3 + 2p_6\right) - \frac{\phi''\phi'}{\hat{\phi}^{3/2}}\left\{\left(\frac{1}{2}p_3 + \frac{1}{4}p_2\right)\gamma^2\right. \\
&\left. - (2p_7 + p_4)\right\} - \frac{\varepsilon\phi'\phi''}{\hat{\phi}^{1/2}}(8p'_7 + 8p'_5 - p_3) \\
&+ \frac{\gamma\phi'^3}{\hat{\phi}^{5/2}}(2p_8 - p_4) - \frac{4\varepsilon\phi'^3}{\hat{\phi}^{3/2}}\left(\frac{p'_{10}}{3} - p'_5\right) \\
&- \frac{8\varepsilon^2\phi\phi'^3}{\hat{\phi}^{3/2}}\left(\frac{1}{6}p'_{11} - p'_5\right)
\end{aligned}
\tag{24.60b}
$$

$$
\begin{aligned}
A_2 = {} & \frac{\gamma\phi''}{\hat{\phi}^{1/2}}\left(4 + \frac{3}{2}p_3 + \frac{1}{2}p_2 - \frac{3}{8}p_1\right) \\
&+ 3\frac{\phi'^2}{\hat{\phi}^{3/2}}\left(\frac{1}{2}p_4 - \gamma^2\frac{p_2}{4}\right) \\
&- \frac{4\varepsilon\phi'^2}{\hat{\phi}^{1/2}}\left(3p'_5 - \frac{1}{4}p_2\right)
\end{aligned}
\tag{24.60c}
$$

$$
A_3 = \frac{\gamma\phi'}{\hat{\phi}^{1/2}}\left(p_2 - \frac{1}{2}p_1\right) \tag{24.60d}
$$

$$
A_4 = \hat{\phi}^{1/2}(p_1 + 16) \tag{24.60e}
$$

The nonrelativistic forms of these are as follows:

$$A'_0 \coloneqq A_0(\varepsilon \to 0) = \frac{\phi^{(4)}}{\phi^{1/2}}(p_6 - 1) + \frac{\phi''^2}{\phi^{3/2}}\left(1 - \frac{p_3}{4} + p_7\right) + \frac{\phi'\phi'''}{\phi^{3/2}}\left(p_7 - \frac{1}{2}P_6\right) - \frac{\phi'^2\phi''}{\phi^{5/2}}\left(\frac{1}{4}p_4 + \frac{3}{2}p_7 - 3p_8\right) - \frac{5}{2}\frac{\phi'^4}{\phi^{7/2}}p_8$$

$$A'_1 \coloneqq A_1(\varepsilon \to 0) = \frac{\phi'''}{2\phi^{1/2}}(p_3 + 4p_6) - \frac{\phi''\phi'}{\phi^{3/2}}\left(\frac{1}{2}p_3 + \frac{1}{4}p_2 - 2p_7 - p_4\right) + \frac{\phi'^3}{\phi^{5/2}}(2p_8 - p_4)$$

$$A'_2 \coloneqq A_2(\varepsilon \to 0) = \frac{\phi''}{\phi^{1/2}}\left(4 + \frac{3}{2}p_3 + \frac{1}{2}p_2 - \frac{3}{8}p_1\right) + 3\frac{\phi'^2}{\phi^{3/2}}\left(\frac{1}{2}p_4 - \frac{p_2}{4}\right)$$

$$A'_3 \coloneqq A_3(\varepsilon \to 0) = \frac{\phi'}{\phi^{1/2}}\left(p_2 - \frac{1}{2}p_1\right)$$

$$A'_4 \coloneqq A_4(\varepsilon \to 0) = \phi^{1/2}(p_1 + 16)$$

General Nonrelativistic Expression

$$32\phi_o^{1/2}C_s = \int \phi^{1/2}\Bigg[h'^4(16 + p_1) + h'^3 h\frac{\phi'}{\phi}(-p_1 + 2p_2) + h'^2h^2\left\{\frac{\phi''}{\phi}\left(8 - \frac{3}{4}p_1 + p_2 + 3p_3\right) + \left(\frac{\phi'}{\phi}\right)^2\left(-\frac{3}{2}p_2 + 3p_4\right)\right\} + h'h^3\left\{\frac{\phi'''}{\phi}(p_3 + 4p_6) + \frac{\phi''\phi'}{\phi^2}\left(-\frac{1}{2}p_2 - p_3 + 2p_4 + 4p_7\right) + \left(\frac{\phi'}{\phi}\right)^3(-2p_4 + 4p_8)\right\}$$

$$+h^4\left\{\frac{\phi^{(4)}}{\phi}(p_6-1)+\frac{\phi'''\phi'}{\phi^2}\left(-\frac{1}{2}p_6+p_7\right)+\left(\frac{\phi''}{\phi}\right)^2\left(1-\frac{1}{4}p_3+p_7\right)-\frac{5}{2}\left(\frac{\phi'}{\phi}\right)^4 p_8+\frac{\phi''\phi'^2}{\phi^3}\left(-\frac{1}{4}p_4-\frac{3}{2}p_7+3p_8\right)\right\}\Bigg]dz$$

$$-\left[\phi^{1/2}\left\{p_1hh'^3+p_2\frac{\phi'}{\phi}h^2h'^2+p_3\frac{\phi''}{\phi}h^3h'+p_4\frac{\phi'^2}{\phi^2}h^3h'+p_5\left(\frac{\phi'}{\phi}\right)^3h^4+p_6\frac{\phi'''}{\phi}h^4+p_7\frac{\phi''\phi'}{\phi^2}h^4\right\}\right]_{z_o}^{z_i} \tag{24.61}$$

Numerous forms are to be found in the literature, all of which correspond to various choices of the coefficients p_i. For example, eliminating all terms involving $h'(z)$ from Eq. (24.61), we find

$$32\phi_0^{1/2}C_s=\int\phi^{1/2}\left(-\frac{\phi'''\phi'}{2\phi^2}+2\frac{\phi''^2}{\phi^2}+5\left(\frac{\phi'}{\phi}\right)^4-5\frac{\phi''\phi'^2}{\phi^3}\right)h^4dz \tag{24.62}$$

which may also be written in terms of

$$\psi(z)\coloneqq\phi'(z)/\phi(z) \tag{24.63}$$

as

$$64\phi_o^{1/2}C_s=\int\phi^{1/2}(4\psi'^2+3\psi^4-5\psi^2\psi'-\psi\psi'')h^4dz \tag{24.64}$$

24.3.2 Magnetic case (ϕ = const, B≠0)

General Relativistic Case

$$\begin{aligned}32C_s=\int\Bigg\{&h'^4(16+p_1)+h^2h'^2\frac{\eta^2B^2}{\hat{\phi}}\left(8-\frac{3}{4}p_1+3p_5\right)\\&+h^3h'\frac{\eta^2BB'}{\hat{\phi}}(2p_5+4p_{10})+h^4\frac{\eta^4B^4}{\hat{\phi}^2}\left(1-\frac{1}{4}p_5\right)\\&+h^4\frac{\eta^2BB''}{\hat{\phi}}(-4+p_{10})+h^4\frac{\eta^2B'^2}{\hat{\phi}}p_{10}\Bigg\}dz\\&-\left[p_1hh'^3+p_5\frac{\eta^2B^2}{\hat{\phi}}h^3h'+p_{10}\frac{\eta^2BB'}{\hat{\phi}}h^4\right]_{z_o}^{z_i}\end{aligned} \tag{24.65}$$

All derivatives of h can be eliminated by writing $p_1 = -16$, $p_5 = -20/3$ and $p_{10} = 10/3$, giving

$$C_s = \frac{1}{48}\int \left(5\frac{\eta^2 B'^2}{\hat{\phi}} - \frac{\eta^2 BB''}{\hat{\phi}} + 4\frac{\eta^4 B^4}{\hat{\phi}^2}\right)h^4 dz \tag{24.66}$$

Other useful forms are as follows:

$$C_s = \int \left\{ \left(\frac{\eta^4 B^4}{32\hat{\phi}^2} - \frac{\eta^2 BB''}{8\hat{\phi}}\right)h^4 + \frac{\eta^2 B^2}{4\hat{\phi}}h^2 h'^2 + \frac{1}{2}h'^4 \right\} dz \tag{24.67a}$$

$$= \int \left\{ \left(\frac{\eta^4 B^4}{32\hat{\phi}^2} - \frac{\eta^2 BB''}{8\hat{\phi}}\right)h^4 + \frac{5\eta^2 B^2}{8\hat{\phi}}h^2 h'^2 \right\} dz \tag{24.67b}$$

$$= \int \left\{ \left(\frac{3\eta^4 B^4}{32\hat{\phi}^2} + \frac{\eta^2 B'^2}{8\hat{\phi}}\right)h^4 - \frac{\eta^2 B^2}{8\hat{\phi}}h^2 h'^2 \right\} dz \tag{24.67c}$$

$$= \int \left\{ \frac{\eta^4 B^4}{16\hat{\phi}^2}h^4 + (hB' + h'B)^2\frac{\eta^2 h^2}{8\hat{\phi}} + \frac{\eta^2 B^2}{8\hat{\phi}}h^2 h'^2 \right\} dz \tag{24.67d}$$

$$= \frac{1}{2}\int \left\{ \left(h'^2 - \frac{\eta^2 B^2}{4\hat{\phi}}h^2\right)^2 + \frac{\eta^2 h^2}{4\hat{\phi}}(2Bh' + B'h)^2 \right\} dz \tag{24.67e}$$

24.3.3 Scherzer's Theorem

One of the aims of the designers of the first electron lenses was to find combinations of lens geometry and excitation for which the aberration coefficients, especially C_s, were small, preferably zero. It was not, however, long before Otto Scherzer (1936b) demonstrated that the formula for C_s can be written as the integral of a sum of squared terms, so that unless all these vanish, we can at best find the lens that corresponds to a minimum value of C_s. Scherzer's expression was nonrelativistic but Rose (1967/8) has since derived the relativistic expression[3]:

$$C_s = \frac{1}{32}\int_{z_o}^{z_i} \left(\frac{\hat{\phi}}{\hat{\phi}_o}\right)^{1/2} \overline{C} h^4 dz$$

where

[3] The expression in Rose (1967/8) and reproduced in the first edition of this book contained an error, pointed out by Preikszas and Rose (1995). Equation 24.68 is essentially the same as Eq. (13) of Preikszas and Rose (1995) and Eq. (8.66) of Rose (2012), apart from notation.

$$\begin{aligned}\overline{C} = &\frac{\phi'^2}{\hat{\phi}^2}\left(\frac{h'}{h}+\frac{5\gamma}{6}\frac{\phi'}{\hat{\phi}}\right)^2 + \frac{3}{2}\left(\frac{\phi''}{\hat{\phi}}+\frac{\phi'}{\hat{\phi}}\frac{h'}{h}-\gamma\frac{\phi'^2}{\hat{\phi}^2}\right)^2 + \frac{2\gamma^2-1}{36}\frac{\phi'^4}{\hat{\phi}^4}\\ &+\frac{\phi'^2}{\hat{\phi}^2}\left(\gamma\frac{h'}{h}+\frac{2+3\gamma^2}{6}\frac{\phi'}{\hat{\phi}}\right)^2\\ &+\left(\gamma\frac{\phi''}{\hat{\phi}}+\gamma\frac{\phi'}{\hat{\phi}}\frac{h'}{h}-\frac{3+2\gamma^2}{4}\frac{\phi'^2}{\hat{\phi}^2}+\frac{\eta^2B^2}{\hat{\phi}}\right)^2\\ &+\frac{4\eta^2B^2}{\hat{\phi}}\left[\left(\frac{h'}{h}+\frac{B'}{B}-\frac{3\gamma}{4}\frac{\phi'}{\hat{\phi}}\right)^2+\left(\frac{h'}{h}+\frac{\gamma}{2}\frac{\phi'}{\hat{\phi}}\right)^2\right]\\ &+\frac{21+2\gamma^2}{8}\frac{\eta^2B^2}{\hat{\phi}}\frac{\phi'^2}{\hat{\phi}^2}+\frac{\eta^4B^4}{\hat{\phi}^2}\end{aligned} \tag{24.68}$$

This expression does not however collapse to Scherzer's formula in the nonrelativistic limit. There is thus more than one positive-definite form of the nonrelativistic integrand; it is not known whether this is also true of the relativistic integrand.

Scherzer's original formula is given by Eq. (24.58) if we set

$$\begin{aligned}&p_1=-16;\ \ p_2=-8;\ \ p_3=-4;\ \ p_4=-5/2;\ \ p_5=-4;\\ &p_6=1;\ \ p_7=1/2;\ \ p_8=-13/8;\ \ p_9=-4;\ \ p_{10}=4\end{aligned} \tag{24.69}$$

The nonrelativistic version of Eq. (24.68) is obtained by replacing p_8 in Eq. (24.69) by $-43/24$ and p_9 by -2; the other values of p_j are unaltered. The electrostatic part of the relativistic expression given by Rose corresponds to the following values of the p_i given in Eqs. (24.59, 24.60):

$$\begin{aligned}&p_1=-16; \quad p_2=-8; \quad p_3=-4; \quad p_4=-5/2; \quad p'_5=1;\\ &p_6=1; \quad p_7=1/2; \quad p_8=-43/24; \quad p'_7=0; \quad p'_{10}=3;\\ &p'_{11}=6\end{aligned} \tag{24.70}$$

For magnetic lenses, the appropriate formula is (24.67d). We note in passing that the list contains two positive-definite forms of the integrand, (24.67d and e), the latter recommended by Lenc (1992).

This result has had an immense influence on electron optical studies. It is true provided that the derivations of the various expressions are correct, which requires that the lens be round, static and free of space charge; furthermore, $\phi(z)$ and its derivatives must be continuous and the object and image must be real. Scherzer himself (1947) proposed methods of correcting spherical aberration by relaxing one or other of these conditions: the use of components

with lower symmetry, such as quadrupoles and octopoles; the introduction of space charge or a potential discontinuity; excitation of the electrodes of an 'electrostatic' lens at high frequency. We shall return to these various possibilities in Chapter 41 of Volume 2. Meanwhile, Glaser (who constantly sought loopholes in Scherzer's proof, see HdP p.227 footnote and GdE, p.677 n.163) had the ingenious idea of seeking a magnetic lens for which $C_s = 0$ by solving the differential equation obtained by setting the integrand in C_s (in the form 24.66) equal to zero (Glaser, 1940a; recalled in Lenz, 1982b and Hawkes, 1986); this failed because the resulting field distribution proved to be incapable of producing a real image of a real object (Rebsch, 1940) but was found useful in β-ray spectroscopy (Siegbahn, 1946). Recknagel (1941) analysed the electrostatic lens integrand in the same way. Rather later, Tretner established the minimum value of C_s as a function of the various lens parameters. His work and the complementary analysis of Moses is examined in Part VII. Fresh attempts to overturn Scherzer's result continue to appear (e.g., Nomura (2008), refuted by Hawkes (2009c) and Garg (1982), Scherzer's rebuttal of which (1982) ended with the word 'Sorry!')

24.3.4 Thin-Lens Approximation

If the lens is assumed to be thin, in the sense discussed in Section 17.4, each of the aberration coefficients collapses to a simpler form. For the spherical aberration, we set out from (24.66) in the magnetic case. An expression may likewise be derived for electrostatic lenses (Riedl, 1937) but has been little used in practice, no doubt because the potential distribution in the latter is rarely as narrow as the field distribution in many magnetic lenses. Hawkes (1987) shows that the thin-lens approximations to the various coefficients explain the empirical findings of Renau and Heddle (1986a,b, 1987). The corresponding expressions are listed at the end of Chapter 25, Asymptotic Aberration Coefficients.

Eq. (24.66) gives straightforwardly

$$C_s = z_o^4 \left\{ \frac{\eta^4}{12\hat{\phi}^2} \int_{-\infty}^{\infty} B^4(z)dz + \frac{\eta^2}{8\hat{\phi}} \int \left(\frac{dB}{dz}\right)^2 dz \right\} \tag{24.71}$$

in which z_o is the distance of the object from the lens centre but it is usual to rewrite this in terms of dimensionless quantities by scaling distances with respect to some characteristic length of the lens, such as the gap between the polepieces, S. Writing

$$z/S =: \zeta \tag{24.72}$$

and setting $B(z) = B_0 b(z)$, where B_0 is the maximum value of $B(z)$, Eq. (24.71) becomes

$$\frac{C_s}{S} = S^4\zeta_o^4\left(\frac{\eta^4 B_0^4}{12\hat{\phi}^2}\tau_4 + \frac{\eta^2 B_0^2}{8S^2\hat{\phi}}\tau_0\right) \tag{24.73}$$

in which

$$\begin{aligned}\zeta_o &= \frac{z_o}{S}\\ \tau_4 &:= \int_{-\infty}^{\infty} b^4(\zeta)d\zeta\\ \tau_0 &:= \int_{-\infty}^{\infty}\left(\frac{db}{d\zeta}\right)^2 d\zeta\end{aligned} \tag{24.74}$$

in agreement with the expression given by Glaser (1952, Eq.124.7) for high magnification, $z_o = f$. Approximate expressions for (24.73) as a function of a parameter characterizing the lens geometry have been derived by Der-Shvarts (1970) and Der-Shvarts and Makarova (1972, 1973); these are analysed in Hawkes (1980b) and at greater length in Hawkes et al. (1995), where formulae proposed by Crewe (1991a–c) and Kanaya et al. (1966) are also examined. Several early papers discuss these thin and/or weak lens approximations; see, for example, Scherzer (1936a), Rebsch and Schneider (1937), Gratsiatos (1937), Voit (1939) and Marschall (1939).

24.4 Coma (Terms Linear in x_o, y_o)

For lenses such as objectives and probe-forming elements, in which the rays are inclined to the axis at a relatively steep angle, the coma is the next most important aberration after the spherical aberration. If we consider a pencil of rays intersecting the aperture around a circle of radius r_a, the corresponding aberration figure will be given by

$$\Delta u_i - 2(K + \mathrm{i}k)r_a^2 r_o \mathrm{e}^{\mathrm{i}\varphi_o} = (K - \mathrm{i}k)r_a^2 r_o \mathrm{e}^{\mathrm{i}(2\varphi_a - \varphi_o)} \tag{24.75}$$

in which we have transferred the term independent of φ_a to the left-hand side since it merely shifts the image point by a distance $2\sqrt{K^2 + k^2}r_a^2 r_o$ along a line inclined at an angle arctan (k/K) to the radius vector of the Gaussian image point. The term on the right-hand side causes the image point to rotate around a circle, centred on the shifted image point, of radius $\sqrt{K^2 + k^2}r_a^2 r_o$. The circle is described twice as φ_a varies from zero to 2π. This behaviour may also be seen by multiplying each side of Eq. (24.75) by its complex conjugate, which gives

$$\begin{aligned}&\{\Delta x - 2r_a^2(Kx_o - ky_o)\}^2 + \{\Delta y - 2r_a^2(Ky_o + kx_o)\}^2\\ &\quad = (K^2 + k^2)r_a^4 r_o^2\end{aligned} \tag{24.76}$$

This clearly represents a circle of radius $(K^2 + k^2)^{1/2}r_a^2 r_o$ centred on the point $(2r_a^2(Kx_o - ky_o),\ \ 2r_a^2(Ky_o + kx_o))$ The tangents to this circle from the origin are inclined to

the line joining the origin to the centre at an angle ψ and sin $\psi = 1/2$ hence $\psi = 30°$. For different values of r_a, we obtain a family of circles, the centres of which lie on a straight line, all tangent to the pair of straight lines inclined to the line of centres at $\pm 30°$ (Fig. 24.6). A beam filling the aperture will therefore generate a comet-shaped aberration figure, from which the aberration takes its name.

Coma is characterized by two coefficients, K and k; if $k = 0$, the line of centres passes through the Gaussian image point and the origin in the image plane: the coma is *radial* and is exactly analogous to that of glass lenses; K is thus known as the *isotropic coma coefficient*. If, on the other hand, $K = 0$ but $k \neq 0$, then the line of centres is perpendicular to the line joining the origin to the Gaussian image point: the coma is *sagittal* (or tangential) and has no analogue in light optics; k is known as the *anisotropic coma coefficient*, and is peculiar to magnetic lenses.

Like every aberration, the coma may be interpreted in terms of the corresponding distortion of the surface $S = \text{const}$; we have

$$-S^I/W = r_a^3 r_o(K \cos\varphi + k \sin\varphi) = r_a^3 r_o\sqrt{K^2 + k^2}\cos(\varphi - \varphi_k),\ \tan\varphi_k := k/K$$

and the surface S is therefore partly on the image side of the reference sphere S_o, partly on the object side. It is not so easy to relate the distortion of S to the aberration figure as in the case of spherical aberration; Fig. 24.7 shows the difference between S and S_o.

The coefficients K and k may be written in numerous forms. The relativistic expressions for mixed lenses (24.32 with 24.3) in terms of object and aperture coordinates are

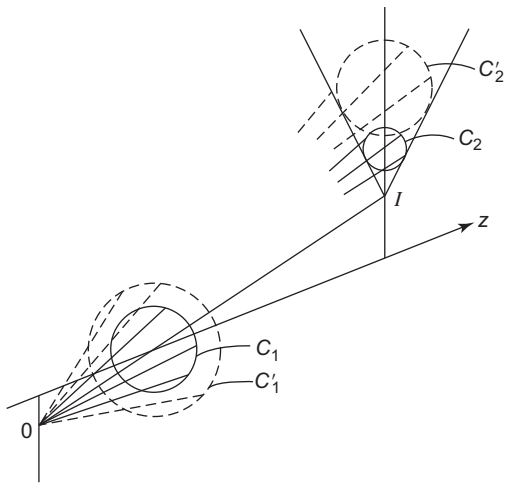


Figure 24.6

Coma. Rays from an object point O intersect the aperture plane around concentric circles C_1, C'_1 and subsequently intersect the image plane around non-concentric circles C_2, C'_2.

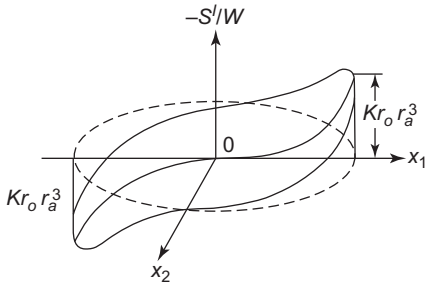


Figure 24.7
Coma. The distance between S and S_o as measured by $-S'/W$. The axes are such that x_1 corresponds to $\varphi - \varphi_k = 0$ and x_2 to $\varphi - \varphi_k = \pi/2$.

$$K = \frac{1}{W_s}\int_{z_o}^{z_i}(L_1 s t^3 + L_2(st)'tt' + L_3 s' t'^3)dz$$
$$k = \int_{z_o}^{z_i}(Pt^2 + Qt'^2)dz \tag{24.77}$$

By substituting $x_a = x'_o/t'_o - x_o s'_o/t'_o$ and likewise for y_a in the complete expressions for Δx_i, Δy_i and using the relations $g = s - s'_o t/t'_o$, $h = t/t'_o$, it is readily shown that the coma coefficients have the same form as Eq. (24.77), namely

$$K = \frac{1}{W_g}\int(L_1 g h^3 + L_2(gh)'hh' + L_3 g' h'^3)dz$$
$$k = \int(Ph^2 + Qh'^2)dz \tag{24.78}$$

in terms of position and gradient in the object plane; note that both the spherical aberration and the coma contribute when making the substitutions. In the following expressions, therefore, s and t may be replaced by g and h at will.

For *electrostatic lenses*, the general relativistic expression for K is

$$\begin{aligned} 32\hat{\phi}_o^{1/2}K = \frac{1}{t'_o}\int_{z_o}^{z_i}\Big\{ & A_0 s t^3 + \frac{1}{2}A_1\left(3s't^2 + st'^2\right)t \\ & + A_2(st + s't')tt' + \frac{1}{2}A_3\left(s't^2 + 3st'^2\right)t' \\ & + A_4 s' t'^3\Big\}dz \\ & - \frac{p_1}{4}\hat{\phi}_o^{1/2}t'_o\left(t'^2_i - t'^2_o\right) \end{aligned} \tag{24.79}$$

where the A_i are given by Eq. (24.60). The corresponding nonrelativistic form is obtained by setting A'_i Eq. (24.60') in place of A_i.

For magnetic lenses, we find

$$
\begin{aligned}
K = \frac{1}{t'_o}\int_{z_o}^{z_i} \Bigg\{ & st^3\left(\frac{\eta^4 B^4}{32\hat{\phi}^2} - \frac{\eta^2 BB''}{8\hat{\phi}} + p_1\frac{\eta^2 B'^2}{\hat{\phi}} + p_1\frac{\eta^4 BB''}{\hat{\phi}}\right.\\
& \left. - p_2\frac{\eta^4 B^4}{\hat{\phi}^2} - p_3\frac{\eta^4 B^4}{4\hat{\phi}^2}\right)\\
& + \frac{\eta^2 BB'}{\hat{\phi}} s't^3(p_1 + 2p_3)\\
& + \frac{\eta^2 BB'}{\hat{\phi}} st^2t'(3p_1 + 2p_2)\\
& + \frac{\eta^2 B^2}{\hat{\phi}} s't^2t'\left(\frac{1}{8} + p_2 - \frac{1}{2}p_4 + 3p_3\right)\\
& + \frac{\eta^2 B^2}{\hat{\phi}} stt'^2\left(\frac{1}{8} + 2p_2 - \frac{1}{4}p_4 - \frac{3}{4}p_5\right)\\
& + s't'^3\left(\frac{1}{2} + p_4 + p_5\right)\Bigg\} dz - \frac{p_5}{t'_o}\left[st'^3\right]_{z_o}^{z_i} - \frac{p_4}{t'_o}\left[s'tt'^2\right]_{z_o}^{z_i}
\end{aligned}
\tag{24.80}
$$

Eliminating the term in t'^2 from k gives

$$
k = \frac{1}{16}\int\left(\frac{\eta B''}{\hat{\phi}^{1/2}} + \frac{2\eta^3 B^3}{\hat{\phi}^{3/2}}\right)t^2 dz \tag{24.81}
$$

(The general form of k, set out in Hawkes (1980b), is of less interest.)

24.4.1 Thin-Lens Formulae

Some care is needed in deriving the thin-lens formula for K. Choosing the p_i in such a way that terms involving the discontinuous function t' vanish in the integrand, we apparently find that $K = 0$; the presence of this discontinuity at z_a means, however, that the integrated term in $[s'tt'^2]_{z_o}^{z_i}$ is not zero, even though $s'tt'^2$ vanishes at the endpoints. We now have

$$
\begin{aligned}
\frac{1}{t'_o}\left[s'tt'^2\right]_{z_o}^{z_i} &= \frac{1}{t'_o}\left[s'tt'^2\right]_{z_o}^{z_{a-}} + \frac{1}{t'_o}\left[s'tt'^2\right]_{z_{a+}}^{z_i}\\
&= -t'^2_o + t'^2_i\\
&= \frac{1}{z_o^2}\left(\frac{1}{M^2} - 1\right)
\end{aligned}
\tag{24.82}
$$

and with $p_4 = -\frac{1}{2}$

$$K \to \frac{1}{2}\frac{1}{z_o^2}\left(\frac{1}{M^2} - 1\right) \tag{24.83}$$

or if the aberrations are expressed in terms of position and gradient at the object plane

$$K \to \frac{1}{2}\left(\frac{1}{M^2} - 1\right) \tag{24.84}$$

This integrated term is neglected by Glaser (1952, eq. 124.30), who thus finds $K \to 0$.

For the anisotropic coefficient, we find

$$k \to \frac{\eta^3 B_0^3 S}{16\hat{\phi}^{3/2}}\tau_3 + \frac{\eta B_0 S}{8\hat{\phi}^{1/2}}\tau_1\left(t_o'^2 + t_i'^2\right) \tag{24.85}$$

in which

$$\begin{aligned} \tau_1 &\coloneqq \int_{-\infty}^{\infty} b(\zeta)d\zeta \\ \tau_3 &\coloneqq \int_{-\infty}^{\infty} b^3(\zeta)d\zeta \end{aligned} \tag{24.86}$$

(cf. 24.74). Here too, we do not agree with Glaser (1952, Eq.124.31); for further discussion, see Hawkes (1980b).

24.5 Astigmatism and Field Curvature (Terms Linear in x_a, y_a)

These aberrations are usually the least important of the third-order defects of electron lenses. The corresponding aberration figure is obtained as usual by considering rays from an object point that intersect the aperture plane around a circle of radius r_a. We have

$$\Delta u_i = (A + \mathrm{i}a)r_o^2 r_a \mathrm{e}^{\mathrm{i}(2\varphi_o - \varphi_a)} + Fr_0^2 r_a \mathrm{e}^{\mathrm{i}\varphi_a} \tag{24.87}$$

The term in F resembles a defocus term; returning to the paraxial approximation, $u = u_o s + u_a t$, we see that in a plane close to the image plane z_i, $z = z_i + \zeta$, say, we have $u(z_i + \zeta) \approx u(z_i) + \zeta u_a t_i'$. Hence $u(z_i + \zeta) - u(z_i) \approx \zeta u_a t_i'$. The term in F therefore resembles the blur caused by a defocus ζ, $\zeta \propto Fr_o^2$. After reading Chapter 31, Parasitic Aberrations, it will be apparent that the term in A and a is the third-order analogue of the paraxial astigmatism of imperfectly round lenses.

Eq. (24.87) is the parametric representation of a tilted ellipse, as we can see by suitably rotating the coordinate frame. With an angle of rotation χ, we have

$$\begin{aligned} \Delta u_i \mathrm{e}^{-\mathrm{i}\chi} &= (A + \mathrm{i}a)r_o^2 r_a \mathrm{e}^{\mathrm{i}(2\varphi_o - \varphi_a - \chi)} + Fr_o^2 r_a \mathrm{e}^{\mathrm{i}(\varphi_a - \chi)} \\ &= \sqrt{A^2 + a^2}\, r_o^2 r_a \mathrm{e}^{\mathrm{i}(2\varphi_o - \varphi_a + \varphi_A - \chi)} + Fr_o^2 r_a \mathrm{e}^{\mathrm{i}(\varphi_a - \chi)} \end{aligned} \tag{24.88}$$

where $A + \mathrm{i}a =: \sqrt{A^2 + a^2}\,\exp(\mathrm{i}\varphi_A)$. On choosing χ so that the exponents are equal and opposite, for which

$$\chi = \varphi_o + \frac{1}{2}\varphi_A \tag{24.89}$$

we have

$$\begin{aligned}\Delta\bar{x}_i + \mathrm{i}\Delta\bar{y}_i &\coloneqq \Delta\bar{u}_i \coloneqq \Delta u_i \mathrm{e}^{-\mathrm{i}\chi} \\ &= F r_o^2 r_a \mathrm{e}^{\mathrm{i}(\varphi_a - \varphi_o - \varphi_A/2)} + \sqrt{A^2 + a^2}\, r_o^2 r_a \mathrm{e}^{-\mathrm{i}(\varphi_a - \varphi_o - \varphi_A/2)} \\ &= \left(F + \sqrt{A^2 + a^2}\right) r_o^2 r_a \cos\left(\varphi_a - \varphi_o - \varphi_A/2\right) \\ &\quad + \mathrm{i}\left(F - \sqrt{A^2 + a^2}\right) r_o^2 r_a \sin\left(\varphi_a - \varphi_o - \varphi_A/2\right)\end{aligned} \tag{24.90}$$

Hence

$$\left(\frac{\Delta\bar{x}_\mathrm{i}}{a_1}\right)^2 + \left(\frac{\Delta\bar{y}_\mathrm{i}}{a_2}\right)^2 = 1, \quad a_{1,2} = r_o^2 r_a \left(F \pm \sqrt{A^2 + a^2}\right) \tag{24.91}$$

For a fixed object point, therefore, the beam that fills the aperture $\{x_a, y_a | x_a^2 + y_a^2 \leq r_A^2\}$ occupies an ellipse in the image plane, centred on the Gaussian image point, with semi-axes a_1 and a_2. The major and minor axes are inclined to the line joining the origin to the image point. Some typical aberration figures are shown in Fig. 24.8.

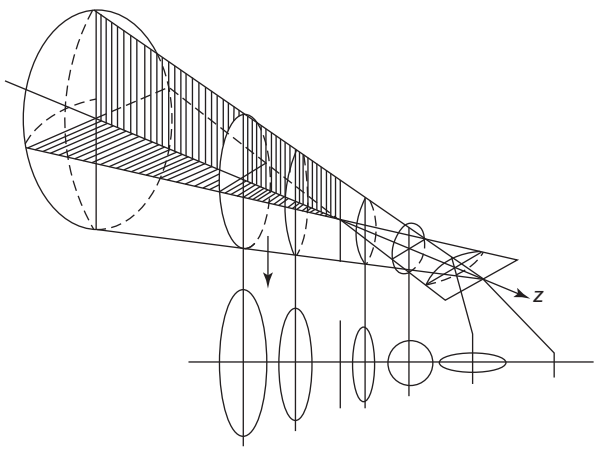


Figure 24.8
Astigmatism. Cross-section of an initially circular beam at various planes in the neighbourhood of the Gaussian image plane. To simplify the drawing, the ellipses are shown vertical or horizontal relative to the principal ray.

The aberration characterized by F is known as the *field curvature* while the coefficients A and a describe the *isotropic* and *anisotropic astigmatism*. The last vanishes in the absence of any magnetic field. The third-order astigmatism is often known as the Seidel astigmatism (as in light optics), to distinguish it from the parasitic paraxial astigmatism, which we shall meet in Chapter 31.

The field curvature is so called from the observation that it vanishes if the image is studied not in the Gaussian image plane but on a curved surface tangent to the latter. In order to see why this is so, we now enquire how the foregoing reasoning would be modified if the object and image points lay not on planes but on curved surfaces, tangent to the object and image planes at the axis. We examine only the case of spherically curved surfaces, S_o and S_i in Fig. 24.9.

We consider a ray from the object point $P_o(x_o, y_o)$, intersecting S_o at $\overline{P}_o\left(\overline{x}_o, \overline{y}_o\right)$, S_i at $\overline{P}_i\left(\overline{x}_i, \overline{y}_i\right)$ and the Gaussian plane at $P_i(x_\mathrm{i}, y_i)$. The distances ζ_o and ζ_i are the sagittae of the chords through $\overline{P}_o$ and $\overline{P}_\mathrm{i}$ respectively, normal to the optic axes, and are hence given by

$$\begin{aligned}\zeta_o &= \frac{\overline{x}_o^2 + \overline{y}_o^2}{2\rho_o} \approx \frac{x_o^2 + y_o^2}{2\rho_o} = \frac{r_o^2}{2\rho_o}\\ \zeta_\mathrm{i} &= -\frac{\overline{x}_\mathrm{i}^2 + \overline{y}_\mathrm{i}^2}{2\rho_\mathrm{i}} \approx -\frac{x_\mathrm{i}^2 + y_\mathrm{i}^2}{2\rho_\mathrm{i}} = -\frac{M^2}{2\rho_\mathrm{i}}\left(x_o^2 + y_o^2\right) = -\frac{M^2 r_o^2}{2\rho_\mathrm{i}}\end{aligned} \tag{24.92}$$

The sign convention is such that ρ_o and ρ_i are regarded as positive if $z(C_o) > z_o$ and $z(C_i) < z_i$, as is the case in Fig. 24.9; ζ_o clearly has the same sign as ρ_o and ζ_i the opposite sign to ρ_i.

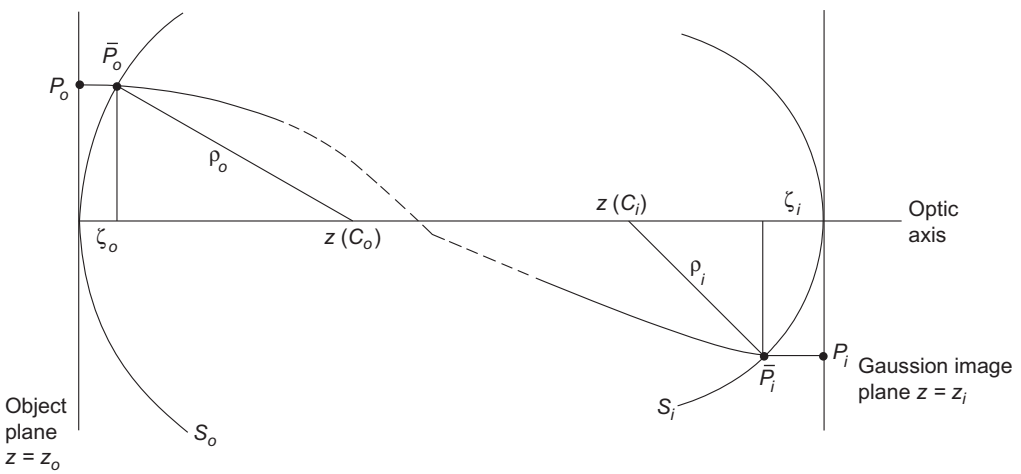


Figure 24.9
Field curvature. Notation used when the object and image surfaces may be curved.

In the Gaussian approximation, we have

$$\begin{aligned}\overline{x}_o &= x_o s(z_o + \zeta) + x_a t(z_o + \zeta)\\ &= x_o + \frac{s'_o}{2\rho_o} x_o r_o^2 + \frac{t'_o}{2\rho_o} x_a r_o^2\\ \overline{y}_o &= y_o + \frac{s'_o}{2\rho_o} y_o r_o^2 + \frac{t'_o}{2\rho_o} y_a r_o^2\end{aligned} \tag{24.93}$$

and similarly

$$\begin{aligned}\overline{x}_i &= x_o s\left(z_i + \zeta_i\right) + x_a t\left(z_i + \zeta_i\right)\\ &\approx x_i + x_o s'_i \zeta_i + x_a t'_i \zeta_i\\ &= x_i - \frac{M^2 s'_i}{2\rho_i} x_o r_o^2 - \frac{M^2 t'_i}{2\rho_i} x_a r_o^2\\ \overline{y}_i &= y_i - \frac{M^2 s'_i}{2\rho_i} y_o r_o^2 - \frac{M^2 t'_i}{2\rho_i} y_a r_o^2\end{aligned} \tag{24.94}$$

Thus

$$\begin{aligned}\Delta\overline{x}_i &:= \frac{\overline{x}_i}{M} - \overline{x}_o = \frac{x_i}{M} - x_o - \left(\frac{M s'_i}{2\rho_i} + \frac{s'_o}{2\rho_o}\right) x_o r_o^2\\ &\quad - \left(\frac{M t'_i}{2\rho_i} + \frac{t'_o}{2\rho_o}\right) x_a r_o^2\\ \Delta\overline{y}_i &= \Delta y_i - \left(\frac{M s'_i}{2\rho_i} + \frac{s'_o}{2\rho_o}\right) y_o r_o^2 - \left(\frac{M t'_i}{2\rho_i} + \frac{t'_o}{2\rho_o}\right) y_a r_o^2\end{aligned} \tag{24.95}$$

Returning to Eq. (24.87), and considering first the terms in F only, we have

$$\Delta\overline{x}_{\mathrm{i}} = -\left(\frac{M s'_{\mathrm{i}}}{2\rho_{\mathrm{i}}} + \frac{s'_o}{2\rho_o}\right) x_o r_o^2 + \left(F - \frac{M t'_{\mathrm{i}}}{2\rho_{\mathrm{i}}} - \frac{t'_o}{2\rho_o}\right) x_a r_o^2 \tag{24.96}$$

with a similar expression for $\Delta\overline{y}_i$. The term in $x_o r_o^2$ is a distortion and will be dealt with in the next section. For a plane image surface, $\rho_i \to \infty$, the last term vanishes if ρ_o is chosen so that

$$\rho_o = \frac{t'_o}{2F} \tag{24.97}$$

If now we retain all the terms of Eq. (24.87), again postponing discussion of the term in $x_o r_o^2$, we obtain expressions for $\Delta\overline{x}_i$ and $\Delta\overline{y}_i$ identical with those given by Eq. (24.90)

except that $F - Mt_i'/2\rho_i - t_o'/2\rho_o$ replaces F. The axes of the ellipse corresponding to particular values of x_o, y_o and are

$$a_{1,2} = r_A r_o^2 \left\{ F - \frac{Mt_i'}{2\rho_i} - \frac{t_o'}{2\rho_o} \pm \left(A^2 + a^2\right)^{1/2} \right\} \tag{24.98}$$

If ρ_i takes the value $\rho_i^{(1)}$ such that

$$\frac{Mt_i'}{2\rho_i^{(1)}} = F - \frac{t_o'}{2\rho_o} + \left(A^2 + a^2\right)^{1/2} \tag{24.99}$$

the semi-axis a_1 is zero and the ellipse then collapses to a line of length $4(A^2 + a^2)^{1/2} r_o^2 r_A$; likewise, for $\rho_i = \rho_i^{(2)}$ such that

$$\frac{Mt_i'}{2\rho_i^{(2)}} = F - \frac{t_o'}{2\rho_o} - \left(A^2 + a^2\right)^{1/2} \tag{24.100}$$

a_2 is zero and the ellipse collapses to a line of the same length. These lines are known as the tangential or meridional and sagittal line foci and the corresponding values of ρ_i as the tangential and sagittal field curvatures. For an intermediate value $\overline{\rho}_i$ of ρ_i, the ellipse becomes a circle; here, $|a_1| = |a_2|$ or

$$\frac{Mt_i'}{2\overline{\rho}_i} = F - \frac{t_o'}{2\rho_o}$$

or

$$\frac{1}{\overline{\rho}_i} = \frac{1}{2}\left(\frac{1}{\rho_i^{(1)}} + \frac{1}{\rho_i^{(2)}}\right) \tag{24.101}$$

This is sometimes known as the mean field curvature and the corresponding circle is again called a 'circle of least confusion'; it is essential to mention that the confusion caused by A, a and F is intended.

We have not mentioned ρ_o in the foregoing reasoning since microscope objects and microprobe targets are usually (assumed to be) flat; it is, however, necessary to consider it in some devices, especially those involving electron emission from surfaces.

Finally, we note that to each ρ_i corresponds a value of ζ_i. Writing Eq. (24.92)

$$\zeta_i^{(1,2)} \coloneqq -\frac{M^2 r_o^2}{2\rho_i^{(1,2)}} \tag{24.102}$$

we have

$$\begin{aligned}\Delta\zeta &:= \zeta_{\mathrm{i}}^{(2)} - \zeta_{\mathrm{i}}^{(1)} = -\frac{M^2 r_o^2}{2}\left(\frac{1}{\rho_{\mathrm{i}}^{(2)}} - \frac{1}{\rho_{\mathrm{i}}^{(1)}}\right) \\ &= \frac{2Mr_o^2}{t_i'}\sqrt{A^2 + a^2}\end{aligned} \tag{24.103}$$

Similarly,

$$\overline{\zeta} := -\frac{M^2 r_o^2}{2\overline{\rho}_{\mathrm{i}}} = \frac{Mr_o^2}{t_{\mathrm{i}}'}\left(F - \frac{t_o'}{\rho_{\mathrm{i}}}\right) \tag{24.104}$$

which vanishes if $F = 0$ and $\rho_{\mathrm{i}} \to \infty$: the circle of least confusion due to astigmatism alone lies in the Gaussian image plane. The quantity $\Delta\zeta$ is known as the astigmatic difference. Fig. 24.10 illustrates the geometrical meanings of $\zeta_i^{(1,2)}$ and $\Delta\zeta$.

In terms of the distortion of the surfaces $S =$ constant, the astigmatism and field curvature have different effects. We have $-S^I/W = \frac{1}{2} r_a^2 r_o^2 (A \cos 2\varphi + a \sin 2\varphi)$ and $-S^I/W = \frac{1}{2} F r_a^2 r_o^2$. Thus field curvature, like spherical aberration, shifts the surface S uniformly away from the reference sphere S_o towards the image plane ($F > 0$) by an amount that now varies with the distance of the object point from the axis.

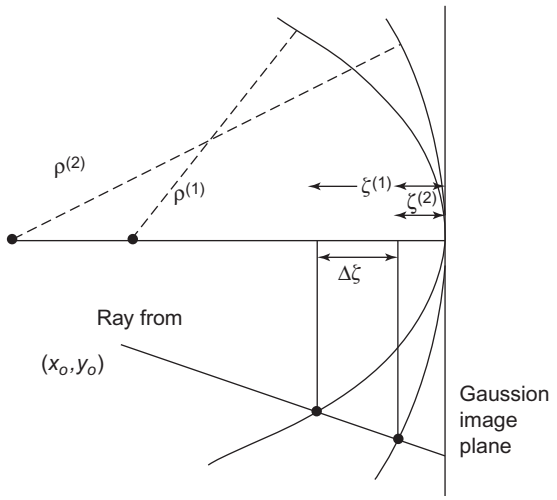


Figure 24.10
Notation used in connection with a curved image surface.

For the astigmatism, we write $-S^I/W = \frac{1}{2} r_a^2 r_o^2 \sqrt{A^2 + a^2} \cos(2\phi - \phi_A)$, $\tan \phi_A := a/A$ and we see that the distance from S_o to S is now described by a saddle-shaped surface (Fig. 24.11), S being shifted towards the image around $\varphi - \varphi_A/2 = 0, \pi$ and away from it in the neighbourhood of $\varphi - \varphi_A/2 = \pi/2, 3\pi/2$. The formation of line foci is readily understood.

The coefficients F, A and a are given by the following formulae:

$$F = \frac{1}{W_s} \int_{z_o}^{z_i} \left\{ 2L_1 s^2 t^2 + L_2 (st' + s't)^2 + 2L_3 s'^2 t'^2 + R(st' - s't)^2 \right\} dz \tag{24.105}$$

$$A = \frac{1}{W_s} \int_{z_o}^{z_i} \left\{ L_1 s^2 t^2 + 2L_2 ss' tt' + L_3 s'^2 t'^2 - R(st' - s't)^2 \right\} dz \tag{24.106}$$

$$a = 2 \int_{z_o}^{z_i} (Pst + Qs't') dz \tag{24.107}$$

Replacing s by g and t by h, we obtain the expressions in terms of position and gradient in the object plane; spherical aberration, coma and the expressions for F and A now all contribute.

The structures of F and A are very similar; we see that

$$\begin{aligned} F - 2A &= W_s \int_{z_o}^{z_i} \frac{L_2 + 3R}{\hat{\phi}} dz \\ &= W_s \int_{z_o}^{z_i} \frac{\gamma\phi'' + 4\eta B^2}{8\hat{\phi}^{3/2}} dz \end{aligned} \tag{24.108}$$

which is a function of $\phi(z)$ and $B(z)$ only and does not contain $s(z)$ or $t(z)$ except in the Wronskian. This quantity is known as the *Petzval coefficient* (Petzval, 1843); it can be

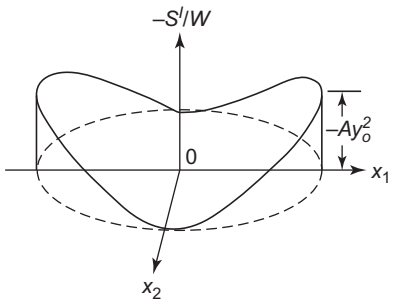


Figure 24.11
The effect of isotropic astigmatism on the wave surface.

shown that it is the only linear combination of the third-order aberration coefficients that possesses this property.

For *electrostatic lenses* the general relativistic expressions for F and A are

$$\begin{aligned} 32\hat{\phi}_o^{1/2}A = \frac{1}{t'_o}\int \Big\{ & A_0 s^2 t^2 + A_1 st(st)' + A_3 s't'(st)' \\ & + A_4 s'^2 t'^2 + \frac{1}{2}A_5\left(s^2 t'^2 + s'^2 t^2\right) + A_6 ss'tt' \Big\} dz \\ & - \frac{1}{2}p_1\left[\hat{\phi}^{1/2} ss't'^2\right] + \frac{p'_2 - p_2}{4}\left[\frac{\gamma\phi'}{\hat{\phi}^{1/2}} s^2 t'^2\right] \end{aligned} \tag{24.109}$$

$$\begin{aligned} 32\hat{\phi}_o^{1/2}F = \frac{2}{t'_o}\int \Big\{ & A_o s^2 t^2 + A_1 st(st)' + A_3 s't'(st)' \\ & + A_4 s'^2 t'^2 + \frac{1}{4}A_6(st' + s't)^2 + A_5 ss'tt' \Big\}\, dz \\ & - p_1\left[\hat{\phi}^{1/2} ss't'^2\right] - \frac{p'_2 + p_2}{4}\left[\frac{\gamma\phi'}{\hat{\phi}^{1/2}} s^2 t'^2\right] \end{aligned} \tag{24.110}$$

in which

$$\begin{aligned} A_5 \coloneqq & \frac{\gamma\phi''}{\hat{\phi}^{1/2}}\left\{p_3 + \frac{1}{2}\left(p_2 - p'_2\right) - \frac{1}{4}p_1\right\} \\ & + \frac{\phi'^2}{\hat{\phi}^{3/2}}\left\{p_4 - \frac{3}{4}\gamma^2\left(p_2 - p'_2\right)\right\} \\ & - \frac{\varepsilon\phi'^2}{\hat{\phi}^{1/2}}\left\{8p'_5 - \left(p_2 - p'_2\right)\right\} \end{aligned} \tag{24.111}$$

$$\begin{aligned} A_6 \coloneqq & \frac{\gamma\phi''}{\hat{\phi}^{1/2}}\left\{8 + 2p_3 + \frac{1}{2}\left(p_2 + p'_2\right) - \frac{1}{2}p_1\right\} \\ & + \frac{\phi'^2}{\hat{\phi}^{3/2}}\left\{2p_4 - \frac{3}{4}\gamma^2\left(p_2 + p'_2\right)\right\} \\ & - \frac{\varepsilon\phi'^2}{\hat{\phi}^{1/2}}\left\{16p'_5 - \left(p_2 + p'_2\right)\right\} \end{aligned} \tag{24.112}$$

For *magnetic lenses*, we have

$$
\begin{aligned}
A = \frac{1}{t'_o}\int_{z_o}^{z_i}\Bigg[& s^2t^2\left\{\frac{\eta^4B^4}{\hat{\phi}^2}\left(\frac{1}{32}-\frac{1}{4}p_2-\frac{1}{4}p_5\right)\right. \\
& \left.+\frac{\eta^2BB''}{\hat{\phi}}\left(p_1-\frac{1}{8}\right)+\frac{\eta^2B'^2}{\hat{\phi}}p_1\right\} \\
& +\frac{2\eta^2BB'}{\hat{\phi}}st\left\{s't(p_1+p_5)+st'(p_1+p_2)\right\} \\
& +\frac{\eta^2B^2}{\hat{\phi}}\left\{s'^2t^2\left(p_5-\frac{1}{4}p_3\right)+s^2t'^2\left(p_2-\frac{1}{4}p_4\right)\right. \\
& +2ss'tt'\left(\frac{1}{8}+p_5+p_2-\frac{1}{4}p_3-\frac{1}{4}p_4\right) \\
& \left.+s'^2t'^2\left(\frac{1}{2}+p_3+p_4\right)\right\}\Bigg]dz \\
& -\frac{\eta^2t'_o}{8\hat{\phi}}\int B^2dz \\
& -\frac{1}{t'_0}\left[p_4ss't'^2+p_3s'^2tt'\right]_{z_o}^{z_i}
\end{aligned}
\tag{24.113}
$$

$$
F = 2A + \frac{\eta^2t'_o}{2\hat{\phi}}\int B^2dz \tag{24.114}
$$

using Eq. (24.108), and

$$
a = \frac{1}{8}\int_{z_o}^{z_i}\left(2\frac{\eta^3B^3}{\hat{\phi}^{3/2}}+\frac{\eta B''}{\hat{\phi}^{1/2}}\right)st\,dz+\frac{1}{8}\left[\frac{\eta B}{\hat{\phi}^{1/2}}st'\right]_{z_o}^{z_i} \tag{24.115}
$$

Lenz (1986) has pointed out a curious feature of the Petzval coefficient. Eq. (24.114) shows that astigmatism and field curvature cannot both vanish. If, however, we evaluate $F - 2A$ in the fixed (not rotating) coordinate system, additional terms appear, showing that the 'fixed' Petzval coefficient can vanish provided that the object plane lies within the magnetic field. Lenz illustrates this by considering a uniform field, $B(z) = \text{const}$.

24.5.1 Thin-Lens Formulae

For magnetic lenses, the isotropic and anisotropic astigmatism both vanish,

$$
A \to 0, \quad a \to 0 \tag{24.116}
$$

and the field curvature is hence equal to the Petzval coefficient:

$$F \to \frac{\eta^2 t'_o}{2\hat{\phi}} \int B^2 dz = \frac{\eta^2 B_0^2 \tau_2}{2\hat{\phi}\zeta_o} \tag{24.117}$$

Glaser (1952, Eqs.124.29,31) also finds $a \to 0$ but his expression for A, and hence that for F, does not agree with ours. Der-Shvarts and Makarova (1973) find the same thin-lens expression for F but do not agree that $A \to 0$. (Note that these authors do not adopt the definitions of field curvature and astigmatism employed here.)

24.6 Distortion (Terms in x_o and y_o only)

These aberrations are of most importance in lenses in which the rays are comparatively far from the axis at the object plane. They are usually negligible in microscope objectives, therefore, but are the dominant geometrical defects of projectors. They can be kept acceptably small either by careful design of the column of a conventional instrument or by employing a very compact type of lens, which we describe in Part VII. As their name indicates, distortions do not blur the image but merely destroy the proportionality between image and object coordinates.

From Eq. (24.36), we see that

$$\begin{aligned} \Delta x_\mathrm{i} &= r_o^2 (D x_o - d y_o) \\ \Delta y_\mathrm{i} &= r_o^2 (D y_o + d x_o) \end{aligned} \tag{24.118}$$

and to avoid repetition, we modify these expressions to include the terms that arise if the object and image surfaces are curved (spherical), as in Section 24.5; this gives

$$\Delta \overline{x}_\mathrm{i} = \left\{ \left(D - \frac{M s'_\mathrm{i}}{2\rho_\mathrm{i}} - \frac{s'_o}{2\rho_o} \right) x_o - d y_o \right\} r_o^2 \tag{24.119a}$$

$$\Delta \overline{y}_\mathrm{i} = \left\{ \left(D - \frac{M s'_\mathrm{i}}{2\rho_\mathrm{i}} - \frac{s'_o}{2\rho_o} \right) y_o + d x_o \right\} r_o^2 \tag{24.119b}$$

or

$$\Delta \overline{u}_\mathrm{i} = \left(\tilde{D} + \mathrm{i}d \right) r_o^2 \, \mathrm{e}^{\mathrm{i}\phi_o}$$

with $\tilde{D} = D - M s'_\mathrm{i}/2\rho_\mathrm{i} - s'_o/2\rho_o$ Writing

$$\Delta \overline{u}_\mathrm{i} =: \mathrm{e}^{\mathrm{i}\varphi_o} \left(\Delta r_o + \mathrm{i} r_o \Delta \varphi_o \right) \tag{24.120}$$

we see that

$$\Delta r_o = \tilde{D} r_o^3 \quad \Delta \varphi_o = d r_o^2 \tag{24.121}$$

so that $\tilde{D}$ shifts the image point radially relative to its Gaussian position whereas d shifts it azimuthally. Consider now a square grid in the object plane, $x_o = m\tilde{x}$ for all y_o and $y_o = n\tilde{y}$ for all x_o, m, $n = 0, 1, \ldots$ (Fig. 24.12A). For $d = 0$, the grid will be shrunk, as shown in Fig. 24.12B if $\tilde{D} < 0$; this is known as *barrel distortion.* If $\tilde{D} > 0$, the grid is distended as in Fig. 24.12C; we then speak of *pincushion distortion.*

The coefficient D is the *isotropic distortion coefficient*. If $D = 0$ but $d \neq 0$, the grid is warped as shown in Fig. 24.12D. The coefficient d is known as the *anisotropic distortion*

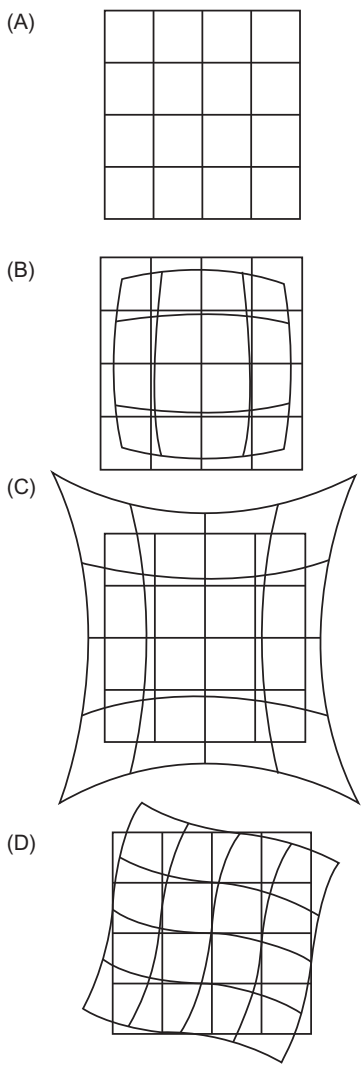


Figure 24.12
Distortion. (A) Square grid in the object plane. (B) Image with barrel distortion. (C) Image with pincushion distortion. (D) Image with anisotropic or 'spiral' distortion.

coefficient and vanishes in the absence of magnetic fields. It is often referred to as the *spiral distortion* or occasionally (Sturrock, 1955) as the *pocket-handkerchief distortion.*

In terms of the shape of the surfaces $S = \text{const}$, the distortion has a very simple effect. We have

$$-S'/W = r_a r_o^3 (D\cos\varphi + d\sin\varphi) = r_a r_o^3 \sqrt{D^2 + d^2}\cos(\varphi - \varphi_D)$$

with $\tan\varphi_D = d/D$ and the distance between S_o and S is therefore represented by a plane. The sphere S_o may thus be pictured as being shifted bodily so that the Gaussian image point is shifted by an amount that is constant for each object point but different ($\propto r_o^3$) for object points at varying distances from the axis. This is exactly the behaviour we have already described as characteristic of distortion. The coefficients D and d are given in general by

$$\begin{aligned} D &= \frac{1}{W_s}\int_{z_o}^{z_i}\left\{L_1 s^3 t + L_2 s s'(st)' + L_3 s'^3 t'\right\} dz \\ d &= \int_{z_o}^{z_i}\left(Ps^2 + Qs'^2\right) dz \end{aligned} \tag{24.122}$$

For *electrostatic lenses*, the general expression is

$$\begin{aligned} 32\hat{\phi}_o^{1/2} D = \frac{1}{t'_o}\int_{z_o}^{z_i}\Bigg\{ & A_0 s^3 t + \frac{1}{2}A_1 s\left(3s^2 t' + s'^2 t\right) + A_2 s s'(st + s't') \\ & + \frac{1}{2}A_3 s'\left(s^2 t' + 3s'^2 t\right) + A_4 s'^3 t'\Bigg\} dz \\ & - \Bigg[\frac{3}{4}p_1\hat{\phi}^{1/2} s s'^2 t' + \frac{1}{2}p_2\frac{\gamma\phi'}{\hat{\phi}^{1/2}} s^2 s' t' \\ & + \left(\frac{1}{4}p_3\frac{\gamma\phi''}{\hat{\phi}^{1/2}} + \frac{1}{4}p_4\frac{\phi'^2}{\hat{\phi}^{3/2}} - 2p'_5\frac{\varepsilon\phi'^2}{\hat{\phi}^{1/2}}\right) s^3 t'\Bigg]_{z_o}^{z_i} \end{aligned} \tag{24.123}$$

For *magnetic lenses*,

$$\begin{aligned} D = \frac{1}{t'_o}\int\Bigg\{ & s^3 t\left(\frac{\eta^4 B^4}{32\hat{\phi}^2} - \frac{\eta^2 B B''}{8\hat{\phi}} + \frac{p_1\eta^2 B'^2}{\hat{\phi}}\right. \\ & \left. + p_1\frac{\eta^2 B B''}{\hat{\phi}} - (p_2 + p_3)\frac{\eta^4 B^4}{4\hat{\phi}^2}\right) \\ & + \frac{\eta^2 B B'}{\hat{\phi}} s^3 t'(p_1 + 2p_3) \end{aligned}$$

$$
\begin{aligned}
&+ \frac{\eta^2 BB'}{\hat{\phi}} s^2 s' t (3p_1 + 2p_2) \\
&+ \frac{\eta^2 B^2}{\hat{\phi}} s^2 s' t' \left(\frac{1}{8} + p_2 - \frac{1}{2} p_4 + 3p_3 \right) \\
&+ \frac{\eta^2 B^2}{\hat{\phi}} s s'^2 t \left(\frac{1}{8} + 2p_2 - \frac{1}{4} p_4 - \frac{3}{4} p_5 \right) \\
&+ s'^3 t' \left(\frac{1}{2} + p_4 + p_5 \right) \Bigg\} dz \\
&- \frac{1}{t'_o} \left[p_5 s'^3 t + p_4 s s'^2 t' + p_3 \frac{\eta^2 B^2}{\hat{\phi}} s^3 t' \right]_{z_o}^{z_i}
\end{aligned} \tag{24.124}
$$

$$
\begin{aligned}
d = &\frac{1}{16 \hat{\phi}^{1/2}} \int_{z_o}^{z_i} \left(\eta B'' + \frac{2\eta^3 B^3}{\hat{\phi}} \right) s^2 \, dz \\
&+ \left[\frac{\eta B}{4 \hat{\phi}^{1/2}} s s' - \frac{\eta B'}{8 \hat{\phi}^{1/2}} s^2 \right]_{z_o}^{z_i}
\end{aligned} \tag{24.125}
$$

24.6.1 Thin-Lens Formulae

For magnetic lenses, we find

$$
\begin{aligned}
D &= 0 \\
d &= \frac{\eta B_o \tau_1}{4 \hat{\phi}^{1/2} S \zeta_o^2}
\end{aligned} \tag{24.126}
$$

Glaser (1952, 124.30) likewise finds that the isotropic distortion coefficient vanishes but, unlike us, he also finds $d = 0$ (123.31); the origin of this disagreement lies in his use of a partially integrated expression.

24.7 The Variation of the Aberration Coefficients with Aperture Position

The aberration coefficients clearly vary with the position of the object and aperture. In the case of the *real* aberration coefficients, with which we are concerned here, it is principally the aperture dependence that is important, since objective and probe-forming lenses are mostly operated with the specimen close to a focus. Furthermore, it is not possible to enunciate

general laws about the dependence of the aberration coefficients on *object* position when the latter lies within the field owing to the fact that z_o is the lower limit of integration in S^I_{oi}. Their dependence on the *aperture* position can, however, be studied as we now show, and it will transpire that some aberration coefficients can be made to vanish by choosing z_a suitably.

Each of the integrands appearing in the aberration coefficients consists of one or more terms of the form $f(z)s^p s'^q t^r t'^s$, $p+q=m$, $r+s=3-m$, in which, we recall, $s(z)$ and $t(z)$ correspond to some particular aperture position. For this choice, we have (24.36)

$$\Delta x_i = \begin{pmatrix} x_o \\ x_a \\ y_o \end{pmatrix}^T \begin{pmatrix} D & K & 2A & a \\ F-A & C & 2K & 2k \\ -d & -k & -a & 0 \end{pmatrix} \begin{pmatrix} r_o^2 \\ r_a^2 \\ V \\ v \end{pmatrix} \tag{24.127}$$

with a similar expression for Δy_i ($x \to y$, $y \to -x$). For another choice of aperture position, we should have

$$\Delta \bar{x}_i = \begin{pmatrix} x_o \\ \bar{x}_a \\ y_o \end{pmatrix}^T \begin{pmatrix} \bar{D} & \bar{K} & 2\bar{A} & \bar{a} \\ \bar{F}-\bar{A} & \bar{C} & 2\bar{K} & 2\bar{k} \\ -\bar{d} & -\bar{k} & -\bar{a} & 0 \end{pmatrix} \begin{pmatrix} r_o^2 \\ \bar{r}_a^2 \\ \bar{V} \\ \bar{v} \end{pmatrix} \tag{24.128}$$

with

$$\begin{aligned} \bar{r}_a^2 &:= \bar{x}_a^2 + \bar{y}_a^2 \\ \bar{V} &:= x_o \bar{x}_a + y_o \bar{y}_a \\ \bar{v} &:= x_o \bar{y}_a - \bar{x}_a y_o \end{aligned} \tag{24.129}$$

In Eq. (24.128), the coefficients must have the same overall structure as those of Δx_i but contain $\bar{s}$ and $\bar{t}$ satisfying $\bar{s}(z_o) = \bar{t}(\bar{z}_a) = 1, \bar{s}(\bar{z}_a) = \bar{t}(z_o) = 0$. Setting

$$\begin{aligned} \bar{s}(z) &= s(z) + \sigma t(z) \\ \bar{t}(z) &= \tau t(z) \end{aligned} \tag{24.130}$$

we see that

$$\begin{aligned} \sigma &= -\frac{s(\bar{z}_a)}{t(\bar{z}_a)} \\ \tau &= \frac{1}{t(\bar{z}_a)} \end{aligned} \tag{24.131}$$

and

$$\begin{aligned} \bar{s}\bar{t}' - \bar{s}'\bar{t} &= \tau(st' - s't) \\ W_{\bar{s}} &= \tau W_s \end{aligned} \tag{24.132}$$

From the expressions for A, C, …, k, it is a simple matter to show that

$$\begin{pmatrix} \overline{C} \\ \overline{K} \\ \overline{A} \\ \overline{F} \\ \overline{D} \\ \overline{k} \\ \overline{a} \\ \overline{d} \end{pmatrix} = \begin{pmatrix} \tau^3 & 0 & 0 & 0 & 0 & & & \\ \tau^2\sigma & \tau^2 & 0 & 0 & 0 & & & \\ \tau\sigma^2 & 2\tau\sigma & \tau & 0 & 0 & & 0 & \\ 2\tau\sigma^2 & 4\tau\sigma & 0 & \tau & 0 & & & \\ \sigma^3 & 3\sigma^2 & 2\sigma & \sigma & 1 & & & \\ & & & & & \tau^2 & 0 & 0 \\ & & 0 & & & 2\sigma\tau & \tau & 0 \\ & & & & & \sigma^2 & \sigma & 1 \end{pmatrix} \begin{pmatrix} C \\ K \\ A \\ F \\ D \\ k \\ a \\ d \end{pmatrix} \tag{24.133}$$

A number of general conclusions about the various aberrations can be drawn from Eq. (24.133). Disappointingly, there is no privileged aperture position so far as the spherical aberration is concerned: Eq. (24.133) merely confirms the obvious result that since the aberration is expressed in terms of aperture coordinates, the closer the aperture is to the image, the smaller will the coefficient be — but x_a and y_a will be large!

The expression for $\overline{K}$ shows that if σ can be chosen so that

$$K + \sigma C = 0 \tag{24.134}$$

the coma will vanish. The existence of a coma-free point has long been known in optics: see Herzberger (1931, 1958) for extensive discussion and Czapski and Eppenstein (1924) for the earlier history.

The isotropic astigmatism vanishes if

$$A + 2\sigma K + \sigma^2 C = 0 \tag{24.135}$$

There will be two astigmatism-free aperture positions if $K^2 > AC$, one if it so happens that $K^2 = AC$ (the Finsterwalder condition, see Finsterwalder 1892), and otherwise none. In the case $K^2 = AC$, we have $\sigma = -K/C$, so that the coma too vanishes. Similar reasoning can be applied, not very profitably, to the field curvature and distortion.

Turning to the anisotropic aberrations, we see that the coma (k) cannot be eliminated by choosing the aperture position suitably; the astigmatism vanishes if $\sigma = -a/2k$, and the distortion if $\sigma = \{-a \pm (a^2 - 4dk)\}/2k$. If $a^2 = 4dk$, astigmatism and coma vanish simultaneously for $\sigma = -a/2k$.

Of these results, only the existence of a coma-free aperture position has found any practical application, notably in aberration corrector design (Chapter 41 of Volume 2).

We perceive that relations identical with Eq. (24.133) are obtained if, instead of replacing $\overline{s}$ and $\overline{t}$ by s and t, as we have done, we write $\overline{x}_a \coloneqq x(\overline{z}_a) = x_o s(\overline{z}_a) + x_a t(\overline{z}_a)$ and use this to replace x_a by $\overline{x}_a$ and likewise y_a by $\overline{y}_a$ in the expressions for Δx_i and Δy_i. Although this

leads to the same result with less effort, it is not immediately obvious that the boundary conditions, $\Delta x = \Delta y = 0$ in the object and aperture planes, are satisfied. In fact, both techniques are legitimate, as the agreement between the results confirms.

24.8 Reduced Coordinates[4]

For reference purposes, we reproduce the expression for each aberration coefficient in terms of reduced coordinates (Section 15.3.1), as derived by Sturrock (1951c,d, 1955). Note that not only is $v(z)$ related to $u(z)$ by Eq. (15.40) but that the paraxial solutions satisfy the reduced equation (15.38). This has been a pitfall in the past (Glaser, 1952, p.676; Kuyatt, 1978).

The forms of the coefficients A, C, ... k listed in Sturrock (1955) are as follows; these are a generalization of the earlier expressions, for electrostatic and magnetic lenses separately, to be found in Sturrock (1951c). The functions $\sigma(z)$ and $\tau(z)$ are the solutions of the reduced paraxial equations satisfying the boundary conditions $\sigma(z_o) = \tau(z_a) = 1$, $\sigma(z_a) = \tau(z_o) = 0$, and W_σ is the corresponding Wronskian, $W_\sigma = \tau'(z_o) = -\sigma'(z_a)$.

$$A + \mathrm{i}a = \frac{1}{16}\left[\left(\frac{\hat{\phi}_o\hat{\phi}_a}{\hat{\phi}^2}\right)^{1/4}\left(4\sigma'\tau' - \gamma W_\sigma\frac{\phi'}{\hat{\phi}} + 2\mathrm{i}W_\sigma\frac{\eta B}{\hat{\phi}^{1/2}}\right)\right]_o^i$$

$$- \frac{4}{W_\sigma}\int_o^i\left(\frac{\hat{\phi}_o\hat{\phi}_a}{\hat{\phi}^2}\right)^{1/4}\Big\{\sigma\tau(P\sigma'\tau' + Q\sigma\tau) - W_\sigma^2 T$$

$$+ \mathrm{i}W_\sigma(R\sigma'\tau' + S\sigma\tau)\Big\}dz$$

$$C = -\frac{4}{W_\sigma}\int\left(\frac{\hat{\phi}_a^3}{\hat{\phi}_o\hat{\phi}^2}\right)^{1/4}\tau^2\left(P\tau'^2 + Q\tau^2\right)dz$$

$$D + \mathrm{i}d = -\left[\left(\frac{\hat{\phi}_o}{\hat{\phi}}\right)^{1/2}\left(-\frac{3}{8}\sigma'^2 + \frac{\gamma}{16}\frac{\phi'}{\hat{\phi}}\sigma\sigma'\right.\right.$$

$$\left.\left. - \frac{1}{16}\left(\frac{\gamma^2}{4}\frac{\phi'^2}{\hat{\phi}^2} + \frac{\gamma}{2}\frac{\phi''}{\hat{\phi}} + \frac{\eta^2B^2}{\hat{\phi}}\right)\sigma^2\right.\right.$$

[4] For those unfamiliar with Gaussian units, a c.g.s. system employing both electromagnetic units (e.m.u.) and electrostatic units (e.s.u.), we note that Sturrock's units may be converted to SI by writing Φ[Sturrock] $\rightarrow$ $2\varepsilon\phi$[SI]; p[St] $\rightarrow$ $2\sqrt{\varepsilon\phi(1+\varepsilon\phi)} = 2\left(\varepsilon\hat{\phi}\right)^{1/2}$[SI]; H[St] $\rightarrow$ $(1/100)(e/m_0c)B$[SI]. Note that when converting Sturrock's formulae, centimetres must be replaced by metres, particularly in derivatives. The presence of the factor (1/100) above is an instance of this.

$$
\begin{aligned}
&\quad - \frac{\mathrm{i}}{8}\frac{\eta B}{\hat{\phi}^{1/2}}\sigma\sigma' + \frac{\mathrm{i}}{16}\frac{\eta B'}{\hat{\phi}^{1/2}}\sigma^2\Bigg]_o^i \\
&\quad - \frac{2}{W_\sigma}\int_o^i \left(\frac{\hat{\phi}_o}{\hat{\phi}}\right)^{1/2}\Big\{2(P\sigma'\tau' + Q\sigma\tau)\sigma^2 - W_\sigma P\sigma\sigma' \\
&\quad + \mathrm{i}W_\sigma(R\sigma'^2 + S\sigma^2)\Big\}\, dz \\
F &= \frac{1}{2}\left[\left(\frac{\hat{\phi}_o\hat{\phi}_a}{\hat{\phi}^2}\right)^{1/4}\sigma' t'\right]_o^{\mathrm{i}} \\
&\quad - \frac{8}{W_\sigma}\int_o^{\mathrm{i}}\left(\frac{\hat{\phi}_o\hat{\phi}_a}{\hat{\phi}^2}\right)^{1/4}\sigma\tau(P\sigma'\tau' + Q\sigma\tau)\, dz \\
&\quad - 2W_\sigma\int_o^{\mathrm{i}}\left(\frac{\hat{\phi}_o\hat{\phi}_a}{\hat{\phi}^2}\right)^{1/4}(P + 2T)\, dz \qquad (24.136) \\
K + \mathrm{i}k &= \frac{1}{8}\left[\left(\frac{\hat{\phi}_a}{\hat{\phi}}\right)^{1/2}\tau'^2\right]_o^{\mathrm{i}} \\
&\quad - \frac{2}{W_\sigma}\int_o^{\mathrm{i}}\left(\frac{\hat{\phi}_a}{\hat{\phi}}\right)^{1/2}\left\{2\left(P\tau'^2 + Q\tau^2\right)\sigma\tau - W_\sigma P\tau\tau'\right\} dz \\
&\quad - 2i\int_o^{\mathrm{i}}\left(\frac{\hat{\phi}_a}{\hat{\phi}}\right)^{1/2}\left(R\tau'^2 + S\tau^2\right) dz
\end{aligned}
$$

in which

$$
\begin{aligned}
P(z) &:= \frac{3}{128}\left(\frac{\phi'}{\hat{\phi}}\right)^2\left(1 + \frac{4}{3}\varepsilon\hat{\phi}\right) + \frac{\eta^2 B^2}{32\hat{\phi}} \\
Q(z) &:= -\frac{59}{1024}\left(1 + \frac{234}{59}\varepsilon\hat{\phi} + \frac{120}{59}\varepsilon^2\hat{\phi}^2\right)\left(\frac{\phi'}{\hat{\phi}}\right)^4 \\
&\quad + \frac{65}{1024}\left(1 + \frac{12}{13}\varepsilon\hat{\phi}\right)\frac{\gamma\phi'^2\varphi''}{\hat{\phi}^3}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{5}{256}\left(1+\frac{8}{5}\varepsilon\hat{\phi}\right)\left(\frac{\phi''}{\hat{\phi}}\right)^2 \\
&- \frac{27}{512}\left(1+\frac{8}{3}\varepsilon\hat{\phi}\right)\frac{\phi'^2\eta^2B^2}{\hat{\phi}^3} \\
&- \frac{3\gamma}{256}\frac{\eta^2B^2\phi''}{\hat{\phi}^2} + \frac{9\gamma}{128}\frac{\eta^2BB'\phi'}{\hat{\phi}^2} \\
&- \frac{3}{128}\frac{\eta^4B^4}{\hat{\phi}^2} - \frac{1}{32}\frac{\eta^2B'^2}{\hat{\phi}} \\
R(z) &:= -\frac{\eta B}{16\hat{\phi}^{1/2}} \\
S(z) &:= -\frac{5}{256}\left(1+\frac{12\varepsilon\hat{\phi}}{5}\right)\frac{\phi'^2\eta B}{\hat{\phi}^{5/2}} \\
&- \frac{\gamma}{32}\frac{\phi''\eta B}{\hat{\phi}^{3/2}} + \frac{\gamma}{32}\frac{\phi'\eta B'}{\hat{\phi}^{3/2}} - \frac{3}{64}\frac{\eta^3B^3}{\hat{\phi}^{3/2}} \\
T(z) &:= -\frac{5}{256}\left(1+\frac{12}{5}\varepsilon\hat{\phi}\right)\left(\frac{\phi'}{\hat{\phi}}\right)^2 - \frac{3}{64}\frac{\eta^2B^2}{\hat{\phi}}
\end{aligned}
\tag{24.137}
$$

Groves (2015) revives the use of such units (which he refers to as 'natural units').

24.9 Seman's Transformation of the Characteristic Function

We have seen how useful is the technique introduced by Seman for performing partial integration on aberration coefficients in a systematic fashion. Originally, however, it was not the individual coefficients but the perturbation characteristic S^I to which the technique was applied, thus generating a set of coefficients all possessing similar characteristics, such as absence of high order derivatives of $\phi(z)$ or $B(z)$. Seman considered electrostatic and magnetic round lenses. The idea behind his method is to be found in two early notes (Seman, 1951, 1954) and a full account appeared shortly afterwards (Seman, 1955a–c, 1958b). We illustrate the power and simplicity of the technique in the case of orthogonal systems consisting of suitably orientated magnetic and electrostatic quadrupoles and round electrostatic lenses. Only the main steps in the reasoning are presented here; for additional details, see Hawkes (1966/7b). This discussion is limited to the nonrelativistic approximation but the relativistic case has been explored in detail for round electrostatic lenses (see Hawkes, 1977a).

It is convenient to introduce the vectors $\boldsymbol{x} \coloneqq (x, y)$, $\tilde{\mathbf{x}} \coloneqq (x, -y)$, whereupon $M^{(2)}$ and $M^{(4)}$ may be written

$$\begin{aligned}
M^{(2)} &= -\frac{\phi''}{8\phi^{1/2}}\boldsymbol{x}^2 + \frac{1}{4}Q(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) + \frac{1}{2}\phi^{1/2}\boldsymbol{x}'^2 \\
M^{(4)} &= \left(\frac{\phi^{(4)}}{128\phi^{1/2}} - \frac{\phi''^2}{128\phi^{3/2}} - \frac{1}{2}O\right)\boldsymbol{x}^4 \\
&\quad + \left(-\frac{p_2^2}{32\phi^{3/2}} + O\right)(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})^2 \\
&\quad + \left(\frac{p_2\phi''}{32\phi^{3/2}} - \frac{p_2''}{48\phi^{1/2}} + \frac{\eta Q_2''}{48}\right)\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) \\
&\quad - \frac{\phi''}{16\phi^{1/2}}\boldsymbol{x}^2\boldsymbol{x}'^2 + \frac{p_2}{8\phi^{1/2}}(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})\boldsymbol{x}'^2 - \frac{\phi^{1/2}}{8}\boldsymbol{x}'^4 \\
&\quad + \frac{\eta Q_2'}{16}\left\{(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})\boldsymbol{x}^{2'} - \boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})'\right\}
\end{aligned} \tag{24.138}$$

in which

$$\begin{aligned}
Q &\coloneqq \frac{p_2}{\phi^{1/2}} - 2\eta Q_2 \\
O &\coloneqq \frac{1}{24}\left(\frac{p_4}{\phi^{1/2}} - 2\eta Q_4\right)
\end{aligned} \tag{24.139}$$

and $\boldsymbol{x}^2$ and $\boldsymbol{x}^4$ denote $(\boldsymbol{x}\cdot\boldsymbol{x})$ and $(\boldsymbol{x}\cdot\boldsymbol{x})^2$ respectively.

The individual terms of $M^{(4)}$ fall into groups, which have the following 'dimensions':

$$\begin{aligned}
&\left[\phi^{1/2}\right]\left[\boldsymbol{x}^2\right]^2\left[d/dz\right]^4 \\
&\left[p_4/\phi^{1/2} \quad \text{or} \quad Q_4\right]\left[\boldsymbol{x}^2\right]^2 \quad \text{and} \quad \left[p_4/\phi^{1/2} \quad \text{or} \quad Q_4\right][\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}]^2 \\
&\left[p_2/\phi^{1/2} \quad \text{or} \quad Q_2\right]\left[\boldsymbol{x}^2\right][\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}]\left[d/dz\right]^2 \\
&\left[\phi^{-1/2}\right]\left[p_2/\phi^{1/2} \quad \text{or} \quad Q_2\right][\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}]^2
\end{aligned} \tag{24.140}$$

Since $S^I = \int M^{(4)}\,dz$, the terms of S^I are generated by quantities with one lower power of d/dz than Eq. (24.140); the only terms in which d/dz survives are therefore

$$\begin{aligned}
&[\phi^{1/2}]\left[\boldsymbol{x}^2\right]^2\left[d/dz\right]^3 \\
&\left[p_2/\phi^{1/2} \quad \text{or} \quad Q_2\right][\boldsymbol{x}^2][\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}]\left[d/dz\right]
\end{aligned} \tag{24.141}$$

These generate 16 terms possessing the appropriate dimensions, eight involving ϕ alone ($r_1 - r_8$), three each involving p_2 alone ($e_1 - e_3$) and Q_2 alone ($m_1 - m_3$) and two mixed terms (r_9 and r_{10}):

$$
\begin{aligned}
r_1 &= \phi^{1/2}(\boldsymbol{x}^2)'\boldsymbol{x}'^2 & r_6 &= \frac{\phi'''}{\phi^{1/2}}\boldsymbol{x}^4 \\
r_2 &= \frac{\phi'}{\phi^{1/2}}\left\{(\boldsymbol{x}^2)'\right\}^2 & r_7 &= \frac{\phi''\phi'}{\phi^{3/2}}\boldsymbol{x}^4 \\
r_3 &= \frac{\phi'}{\phi^{1/2}}\boldsymbol{x}^2\boldsymbol{x}'^2 & r_8 &= \frac{\phi'^3}{\phi^{5/2}}\boldsymbol{x}^4 \\
r_4 &= \frac{\phi''}{\phi^{1/2}}\boldsymbol{x}^2(\boldsymbol{x}^2)' & r_9 &= \frac{p_2\phi'}{\phi^{3/2}}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) \\
r_5 &= \frac{\phi'^2}{\phi^{3/2}}\boldsymbol{x}^2(\boldsymbol{x}^2)' & r_{10} &= \frac{\eta Q_2\phi'}{\phi}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) \\
e_1 &= \frac{p_2}{\phi^{1/2}}(\boldsymbol{x}^2)'(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & m_1 &= \eta Q_2(\boldsymbol{x}^2)'(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) \\
e_2 &= \frac{p_2}{\phi^{1/2}}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})' & m_2 &= \eta Q_2\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})' \\
e_3 &= \frac{p_2'}{\phi^{1/2}}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & m_3 &= \eta Q'\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})
\end{aligned}
\tag{24.142}
$$

The terms r_1, r_1, ... r_8 are identical with eight of Seman's terms i_j, those not containing the magnetic field.

The refractive index can be altered in form without affecting the eikonal by adding to it expressions that vanish; such expressions are obtained by differentiation of each of the quantities m_j, e_j and r_j and elimination of x'' with the aid of the paraxial equations. We find

$$r_1' + \frac{1}{2}R_{17} + \frac{1}{4}R_{22} + R_{23} - 2R_{24} - E_5 - \frac{1}{2}E_6 + 2M_5 + M_6 = 0$$

$$r_2' + R_{13} - 2R_{15} + 4R_{19} + \frac{3}{2}R_{21} - R_{22} - 4R_{23} = 0$$

$$r_3' + \frac{1}{4}R_{13} - \frac{1}{2}R_{16} - R_{17} + \frac{3}{2}R_{18} + R_{20} - R_{23} = 0$$

$$r_4' + \frac{1}{2}R_3 - R_6 + 2R_9 + R_{13} - R_{14} - 2R_{17} - R_{22} = 0$$

$$r_5' + \frac{1}{2}R_2 - R_7 + 2R_{10} + 2R_{12} - 2R_{13} - 2R_{18} - R_{21} = 0$$

$$r_6' + \frac{1}{2}R_4 - R_5 - 2R_{14} = 0$$

$$r_7' + \frac{3}{2}R_2 - R_3 - R_4 - 2R_{13} = 0$$

$$r_8' + \frac{5}{2}R_1 - 3R_2 - 2R_{12} = 0$$

$$\begin{aligned}
&r_9' - 2R_6 + 3R_7 - 2R_8 - 2R_{15} - 2R_{16} = 0\\
&r_{10}' - R_9 + R_{10} - R_{11} - R_{19} - R_{20} = 0\\
&e_1' + R_6 + 2R_{15} - 2E_1 - 2E_3 - 4E_5 - 2E_6 + 4E_8 = 0\\
&e_2' + R_6 + 2R_{16} - 2E_2 - 2E_4 + 4E_5 - 4E_6 + 4M_8 = 0\\
&e_3' + R_8 - 2E_1 - 2E_2 - 2E_7 = 0\\
&m_1' + \frac{1}{2}R_9 + \frac{1}{2}R_{19} - E_8 - M_1 + 2M_3 - 2M_5 - M_6 = 0\\
&m_2' + \frac{1}{2}r_9 + \frac{1}{2}R_{20} - M_2 + 2M_4 + 2M_5 - 2M_6 - M_8 = 0\\
&m_3' - M_1 - M_2 - M_7 = 0
\end{aligned} \tag{24.143}$$

in which

$$R_1 = \frac{\phi'^4}{\phi^{7/2}}\boldsymbol{x}^4 \qquad R_{13} = \frac{\phi''\phi'}{\phi^{3/2}}\boldsymbol{x}^2(\boldsymbol{x}^2)'$$

$$R_2 = \frac{\phi'^2\phi''}{\phi^{5/2}}\boldsymbol{x}^4 \qquad R_{14} = \frac{\phi'''}{\phi^{1/2}}\boldsymbol{x}^2(\boldsymbol{x}^2)'$$

$$R_3 = \frac{\phi''^2}{\phi^{3/2}}\boldsymbol{x}^4 \qquad R_{15} = \frac{p_2\phi'}{\phi^{3/2}}\boldsymbol{x}^{2'}(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})$$

$$R_4 = \frac{\phi'''\phi'}{\phi^{3/2}}\boldsymbol{x}^4 \qquad R_{16} = \frac{p_2\phi'}{\phi^{3/2}}(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})'$$

$$R_5 = \frac{\phi^{(4)}}{\phi^{1/2}}\boldsymbol{x}^4 \qquad R_{17} = \frac{\phi''}{\phi^{1/2}}\boldsymbol{x}^2\boldsymbol{x}'^2$$

$$R_6 = \frac{p_2\phi''}{\phi^{3/2}}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) \qquad R_{18} = \frac{\phi'^2}{\phi^{3/2}}\boldsymbol{x}^2\boldsymbol{x}'^2$$

$$R_7 = \frac{p_2\phi'^2}{\phi^{5/2}}x^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) \qquad R_{19} = \eta\frac{Q_2\phi'}{\phi}(\boldsymbol{x}^2)'\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})$$

$$R_8 = \frac{p_2'\phi'}{\phi^{3/2}}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) \qquad R_{20} = \eta\frac{Q_2\phi'}{\phi}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})'$$

$$
\begin{aligned}
R_9 &= \eta\frac{Q_2\phi''}{\phi}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & R_{21} &= \frac{\phi'^2}{\phi^{3/2}}\{(\boldsymbol{x}^2)'\}^2\\
R_{10} &= \eta\frac{Q_2\phi'^2}{\phi^2}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & R_{22} &= \frac{\phi''}{\phi^{1/2}}\{(\boldsymbol{x}^2)'\}^2\\
R_{11} &= \eta\frac{Q_2'\phi'}{\phi}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & R_{23} &= \frac{\phi'}{\phi^{1/2}}\boldsymbol{x}'^2(\boldsymbol{x}^2)'\\
R_{12} &= \frac{\phi'^3}{\phi^{5/2}}\boldsymbol{x}^2(\boldsymbol{x}^2)' & R_{24} &= \phi^{1/2}\boldsymbol{x}'^4
\end{aligned}
$$

$$
\begin{aligned}
E_1 &= \frac{p_2'}{\phi^{1/2}}(\boldsymbol{x}^2)'(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & M_1 &= \eta Q_2'(\boldsymbol{x}^2)'(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})\\
E_2 &= \frac{p_2'}{\phi^{1/2}}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})' & M_2 &= \eta Q_2'\mathbf{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})'\\
E_3 &= \frac{p_2^2}{\phi^{3/2}}(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})^2 & M_3 &= \eta^2\frac{Q_2^2}{\phi^{1/2}}(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})^2\\
E_4 &= \frac{p_2^2}{\phi^{3/2}}\boldsymbol{x}^4 & M_4 &= \eta^2\frac{Q_2^2}{\phi^{1/2}}\boldsymbol{x}^4\\
E_5 &= \frac{p_2}{\phi^{1/2}}\boldsymbol{x}'^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & M_5 &= \eta Q_2\boldsymbol{x}'^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})\\
E_6 &= \frac{p_2}{\phi^{1/2}}(\boldsymbol{x}^2)'(\mathbf{x}\cdot\tilde{\boldsymbol{x}}) & M_6 &= \eta^2 Q_2(\boldsymbol{x}^2)'(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})'\\
E_7 &= \frac{p_2''}{\phi^{1/2}}\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}) & M_7 &= \eta Q_2'\boldsymbol{x}^2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})\\
E_8 &= \eta\frac{p_2Q_2}{\phi}(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})^2 & M_8 &= \eta\frac{p_2Q_2}{\phi}\boldsymbol{x}^4
\end{aligned}
\tag{24.144}
$$

Each of the identities (24.143) is now weighted by an arbitrary multiplier, ρ_j, ε_j and μ_j, and added to the expression for $\int M^{(4)}\,dz$ which has the primitive form

$$
\begin{aligned}
S^I &= \int M^{(4)}dz\\
&= \frac{1}{2}\int O\{-\boldsymbol{x}^4 + 2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})^2\}dz\\
&\quad + \frac{1}{128}\int\Big(-R_3 + R_5 + 4R_6 - 8R_{17} - 16R_{24}\\
&\quad - 4E_3 + 16E_5 - \frac{8}{3}E_7 + 8M_1 - 8M_2 + \frac{8}{3}M_7\Big)dz
\end{aligned}
\tag{24.145}
$$

The general form of S^I is thus as follows:

$$
\begin{aligned}
&128S^I - 64\int O\Big\{-\boldsymbol{x}^4 + 2(\boldsymbol{x}\cdot\tilde{\boldsymbol{x}})^2\Big\}\}dz \\
&= \left[\sum_{j=1}^{10}\rho_j r_j + \sum_{j=1}^{3}\varepsilon_j e_j + \sum_{j=1}^{3}\mu_j m_j\right] \\
&\quad + \int\left\{\frac{5}{2}\rho_8 R_1 + \left(\frac{1}{2}\rho_5 + \frac{3}{2}\rho_7 - 3\rho_8\right)R_2 + \left(-1 + \frac{1}{2}\rho_4 - \rho_7\right)R_3\right. \\
&\quad + \left(\frac{1}{2}\rho_6 - \rho_7\right)R_4 + (1-\rho_6)R_5 + (4 - \rho_4 - 2\rho_9 + \varepsilon_1 + \varepsilon_2)R_6 \\
&\quad + (-\rho_5 + 3\rho_9)R_7 + (-2\rho_9 + \varepsilon_3)R_8 \\
&\quad + \left(2\rho_4 - \rho_{10} + \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2\right)R_9 + (2\rho_5 + \rho_{10})R_{10} - \rho_{10}R_{11} \\
&\quad + 2(\rho_5 - \rho_8)R_{12} \\
&\quad + \left(\rho_2 + \frac{1}{4}\rho_3 + \rho_4 - 2\rho_5 - 2\rho_7\right)R_{13} + (-\rho_4 - 2\rho_6)R_{14} \\
&\quad + 2(-\rho_2 - \rho_9 + \varepsilon_1)R_{15} + \left(-\frac{1}{2}\rho_3 - 2\rho_9 + 2\varepsilon_2\right)R_{16} \\
&\quad + \left(-8 + \frac{1}{2}\rho_1 - \rho_3 - 2\rho_4\right)R_{17} + \left(\frac{3}{2}\rho_3 - 2\rho_5\right)R_{18} \\
&\quad + \left(4\rho_2 - \rho_{10} + \frac{1}{2}\mu_1\right)R_{19} + \left(\rho_3 - \rho_{10} + \frac{1}{2}\mu_2\right)R_{20} + \left(\frac{3}{2}\rho_2 - \rho_5\right)R_{21} \\
&\quad + \left(\frac{1}{4}\rho_1 - \rho_2 - \rho_4\right)R_{22} + (\rho_1 - 4\rho_2 - \rho_3)R_{23} + (-16 - 2\rho_1)R_{24} \\
&\quad + 2(-\varepsilon_1 - \varepsilon_3)E_1 + 2(-\varepsilon_2 - \varepsilon_3)E_2 + (-4 - 2\varepsilon_1)E_3 \\
&\quad - 2\varepsilon_2 E_4 + (16 - \rho_1 - 4\varepsilon_1 + 4\varepsilon_2)E_5 + \left(-\frac{1}{2}\rho_1 - 2\varepsilon_1 - 4\varepsilon_2\right)E_6 \\
&\quad + \left(-\frac{8}{3} - 2\varepsilon_3\right)E_7 + (4\varepsilon_1 - \mu_1)E_8 \\
&\quad + (8 - \mu_1 - \mu_3)M_1 + (-8 - \mu_2 - \mu_3)M_2 + 2\mu_1 M_3 \\
&\quad + 2\mu_2 M_4 + (2\rho_1 - 2\mu_1 + 2\mu_2)M_5 \\
&\quad + (\rho_1 - \mu_1 - 2\mu_2)M_6 + \left(\frac{8}{3} - \mu_3\right)M_7 \\
&\quad \left. + (4\varepsilon_2 - \mu_2)M_8\}\right\}dz
\end{aligned}
\tag{24.146}
$$

From this very general expression, many useful forms of S^I and hence of the aberration coefficients can be derived. These are exhaustively analysed in Hawkes (1966/7b) and we say no more about them here. Before leaving this topic, we should, however, mention that not quite all possible forms of the eikonal or aberration coefficients emerge from the foregoing expression. Further transformations, which are in practice very often simplifications, can be made by recalling that the Wronskian is a constant. Thus $\phi^{1/2}(st' - s't)$ or its square may be differentiated, weighted and added to the aberration integrals; this frequently enables us to eliminate otherwise persistent but unwanted terms.

24.10 Fifth-Order Aberrations

We list here a set of formulae for the fifth-order geometrical aberration coefficients of round lenses. Such a list was published by Hawkes (1965) but a more convenient form has been derived by Liu and it is these that are reproduced here (Liu, 2004a). They include a few corrections provided by the author and are hence are not quite identical with those in the original publication. The notation is defined by the following expression for the aberration in the Gaussian image plane, referred back to the object plane as usual:

$$\begin{aligned}
\Delta\boldsymbol{r_5} &= (C_5\boldsymbol{r}'_o + c_5\boldsymbol{r}'^*_o)(\boldsymbol{r}'_o\cdot\boldsymbol{r}'_o)^2 \\
&+ (K_{51}\boldsymbol{r_o} + k_{51}\boldsymbol{r_o}^*)(\boldsymbol{r}'_o\cdot\boldsymbol{r}'_o)^2 + (K_{52}\boldsymbol{r}'_o + k_{52}\mathbf{r}'^*_o)(\boldsymbol{r}'_o\cdot\boldsymbol{r}'_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}'_o) \\
&+ (S_{51}\boldsymbol{r_o} + s_{51}\boldsymbol{r_o}^*)(\boldsymbol{r}'_o\cdot\boldsymbol{r}'_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}'_o) + (S_{52}\boldsymbol{r}'_o + s_{52}\boldsymbol{r}'^*_o)(\boldsymbol{r}'_o\cdot\boldsymbol{r}'_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}_o) \\
&+ (S_{53}\boldsymbol{r}'_o + s_{53}\boldsymbol{r}'^*_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}'_o)^2 \\
&+ (T_{51}\boldsymbol{r_o} + t_{51}\boldsymbol{r_o}^*)(\boldsymbol{r}'_o\cdot\boldsymbol{r}'_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}_o) + (T_{52}\boldsymbol{r_o} + t_{52}\boldsymbol{r_o}^*)(\boldsymbol{r}_o\cdot\boldsymbol{r}'_o)^2 \\
&+ (T_{53}\boldsymbol{r}'_o + t_{53}\boldsymbol{r}'^*_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}'_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}_o) \\
&+ (A_{51}\boldsymbol{r_o} + a_{51}\boldsymbol{r_o}^*)(\boldsymbol{r}_o\cdot\boldsymbol{r}'_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}_o) + (A_{52}\boldsymbol{r}'_o + a_{52}\boldsymbol{r}'^*_o)(\boldsymbol{r}_o\cdot\boldsymbol{r}_o)^2 \\
&+ (D_5\boldsymbol{r_o} + d_5\boldsymbol{r_o}^*)(\boldsymbol{r}_o\cdot\boldsymbol{r}_o)^2
\end{aligned} \tag{24.147}$$

The coefficients denoted by capital letters are isotropic aberrations, those denoted by lowercase letters are anisotropic. C_5 and c_5 characterize the fifth-order spherical aberration; K_5 and k_5 are coma-like; A_5 and a_5 resemble astigmatism and field curvature and D_5 and d_5 are distortions. The others, which depend on (angle)^3.(position)^2 and (angle)^2.position)^3, S, s and T, t respectively, have no simple analogues in third-order. $\boldsymbol{r} = (x,y)$ and $\boldsymbol{r}^* = (-y,x)$

Each of the analytical expressions for the aberration coefficients can be written in the unified form,

$$X = -\frac{1}{\phi_o^{1/2}}\int_{z_o}^{z_i}\left\{X_i - \frac{1}{\phi_o^{1/2}}(X_{c0} + X_{c1}h^2 + X_{c2}gh + X_{c3}g^2)\right\}dz \tag{24.148}$$

where X stands for each individual aberration coefficient and is a function of F_{lmn} and F^*_{lmn} (24.174–24.176).

24.10.1 Isotropic Aberration Coefficients

$$
\begin{aligned}
C_{5i} &= 6F_{030} \\
C_{5c0} &= -12(F_{011}\varepsilon_{020} - F_{020}\varepsilon_{011} - 4F_{020,o}F_{020}) \\
C_{5c1} &= -3\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{011}^2 + F_{010}^{*2}) \\
C_{5c2} &= 24\left(\frac{\phi}{\phi_o}\right)^{1/2}F_{020}F_{011} \\
C_{5c3} &= -48\left(\frac{\phi}{\phi_o}\right)^{1/2}F_{020}^2
\end{aligned}
\tag{24.149}
$$

$$
\begin{aligned}
K_{51i} &= F_{021} \\
K_{51c0} &= -8F_{020}\varepsilon_{110} - 8F_{000}\varepsilon_{020} + F_{011}\varepsilon_{011} + 5F_{010}^{*}\varepsilon_{010}^{*} \\
&\quad + 4F_{011,o}F_{020} + 8F_{020,o}F_{011} \\
K_{51c1} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[2(F_{011}(F_{110} + F_{002}) + F_{010}^{*}F_{001}^{*}\right] \\
K_{51c2} &= \left(\frac{\phi}{\phi_o}\right)^{1/2}\left[8F_{020}(F_{110} + F_{002}) + 3F_{011}^2 - F_{001}^{*2}\right] \\
K_{51c3} &= -12\left(\frac{\phi}{\phi_o}\right)^{1/2}F_{020}F_{011}
\end{aligned}
\tag{24.150}
$$

$$
\begin{aligned}
K_{52i} &= 4K_{51i} \\
K_{52c0} &= 4[4F_{020}\varepsilon_{110} - 2(2F_{110} + F_{002})\varepsilon_{020} - F_{011}\varepsilon_{011} \\
&\quad + 6F_{020}\varepsilon_{002} - F_{010}^{*}\varepsilon_{010}^{*} + 8F_{011,o}F_{020} + 4F_{020,o}F_{011}] \\
K_{52c1} &= 4K_{51c1} \\
K_{52c2} &= 4K_{51c2} \\
K_{52c3} &= 4K_{51c3}
\end{aligned}
\tag{24.151}
$$

$$
\begin{aligned}
S_{51i} &= 2F_{012} \\
S_{51c0} &= 2(4F_{011}\varepsilon_{110} + 4F_{020}\varepsilon_{101} - 4F_{101}\varepsilon_{020} - 3F_{002}\varepsilon_{011} \\
&\quad + F_{011}\varepsilon_{002} + 3F_{001}^{*}\varepsilon_{010} + F_{010}^{*}\varepsilon_{001} + 4F_{002,o}F_{020} \\
&\quad + 4F_{011,o}F_{011} + 4F_{020,o}F_{002} - 4F_{010,o}^{*}F_{010}^{*}) \\
S_{51c1} &= -4\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[F_{101}F_{011} + (2F_{110} + F_{002})F_{002} - F_{100}^{*}F_{010}^{*}\right]
\end{aligned}
$$

$$
\begin{aligned}
S_{51c2} &= 4\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[4F_{101}F_{020} + 2(F_{110} + 2F_{002})F_{011} - 3F^{*}_{010}F^{*}_{001}\right] \\
S_{51c3} &= -8\left(\frac{\phi}{\phi_o}\right)^{1/2}(F^{2}_{011} + 2F_{020}F_{002} - F^{*2}_{010})
\end{aligned}
\tag{24.152}
$$

$$
\begin{aligned}
S_{52i} &= 4F_{120} \\
S_{52c0} &= -2F_{011}\varepsilon_{110} + 12F_{020}\varepsilon_{101} - 4F_{101}\varepsilon_{020} - 2F_{110}\varepsilon_{011} \\
&\quad - 5F^{*}_{001}\varepsilon^{*}_{010} + 3F^{*}_{010}\varepsilon^{*}_{001} + 8F_{020,o}F_{110} + 24F_{110,o}F_{020} \\
&\quad + 2F_{011,o}F_{011} + 18F^{*}_{010,o}F^{*}_{010} \\
S_{52c1} &= -2\left(\frac{\phi}{\phi_o}\right)^{1/2}(4F^{2}_{110} + 2F_{101}F_{011} + 6F^{*}_{100}F^{*}_{010} + F^{*2}_{001}] \\
S_{52c2} &= 16\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{101}F_{020} + F_{110}F_{011} + F^{*}_{010}F^{*}_{001}] \\
S_{52c3} &= -2\left(\frac{\phi}{\phi_o}\right)^{1/2}(16F_{110}F_{020} + F^{2}_{110} + 9F^{*2}_{010})
\end{aligned}
\tag{24.153}
$$

$$
\begin{aligned}
S_{53i} &= S_{51i} \\
S_{53c0} &= 4[4F_{020}\varepsilon_{101} - (2F_{110} + F_{002})\varepsilon_{011} + F_{011}\varepsilon_{002} \\
&\quad - F^{*}_{010}\varepsilon^{*}_{001} + 4F_{002,o}F_{020} + 2F_{011,o}F_{011} \\
&\quad - 2F^{*}_{010,o}F^{*}_{010}] \\
S_{53c1} &= S_{51c1} \\
S_{53c2} &= S_{51c2} \\
S_{53c3} &= S_{51c3}
\end{aligned}
\tag{24.154}
$$

$$
\begin{aligned}
T_{51i} &= F_{111} \\
T_{51c0} &= 16F_{020}\varepsilon_{200} + 4F_{110}\varepsilon_{110} + F_{011}\varepsilon_{101} - 3F_{101}\varepsilon_{011} \\
&\quad + 7F^{*}_{010}\varepsilon^{*}_{100} + 3F^{*}_{100}\varepsilon^{*}_{010} - F^{*}_{001}\varepsilon^{*}_{001} + 2F_{011,o}F_{110} \\
&\quad + 4F_{101,o}F_{020} + 4F_{110,o}F_{011} + 2F_{011,o}F_{002} \\
&\quad + 2F^{*}_{001,o}F^{*}_{010} + 3F^{*}_{010,o}F^{*}_{001} \\
T_{51c1} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[4F_{011}F_{200} + 2(F_{002} + 3F_{110})F_{101} + 5F^{*}_{100}F^{*}_{001}\right] \\
T_{51c2} &= \left(\frac{\phi}{\phi_o}\right)^{1/2}[4(F_{110} + 2F_{002})F_{110} + 16F_{200}F_{020} + 6F_{101}F_{011} \\
&\quad + 2F^{*}_{100}F^{*}_{010} + 3F^{*2}_{001}] \\
T_{51c3} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[4F_{101}F_{020} + 2(3F_{110} + F_{002})F_{011} + 5F^{*}_{010}F^{*}_{001}\right]
\end{aligned}
\tag{24.155}
$$

$$
\begin{aligned}
T_{52i} &= 3F_{003} \\
T_{52c0} &= 4(2F_{011}\varepsilon_{101} - F_{101}\varepsilon_{011} - F_{002}\varepsilon_{002} - 2F^{*}_{010}\varepsilon^{*}_{100} \\
&\quad + F^{*}_{100}\varepsilon^{*}_{010} + F^{*}_{001}\varepsilon^{*}_{001} + 2F_{002,o}F_{011} + 2F_{011,o}F_{002} \\
&\quad - 2F^{*}_{001,o}F^{*}_{010} - F^{*}_{010,o}F^{*}_{001}) \\
T_{52c1} &= -12\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{002}F_{101} - F^{*}_{001}F^{*}_{100}) \\
T_{52c2} &= 12\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{101}F_{011} + F^{2}_{002} - F^{*}_{100}F^{*}_{010} - F^{*2}_{001}) \\
T_{52c3} &= -12\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{011}F_{002} - F^{*}_{010}F^{*}_{001})
\end{aligned}
\tag{24.156}
$$

$$
\begin{aligned}
T_{53i} &= 2T_{51i} \\
T_{53c0} &= 2[16F_{020}\varepsilon_{002} - 2(2F_{110} + F_{002})\varepsilon_{110} + 3F_{011}\varepsilon_{101} \\
&\quad - F_{101}\varepsilon_{011} - 2F_{110}\varepsilon_{002} + F^{*}_{010}\varepsilon^{*}_{100} - 3F^{*}_{100}\varepsilon^{*}_{010} \\
&\quad - F^{*}_{001}\varepsilon^{*}_{001} + 2F_{011,o}F_{110} + 4F_{101,o}F_{020} \\
&\quad + 2(2F_{110,o} + F_{002,o})F_{011} + 4F^{*}_{001,o}F^{*}_{010} + F^{*}_{010,o}F^{*}_{001}] \\
T_{53c1} &= 2T_{51c1} \\
T_{53c2} &= 2T_{51c2} \\
T_{53c3} &= 2T_{51c3}
\end{aligned}
\tag{24.157}
$$

$$
\begin{aligned}
A_{51i} &= 2F_{102} \\
A_{51c0} &= 2[8F_{011}\varepsilon_{200} - 2F_{101}\varepsilon_{110} + (2F_{110} + F_{002})\varepsilon_{101} \\
&\quad - 3F_{101}\varepsilon_{002} + 2F^{*}_{100}\varepsilon^{*}_{001} + 2F_{002,o}F_{110} + 2F_{101,o}F_{011} \\
&\quad + 2(F_{110,o} + F_{002,o})F_{002} - 2F^{*}_{100,o}F^{*}_{010}] \\
A_{51c1} &= -8\left(\frac{\phi}{\phi_o}\right)^{1/2}(F^{2}_{101} + 2F_{200}F_{002} - F^{*2}_{100}) \\
A_{51c2} &= 4\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[2(2F_{002} + F_{110})F_{101} + 4F_{200}F_{011} - 3F^{*}_{100}F^{*}_{001}\right] \\
A_{51c3} &= -4\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[F_{101}F_{011} + (2F_{110} + F_{002})F_{002} - F^{*}_{100}F^{*}_{010}\right]
\end{aligned}
\tag{24.158}
$$

$$
\begin{aligned}
A_{52i} &= 2F_{210} \\
A_{52c0} &= 8F_{011}\varepsilon_{200} - 2F_{101}\varepsilon_{110} - 2F_{110}\varepsilon_{101} + F^{*}_{001}\varepsilon^{*}_{100} \\
&\quad - 3F^{*}_{100}\varepsilon^{*}_{001} + F_{110,o}F_{110} + 2F_{101,o}F_{011} \\
&\quad + 6F^{*}_{100,o}F^{*}_{010} + F^{*}_{001,o}F^{*}_{001} \\
A_{52c1} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2}(16F_{200}F_{110} + F^{2}_{101} + 9F^{*2}_{100}) \\
A_{52c2} &= 8\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{110}F_{101} + F_{200}F_{011} + F^{*}_{100}F^{*}_{001}) \\
A_{52c3} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2}(4F^{2}_{110} + 2F_{101}F_{011} + 6F^{*}_{100}F^{*}_{010} + F^{*2}_{001})
\end{aligned}
\tag{24.159}
$$

$$
\begin{aligned}
D_{5i} &= F_{201} \\
D_{5c0} &= 8(F_{110} + F_{002})\varepsilon_{200} - 3F_{101}\varepsilon_{101} + F^{*}_{100}\varepsilon^{*}_{100} \\
&\quad + 2F_{101,o}F_{110} + 2F_{101,o}F_{002} + F^{*}_{100,o}F^{*}_{001} \\
D_{5c1} &= -12\left(\frac{\phi}{\phi_o}\right)^{1/2}F_{200}F_{101} \\
D_{5c2} &= \left(\frac{\phi}{\phi_o}\right)^{1/2}\left[8(F_{110} + F_{002})F_{200} + 3F^{2}_{101} - F^{*2}_{100}\right] \\
D_{5C3} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[2(F_{110} + F_{002})F_{101} + F^{*}_{100}F^{*}_{001}\right]
\end{aligned}
\tag{24.160}
$$

24.10.2 Anisotropic Aberration Coefficients

$$
\begin{aligned}
c_{5i} &= 0 \\
c_{5c0} &= -4(F^{*}_{010}\varepsilon_{020} + F_{020}\varepsilon^{*}_{010}) \\
c_{5c1} &= 0 \\
c_{5c2} &= 0 \\
c_{5c3} &= 0
\end{aligned}
\tag{24.161}
$$

$$
\begin{aligned}
k_{51i} &= 5F^{*}_{020} \\
k_{51c0} &= -8F^{*}_{001}\varepsilon_{020} + 3F^{*}_{010}\varepsilon_{011} - 7F_{011}\varepsilon^{*}_{010} + 12F_{020}\varepsilon^{*}_{001} \\
&\quad + 36F^{*}_{010,o}F_{020} + 24F_{020,o}F^{*}_{010} \\
k_{51c1} &= -5\left(\frac{\phi}{\phi_o}\right)^{1/2}(2F_{110}F^{*}_{010} + F_{011}F^{*}_{001}) \\
k_{51c2} &= 20\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{011}F^{*}_{010} + F_{020}F^{*}_{001}) \\
k_{51c3} &= -60\left(\frac{\phi}{\phi_o}\right)^{1/2}F_{020}F^{*}_{010}
\end{aligned}
\tag{24.162}
$$

$$
\begin{aligned}
k_{52i} &= -\frac{4}{5}k_{51i} \\
k_{52c0} &= 4(-F^*_{010}\varepsilon_{011} + F_{011}\varepsilon^*_{010} - 4F_{020}\varepsilon^*_{001} - 8F^*_{010,o}F_{020} \\
&\quad - 4F_{020,o}F^*_{010}) \\
k_{52c1} &= -\frac{4}{5}k_{51c1} \\
k_{52c2} &= -\frac{4}{5}k_{51c2} \\
k_{52c3} &= -\frac{4}{5}k_{51c3}
\end{aligned}
\tag{24.163}
$$

$$
\begin{aligned}
s_{51i} &= 4F^*_{011} \\
s_{51c0} &= 2[4F^*_{010}\varepsilon_{110} - 4F^*_{100}\varepsilon_{020} - 3F^*_{001}\varepsilon_{011} + 3F^*_{010}\varepsilon_{002} \\
&\quad + 12F_{020}\varepsilon^*_{100} - (4F_{110} + 5F_{002})\varepsilon^*_{010} + F_{011}\varepsilon^*_{001} \\
&\quad + 12F^*_{001,o}F_{020} + 8F^*_{010,o}F_{011} + 8F_{011,o}F^*_{010} \\
&\quad + 4F_{020,o}F^*_{001}] \\
s_{51c1} &= -8\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[F_{011}F^*_{100} + F_{101}F^*_{010} + (2F_{110} + F_{002})F^*_{001}\right] \\
s_{51c2} &= 8\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[4F_{020}F^*_{100} + 2(F_{110} + 2F_{002})F^*_{010} + 3F_{011}F^*_{001}\right] \\
s_{51c3} &= -32\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{011}F^*_{010} + F_{020}F^*_{001})
\end{aligned}
\tag{24.164}
$$

$$
\begin{aligned}
s_{52i} &= -\frac{1}{4}s_{51i} \\
s_{52c0} &= -6F^*_{010}\varepsilon_{110} - 4F^*_{100}\varepsilon_{020} + F^*_{001}\varepsilon_{011} - 12F_{020}\varepsilon^*_{100} \\
&\quad - 2(F_{110} - 2F_{002})\varepsilon^*_{010} - F_{011}\varepsilon^*_{001} - 4F^*_{001,o}F_{020} \\
&\quad - 6F^*_{010,o}F_{011} - 2F_{011,o}F^*_{010} - 4F_{020,o}F^*_{001} \\
s_{52c1} &= -\frac{1}{4}s_{51c1} \\
s_{52c2} &= -\frac{1}{4}s_{51c2} \\
s_{52c3} &= -\frac{1}{4}s_{51c3}
\end{aligned}
\tag{24.165}
$$

$$
\begin{aligned}
s_{53i} &= -\frac{1}{2}s_{51i} \\
s_{53c0} &= 4[-F^*_{010}\varepsilon_{002} - 4F_{020}\varepsilon^*_{100} + (2F_{110} + F_{002})\varepsilon^*_{010} \\
&\quad - F_{011}\varepsilon^*_{001} - 4F^*_{001,o}F_{020} - 2F^*_{010,o}F_{011} \\
&\quad - 2F_{011,o}F^*_{010}] \\
s_{53c1} &= -\frac{1}{2}s_{51c1} \\
s_{53c2} &= -\frac{1}{2}s_{51c2} \\
s_{53c3} &= -\frac{1}{2}s_{51c3}
\end{aligned}
\tag{24.166}
$$

$$
\begin{aligned}
t_{51i} &= 3F^*_{110} \\
t_{51c0} &= -2F^*_{001}\varepsilon_{110} + 3F^*_{010}\varepsilon_{101} - 3F^*_{100}\varepsilon_{011} + 3F_{011}\varepsilon^*_{100} \\
&\quad - 3F_{101}\varepsilon^*_{010} - 2F_{110}\varepsilon^*_{001} + 6F^*_{010,o}F_{110} + 12F^*_{100,o}F_{020} \\
&\quad + 2F^*_{001,o}F_{011} + 12F_{110,o}F^*_{010} + F_{011,o}F^*_{001} \\
t_{51c1} &= -3\left(\frac{\phi}{\phi_o}\right)^{1/2}(6F_{110}F^*_{100} + 4F_{200}F^*_{010} + F_{101}F^*_{001}) \\
t_{51c2} &= 12\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{011}F^*_{100} + F_{101}F^*_{010} + F_{110}F^*_{001}) \\
t_{51c3} &= -3\left(\frac{\phi}{\phi_o}\right)^{1/2}(4F_{020}F^*_{100} + 6F_{110}F^*_{010} + F_{011}F^*_{001})
\end{aligned}
\tag{24.167}
$$

$$
\begin{aligned}
t_{52i} &= 3F^*_{002} \\
t_{52c0} &= 4(2F^*_{010}\varepsilon_{101} - F^*_{100}\varepsilon_{011} - F^*_{001}\varepsilon_{002} + 2F_{011}\varepsilon^*_{100} \\
&\quad - F_{101}\varepsilon^*_{010} - F_{002}\varepsilon^*_{001} + 2F^*_{001,o}F_{011} + F^*_{010,o}F_{002} \\
&\quad + 2F^*_{002,o}F^*_{010} + F_{011,o}F^*_{001} \\
t_{52c1} &= -12\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{002}F^*_{100} + F_{101}F^*_{001}) \\
t_{52c2} &= 12\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{011}F^*_{100} + F_{101}F^*_{010} + 2F_{002}F^*_{001}) \\
t_{52c3} &= -12\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{002}F^*_{010} + F_{011}F^*_{001})
\end{aligned}
\tag{24.168}
$$

$$
\begin{aligned}
t_{53i} &= -\frac{2}{3}(t_{51i} + t_{52i}) \\
t_{53c0} &= 2[-3F^*_{010}\varepsilon_{101} - F^*_{100}\varepsilon_{011} + F^*_{001}\varepsilon_{002} - 7F_{011}\varepsilon^*_{100} \\
&\quad + 3F_{101}\varepsilon^*_{010} + F_{002}\varepsilon^*_{001} - 2F^*_{010,o}F_{110} \\
&\quad - 4F^*_{100,o}F_{020} - 4F^*_{001,o}F_{011} - 2F^*_{010,o}F_{002} \\
&\quad - 2(2F_{110,o} + F_{002,o})F^*_{010} - F_{011,o}F^*_{001}] \\
t_{53c1} &= -\frac{2}{3}(t_{51c1} + t_{52c1}) \\
t_{53c2} &= -\frac{2}{3}(t_{51c2} + t_{52c2}) \\
t_{53c3} &= -\frac{2}{3}(t_{51c1} + t_{51c2})
\end{aligned}
\tag{24.169}
$$

$$
\begin{aligned}
a_{51i} &= 2F^*_{101} \\
a_{51c0} &= 2[8F^*_{010}\varepsilon_{200} - 2F^*_{100}\varepsilon_{110} - 3F^*_{100}\varepsilon_{002} + (2F_{110} \\
&\quad + F_{002})\varepsilon^*_{100} - 2F_{101}\varepsilon^*_{001} + 2F^*_{001,o}F_{110} + 2F^*_{100,o}F_{011} \\
&\quad + F^*_{001,o}F_{002} + 2F_{101,o}F^*_{010} + (F_{110,o} + F_{002,o})F^*_{001}] \\
a_{51c1} &= -16\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{101}F^*_{100} + F_{200}F^*_{001}) \\
a_{51c2} &= 4\left(\frac{\phi}{\phi_o}\right)^{1/2}\left[2(F_{110} + 2F_{002})F^*_{100} + 4F_{200}F^*_{010} + 3F_{101}F^*_{001}\right] \\
a_{51c3} &= -4\left(\frac{\phi}{\phi_o}\right)^{1/2}(F_{011}F^*_{100} + F_{101}F^*_{010} + (2F_{110} + F_{002})F^*_{001})
\end{aligned}
\tag{24.170}
$$

$$
\begin{aligned}
a_{52i} &= -\frac{1}{2}a_{51i} \\
a_{52c0} &= -8F^*_{010}\varepsilon_{200} - 2F^*_{100}\varepsilon_{110} + F^*_{001}\varepsilon_{101} \\
&\quad - 2(3F_{110} + 2F_{002})\varepsilon^*_{100} + 3F_{101}\varepsilon^*_{001} - 2F^*_{001,o}F_{110} \\
&\quad - 2F^*_{100,o}F_{011} - F^*_{001,o}F_{002} - 2F_{101,o}F^*_{010} \\
&\quad - 2F_{110,o}F^*_{001} \\
a_{52c1} &= -\frac{1}{2}a_{51c1} \\
a_{52c2} &= -\frac{1}{2}a_{51c2} \\
a_{52c3} &= -\frac{1}{2}a_{51c3}
\end{aligned}
\tag{24.171}
$$

$$
\begin{aligned}
d_{5i} &= F^*_{200} \\
d_{5c0} &= 4F^*_{001}\varepsilon_{200} - 3F^*_{100}\varepsilon_{101} - F_{101}\varepsilon^*_{100} + 2F^*_{100,o}F_{110} \\
&\quad + F_{101,o}F^*_{001} \\
d_{5c1} &= -12\left(\frac{\phi}{\phi_o}\right)^{1/2} F_{200}F^*_{100} \\
d_{5c2} &= 4\left(\frac{\phi}{\phi_o}\right)^{1/2} (F_{101}F^*_{100} + F_{200}F^*_{001}) \\
d_{5c3} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2} (2F_{110}F^*_{100} + F_{101}F^*_{001})
\end{aligned}
\tag{24.172}
$$

In these equations we have written

$$
\begin{aligned}
\varepsilon_{lmn} &= \int_{z_o}^{z} F_{lmn}dz \begin{pmatrix} l,m,n = 0,1,2 & l+m+n = 2 \\ l,m,n = 0,1,2,3 & l+m+n = 3 \end{pmatrix} \\
\varepsilon^*_{lmn} &= \int_{z_o}^{z} F^*_{lmn}dz \begin{pmatrix} l,m,n = 0,1 & l+m+n = 1 \\ l,m,n = 0,1,2 & l+m+n = 2 \end{pmatrix}
\end{aligned}
\tag{24.173}
$$

The quantities F_{lmn} are as follows:

$$
\begin{aligned}
F_{200} &= F_{LMN}(g^4, g^2g'^2, g'^4) \\
F_{110} &= 2F_{LMN}\left(h^2g^2, \frac{h^2g'^2 + h'^2g^2}{2}, h'^2g'^2\right) - R\phi_o \\
F_{101} &= 4F_{LMN}(hg^3, \frac{hgg'^2 + h'g'g^2}{2}, h'g'^3) \\
F_{020} &= F_{LMN}(h^4, h^2h'^2, h'^4) \\
F_{011} &= 4F_{LMN}\left(h^3g, \frac{hgh'^2 + h'g'h^2}{2}, h'^3g'\right) \\
F_{002} &= 4F_{LMN}(h^2g^2, hgh'g', h'^2g'^2) + R\phi_o
\end{aligned}
\tag{24.174}
$$

$$
\begin{aligned}
F^*_{100} &= F_{PQ}(g^2, g'^2) \\
F^*_{010} &= F_{PQ}(h^2, h'^2) \\
F^*_{001} &= 2F_{PQ}(gh, g'h')
\end{aligned}
\tag{24.175}
$$

$$
\begin{aligned}
F_{300} &= F_L(g^6, g^4 g'^2, g^2 g'^4, g'^6) \\
F_{210} &= 3F_L\left[g^4 h^2, \frac{g^2(g^2 h'^2 + 2g'^2 h^2)}{3}, \frac{g'^2(g'^2 h^2 + 2g^2 h'^2)}{3}, g'^4 h'^2\right] \\
&\quad + F_N(g^2, g'^2) \\
F_{201} &= 6F_L\left[g^5 h, \frac{g^3 g'(gh' + 2g'h)}{3}, \frac{g g'^3(g'h + 2gh')}{3}, g'^5 h'\right] \\
F_{120} &= 3F_L\left[g^2 h^4, \frac{h^2(h^2 g'^2 + 2g^2 h'^2)}{3}, \frac{h'^2(g^2 h'^2 + 2g'^2 h^2)}{3}, g'^2 h'^4\right] \\
&\quad + F_N(h^2, h'^2) \\
F_{111} &= 12F_L\left[g^3 h^3, \frac{gh(h^2 g'^2 + ghg'h' + g^2 h'^2)}{3}, \frac{g'h'(g'^2 h^2 + ghh'g' + g^2 h'^2)}{3}, h'^3 g'^3\right] \\
&\quad + 2F_N(gh, g'h') \\
F_{102} &= 12F_L\left[g^4 h^2, \frac{g^2 hg'(g'h + 2gh')}{3}, \frac{gh'g'^2(gh' + 2g'h)}{3}, g'^4 h'^2\right] \\
&\quad - F_N(g^2, g'^2) \\
F_{030} &= F_L(h^6, h^4 h'^2, h^2 h'^4, h'^6) \\
F_{021} &= 6F_L\left[gh^5, \frac{h^3 h'(g'h + 2gh')}{3}, \frac{hh'^3(gh' + 2g'h)}{3}, g'h'^5\right] \\
F_{012} &= 12F_L\left[g^2 h^4, \frac{gh^2 h'(gh' + 2g'h)}{3}, \frac{g'hh'^2(g'h + 2gh')}{3}, h'^4 g'^2\right] - F_N(h^2, h'^2) \\
F_{003} &= 8F_L(g^3 h^3, g^2 h^2 g'h', ghg'^2 h'^2, g'^3 h'^3) - 2F_N(gh, g'h') \\
F^*_{200} &= F_M(g^4, g^2 g'^2, g'^4) \\
F^*_{110} &= 2F_M(h^2 g^2, \frac{h^2 g'^2 + h'^2 g^2}{2}, h'^2 g'^2) - R_3 \phi_o^{3/2} \\
F^*_{101} &= 4F_M(hg^3, \frac{ghg'^2 + g'h'g^2}{2}, h'g'^3) \\
F^*_{020} &= F_M(h^4, h^2 h'^2, h'^4) \\
F^*_{011} &= 4F_M(h^3 g, \frac{ghh'^2 + g'h'h^2}{2}, h'^3 g') \\
F^*_{002} &= 4F_M(g^2 h^2, ghg'h', g'^2 h'^2) + R_3 \phi_o^{3/2}
\end{aligned}
\tag{24.176}
$$

Here,

$$
\begin{aligned}
F_{LMN} &= -\frac{L_1}{4}x_1 - \frac{L_2}{4}x_2 - \frac{L_3}{4}x_3 \\
F_{PQ} &= \phi_o^{1/2}(-Px_1 - Qx_2) \\
F_L &= -L_4x_1 - L_5x_2 - L_6x_3 - L_7x_4 \\
F_M &= \phi_o^{1/2}(-M_1x_1 - M_2x_2 - M_3x_3) \\
F_N &= \phi_o(-N_1x_1 - N_2x_2)
\end{aligned}
\tag{24.177}
$$

in which

$$
\begin{aligned}
L_4 &= -\frac{1}{9216\phi^{5/2}}\left[9\eta^6B^6 + 9\eta^4\phi''B^4 - 9\eta^2(\phi''^2 - \phi\phi^{(4)})B^2\right. \\
&\quad \left. - 12\eta^2\phi^2B^{(4)}B - 9\phi''^3 + 9\phi\phi''\phi^{(4)} - 2\phi^2\phi^{(6)}\right] \\
L_5 &= -\frac{3\eta^4B^4 + 2\eta^2\phi''B^2 - \phi''^2 + \phi\phi^{(4)}}{256\phi^{3/2}} \\
L_6 &= -\frac{3\eta^2B^2 + \phi''}{64\phi^{1/2}} \\
L_7 &= -\frac{\phi^{1/2}}{16} \\
M_1 &= -\frac{\eta[9\eta^4B^5 + 6\eta^2\phi''B^3 - 3(\phi''^2 - \phi\phi^{(4)})B - 2\phi^2B^4]}{768\phi^{5/2}} \\
M_2 &= -\frac{\eta B(3\eta^2B^2 + \phi'')}{32^{3/2}} \\
M_3 &= -\frac{3\eta B}{16\phi^{1/2}} \\
N_1 &= -\frac{\eta^2B^2(3\eta^2B^2 + \phi\phi'')}{64\phi^{5/2}} \\
N_2 &= -\frac{3\eta^2B^2}{16\phi^{3/2}} \\
R_3 &= -\frac{\eta^3B^3}{16\phi^{5/2}}
\end{aligned}
\tag{24.178}
$$

For the definitions of L_1, L_2, L_3, P, Q and R, see Eq. (24.3). Finally, we note that $F_{lmn,o}$ denotes the value of F_{lmn} at the object plane; for F_{200}, for example, we have

$$F_{200,o} = F_{200}|_{z=z_o} = -\frac{1}{4}L|_{z=z_o} = -\frac{1}{128\phi_o^{1/2}}\left\{\frac{(\phi_o'' + \eta^2 B_o^2)^2}{\phi_o} - \phi_o^{(4)} - 4\eta^2 B_o B_o''\right\} \qquad (24.179)$$

Related formulae. The fifth-order aberration coefficients have also been studied by U [Russian transliteration of Wu] (1957), Ximen (1990a, b, 1991), Li and Ni (1988), Li et al. (1995, 2002), Mu et al. (1999) and Liu (2002). Liu has also performed an extremely thorough investigation of the chromatic aberrations (Liu, 2007) in which the role of MATHEMATICA is presented in detail and the correctness of the formulae is verified by comparison with results obtained with the aid of differential algebra (Section 34.8); his formulae are reproduced at the end of Section 26.1. The work of Li and Ni, and Li et al. and Mu et al. on the model field (36.92) must be read with caution: we have seen that there are two contributions to the fifth-order coefficients, one arising from $M^{\mathrm{II}} = \mathrm{M}^{(6)}$, the other from a combination of the third-order solution and the paraxial trajectories (22.38−22.44); these authors retain only the $M^{(6)}$ term. Ximen and Liu (1997) have calculated these terms separately for Glaser's bell-shaped field (36.8) and found that the $M^{(6)}$ term is dominant.

The values of the axial aberration coefficients for a uniform magnetic field have been calculated by ray tracing, by solution of the exact ray equation and by evaluating the formulae listed above (Lencová et al., 2008). The agreement was exact. This comparison drew attention to the need for care when defining the aberrations: when expressed in terms of the tangent of the angle at the object plane (α) instead of the angle itself, different values are obtained (Rose, 1968). If we write

$$C_s\alpha^3 + (C_5 + \mathrm{i}c_5)\alpha^5 \equiv \overline{C}_s \tan^3\alpha + (\overline{C}_5 + \mathrm{i}\overline{c}_5)\tan^5\alpha$$

we see that

$$C_s = \overline{C}_s, \quad C_5 = \overline{C}_5 + \overline{C}_s, \quad c_5 = \overline{c}_5$$

as found by Lencová et al.

CHAPTER 25

Asymptotic Aberration Coefficients

For most lenses, projectors and condensers in particular, the 'object' is in fact an intermediate image of the specimen or the source created by the lenses upstream, so that the entire lens field participates in the image formation. We are thus interested in the coordination between incident and emergent *asymptotes*, as explained in Chapter 16, Gaussian Optics of Rotationally Symmetric Systems: Asymptotic Image Formation, and the corresponding aberrations are hence said to be *asymptotic*. They were first thoroughly studied by Lenz (1956, 1957), using the trajectory method; the correct form of the characteristic function is to be found in Sturrock (1955) and this was later employed by Hawkes (1968, 1970b,c) to explore them in more detail. Just as for the real aberrations, we must distinguish between aberrations expressed in terms of position and gradient at the (asymptotic) object and in terms of position in the (asymptotic) object plane and *real* aperture plane (or the entrance pupil). The former are generally more useful and we deal with them first.

In order to introduce the boundary conditions $x_o^{(1)} = y_o^{(1)} = x_o'^{(1)} = y_o'^{(1)} = 0$ directly, we need the relation between $p^{(1)}$, $q^{(1)}$, $x'^{(1)}$ and $y'^{(1)}$; to the paraxial approximation, we have

$$p \coloneqq p^{(p)} = \hat{\phi}^{1/2} x' \quad q \coloneqq q^{(p)} = \hat{\phi}^{1/2} y' \tag{25.1}$$

while to the third-order approximation,

$$\begin{aligned} p &\coloneqq p^{(p)} + p^{(1)} = \hat{\phi}^{1/2} x' \left(1 - \frac{x'^2 + y'^2}{2}\right) \\ q &\coloneqq q^{(p)} + q^{(1)} = \hat{\phi}^{1/2} y' \left(1 - \frac{x'^2 + y'^2}{2}\right) \end{aligned} \tag{25.2}$$

Setting $x' = x'^{(p)} + x'^{(1)}$ and $y' = y'^{(p)} + y'^{(1)}$ we obtain

$$\begin{aligned} p^{(1)} &= \hat{\phi}^{1/2} \left\{ x'^{(1)} - \frac{1}{2} x'^{(p)} \left(x'^{(p)2} + y'^{(p)2} \right) \right\} \\ q^{(1)} &= \hat{\phi}^{1/2} \left\{ y'^{(1)} - \frac{1}{2} y'^{(p)} \left(x'^{(p)2} + y'^{(p)2} \right) \right\} \end{aligned} \tag{25.3}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00025-0
© 2018 Elsevier Ltd. All rights reserved.

Thus if we impose the condition $x'^{(1)} = y'^{(1)} = 0$ in the object plane, the quantities $p_o^{(1)}$ and $q_o^{(1)}$ must take the values

$$\begin{aligned} p_o^{(1)} &= -\frac{1}{2}\hat{\phi}_o^{1/2} x'_o\left(x'^2_o + y'^2_o\right) \\ q_o^{(1)} &= -\frac{1}{2}\hat{\phi}_o^{1/2} y'_0\left(x'^2_o + y'^2_o\right) \end{aligned} \tag{25.4}$$

in that plane, where we have dropped the index (p). Substituting this into the expression for S^I_{o2} (22.23), we find

$$\begin{aligned} \Delta S^I_{o2} = p_2^{(1)} \cdot \Delta x_2 + q_2^{(1)} \cdot \Delta y_2 - \left(x_2^{(1)} \cdot \Delta p_2 + y_2^{(1)} \cdot \Delta q_2\right) \\ + \frac{1}{2}\hat{\phi}_o^{1/2}\left(x'_o \cdot \Delta x_o + y'_o \cdot \Delta y_o\right)\left(x'^2_o + y'^2_o\right) \end{aligned} \tag{25.5}$$

Substituting $x(z) = x_o G(z) + x'_o H(z)$, $y(z) = y_o G(z) + y'_o H(z)$ from (16.1), with

$$\lim_{z \to -\infty} G(z) = 1 \qquad \lim_{z \to -\infty} H(z) = z - z_o \tag{25.6}$$

we find

$$\begin{aligned} x^{(1)}(z_2) &= \frac{1}{\hat{\phi}_o^{1/2}}\left(\tilde{H}_2 \frac{\partial S^I_{o2}}{\partial x_o} - \tilde{G}_2 \frac{\partial S^I_{o2}}{\partial x'_o}\right) - \frac{1}{2}\tilde{H}_2 x'_o \theta_o^2 \\ y^{(1)}(z_2) &= \frac{1}{\hat{\phi}_o^{1/2}}\left(\tilde{H}_2 \frac{\partial S^I_{o2}}{\partial y_o} - \tilde{G}_2 \frac{\partial S^I_{o2}}{\partial y'_o}\right) - \frac{1}{2}\tilde{H}_2 y'_o \theta_o^2 \end{aligned} \tag{25.7a}$$

and

$$\begin{aligned} x'^{(1)}(z_2) &= \frac{1}{\hat{\phi}_o^{1/2}}\left(\tilde{H'}_2 \frac{\partial S^I_{o2}}{\partial x_o} - \tilde{G}'_2 \frac{\partial S^I_{o2}}{\partial x'_o}\right) - \frac{1}{2}\tilde{H'} x'_o \theta_o^2 + \frac{1}{2} x' \theta_2^2 \\ y'^{(1)}(z_2) &= \frac{1}{\hat{\phi}_o^{1/2}}\left(\tilde{H'} \frac{\partial S^I_{o2}}{\partial y_o} - \tilde{G}'_2 \frac{\partial S^I_{o2}}{\partial y'_o}\right) - \frac{1}{2}\tilde{H'} y'_o \theta_o^2 + \frac{1}{2} y' \theta_2^2 \end{aligned} \tag{25.7b}$$

in which $\tilde{G}_2$ and $\tilde{H}_2$ denote the asymptotes to $G(z)$ and $H(z)$ in image space and we extend (24.33a) by writing

$$\begin{aligned} \theta_o^2 &:= x'^2_o + y'^2_o \qquad V_o = x_o x'_o + y_o y'_o \\ v_o &= x_o y'_o - x'_o y_o \end{aligned} \tag{25.8}$$

These formulae are exactly analogous to Eq. (24.27) and by following the same reasoning as we employed for the real aberrations, aberration integrals involving B, ϕ, G and H similar in appearance to those listed in Chapter 24, The Geometrical Aberrations of Round Lenses, are obtained. Before doing this, it is advantageous to replace the ray $H(z)$, which varies with object position, by the ray $\overline{G}(z)$ introduced in Chapter 16, Gaussian Optics of Rotationally Symmetric Systems: Asymptotic Image Formation, which does not. In this way, we obtain aberration integrals that are characteristic of the distribution $B(z)$ or $\phi(z)$, irrespective of the working conditions. Another choice is often adopted when studying aberrations; it may be possible to eliminate some of these by exploiting symmetry properties and for this, it is helpful to use a first ray that is symmetric and a second ray that is antisymmetric about some symmetry plane (Hawkes, 1987a,b).

The ray $\overline{G}(z)$ is, we recall, defined by the boundary condition (16.1) $\lim_{z\to\infty} \overline{G}(z) = 1$ and hence

$$\begin{aligned} H(z) &= -\frac{f_o}{M} G(z) + f_o \overline{G}(z) \\ &= -f_o \left(mG - \overline{G} \right) \end{aligned} \tag{25.9}$$

where

$$m \coloneqq \frac{1}{M} \tag{25.10}$$

is the *reciprocal magnification* (not to be confused with the demagnification, the term commonly employed to describe a magnification $|M|$ less than unity).

The characteristic function now consists of three branches (Fig. 25.1). We write

$$S^I_{oi} =: S^I(z_o, -\infty) + S^I(-\infty, \infty) + S^I(\infty, z_i) \tag{25.11a}$$

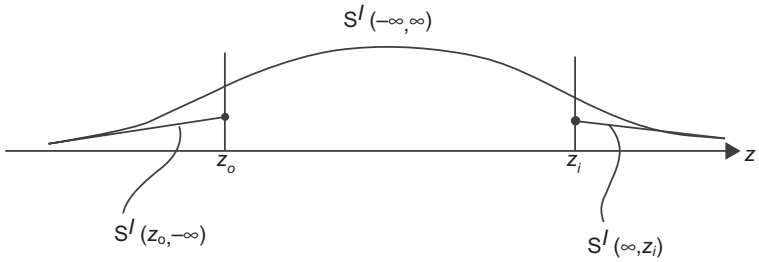


Figure 25.1
Branches of S^I_{oi}. The asymptotic branches $S^I(z_o, -\infty)$ and $S^I(\infty, z_i)$ are straight lines, the asymptotes to $S^I(-\infty, \infty)$.

The contribution $S^I(-\infty, \infty)$ is given by

$$S^I(-\infty, \infty) = \int_{-\infty}^{\infty} M^{(4)} dz \tag{25.11b}$$

into which we must substitute

$$x(z) = x_o G(z) + x'_o H(z) = \left(x_o - f_o x'_o m\right) G(z) + f_o x'_o \overline{G}(z)$$

and similarly for $y(z)$. The other two contributions are 'virtual'. In them, $\phi(z) = \text{const}$ and $B(z) = 0$ and the rays $G(z)$ and $H(z)$ or $\overline{G}(z)$ are replaced by their asymptotes. Thus

$$\begin{aligned} S^I(z_o, -\infty) &= -\frac{\hat{\phi}_o^{1/2}}{8} \theta_o^4 \lim_{z \to -\infty} (z - z_o) \\ &= -\frac{\hat{\phi}_0^{1/2}}{8} \theta_o^4 \left\{ \lim_{z \to -\infty} (z - z_{Fo}) - f_o m \right\} \end{aligned} \tag{25.11c}$$

$$S^I(\infty, z_i) = -\frac{\hat{\phi}_i^{1/2}}{8} \left\{ \frac{r_0^2}{f_i^2} - 2\frac{V_0}{f_i M} \left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/2} + \frac{\theta_o^2}{M^2} \frac{\hat{\phi}_o}{\hat{\phi}_i} \right\}^2 \left\{ \lim_{z \to \infty} (z_{Fi} - z) - f_i M \right\} \tag{25.11d}$$

The reason for introducing z_{Fo} and z_{Fi} will become clearer below but we can anticipate that the convergence of integrals involving G'^4 and $\overline{G}'^4$ will be easier to understand in combination with the limit terms in Eq. (25.11c,d).

Substituting for $x(z)$ and $y(z)$ in $M^{(4)}$, we find

$$\frac{S^I_{oi}}{\hat{\phi}_o^{1/2}} = -\begin{pmatrix} u_o^{*2} \\ u_o^* u_o^{*\prime} \\ u_o^{*\prime 2} \end{pmatrix}^T \begin{pmatrix} E/4 & (D - \mathrm{i}d)/2 & (A - \mathrm{i}a)/4 \\ (D + \mathrm{i}d)/2 & F/2 & (K - \mathrm{i}k)/2 \\ (A + \mathrm{i}a)/4 & (K + \mathrm{i}k)/2 & C/4 \end{pmatrix} \begin{pmatrix} u_o^2 \\ u_o u'_o \\ u'^2_o \end{pmatrix} \tag{25.12}$$

as in Eq. (24.33) or (24.34). The nine magnification-dependent quantities A, C, ... k can now be expressed in terms of nine integrals, six for the isotropic aberration coefficients (i_1–i_6) and three for the anisotropic coefficients (i_7–i_9), that are *independent of the magnification*: they are properties of the lens, determined by its geometry and excitation but unaffected by the working conditions. The dependence of the coefficients A, C, ... k on magnification is expressed by a set of polynomials in reciprocal magnification m in which the coefficients of the various powers of m are simple multiples of the integrals i_j ($j = 1$–9).

25.1 Spherical Aberration

$$C = C_4 m^4 + C_3 m^3 + C_2 m^2 + C_1 m + C_0 \tag{25.13}$$

$$\left.\begin{aligned} C_4 &= i_1 f_o^4 \\ C_3 &= -4 i_2 f_o^4 - \frac{1}{2} r^2 f_o \\ C_2 &= 2(i_3 + i_4) f_o^4 \\ C_1 &= -4 i_5 f_o^4 - \frac{1}{2} f_o \\ C_o &= i_6 f_o^4 \end{aligned}\right\} \tag{25.14}$$

25.2 Coma

$$K = K_3 m^3 + K_2 m^2 + K_1 m + K_0 \tag{25.15}$$

$$\left.\begin{aligned} K_3 &= -i_1 f_o^3 \\ K_2 &= 3 i_2 f_o^3 + \frac{1}{2} r^2 \\ K_1 &= -(i_3 + i_4) f_o^3 \\ K_0 &= i_5 f_o^3 \end{aligned}\right\} \tag{25.16}$$

$$k = k_2 m^2 + k_1 m + k_0 \tag{25.17}$$

$$\left.\begin{aligned} k_2 &= i_7 f_o^2 \\ k_1 &= -i_8 f_o^2 \\ k_0 &= i_9 f_o^2 \end{aligned}\right\} \tag{25.18}$$

25.3 Astigmatism and Field Curvature

$$A = A_2 m^2 + A_1 m + A_0 \tag{25.19}$$

$$F = F_2 m^2 + F_1 m + F_0 \tag{25.20}$$

$$\left.\begin{aligned} A_2 &= \frac{1}{2}F_2 = i_1 f_o^2 \\ A_1 &= \frac{1}{2}F_1 = -2i_2 f_o^2 - \frac{r^2}{2f_o} \\ A_0 &= i_3 f_o^2 \quad F_0 = i_4 f_o^2 \end{aligned}\right\} \tag{25.21}$$

$$a = a_1 m + a_0 \tag{25.22}$$

$$\left.\begin{aligned} a_1 &= -2i_7 f_o \\ a_0 &= i_8 f_o \end{aligned}\right\} \tag{25.23}$$

25.4 Distortion

$$D = D_1 m + D_0 \tag{25.24}$$

$$\left.\begin{aligned} D_1 &= -i_1 f_o \\ D_0 &= i_2 f_o + \frac{1}{2f_i^2} \end{aligned}\right\} \tag{25.25}$$

$$d = d_0 = i_7 \tag{25.26}$$

where

$$r := \left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/2} = \frac{f_o}{f_i} \tag{25.27}$$

25.5 Aberration Matrices and the Integrals i_j

More compactly

$$\begin{pmatrix} C \\ K \\ A \\ F \\ D \end{pmatrix} = Q \begin{pmatrix} m^4 \\ m^3 \\ m^2 \\ m \\ 1 \end{pmatrix} \qquad \begin{pmatrix} k \\ a \\ d \end{pmatrix} = q \begin{pmatrix} m^2 \\ m \\ 1 \end{pmatrix} \tag{25.28}$$

with

$$Q = \begin{pmatrix} i_1 f_o^4 & -4i_2 f_o^4 - r^2 f_o/2 & 2(i_3 + i_4) f_o^4 & -4i_5 f_o^4 - f_o/2 & i_6 f_o^4 \\ 0 & -i_1 f_o^3 & 3i_2 f_o^3 + r^2/2 & -(i_3 + i_4) f_0^3 & i_5 f_o^3 \\ 0 & 0 & i_1 f_o^2 & -2i_2 f_o^2 - r/2f_o & i_3 f_o^2 \\ 0 & 0 & 2i_1 f_o^2 & -4i_2 f_o^2 - r/f_o & i_4 f_o^2 \\ 0 & 0 & 0 & -i_1 f_o & i_2 f_o + 1/2f_i^2 \end{pmatrix} \tag{25.29}$$

and

$$q = \begin{pmatrix} i_7 f_o^2 & -i_8 f_o^2 & i_9 f_o^2 \\ 0 & -2i_7 f_o & i_8 f_o \\ 0 & 0 & i_7 \end{pmatrix} \tag{25.30}$$

The integrals $i_1 - i_9$ can as usual be cast into a host of different forms, which can be established without difficulty by the methods explained in Section 24.3. We list here the primitive form, on which no partial integration has been performed, and two particularly simple forms, one for magnetic lenses ($\phi = \text{const}$) and the other for electrostatic lenses ($B = 0$). Another set of (nonrelativistic) integrals for electrostatic lenses, well suited for numerical work, is to be found in Kuyatt et al. (1974).

$$i_1 = \frac{1}{\hat{\phi}_o^{1/2}} \int_{-\infty}^{\infty} \left(L_1 G^4 + 2L_2 G^2 G'^2 + L_3 G'^4\right) dz + \frac{1}{2f_i^3 f_o} \lim_{z \to \infty} (z_{Fi} - z)$$

$$\underset{\varphi = \text{const}}{\to} \int_{-\infty}^{\infty} \Lambda_m G^4 \, dz$$

$$\underset{B = 0}{\to} \int_{-\infty}^{\infty} \Lambda_e G^4 \, dz$$

$$i_2 = \frac{1}{\hat{\phi}_o^{1/2}} \int_{-\infty}^{\infty} \left\{ L_1 G^3 \overline{G} + L_2 G G' (G\overline{G})' + L_3 G'^3 \overline{G}' \right\} dz$$

$$\underset{\varphi = \text{const}}{\to} \int_{-\infty}^{\infty} \Lambda_m G^3 \overline{G} dz - \frac{1}{8f^3}$$

$$\underset{B = 0}{\to} \int_{-\infty}^{\infty} \Lambda_e G^3 \overline{G} dz - \frac{1}{8 f_o f_i^2}$$

$$i_3 = \frac{1}{\hat{\phi}_o^{1/2}} \int_{-\infty}^{\infty} \left(L_1 G^2 \overline{G}^2 + 2L_2 G G' \overline{G}\, \overline{G}' + L_3 G'^2 \overline{G}'^2\right) dz - \frac{1}{8f_o^2} \int_{-\infty}^{\infty} \left(\frac{\hat{\phi}_o}{\hat{\phi}}\right)^{1/2} \frac{\eta^2 B^2}{\hat{\phi}} dz$$

$$\underset{\varphi = \text{const}}{\to} \int_{-\infty}^{\infty} \Lambda_m G^2 \overline{G}^2 - \frac{1}{6f^2} \int_{-\infty}^{\infty} \frac{\eta^2 B^2}{\hat{\phi}} \, dz$$

$$\underset{B - 0}{\to} \int_{-\infty}^{\infty} \Lambda_e G^2 \overline{G}^2 - \frac{\hat{\phi}_o^{1/2}}{24 f_o^2} \int_{-\infty}^{\infty} \frac{\gamma \phi'''}{\hat{\phi}^{3/2}} \, dz$$

$$i_4 = \frac{1}{\hat{\phi}_o^{1/2}} \int_{-\infty}^{\infty} \left\{ 2L_1 G^2 \overline{G}^2 + L_2 (G\,\overline{G})'^2 + 2L_3 G'^2 \overline{G}'^2 \right\} dz$$

$$+ \frac{1}{8f_o^2} \int_{-\infty}^{\infty} \left(\frac{\hat{\phi}_o}{\hat{\phi}} \right)^{1/2} \frac{\eta^2 B^2}{\hat{\phi}} dz$$

$$\underset{\phi = \text{const}}{\rightarrow} 2 \int_{-\infty}^{\infty} \Lambda_m G^2 \overline{G}^2 + \frac{1}{6f^2} \int_{-\infty}^{\infty} \frac{\eta^2 B^2}{\hat{\phi}}\, dz$$

$$\underset{B = 0}{\rightarrow} 2 \int_{-\infty}^{\infty} \Lambda_e G^2 \overline{G}^2 + \frac{\hat{\varphi}_o^{1/2}}{24 f_o^2} \int_{-\infty}^{\infty} \frac{\gamma \varphi''}{\hat{\phi}^{3/2}}\, dz$$

$$i_5 = \frac{1}{\hat{\phi}_o^{1/2}} \int_{-\infty}^{\infty} \left\{ L_1 G \overline{G}^3 + L_2 (G\overline{G})' \overline{G}\, \overline{G}' + L_3 G' \overline{G}'^3 \right\} dz$$

$$\underset{\phi = \text{const}}{\rightarrow} \int_{-\infty}^{\infty} \Lambda_m G\, \overline{G}^3\, dz - \frac{1}{8f^3}$$

$$\underset{B = 0}{\rightarrow} \int_{-\infty}^{\infty} \Lambda_e G \overline{G}^3\, dz - \frac{1}{8f_o^3}$$

$$i_6 = \frac{1}{\hat{\phi}_o^{1/2}} \int_{-\infty}^{\infty} \left(L_1 \overline{G}^4 + 2L_2 \overline{G}^2\, \overline{G}'^2 + L_3 \overline{G}'^4 \right) dz$$

$$+ \frac{1}{2f_o^4} \lim_{z \rightarrow -\infty} (z - z_{Fo})$$

$$\underset{\phi = \text{const}}{\rightarrow} \int_{-\infty}^{\infty} \Lambda_m \overline{G}^4\, dz$$

$$\underset{B = 0}{\rightarrow} \int_{-\infty}^{\infty} \Lambda_e \overline{G}^4\, dz$$

$$i_7 = \int_{-\infty}^{\infty} \left(PG^2 + QG'^2 \right) dz$$

$$\underset{\phi = \text{const}}{\rightarrow} \int_{-\infty}^{\infty} \Lambda_a G^2\, dz$$

$$
\begin{aligned}
i_8 &= 2\int_{-\infty}^{\infty} \left(PG\overline{G} + QG'\overline{G}'\right) dz \\
&\underset{\phi=\text{const}}{\rightarrow} 2\int_{-\infty}^{\infty} \Lambda_a G\overline{G}\, dz \\
i_9 &= 2\int_{-\infty}^{\infty} \left(P\overline{G}^2 + Q\overline{G}'^2\right) dz \\
&\underset{\phi=\text{const}}{\rightarrow} \int_{-\infty}^{\infty} \Lambda_a \overline{G}^2\, dz
\end{aligned}
\tag{25.31}
$$

In these integrals, the functions L_1, L_2, L_3, P and Q are given by Eq. (24.3) and Λ_m, Λ_e and Λ_a by

$$
\begin{aligned}
\Lambda_m &:= \frac{1}{48}\left(\frac{4\eta^4 B^4}{\hat{\phi}^2} + \frac{5\eta^2 B'^2}{\hat{\phi}} - \frac{\eta^2 BB''}{\hat{\phi}}\right) \\
\Lambda_e &:= \frac{1}{192\hat{\phi}_o^{1/2}}\left\{4\left(3+5\varepsilon\hat{\phi}\right)\frac{\phi''^2}{\hat{\phi}^{3/2}} - (3+4\varepsilon\hat{\phi})\frac{\phi'\phi'''}{\hat{\phi}^{3/2}} - 30\left(1+\varepsilon\hat{\phi}\right)\frac{\gamma\phi'^2\phi''}{\hat{\phi}^{5/2}} + 2\left(2\gamma^4+11\gamma^2+2\right)\frac{\phi'^4}{\hat{\phi}^{7/2}}\right. \\
\Lambda_a &:= \frac{1}{16}\left(\frac{\eta B''}{\hat{\phi}^{1/2}} + \frac{2\eta^3 B^3}{\hat{\phi}^{3/2}}\right)
\end{aligned}
\tag{25.32}
$$

(For Λ_m, see Jandeleit and Lenz (1959) and for Λ_e, see Hawkes (1985b).)

For symmetric lenses (or lens combinations), $G(z) = \overline{G}(-z)$ if the origin is set in the symmetry plane, and hence

$$
i_1 = i_6 \quad i_2 = i_5 \quad i_7 = i_9 \tag{25.33}
$$

The asymptotic Petzval coefficient is given by $F - 2A$, which is simply $F_o - 2A_o$ (25.21) or $f_o^2(i_4 - 2i_3)$. Thus

$$
\begin{aligned}
F - 2A &= \int_{-\infty}^{\infty} \left(\frac{\hat{\phi}_o}{\hat{\phi}}\right)^{1/2} \frac{\gamma\phi'' + 4\eta^2 B^2}{8\hat{\phi}} dz \\
&\underset{\phi=\text{const}}{\rightarrow} \int_{-\infty}^{\infty} \frac{\eta^2 B^2}{2\hat{\phi}}\, dz \\
&\underset{B=0}{\rightarrow} \int_{-\infty}^{\infty} \left(\frac{\hat{\phi}_o}{\hat{\phi}}\right)^{1/2} \frac{\gamma\phi''}{8\hat{\phi}}\, dz
\end{aligned}
\tag{25.34}
$$

25.6 Dependence on Object Position or Magnification

Expressions (25.13−25.28) for the aberration coefficients as polynomials in reciprocal magnification m are convenient in that any symmetries in the lens are immediately reflected in the coefficients. For design purposes, it may be convenient to replace m by the object (or image) position, which is easily done with the aid of Eq. (16.25). Again, we might wish to know the magnitudes of the coefficients for a specific value of the magnification (or object position), given those for some other value. The appropriate formulae have been derived from first principles by Ade (1973, 1982). They can also be read off from the polynomial expressions (25.28), see Hawkes (1984a). Thus suppose that C, K, A, F and D are known for some inverse magnification m and that the latter is altered to $m + \mu$, so that for the spherical aberration, for example,

$$C(m+\mu) = \sum_{j=0}^{4} C_j (m+\mu)^j$$

and likewise for the other coefficients. Clearly,

$$C(m+\mu) = \sum_{j=0}^{4} C'_j \mu^j \tag{25.35}$$

where

$$\begin{aligned} C'_4 &= C_4 \\ C'_3 &= 4C_4 m + C_3 \\ C'_2 &= 6C_4 m^2 + 3C_3 m + C_2 \\ C'_1 &= 4C_4 m^3 + 3C_3 m^2 + 2C_2 m + C_1 \\ C'_0 &= C(m) \end{aligned} \tag{25.36}$$

But

$$\begin{aligned} 4C_4 m + C_3 &= 4i_1 f_o^4 m - 4i_2 f_o^4 - \frac{1}{2} r^2 f_o \\ &= -4f_o^4 (i_2 - i_1 m) - \frac{1}{2} r^2 f_o \\ &= -4f_o^3 \left\{ D(m) - \frac{1}{2f_i^2} \right\} - \frac{1}{2} r^2 f_o \\ &= -4f_o^3 D(m) + \frac{3f_o^3}{2f_i^2} \end{aligned} \tag{25.37}$$

The remaining coefficients C'_i and their counterparts for all the other aberrations can likewise be expressed in terms of the known coefficients $A(m) \ldots k(m)$ plus one extra quantity, i_1, with the following results. Recalling that the change in inverse magnification μ corresponds to a shift $\Delta z_o = -f_o \mu$ in object position with the sign convention employed by Ade (1982), these expressions are identical with those derived by him.

$$\begin{pmatrix} C(m+\mu) \\ K(m+\mu) \\ A(m+\mu) \\ F(m+\mu) \\ D(m+\mu) \end{pmatrix} = Q_\mu \begin{pmatrix} \mu^4 \\ \mu^3 \\ \mu^2 \\ \mu \\ 1 \end{pmatrix} \tag{25.38a}$$

$$\begin{pmatrix} k(m+\mu) \\ a(m+\mu) \\ d(m+\mu) \end{pmatrix} = q_\mu \begin{pmatrix} \mu^2 \\ \mu \\ 1 \end{pmatrix} \tag{25.38b}$$

where the elements of Q_μ are as follows:

$$\begin{aligned}
&Q_{\mu 11} = f_o^4 i_1, \quad Q_{\mu 12} = -4f_o^3 D + 3r^2 f_o/2 \\
&Q_{\mu 13} = 2f_o^2(A+F) + 3r^2 f_o m/2, \quad Q_{\mu 14} = -4f_o K + f_o\left(r^2 m^2 - 1\right)/2 \\
&Q_{\mu 15} = C_\mu, \quad Q_{\mu 22} = -f_o^3 i_1, \quad Q_{\mu 23} = 3f_o^2 D - r^2 \\
&Q_{\mu 24} = -f_o(A+F) - r^2 m/2, \quad Q_{\mu 25} = K, \quad Q_{\mu 33} = f_o^2 i_1 \\
&Q_{\mu 34} = -2f_o D + r^2/2f_o, \quad Q_{\mu 35} = A, \quad Q_{\mu 43} = 2f_o^2 i_1 \\
&Q_{\mu 44} = -4f_o D + r^2/f_o, \quad Q_{\mu 45} = F, \quad Q_{\mu 54} = -f_o i_1, \quad Q_{\mu 55} = D \\
&Q_{\mu 21} = Q_{\mu 31} = Q_{\mu 32} = Q_{\mu 41} = Q_{\mu 42} = Q_{\mu 51} = Q_{\mu 52} = Q_{\mu 53} = 0
\end{aligned} \tag{25.39a}$$

and

$$q_\mu = \begin{pmatrix} f_o^2 & -f_o a & k \\ 0 & -2f_o d & a \\ 0 & 0 & d \end{pmatrix} \tag{25.39b}$$

25.7 Dependence on Aperture Position

One other situation remains to be examined[1]: hitherto the aberrations have been expressed in terms of position and gradient in the asymptotic object plane but, in practice, the effect of the real aperture will often be important; in projectors, for example, where the dominant geometrical aberration is distortion, the sign of the coefficient can be changed by varying the aperture position, since the aperture selects different pencils of rays from the beam at the asymptotic object as its position is altered. We therefore consider the situation illustrated in Fig. 25.2, which shows that we shall be concerned with the aberrations of the asymptotic image of an asymptotic object with boundary conditions requiring aberrations of position to vanish in two planes (cf. 24.14, 24.18): the asymptotic aberrations must vanish in the asymptotic object plane and the real aberrations in the real aperture plane. We shall,

[1]The following account is confined to magnetic lenses; it can be extended straightforwardly to electrostatic lenses if needed.

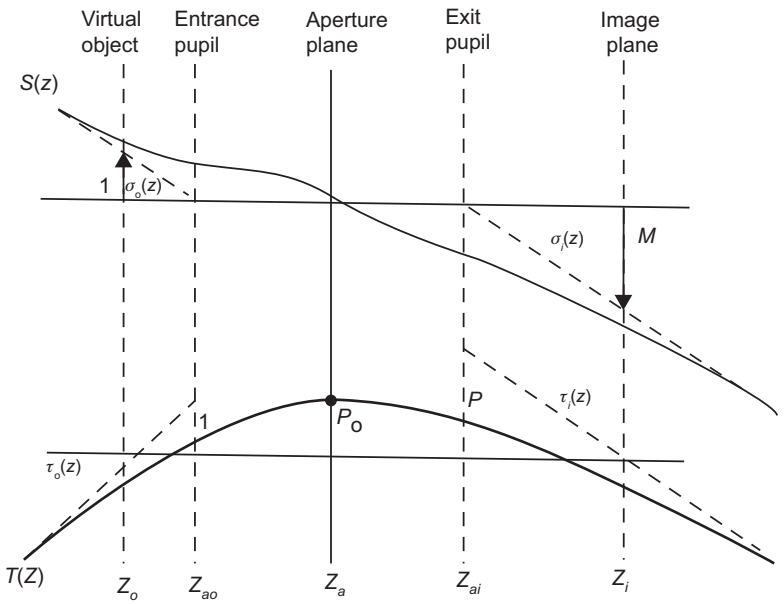


Figure 25.2
The rays $S(z)$ and $T(z)$ have asymptotes $\sigma_o(z)$ and $\tau_o(z)$ in object space and $\sigma_i(z)$ and $\tau(z_i)$ in image space. The object–image magnification M and the pupil magnification N are defined in terms of these asymptotes.

however, find it useful to introduce the *entrance* and *exit pupils* associated with the real aperture; these are the asymptotic images of the latter formed by the part of the lens field before the real aperture (entrance pupil) and the part beyond it (exit pupil). This in turn introduces the important notion of pupil magnification, the magnification between these two pupil planes, which are (asymptotically) conjugate.

Fig. 25.2 shows two paraxial solutions, analogous to $s(z)$ and $t(z)$, which we here denote $S(z)$ and $T(z)$; these satisfy the conditions

$$\lim_{z\to-\infty} S(z) =: \sigma_o(z) = \sigma_o'(z - z_{ao}) \tag{25.40a}$$

$$\lim_{z\to\infty} S(z) =: \sigma_i(z) = \sigma_i'(z - z_{oi}) \tag{25.40b}$$

$$\lim_{z\to-\infty} T(z) =: \tau_o(z) = \tau_o'(z - z_o) \tag{25.40c}$$

$$\lim_{z\to-\infty} T(z) =: \tau_i(z) = \tau_i'(z - z_i) \tag{25.40d}$$

and

$$\begin{aligned} S(z_a) &= 0 \\ \sigma(z_o) &= 1 \end{aligned} \tag{25.41}$$

Clearly, $z = z_{ao}$ and $z = z_{ai}$ are the entrance and exit pupil planes, respectively. From the Wronskian, we see that

$$\tau'_o = -S'_a T_a = M\tau'_i = -\sigma'_o \tau_{ao} \tag{25.42}$$

It is convenient to denote the pupil magnification by N, and we write

$$N = N_o N_i \tag{25.43}$$

where N_o is the magnification between the entrance pupil and the real aperture and N_i that between the latter and the exit pupil:

$$\frac{S'_a}{\sigma'_o} = \frac{1}{N_o} \qquad \frac{\sigma'_i}{S'_a} = \frac{1}{N_i} \tag{25.44}$$

We have not yet fully specified $T(z)$, for we see from Eq. (25.40c) that the object asymptote τ_o intersects the entrance pupil at $\tau_o(z_{\alpha o}) = \tau'_o(z_{ao} - z_o)$ so that from Eq. (25.40–25.42)

$$\tau_{ao} = -\frac{\tau'_o}{\sigma'_o} = \frac{S'_a T_a}{\sigma'_o} \tag{25.45}$$

or

$$\tau_{ao} = \frac{T_a}{N_o} \tag{25.46}$$

By choosing

$$T(z_a) = N_o \tag{25.47}$$

so that

$$\tau_{ao} = 1 \tag{25.48}$$

we express the aberrations in terms of asymptotic object coordinates (in $z = z_o$) and entrance pupil coordinates (in $z = z_{ao}$). (If we had set $T(z_a) = 1$, $\tau_{ao} = 1/N_o$, the aberrations would have been expressed in terms not of entrance pupil coordinates but of real aperture coordinates, which is not so convenient; the important point is that the aberrations do actually vanish in $z = z_a$.) These relations and some simple consequences of them are listed in Table 25.1

Table 25.1: Relations between the asymptotes to $S(z)$ and $T(z)$ and various paraxial properties: pupil magnification, N; magnification, M; focal length, f; and others

$\sigma_o(z) = \sigma'_o(z - z_o) + 1 = \sigma'_o(z - z_{ao})$	$\sigma_i(z) = \sigma'_i(z - z_i) + M = \sigma'_i(z - z_{ai})$
$\tau_o(z) = \tau'_o(z - z_o) = \tau'_o(z - z_{ao}) + 1$	$\tau_i(z) = \tau'_i(z - z_i) = \tau'_i(z - z_{ai}) + N$
$\sigma'_o = -\dfrac{1}{z_{ao} - z_o} = \dfrac{1}{f(m - m_a)}$	$\sigma'_i = \dfrac{M}{z_i - z_{ai}} = \dfrac{\sigma'_o}{N}$
$\tau'_o = \dfrac{1}{z_{ao} - z_o} = -\dfrac{1}{f(m - m_a)}$	$\tau'_i = \dfrac{\tau'_o}{M} = \dfrac{N}{z_{ai} - z_i}$
$T_a = N_o = \dfrac{\sigma'_o}{S'_a}$	$N_i = \dfrac{S'_a}{\sigma'_i}$
$z_i - z_{Fi} = -fM$	$z_o - z_{Fo} = \dfrac{f}{M} = fm$
$N = N_o N_i$	

The general paraxial solution is now

$$\begin{aligned} x(z) &= x_o S(z) + x_a T(z) \\ y(z) &= y_o S(z) + y_a T(z) \end{aligned} \tag{25.49}$$

in which x_a, y_a are position coordinates in the entrance pupil plane. (They should strictly speaking be denoted x_{ao}, y_{ao}, but we drop the second suffix since no confusion should arise.) With the boundary conditions already explained, the asymptotic aberrations at the image plane z_i are given by

$$\begin{aligned} \Delta x_i &= -\frac{1}{\tau'_o}\frac{\partial S^I_{oi}}{\partial x_a} \\ \Delta y_i &= -\frac{1}{\tau'_o}\frac{\partial S^I_{oi}}{\partial y_a} \end{aligned} \tag{25.50}$$

and the various coefficients can hence be written down immediately. Our purpose here is to enquire how they vary when the (real) aperture and (asymptotic) object are shifted, and how this variation compares with the simple polynomial forms obtained earlier.

Let us suppose that the image magnification is initially M_1 and the pupil magnification N_1; we add the label 1 to the other relevant quantities, S_1, T_1 in particular. After moving object and aperture, the magnifications become M_2 and N_2 and the new fundamental paraxial solutions, S_2 and T_2, are linear combinations of their predecessors,

$$\begin{aligned} S_2 &= \lambda_s S_1 + \lambda_t T_1 \\ T_2 &= \mu_s S_1 + \mu_t T_1 \end{aligned} \tag{25.51}$$

By considering the relations between the asymptotes to these rays and the object and aperture positions (Table 25.1), it is easily seen that

$$\begin{aligned}\lambda_s &= \frac{z_{o1}-z_{ao2}}{z_{o2}-z_{ao2}} = \frac{m_1-n_2}{m_2-n_2}\\ \lambda_t &= \frac{z_{ao1}-z_{ao2}}{z_{o2}-z_{ao2}} = \frac{n_1-n_2}{m_2-n_2}\\ \mu_s &= \frac{z_{o2}-z_{o1}}{z_{o2}-z_{ao2}} = \frac{m_2-m_1}{m_2-n_2}\\ \mu_t &= \frac{z_{o2}-z_{ao1}}{z_{o2}-z_{ao2}} = \frac{m_2-n_1}{m_2-n_2}\end{aligned} \tag{25.52}$$

The individual aberration coefficients can be expressed in terms of the integrals

$$\begin{aligned}I_{ij} &\coloneqq \frac{1}{\tau'_o}\int_{-\infty}^{\infty} \Lambda_m(z) S^i T^j\, dz \quad i+j=4,\ \ i,j=0-4\\ I_{00} &\coloneqq \frac{\eta^2}{\hat{\phi}}\int_{-\infty}^{\infty} B^2(z)\, dz\end{aligned} \tag{25.53}$$

in which Eq. (25.32)

$$\Lambda_m(z) \coloneqq \frac{\eta^4 B^4}{12\hat{\phi}^2} + \frac{5}{48}\frac{\eta^2 B'^2}{\hat{\phi}} - \frac{\eta^2 B B''}{48\hat{\phi}} \tag{25.54}$$

and

$$J_{ij} \coloneqq \int_{\infty}^{\infty} \Lambda_a(z) S^i T^j\, dz \quad i+j=2\ \ i,j=0-2 \tag{25.55}$$

in which Eq. (25.32)

$$\Lambda_a(z) \coloneqq \frac{1}{16}\left(\frac{\eta B''}{\hat{\phi}^{1/2}} + \frac{2\eta^3 B^3}{\hat{\phi}^{3/2}}\right) \tag{25.56}$$

For $m = m_1$ and $n = n_1$, we have

$$\begin{aligned}C_1 &= I_{04}^{(1)}\\ K_1 &= I_{13}^{(1)} - \left(1-m_1^2\right)/8f^2(m_1-n_1)^2 && k_1 = J_{02}\\ A_1 &= I_{22}^{(1)} - \frac{1}{6}\tau'_{1o} I_{00} + (1-m_1 n_1)/4f^2(m_1-n_1)^2 && a_1 = 2J_{11}\\ F_1 &= 2I_{22}^{(1)} + \frac{1}{6}\tau'_{1o} I_{00} + (1-m_1 n_1)/2f^2(m_1-n_1)^2\\ D_1 &= I_{31}^{(1)} - 3\left(1-n_1^2\right)/8f^2(m_1-n_1)^2 && d_1 = J_{20}\end{aligned} \tag{25.57}$$

For any other pair of reciprocal magnifications m_2, n_2, similar expressions relate C_2, K_2, … d_2 to $I_{ij}^{(2)}$ and $J_{ij}^{(2)}$. Using Eq. (25.51), it is easy to express the integrals $I_{ij}^{(2)}$ and $J_{ij}^{(2)}$ in terms of $I_{ij}^{(1)}$ and $J_{ij}^{(1)}$ and hence $C_2 \ldots d_2$ in terms of $C_1, \ldots d_1$. The results are as follows.

$$\begin{pmatrix} I_{40}^{(2)} \\ I_{31}^{(2)} \\ I_{22}^{(2)} \\ I_{13}^{(2)} \\ I_{04}^{(2)} \end{pmatrix} = I \begin{pmatrix} I_{40}^{(1)} \\ I_{31}^{(1)} \\ I_{22}^{(1)} \\ I_{13}^{(1)} \\ I_{04}^{(1)} \end{pmatrix} \qquad \begin{pmatrix} J_{20}^{(2)} \\ J_{11}^{(2)} \\ J_{02}^{(2)} \end{pmatrix} = J \begin{pmatrix} J_{20}^{(1)} \\ J_{11}^{(1)} \\ J_{02}^{(1)} \end{pmatrix} \tag{25.58}$$

where

$$I = \frac{m_1 - n_1}{m_2 - n_2} I^*$$

and

$$\begin{aligned}
&I_{11}^* = \lambda_s^4,\ I_{12}^* = 4\lambda_s^3\lambda_t,\ I_{13}^* = 6\lambda_s^2\lambda_t^2,\ I_{14}^* = 4\lambda_s\lambda_t^3,\ I_{15}^* = \lambda_\tau^4 \\
&I_{21}^* = \lambda_s^3\mu_s,\ I_{22}^* = 3\lambda_s^2\lambda_t\mu_s + \lambda_s^3\mu_t,\ I_{23}^* = 3\lambda_s\lambda_t^2\mu_s + 3\lambda_s^2\lambda_t\mu_t,\ I_{24}^* = \lambda_t^3\mu_s + 3\lambda_s\lambda_t^2\mu_t,\ I_{25}^* = \lambda_t^3\mu_t \\
&I_{31}^* = \lambda_s^2\mu_s^2,\ I_{32}^* = 2\left(\lambda_s^2\mu_s\mu_t + \lambda_s\lambda_t\mu_s^2\right),\ I_{33}^* = \lambda_s^2\mu_t^2 + \mu_s^2\lambda_t^2 + 4\lambda_s\lambda_t\mu_s\mu_t \\
&I_{34}^* = 2\left(\lambda_s\lambda_t\mu_t^2 + \mu_s\mu_t\lambda_t^2\right),\ I_{35}^* = \lambda_t^2\mu_t^2 \\
&I_{41}^* = \lambda_s\mu_s^3,\ I_{42}^* = 3\lambda_s\mu_s^2\mu_t + \lambda_t\mu_s^3,\ I_{43}^* = 3\left(\lambda_t\mu_s^2\mu_t + \lambda_s\mu_s\mu_t^2\right) \\
&I_{44}^* = \lambda_s\mu_t^3 + 3\lambda_t\mu_s\mu_t^2,\ I_{45}^* = \lambda_t\mu_t^3 \\
&I_{51}^* = \mu_s^4,\ I_{52}^* = 4\mu_s^3\mu_t,\ I_{53}^* = 6\mu_s^2\mu_t^2,\ I_{54}^* = 4\mu_s\mu_t^3,\ I_{55}^* = \mu_t^4
\end{aligned} \tag{25.59}$$

$$J = \begin{pmatrix} \lambda_s^2 & 2\lambda_s\lambda_t & \lambda_t^2 \\ \lambda_s\mu_s & \lambda_s\mu_t + \lambda_t\mu_s & \lambda_t\mu_t \\ \mu_s^2 & 2\mu_s\mu_t & \mu_t^2 \end{pmatrix} \tag{25.60}$$

This question, the variation of the aberrations when object and aperture (or stop) are shifted, has a long history in light optics, and we draw attention to the work that is closest to the contents of this chapter. The earliest thorough exploration, using the eikonal or characteristic function, is that of T. Smith (1921/22) but owing to the extreme generality of the results, and to the fact that the latter were not expressed in terms of familiar quantities, his findings were frequently referred to but little used. For a more recent account, see Pegis (1961) The problem was solved in language of direct use to the lens designer by a series of members of van Heel's laboratory in Delft (Korringa, 1942; Stephan, 1947; Brouwer, 1957). The full matrix treatment developed by Brouwer renders his work easy to apply and we can derive Eqs (25.59–25.60) straightforwardly by his methods. This requires the notion of aberration matrices, however, and we therefore postpone further discussion to Chapter 27, Aberration Matrices and the

Aberrations of Lens Combinations, where these are introduced. For a very scholarly account of the various formulations of aberration theory with many more references, see Focke (1965) and for a more recent development, Velzel (1987) and Velzel and de Meijere (1988).

25.8 Thin-Lens Approximations

The computations by Renau and Heddle (1986b) of the aberration coefficients of accelerating two-cylinder electrostatic lenses suggested that these are, to a good approximation, related in a simple way. Hawkes (1987) showed that these relations emerge naturally from a study of the thin-lens approximation. Explicitly, Renau and Heddle showed that the quantities

$$\begin{aligned} m_{13} &= -\left(f_o^4/f_i\right)i_6 & m_{23} &= -f_o^3 i_5 - 1/2 \\ m_{14} &= -3f_o^3 i_5 & m_{24} &= -f_o^2 f_i(i_3 + i_4) \\ m_{15} &= -f_o^2 f_i(i_3 + i_4) & m_{25} &= -3f_o f_i^2 i_2 \\ m_{16} &= -f_o f_i^2 i_2 - 1/2 & m_{26} &= -4i_1 \end{aligned} \tag{25.61}$$

satisfy the relations

$$\frac{m_{2j}}{m_{1j}} = \left(\frac{\phi_i}{\phi_o}\right)^{1/4} \qquad j = 3,4,5,6 \tag{25.62}$$

remarkably well. This cannot be true in general but if we use (24.62) to derive thin-lens formulae for the i_j, the ratios do indeed show this behaviour. Setting $g = \phi^{1/4}G$ and $\overline{g} = \phi^{1/4}\overline{G}$ in the integrals for the i_j (25.31) and replacing g in the integrands by $\phi_o^{1/4}$ and $\overline{g}$ by $\phi_i^{1/4}$, we find

$$\begin{aligned} m_{13} &= \phi_i^{1/2} f_o^3 a & m_{23} &= \left(\phi_o\phi_i\right)^{1/4} f_i f_o^2 a - 3/8 \\ m_{14} &= 3\left(\phi_o\phi_i\right)^{1/4} f_i f_o^2 a + 3/8 & m_{24} &= 3\phi_o^{1/2} f_i^2 f_o a \\ m_{15} &= 3\phi_o^{1/2} f_i^2 f_o a & m_{25} &= 3\phi_o^{3/4}\phi_i^{-1/4} f_i^3 a + 3/8 \\ m_{16} &= \phi_o^{3/4}\phi_i^{-1/4} f_i^3 a - 3/8 & m_{26} &= \phi_o\phi_i^{-1/2}\left(f_i^4/f_o\right)a \end{aligned} \tag{25.63}$$

and so

$$\begin{aligned} \frac{m_{23}}{m_{13}} &= s - \frac{3}{8\phi_i^{1/2} f_o^3 a} = s - \frac{3}{8m_{13}} \\ \frac{m_{14}}{m_{24}} &= s^{-1} + \frac{1}{8\phi_i^{1/2} f_o^2 f_i a} \quad \text{or} \quad \frac{m_{24}}{m_{14}} = \frac{s}{1 + 3s/8m_{24}} \\ \frac{m_{25}}{m_{15}} &= s + \frac{1}{8\phi_o^{1/2} f_o f_i^2 a} = s + \frac{3}{8m_{15}} \\ \frac{m_{16}}{m_{26}} &= s^{-1} - \frac{3}{8\phi_i^{1/2} f_o f_i^2 a} \quad \text{or} \quad \frac{m_{26}}{m_{16}} = \frac{s}{1 - 3s/8m_{26}} \end{aligned} \tag{25.64}$$

in which $s = (\phi_i/\phi_o)^{1/4}$ and $a = \int (A_0/\phi)dz$. When the correction terms are negligible, we have $m_{2j}/m_{1j} \approx s$.

The thin-lens approximation also helps to explain Rempfer's mistaken claim that the aberrations of electrostatic lenses are characterized by fewer coefficients than is generally believed (Rempfer, 1985, repeated in Rempfer, 1999a). A note pointing out that her reasoning was faulty (Hawkes, 1999) was not accepted by Rempfer (1999b). After further exchanges (Rempfer, 2000, Hawkes, 2000), however, she was finally convinced by computations demonstrating that her claim was correct only in the thin-lens approximation (Hawkes and Lencová, 2002).

CHAPTER 26

Chromatic Aberrations

26.1 Real Chromatic Aberrations

Electron guns do not furnish rigorously monoenergetic beams of particles and the voltage and current supplies needed to excite electron lenses are never perfectly stable. The spread of energies in a beam is usually further broadened by its passage through a specimen, for inelastically scattered electrons will emerge with diminished energy. The initial energy spread is typically a few electronvolts for thermionic emitters and about 1 eV for field emitters; power supplies can be stabilized to within a few parts in a million (better than 1 V at 100 kV) and although the specimen energy losses may reach tens or hundreds of electronvolts for moderately thick specimens, the mean loss is often much smaller.

In a column of lenses, the image plane conjugate to some fixed plane will therefore inevitably fluctuate, retreating towards the fixed plane if the electron energy falls or the lens strength increases and *vice versa.* Since fluctuations of lens strength may be rapid and the energy of the beam will be spread over a finite if narrow range, these phenomena must be expected to blur the image. The resulting effects are known as the *chromatic aberrations* and are characterized by a set of *chromatic aberration coefficients.* The latter are conveniently calculated by perturbation theory, taking $M^{(2)}$ as the unperturbed refractive index and writing

$$M^{(P)} = \frac{\partial M^{(2)}}{\partial \phi}\Delta\phi + \frac{\partial M^{(2)}}{\partial B}\Delta B \tag{26.1}$$

From Eq. (15.24), we see that

$$\begin{aligned}\frac{\partial M^{(2)}}{\partial \phi} &= \frac{\phi''}{16\hat{\phi}^{3/2}}\left(X^2+Y^2\right) + \frac{\gamma}{4\hat{\phi}^{1/2}}\left(X'^2+Y'^2\right)\\ \frac{\partial M^{(2)}}{\partial B} &= -\frac{1}{2}\eta(XY'-X'Y)\end{aligned} \tag{26.2}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00026-2
© 2018 Elsevier Ltd. All rights reserved.

so that introducing rotating coordinates (15.9),

$$
\begin{aligned}
M^{(c)} &:= M^{(P)} = A_1\left(x^2+y^2\right) + A_2\left(x'^2+y'^2\right) + A_3(xy'-x'y)\\
A_1 &= \frac{\phi'' + \gamma\eta^2B^2}{16\hat{\phi}^{3/2}}\Delta\phi - \frac{\eta^2 B}{4\hat{\phi}^{1/2}}\Delta B\\
A_2 &= \frac{\gamma}{4\hat{\phi}^{1/2}}\Delta\phi\\
A_3 &= \frac{1}{4}\eta B\left(\frac{\gamma}{\hat{\phi}}\Delta\phi - 2\frac{\Delta B}{B}\right)
\end{aligned}
\tag{26.3}
$$

With $x = x_o s + x_a t$, $y = y_o s + y_a t$, we find

$$
\begin{aligned}
M^{(c)} &= a_1\left(x_o^2+y_o^2\right) + a_2\left(x_a^2+y_a^2\right) + a_3(x_o x_a + y_o y_a) + a_4(x_o y_a - x_a y_o)\\
a_1 &= A_1 s^2 + A_2 s'^2\\
a_2 &= A_1 t^2 + A_2 t'^2\\
a_3 &= 2(A_1 st + A_2 s't')\\
a_4 &= A_3(st' - s't)
\end{aligned}
\tag{26.4}
$$

Using Eq. (24.29), we obtain the following expressions in a general plane:

$$
\begin{aligned}
x^{(c)} &= x_o s + x_a t + \frac{t}{W_s}\left(2x_o\int_{z_a} a_1 dz + x_a\int_{z_a} a_3 dz + y_a\int_{z_a} a_4 dz\right) - \frac{s}{W_s}\left(2x_a\int_{z_o} a_2 dz + x_o\int_{z_o} a_3 dz - y_o\int_{z_o} a_4 dz\right)\\
y^{(c)} &= y_o s + y_a t + \frac{t}{W_s}\left(2y_o\int_{z_a} a_1 dz + y_a\int_{z_o} a_3 dz - x_a\int_{z_a} a_4 dz\right) - \frac{s}{W_s}\left(2y_a\int_{z_o} a_2 dz + y_o\int_{z_o} a_3 dz + x_o\int_{z_o} a_4 dz\right)
\end{aligned}
\tag{26.5}
$$

and in the image plane

$$
\begin{aligned}
\Delta x^{(c)} &:= \frac{x_i^{(c)} - Mx_o}{M} = -\frac{1}{W_s}\left(2x_a\int_{z_o}^{z_i} a_2 dz + x_o\int_{z_o}^{z_i} a_3 dz - y_o\int_{z_o}^{z_i} a_4 dz\right)\\
\Delta y^{(c)} &:= \frac{y_i^{(c)} - My_o}{M} = -\frac{1}{W_s}\left(2y_a\int_{z_o}^{z_i} a_2 dz + y_o\int_{z_o}^{z_i} a_3 dz + x_o\int_{z_o}^{z_i} a_4 dz\right)\\
\Delta u^{(c)} &:= \Delta x(c) + \mathrm{i}\Delta y^{(c)} = -\frac{1}{W_s}\left(2u_a\int_{z_o}^{z_i} a_2 dz + u_o\int_{z_o}^{z_i} a_3 dz + iu_o\int_{z_o}^{z_i} a_4 dz\right)
\end{aligned}
\tag{26.6}
$$

The integrals may be rewritten in the convenient forms

$$\begin{aligned}
\int_{z_o}^{z_i} a_2\, dz &= \int_{z_o}^{z_i} \Delta\phi \left\{ \frac{\gamma\eta^2 B^2}{8\hat{\phi}^{3/2}} + \frac{\gamma\phi'^2\left(3+2\varepsilon\hat{\phi}\right)}{16\hat{\phi}^{5/2}} \right\} t^2 dz - \int_{z_o}^{z_i} \Delta B \frac{\eta^2 B}{4\hat{\phi}^{1/2}} t^2 dz \\
\int_{z_o}^{z_i} a_3\, dz &= \int_{z_o}^{z_i} \gamma\Delta\phi \left\{ \frac{\eta^2 B^2}{4\hat{\phi}^{3/2}} + \frac{\phi'^2\left(3+2\varepsilon\hat{\phi}\right)}{8\hat{\phi}^{5/2}} \right\} st\, dz - \int_{z_o}^{z_i} \Delta B \frac{\eta^2 B}{2\hat{\phi}^{1/2}} st\, dz + \Delta\phi \left(\frac{\gamma_i s_i t_i'}{\hat{\phi}_i^{1/2}} - \frac{\gamma_o t_o'}{\hat{\phi}_o^{1/2}} \right) \\
\int_{z_o}^{z_i} a_4\, dz &= \hat{\phi}^{1/2}(st' - s't) \int_{z_o}^{z_i} \frac{\eta B}{4\hat{\phi}^{1/2}} \left(\frac{\gamma}{\hat{\phi}} \Delta\phi - 2\frac{\Delta B}{B} \right) dz
\end{aligned} \tag{26.7}$$

In practice, we take $\Delta\phi$ to be a variation of the accelerating voltage ϕ_0, and ΔB a variation of the maximum magnetic field B_0. Thus if $B(z) = B_0 b(z)$, $\Delta B = b(z)\Delta B_0 = B\Delta B_0/B_0$, provided that $b(z)$ does not alter; this is true if the metal of the lens is not saturated. We can then apply Ampère's circuital theorem (Fig. 36.2), so that for a lens of traditional design, with a winding of N turns carrying a current I,

$$\int_{-\infty}^{\infty} B(z)\, dz = B_0 \int_{-\infty}^{\infty} b(z)\, dz = \mu_o NI \tag{26.8}$$

and so

$$\frac{\Delta B_0}{B_0} = \frac{\Delta(NI)}{NI} \tag{26.9}$$

In the magnetic case ($\phi \equiv \phi_0$), we write

$$\begin{aligned}
\Delta x^{(c)} &\coloneqq -\left(\tilde{C}_c x_a + \tilde{C}_D x_o - \tilde{C}_\theta y_o\right)\left(\frac{\gamma\Delta\phi_0}{\hat{\phi}_0} - 2\frac{\Delta B_0}{B_0} \right) \\
\Delta y^{(c)} &\coloneqq -\left(\tilde{C}_c y_a + \tilde{C}_D y_o + \tilde{C}_\theta x_o\right)\left(\frac{\gamma\Delta\phi_0}{\hat{\phi}_0} - 2\frac{\Delta B_0}{B_0} \right) \\
\Delta u^{(c)} &= -\left\{\tilde{C}_c u_a + \left(\tilde{C}_D + \mathrm{i}\tilde{C}_\theta\right) u_o\right\}\left(\frac{\gamma\Delta\phi_0}{\hat{\phi}_0} - 2\frac{\Delta B_0}{B_0} \right)
\end{aligned} \tag{26.10}$$

where

$$\begin{aligned}
\tilde{C}_c &= \frac{1}{t_o'} \int_{z_o}^{z_i} \frac{\eta^2 B^2 t^2}{4\hat{\phi}_0} dz \\
\tilde{C}_D &= \frac{1}{t_o'} \int_{z_o}^{z_i} \frac{\eta^2 B^2 st}{4\hat{\phi}_0} dz \\
\tilde{C}_\theta &= \int_{z_o}^{z_i} \frac{\eta B}{4\hat{\phi}_0^{1/2}} dz
\end{aligned} \tag{26.11}$$

In the electrostatic case, $\Delta B \equiv 0$ and we write

$$\Delta u^{(c)} = \Delta x^{(c)} + \mathrm{i}\Delta y^{(c)} = -\left(\tilde{C}_c u_a + \tilde{C}_D u_o\right)\left(\frac{\Delta\phi_0}{\hat{\phi}_0}\right) \tag{26.12}$$

where

$$\begin{aligned}\tilde{C}_c &= \frac{\hat{\phi}_0^{1/2}}{t_o'}\int_{z_o}^{z_i}\frac{\gamma\phi'^2\left(3+2\varepsilon\hat{\phi}\right)}{8\hat{\phi}^{5/2}}t^2\,dz \\ \tilde{C}_D &= \frac{\hat{\phi}_0^{1/2}}{t_o'}\int_{z_o}^{z_i}\frac{\gamma\phi'^2\left(3+2\varepsilon\hat{\phi}\right)}{8\hat{\phi}^{5/2}}st\,dz + \frac{\hat{\phi}_0}{4}\left(\frac{\gamma_i}{\hat{\phi}_i}-\frac{\gamma_o}{\hat{\phi}_0}\right)\end{aligned} \tag{26.13}$$

On measuring the variations in accelerating voltage or energy spread in terms not of ϕ_0 but of $\hat{\phi}_0$ in the magnetic case, so that $\Delta\hat{\phi}_0 = \gamma\Delta\phi_0$, and replacing x_a, y_a by x_o', y_o', we find the following expressions, which are the most common forms of the chromatic aberration coefficients:

$$\Delta u^{(c)} = -\left\{C_c u_o' + (C_D + \mathrm{i}C_\theta)u_o\right\}\Delta\left(\frac{\Delta\hat{\phi}_0}{\hat{\phi}_0} - 2\frac{\Delta B_0}{B_0}\right) \tag{26.14}$$

with

$$\begin{aligned}C_c &= \int_{z_o}^{z_i}\frac{\eta^2B^2}{4\hat{\phi}_0}h^2\,dz \\ C_D &= \int_{z_o}^{z_i}\frac{\eta^2B^2}{4\hat{\phi}_0}gh\,dz \\ C_\theta &= \int_{z_o}^{z_i}\frac{\eta B}{4\hat{\phi}_0^{1/2}}dz\end{aligned} \tag{26.15}$$

It is not convenient to use $\Delta\hat{\phi}_0/\hat{\phi}_0$ in the electrostatic case, and the counterparts of Eq. (26.14–26.15) are then

$$\Delta u^{(c)} = -\left(C_c u_o' + C_D u_o\right)\frac{\Delta\phi_0}{\hat{\phi}_0} \tag{26.16}$$

with

$$\begin{aligned}C_c &= \hat{\phi}_0^{1/2}\int_{z_o}^{z_i}\frac{\gamma\phi'^2\left(3+2\varepsilon\hat{\phi}\right)h^2}{8\hat{\phi}^{5/2}}dz \\ C_D &= \hat{\phi}_0^{1/2}\int_{z_o}^{z_i}\frac{\gamma\phi'^2\left(3+2\varepsilon\hat{\phi}\right)gh}{8\hat{\phi}^{5/2}}dz + \frac{\hat{\phi}_0}{4}\left(\frac{\gamma_i}{\hat{\phi}_i}-\frac{\gamma_o}{\hat{\phi}_o}\right)\end{aligned} \tag{26.17}$$

The chromatic aberrations are therefore of two kinds: an axial or 'aperture' aberration, characterized by C_c, which is invariably known simply as the *chromatic aberration coefficient*, and a distortion, characterized by $C_D + \mathrm{i}C_\theta$. The term in C_c, like the spherical aberration, creates an aberration disc, the radius of which in the image plane is the same for all points in the object plane. This radius, $M\Delta r^{(c)}$, is given by

$$M\Delta r^{(c)} = M\left(\Delta x^{(c)2} + \Delta y^{(c)2}\right)^{1/2}$$

or referred back to object space,

$$\Delta r^{(c)} = C_c\left(x_o'^{\,2} + y_o'^{\,2}\right)^{1/2}\Delta_c = C_c\theta_o\Delta_c$$

or

$$\Delta r^{(c)} = \tilde{C}_c\left(x_a^2 + y_a^2\right)^{1/2}\Delta_c = \tilde{C}_c r_a\Delta_c \tag{26.18}$$

where $\theta_o := (x_o'^{\,2} + y_o'^{\,2})^{1/2}$, $r_a := (x_a^2 + y_a^2)^{1/2}$ and Δ_c is the appropriate term in $\Delta\phi_0$ or $\Delta\hat{\phi}_0$ and ΔB_0. (Note that the radius of the aberration disc is the same for a variation of the form $\phi_0 \to \phi_0 + \Delta\phi_0$, and likewise for B_0, as it is for a variation $\phi_0 \to \phi_0 - \Delta\phi_0$. Some thought must therefore be given to the meaning of $\Delta\phi_0$ when calculating the size of the disc. Cf. Orloff, 1983b.) The current distribution in the aberration disc will be determined by the energy distribution as well as the energy spread. This point is examined by Barth and Nykerk (1999).

From Eqs (26.15, 26.17), we see that, again like C_s, C_c is always positive with the sign convention we have adopted (minus signs have been inserted in Eqs (26.14, 26.16) purely for convenience, since $\Delta\phi_0$ and ΔB_0 are usually oscillatory fluctuations about ϕ_0 and B_0 rather than steady drifts, though the latter are not of course excluded if present). This result too was first obtained by Scherzer (1936b). Methods of correcting or eliminating this chromatic aberration will be discussed at the same time as those for combating spherical aberration in Chapter 41 of Volume 2. In practice, however, it is more usual and certainly easier to reduce the energy spread of the incident beam by means of a monochromator than to incorporate a chromatic-aberration corrector in the instrument.

The chromatic aberration C_c is of most concern in objective lenses, where the angles are comparatively large, while the distortions characterized by C_D and C_θ mainly affect projectors, where the angles are small but the rays are farther from the axis. The isotropic chromatic distortion coefficient, C_D, is commonly known as the *chromatic aberration of magnification*, since we may write

$$u_i = M\left\{1 - C_D\left(\frac{\Delta\hat{\phi}_0}{\hat{\phi}_0} - 2\frac{\Delta B_0}{B_0}\right)\right\}u_o \tag{26.19}$$

in the magnetic case, and similarly for electrostatic lenses.

The *anisotropic chromatic distortion coefficient*, C_θ, is often known as the *chromatic aberration of rotation*, since it causes a small change in the image rotation in magnetic lenses. We note that C_θ is equal to half the image rotation Eq. (15.27).

A slightly different way of considering the chromatic aberrations yields forms for the coefficients that can save a lot of calculation. We now denote the paraxial solution corresponding to accelerating voltage $\hat{\phi}_0$ by $w(z, \hat{\phi}_0)$, so that

$$w\left(z, \hat{\phi}_0\right) = w_o g\left(z, \hat{\phi}_0\right) + w_o' h\left(z, \hat{\phi}_0\right) \tag{26.20}$$

In the image plane z_i for $\hat{\phi}_0$, we have $h(z_i, \hat{\phi}_0) = 0$. Hence

$$\begin{aligned} w\left(z_i, \hat{\phi}_0 + \Delta\hat{\phi}_0\right) &= w_o g\left(z_i, \hat{\phi}_0 + \Delta\hat{\phi}_0\right) + w_o' h\left(z_i, \hat{\phi}_0 + \Delta\hat{\phi}_0\right) \\ &\approx w_o\left(M + \frac{\partial g_i}{\partial\hat{\phi}_0}\Delta\hat{\phi}_0\right) + \Delta\hat{\phi}_0 w_o' \frac{\partial h_i}{\partial\hat{\phi}_0} \end{aligned}$$

In the fixed coordinate system, this is equivalent to

$$\begin{aligned} u\left(z_{i,}\hat{\phi}_0 + \Delta\hat{\phi}_0\right) &= w\left(z_i, \hat{\phi}_0 + \Delta\hat{\phi}_0\right) \exp\left\{\frac{\mathrm{i}\eta}{2\left(\hat{\phi}_0 + \Delta\hat{\phi}_0\right)^{1/2}}\int B\, dz\right\} \\ &= \exp(\mathrm{i}\theta_i)\left[M w_o + \Delta\hat{\phi}_0\left\{w_o\left(\frac{\partial g_i}{\partial\hat{\phi}_0} - \frac{M\mathrm{i}\theta_i}{2\hat{\phi}_0}\right) + w_o'\frac{\partial h_i}{\partial\hat{\phi}_0}\right\}\right] \end{aligned}$$

which must have the same form as

$$u\left(z_i, \hat{\phi}_0 + \Delta\hat{\phi}_0\right) = \exp(\mathrm{i}\theta_i)\left[M w_o - M\frac{\Delta\hat{\phi}_0}{\hat{\phi}_0}\left\{C_c w_o' + (C_D + \mathrm{i}C_\theta) w_o\right\}\right]$$

so that

$$\begin{aligned} C_c &= -\frac{\hat{\phi}_0}{M}\frac{\partial h_i}{\partial\hat{\phi}_0} \\ C_D &= -\frac{\hat{\phi}_0}{M}\frac{\partial g_i}{\partial\hat{\phi}_0} \\ C_\theta &= \frac{1}{2}\theta_i \end{aligned} \tag{26.21}$$

This form is most useful when the lens is operating at high magnification, $z_o \to z_{Fo}$. Since in general

$$x_i = Mx_o - MC_c x'_o \frac{\Delta\hat{\phi}_0}{\hat{\phi}_0} = Mx_o + \hat{\phi}_0 x'_o \frac{\partial h_i}{\partial\hat{\phi}_0}\frac{\Delta\hat{\phi}_0}{\hat{\phi}_0}$$

a ray intersecting the axis at $z = z_i$ leaves the object plane at $x_o = (-1/M)x'_o\Delta\hat{\phi}_0(\partial h_i/\partial\hat{\phi}_0)$. But $x_o = -x'_o\Delta\hat{\phi}_o(\partial z_{Fo}/\partial\hat{\phi}_o)$ so that

$$\frac{\partial z_{Fo}}{\partial\hat{\phi}_0} = \frac{1}{M}\frac{\partial h_i}{\partial\hat{\phi}_0}$$

and hence

$$C_c = -\hat{\phi}_0\frac{\partial z_{Fo}}{\partial\hat{\phi}_0} \tag{26.22a}$$

Likewise,

$$C_D = \frac{\hat{\phi}_0}{f_o}\frac{\partial f_o}{\partial\hat{\phi}_0} \tag{26.22b}$$

These expressions are particularly convenient when field models that yield explicit formulae for the cardinal elements are being studied (Lenz, 1982a).

26.2 Asymptotic Chromatic Aberrations

These aberration coefficients can be written down immediately as there is no term analogous to $L_3(x'^2 + y'^2)^2$ (24.2) needing special attention. We find

$$\Delta u^{(c)} = -\left\{C_c u'_o + (C_D + \mathrm{i}C_\theta)u_o\right\}\left(\frac{\gamma\Delta\phi_o}{\hat{\phi}_0} - 2\frac{\Delta B_0}{B_0}\right) \tag{26.23a}$$

or

$$\Delta u^{(c)} = -\left(C_c u'_o + C_D u_o\right)\frac{\Delta\phi_o}{\hat{\phi}_o} \tag{26.23b}$$

for magnetic and electrostatic lenses, respectively, where

$$\begin{aligned}
C_c &= \int_{-\infty}^{\infty}\frac{\eta^2B^2H^2}{4\hat{\phi}_o}dz \quad \text{or} \quad C_c = \hat{\phi}_0^{1/2}\int_{-\infty}^{\infty}\frac{\gamma\left(3+2\varepsilon\hat{\phi}\right)\phi'^2H^2}{8\hat{\phi}^{5/2}}dz \\
C_D &= \int_{-\infty}^{\infty}\frac{\eta^2B^2GH}{4\hat{\phi}_o}dz \quad \text{or} \quad C_D = \hat{\phi}_0^{1/2}\int_{-\infty}^{\infty}\frac{\gamma\left(3+2\varepsilon\hat{\phi}\right)\phi'^2GH}{8\hat{\phi}^{5/2}}dz \\
C_\theta &= \int_{-\infty}^{\infty}\frac{\eta B}{4\hat{\phi}_o^{1/2}}dz \qquad\qquad + \frac{\hat{\phi}_0}{4}\left(\frac{\gamma_i}{\hat{\phi}_i} - \frac{\gamma_o}{\hat{\phi}_0}\right)
\end{aligned} \tag{26.24}$$

These can of course be cast into polynomial form with the aid of Eq. (25.9) (Hawkes, 1980a; Shimoyama, 1982):

$$\begin{pmatrix} C_c \\ C_D \\ C_\theta \end{pmatrix} = \begin{pmatrix} f_o^2 C_2 & -2f_o^2 C_1 & f_o^2 C_0 \\ 0 & -f_o C_2 & f_o C_1 + C_e \\ 0 & 0 & C_\theta \end{pmatrix} \begin{pmatrix} m^2 \\ m \\ 1 \end{pmatrix} \tag{26.25}$$

in which for magnetic lenses,

$$\begin{aligned} C_2 &= \frac{\eta^2}{4\hat{\phi}_0} \int_{-\infty}^{\infty} B^2 G^2 \, dz \qquad C_1 = \frac{\eta^2}{4\hat{\phi}_0} \int_{-\infty}^{\infty} B^2 G\overline{G} \, dz \\ C_0 &= \frac{\eta^2}{4\hat{\phi}_0} \int_{-\infty}^{\infty} B^2 \overline{G}^2 \, dz \qquad C_\theta = \frac{\eta}{4\hat{\phi}_0^{1/2}} \int_{-\infty}^{\infty} B \, dz \\ C_e &= 0 \end{aligned} \tag{26.26}$$

and for electrostatic lenses

$$\begin{aligned} C_2 &= \hat{\phi}_0^{1/2} \int_{-\infty}^{\infty} \frac{\gamma\left(3 + 2\varepsilon\hat{\phi}\right)\phi'^2 G^2}{8\hat{\phi}^{5/2}} dz \\ C_1 &= \hat{\phi}_0^{1/2} \int_{-\infty}^{\infty} \frac{\gamma\left(3 + 2\varepsilon\hat{\phi}\right)\phi'^2 G\overline{G}}{8\hat{\phi}^{5/2}} dz \\ C_0 &= \hat{\phi}_0^{1/2} \int_{-\infty}^{\infty} \frac{\gamma\left(3 + 2\varepsilon\hat{\phi}\right)\phi'^2 \overline{G}^2}{8\hat{\phi}^{5/2}} dz \\ C_e &= \frac{\hat{\phi}_0}{4}\left(\frac{\gamma_i}{\hat{\phi}_i} - \frac{\gamma_o}{\hat{\phi}_o}\right) \end{aligned} \tag{26.27}$$

26.3 Higher Order Chromatic Aberration Coefficients

If the primary chromatic aberrations have been corrected, it may be necessary to study the chromatic aberrations of third order (fourth rank) and of second degree (third rank: linear in position and gradient, quadratic in the $\Delta\phi_0/\phi_0$ or $\Delta B_0/B_0$), if only to be sure that their influence is small. We follow Rose in using *order* to signify the sum of the exponents of position and gradient coordinates while *degree* is a measure of the power of the chromatic parameter Δ_c. For any given aberration, order + degree = *rank*, so that the higher the rank, the less important the aberration will be.

Both these sets of aberrations have been studied by Liu (2007, 2017) and we reproduce his formulae for the various coefficients below.

26.3.1 Third-Order (Fourth-Rank) Aberrations

These are aberrations that depend on third powers of the position and gradient coordinates and are linear in $\Delta\phi_0/\phi_0$ or $\Delta B_0/B_0$. Full details of the calculation that leads to the expressions for the coefficients and of the MATHEMATICA program with which they were obtained are given by Liu (2007, 2017). Here we list only the results. The calculation is performed in the rotating coordinate system and we write

$$\boldsymbol{r} = (x, y), \quad \boldsymbol{r}' = (x', y'), \quad \boldsymbol{r}^* = (-y, x) \tag{26.28}$$

The aberration coefficients are characterized by the following expression for $\Delta\boldsymbol{r}$:

$$\Delta\boldsymbol{r} = \tilde{\Phi}\left(\frac{\Delta\phi_0}{\phi_0}\right) + \tilde{B}\left(\frac{\Delta B_0}{B_0}\right) \tag{26.29}$$

where

$$\begin{aligned}\tilde{\Phi} = {} & (C_\phi \boldsymbol{r}'_o + c_\phi \boldsymbol{r}'^*_o)(\boldsymbol{r}'_o \cdot \boldsymbol{r}'_o) + (K_{1\phi}\boldsymbol{r}_0 + k_{1\phi}\boldsymbol{r}_0^*)(\boldsymbol{r}'_o \cdot \boldsymbol{r}'_o) + (K_{2\phi}\boldsymbol{r}'_0 + k_{2\phi}\boldsymbol{r}'^*_0)(\boldsymbol{r}'_o \cdot \boldsymbol{r}_o) \\ & + (A_\phi \boldsymbol{r}_0 + a_\phi \boldsymbol{r}_0^*)(\boldsymbol{r}'_o \cdot \boldsymbol{r}_o) + (F_\phi \boldsymbol{r}'_0 + f_\phi \boldsymbol{r}'^*_0)(\boldsymbol{r}_o \cdot \boldsymbol{r}_o) + (D_\phi \boldsymbol{r}_0 + d_\phi \boldsymbol{r}_0^*)(\boldsymbol{r}_o \cdot \boldsymbol{r}_o)\end{aligned} \tag{26.30}$$

with a similar expression for $\tilde{B}$ in which B replaces ϕ.

The coefficients can be written in unified form

$$\begin{aligned}\Xi_X &= -\frac{1}{\phi_o^{1/2}}\int_{z_o}^{z_i} \Xi_{Xi} dz - \frac{1}{\phi_o}\int_{z_o}^{z_i} \Xi_{Xc} dz \\ \Xi_{Xc} &= \Xi_{Xco} + \Xi_{X1}h^2 + \Xi_{Xc2}gh + \Xi_{Xc3}hh' + \Xi_{Xc4}g^2 + \Xi_{Xc5}gh' + \Xi_{Xc6}gg' + \Xi_{Xc7}h'^2\end{aligned} \tag{26.31}$$

in which X may denote ϕ or B. In the course of the calculation, the contributions to each coefficient are divided into two groups, 'intrinsic' and 'combined'; this is indicated by the second subscript, i or c.

Intrinsic contribution

Isotropic aberrations

$$\begin{aligned} C_{Xi} &= 4f_{X,0,2,0} \quad K_{1Xi} = f_{X,0,1,1} \quad K_{2Xi} = 2f_{X,0,1,1} \\ A_{Xi} &= 2f_{X,0,0,2} \quad F_{Xi} = 2f_{X,1,1,0} \quad D_{Xi} = f_{X,1,0,1}\end{aligned} \tag{26.32a}$$

Anisotropic aberrations

$$\begin{aligned} c_{Xi} &= 0 \quad k_{1Xi} = 3(f_{X,0,1,0})^* \quad k_{2Xi} = -2(f_{X,0,1,0})^* \\ a_{Xi} &= 2(f_{X,0,0,1})^* \quad f_{Xi} = -(f_{X,0,0,1})^* \quad d_{Xi} = (f_{X,1,0,0})^*\end{aligned} \tag{26.32b}$$

Combined contribution. Each coefficient has at most eight elements; these are listed separately as $C_{Xc0}, \ldots C_{Xc7}$, for example. $C_{Xc} = \sum_j C_{Xcj}$.

26.3.1.1 Isotropic Aberrations

Spherical aberration

$$\begin{aligned}
C_{Xc0} &= -4\varepsilon_{020}f_{X010} + 2\varepsilon_{011}f_{010} + 8f_{020,a}f_{X010} + 12f_{020}\varepsilon_{X001} \\
&\quad - 6f_{011}\varepsilon_{X010} + 24f_{020}f_{X010,o} \\
C_{Xc1} &= -C_X(f_{010})^* - 3\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{X000})^*(f_{010})^* \\
C_{Xc2} &= 0 \\
C_{Xc3} &= 2B_X f_{011} - 3f_{mn2}f_{X001} \\
C_{Xc4} &= C_{Xc6} = C_{Xc7} = 0 \\
C_{Xc5} &= 6f_{mn2}f_{X010} - 8B_X f_{020}
\end{aligned} \tag{26.33}$$

Coma

$$\begin{aligned}
K_{1Xc0} &= -\varepsilon_{011}f_{X001} + 4\varepsilon_{110}f_{X010} + 2f_{011,a}f_{X010} + f_{011}\varepsilon_{X001} - 4f_{002}\varepsilon_{X010} \\
&\quad + 8f_{020}\varepsilon_{X100} + 4f_{020}f_{X001,o} + 4f_{011}f_{X010,o} + 3(\varepsilon_{010})^*(f_{X000})^* + 3(f_{010})^*(\varepsilon_{X000})^* \\
K_{1Xc1} &= -2(\phi\phi_o)^{1/2}K_0 f_{X001} + C_X(f_{001})^* - 2\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{001})^*(f_{X000})^* \\
K_{1Xc2} &= 4(\phi\phi_o)^{1/2}K_0 f_{X010} - 3C_X(f_{010})^* + \left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{010})^*(f_{X000})^* \\
K_{1Xc3} &= 4B_X f_{110} - 2f_{mn3}f_{X001} - 2f_{mn2}f_{X100} \\
K_{1Xc4} &= K_{1Xc6} = K_{1Xc7} = 0 \\
K_{1Xc5} &= -2B_X f_{011} + f_{mn2}f_{X001} + 4f_{mn3}f_{X010}
\end{aligned} \tag{26.34}$$

$$\begin{aligned}
K_{2Xc0} &= -2\varepsilon_{011}f_{X001} + 4\varepsilon_{002}f_{X010} + 4f_{001,a}f_{X010} + 2f_{011}\varepsilon_{X001} - 4f_{002}\varepsilon_{X010} \\
&\quad - 8f_{110}\varepsilon_{X010} + 16f_{020}\varepsilon_{X100} + 8f_{020}f_{X001,o} + 8f_{011}f_{X010,o} - 2(\varepsilon_{010})^*(f_{X000})^* \\
&\quad - 2(f_{010})^*(\varepsilon_{X000})^* \\
K_{2Xc1} &= 2(\phi_o\phi)^{1/2}K_0 f_{X001} + 2\left(\frac{\phi_o}{\phi}\right)^{1/2}M_0 f_{X001} - 2C_X(f_{001})^* \\
K_{2Xc2} &= -4(\phi\phi_o)K_0 f_{X010} + 2C_X(f_{010})^* - 6\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{010})^*(f_{X000})^* \\
K_{2Xc3} &= 4B_X f_{002} - 4f_{mn2}f_{X100} \\
K_{2Xc4} &= 0 \\
K_{2Xc5} &= -4B_X f_{001} - 2f_{mn2}f_{X001} + 4f_{mn3}f_{X010} \\
K_{2Xc6} &= 4f_{mn2}f_{X010} \\
K_{2Xc7} &= 4\left(\frac{\phi_o}{\phi}\right)^{1/2}N_0 f_{X001}
\end{aligned} \tag{26.35}$$

Astigmatism

$$A_{Xc0} = -2\varepsilon_{002}f_{X001} + 2\left(\frac{\phi_0}{\phi}\right)^{1/2} f_{mn3}f_{X001} + 4\varepsilon_{101}f_{X010} + 4f_{002,o}f_{X010} - 2f_{002}\varepsilon_{X001}$$
$$-4f_{101}\varepsilon_{X010} + 8f_{011}\varepsilon_{X100} + 4f_{011}f_{X001.o} + + 4f_{002}f_{X010.o} + 2(\varepsilon_{001})^*(f_{X000})^* \quad (26.36)$$
$$+ 2(f_{001})^*(\varepsilon_{X000})^* - 4(f_{010})^*(f_{X000,o})^* - 2\left(\frac{\phi_0}{\phi}\right)^{1/2} Q_0(f_{X000})^*$$

$$A_{Xc1} = 2\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{100})^*(f_{X000})^* - 8(\phi\phi)^{1/2}K_0f_{X100}$$

$$A_{Xc2} = 6(\phi\phi)^{1/2}K_0f_{X001} - 6\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{001})^*(f_{X000})^*$$

$$A_{Xc3} = 4B_Xf_{101} - 8f_{mn3}f_{X100}$$

$$A_{Xc4} = 2\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{010})^*(f_{X000})^* - 4(\phi\phi)^{1/2}K_0(f_{X010})^*$$

$$A_{Xc5} = 2f_{mn3}f_{X001} - 4B_Xf_{002}$$
$$A_{Xc6} = 4f_{mn3}f_{X010}$$
$$A_{Xc7} = 0$$

Field curvature

$$F_{Xc0} = -2\varepsilon_{110}f_{X001} + 2\varepsilon_{101}f_{X010} + 4f_{110,o}f_{X010} - 2f_{110}\varepsilon_{X001} - 2f_{101}\varepsilon_{X010} + 4f_{011}\varepsilon_{X100}$$
$$+2f_{011}f_{X001,o} + 4f_{110}f_{X010,o} - (\varepsilon_{001})^*(f_{X000})^* - (f_{001})^*(\varepsilon_{X000})^* + 6(f_{010})^*(f_{X000,o})^*$$
$$+4\left(\frac{\phi_o}{\phi}\right)^{1/2} f_{mn2}f_{X100} + 3\left(\frac{\phi_o}{\phi}\right)^{1/2} Q_0(f_{X000})^*$$

$$F_{Xc1} = 8(\phi\phi_o)^{1/2}K_0f_{X100} - 3\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{100})^*(f_{X000})^*$$

$$F_{Xc2} = \left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{001})^*(f_{X000})^* - 6(\phi\phi_o)^{1/2}K_0f_{X001}$$

$$F_{Xc3} = 2B_Xf_{101} - f_{mn1}f_{X001}$$

$$F_{Xc4} = 4(\phi\phi_o)^{1/2}K_0f_{X010} - 3\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{010})^*(f_{X000})^*$$

$$F_{Xc5} = -4B_Xf_{110} + 2f_{mn1}f_{X010} - 4f_{mn2}f_{X100}$$
$$F_{Xc6} = 2f_{mn2}f_{X001}$$
$$F_{Xc7} = 0$$

(26.37)

Distortion

$$D_{Xc0} = -\varepsilon_{101} f_{X001} + 8\varepsilon_{200} f_{X010} + 2f_{101,o} f_{X010} - 3f_{101}\varepsilon_{X001} + 4(f_{002} + f_{110})\varepsilon_{X100} + 2(f_{002} + f_{110}) f_{X001,o} + (\varepsilon_{100})^*(f_{X000})^* + (f_{100})^*(\varepsilon_{X000})^* + (f_{001})^*(f_{X000,o})^* + 4\left(\frac{\phi_o}{\phi}\right)^{1/2} f_{mn3} f_{X100}$$

$$\begin{aligned} D_{Xc1} &= D_{Xc7} = 0 \\ D_{Xc2} &= -C_X(f_{100})^* - \left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{X000})^*(f_{100})^* \\ D_{Xc3} &= 8B_X f_{200} - 2f_{mn1} f_{X100} \\ D_{Xc4} &= -\left(\frac{\phi}{\phi_o}\right)^{1/2}(f_{001})^*(f_{X000})^* \\ D_{Xc5} &= -2B_X f_{101} + f_{mn1} f_{X001} - 4f_{mn3} f_{X100} \\ D_{Xc6} &= 2f_{mn3} f_{X001} \end{aligned} \tag{26.38}$$

26.3.1.2 Anisotropic Aberrations

Spherical aberration

$$\begin{aligned} c_{Xc0} &= -2\varepsilon_{X010}(f_{010})^* - 2f_{X010}(\varepsilon_{010})^* - 4\varepsilon_{020}(f_{X000})^* - 4f_{020}(\varepsilon_{X000})^* \\ c_{Xc1} &= -C_X f_{011} - \left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{010})^* \\ c_{Xc2} &= 4C_X f_{020} + 2\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X010}(f_{010})^* \\ c_{Xc3} &= f_{mn2}(f_{X000})^* - 2B_X(f_{010})^* \\ c_{Xc4} &= c_{Xc5} = c_{Xc6} = c_{Xc7} = 0 \end{aligned} \tag{26.39}$$

Coma

$$k_{1Xc0} = -4\varepsilon_{X010}(f_{001})^* + 3\varepsilon_{X001}(f_{010})^* + 12f_{X010,o}(f_{010})^* + 2f_{X010}(\varepsilon_{001})^* - f_{X001}(\varepsilon_{010})^* + 6f_{X010}(f_{010,o})^* - \varepsilon_{011}(f_{X000})^* - f_{011}(\varepsilon_{X000})^* + 12f_{020}(f_{X000,o})^* + 2\left(\frac{\phi_o}{\phi}\right)^{1/2} f_{mn2}(f_{X000})^*$$

$$k_{1Xc1} = -2C_X f_{110} - 2\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{001})^* - 6\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X100}(f_{010})^* + 4(\phi\phi)^{1/2} K_0(f_{X000})^*$$

$$k_{1Xc2} = C_X f_{011} + 9\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{010})^*$$

$$k_{1Xc3} = 2B_X(f_{001})^*$$

$$k_{1Xc4} = -4\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X010}(f_{010})^*$$

$$k_{1Xc5} = 8\phi_o^{1/2} Q_0 f_{X010} - 6B_X(f_{010})^* + f_{mn2}(f_{X000})^*$$

$$k_{1Xc6} = k_{1Xc7} = 0$$

(26.40)

$$k_{2Xc0} = -2\varepsilon_{X001}(f_{010})^* - 8f_{X010,o}(f_{010})^* - 4f_{X010}(\varepsilon_{001})^* + 2f_{X001}(\varepsilon_{010})^* - 4f_{X010}(f_{010,o})^*$$
$$- 2\varepsilon_{011}(f_{X000})^* - 2f_{011}(\varepsilon_{X000})^* - 8f_{020}(f_{X000,o})^* - 2\left(\frac{\phi_o}{\phi}\right)^{1/2} f_{mn2}(f_{X000})^*$$

$$k_{2Xc1} = -2C_X f_{002} + 4\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X100}(f_{010})^* - 2(\phi\phi_o)^{1/2} K_0(f_{X000})^*$$

$$k_{2Xc2} = 2C_X f_{011} - 6\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{010})^*$$

$$k_{2Xc3} = -4B_X(f_{001})^*$$

$$k_{2Xc4} = 8\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X010}(f_{010})^*$$

$$k_{2Xc5} = 4B_X(f_{010})^* + 2f_{mn2}(f_{X000})^*$$

$$k_{2Xc6} = 0$$

$$k_{2Xc7} = -2\left(\frac{\phi_o}{\phi}\right)^{1/2} N_0(f_{X000})^*$$

(26.41)

Astigmatism

$$
\begin{aligned}
a_{Xc0} &= -2(\varepsilon_{001})^* f_{X001} + 4\frac{\phi_o}{\phi^{1/2}} Q_0 f_{X001} - 2\varepsilon_{X001}(f_{001})^* + 4f_{X010,o}(f_{001})^* + 8\varepsilon_{X100}(f_{010})^* \\
&\quad + 4f_{X001,o}(f_{010})^* - 4\varepsilon_{X010}(f_{1001})^* + 4f_{X010}(\varepsilon_{100})^* + 4f_{X010}(f_{001,o})^* - 2\varepsilon_{002}(f_{X000})^* \\
&\quad - 2f_{002}(\varepsilon_{X000})^* + 4f_{011}(f_{X000,o})^* + 2\left(\frac{\phi_o}{\phi}\right)^{1/2} f_{mn3}(f_{X000})^* \\
a_{Xc1} &= 2\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{100})^* - 2C_X f_{101} \\
a_{Xc2} &= 2C_X f_{002} + 2\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{001})^* - 8\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X100}(f_{010})^* - 4\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X010}(f_{100})^* \\
&\quad + 6(\phi\phi_o)^{1/2} K_0 (f_{X000})^* \\
a_{Xc3} &= 4B_X(f_{100})^* - 16\phi_o^{1/2} Q_o f_{X100} \\
a_{Xc4} &= 0 \\
a_{Xc5} &= 2f_{mn3}(f_{X000})^* - 4B_X(f_{001})^*
\end{aligned}
\tag{26.42}
$$

Field curvature

$$
\begin{aligned}
f_{Xc0} &= (\varepsilon_{001})^* f_{X001} - 2\left(\frac{\phi_o}{\phi}\right)^{1/2} Q_0 f_{X001} + \varepsilon_{X001}(f_{001})^* - 2f_{X010,o}(f_{001})^* \\
&\quad - 4\varepsilon_{X100}(f_{010})^* - 2f_{X001,o}(f_{010})^* - 2\varepsilon_{X010}(f_{100})^* - 6f_{X010}(\varepsilon_{100})^* \\
&\quad - 2f_{X010}(f_{001,o})^* - 2\varepsilon_{110}(f_{X000})^* - 2f_{110}(\varepsilon_{X000})^* - 2f_{011}(f_{X000,o})^* \\
&\quad - 2\left(\frac{\phi_o}{\phi}\right)^{1/2} f_{mn3}(f_{X000})^* \\
f_{Xc1} &= -C_X f_{101} - 3\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{100})^* \\
f_{Xc2} &= 2C_X f_{110} - \left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{001})^* + 4\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X100}(f_{010})^* \\
&\quad + 2\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X010}(f_{100})^* - 2(\phi\phi_o)^{1/2} K_0 (f_{X000})^* \\
f_{Xc3} &= 8\phi_o^{1/2} Q_0 f_{X100} - 6B_X(f_{100})^* + f_{mn1}(f_{X000})^* \\
f_{Xc4} &= 2\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X010}(f_{001})^* \\
f_{Xc5} &= 2B_X(f_{001})^* \\
f_{Xc6} &= f_{Xc7} = 0
\end{aligned}
\tag{26.43}
$$

Distortion

$$
\begin{aligned}
d_{Xc0} &= 4\frac{\phi_o}{\phi^{1/2}}Q_0 f_{X100} + 2\varepsilon_{X100}(f_{001})^* + f_{X001,o}(f_{001})^* - 3\varepsilon_{X001}(f_{100})^* - f_{X001}(\varepsilon_{100})^* \\
&\quad + 2f_{X010}(f_{100,o})^* - \varepsilon_{101}(f_{X000})^* - f_{101}(\varepsilon_{X000})^* + 2f_{110}(f_{X000,o})^* \\
d_{Xc1} &= -4C_X f_{200} + -2\left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X100}(f_{100})^* \\
d_{Xc2} &= C_X f_{101} + \left(\frac{\phi}{\phi_o}\right)^{1/2} f_{X001}(f_{100})^* \\
d_{Xc3} &= d_{Xc7} = 0 \\
d_{Xc4} &= 2(\phi\phi_o)^{1/2} K_0 (f_{X000})^* \\
d_{Xc5} &= -4\phi_o^{1/2} Q_0 f_{X100} - 2B_X(f_{100})^* + f_{mn1}(f_{X000})^* \\
d_{Xc6} &= 2\phi_o^{1/2} Q_0 f_{X001}
\end{aligned}
\tag{26.44}
$$

26.3.1.3 Definitions

The quantities appearing in these formulae are defined as follows:

$$
\begin{aligned}
A_\phi &= \frac{\phi_o}{16\phi^{3/2}}\left(\eta^2 B^2 + \phi''\right) & A_B &= -\frac{\eta^2 B^2}{4\phi^{1/2}} \\
B_\phi &= \phi_o/4\phi^{1/2} & B_B &= 0 \\
C_\phi &= \eta B\phi_o/4\phi & C_B &= -\eta B/2
\end{aligned}
\tag{26.45}
$$

$$K_0 = \frac{\eta^2 B^2}{8\phi^{3/2}}$$

$$L_0 = \frac{1}{32\phi^{3/2}}\left(\eta^4 B^4 + 2\eta^2 B^2\phi'' - 4\eta^2 BB''\phi + \phi''^2 - \phi\phi^{(4)}\right)$$

$$M_0 = \frac{\eta^2 B^2}{8\phi^{1/2}}$$

$$N_0 = \frac{\phi^{1/2}}{2}$$

$$P_0 = \frac{\eta}{16\phi^{3/2}}\left(\eta^2 B^3 + \phi'' B - \phi B''\right)$$

$$Q_0 = \frac{\eta B}{4\phi^{1/2}}$$

$$
\begin{aligned}
L_\phi &= \frac{\phi_o}{64\phi^{3/2}}\left(\eta^4 B^4 - 2\eta^2 B^2 \phi'' - 3\phi''^2 + \frac{\phi\phi^{(4)}}{\phi_0}\right)\\
M_\phi &= \frac{\phi_o}{16\phi^{3/2}}\left(\eta^2 B^2 - \phi''\right)\\
N_\phi &= \frac{\phi_o}{4\phi^{1/2}}\\
L_B &= -\frac{\eta^2 BB''}{8\phi^{1/2}}\\
M_B &= 0\\
N_B &= 0
\end{aligned}
\tag{26.46}
$$

$$
\begin{aligned}
f_{X200} &= -\frac{1}{4}L_X g^4 - \frac{1}{2}M_X g^2 g'^2 - \frac{1}{4}N_X g'^4\\
f_{X110} &= -\frac{1}{2}L_X g^2 h^2 - \frac{1}{2}M_X\left(g^2 h'^2 + g'^2 h^2\right) - \frac{1}{2}N_X g'^2 h'^2 - \phi_o K_X\\
f_{X101} &= -L_X g^3 h - M_X g g'(gh' + g'h) - N_X g'^3 h'\\
f_{X020} &= -\frac{1}{4}L_X h^4 - \frac{1}{2}M_X h^2 h'^2 - \frac{1}{4}N_X h'^4\\
f_{X011} &= -L_X g h^3 - M_X h h'(gh' + g'h) - N_X g' h'^3\\
f_{X002} &= -L_X g^2 h^2 - 2M_X g g' h h' - N_X g'^2 h'^2 + \phi_o K_X\\
(f_{X100})^* &= -\phi_o^{1/2}\left(P_X g^2 + Q_X g'^2\right)\\
(f_{X010})^* &= -\phi_o^{1/2}\left(P_X h^2 + Q_X h'^2\right)\\
(f_{X001})^* &= -2\phi_o^{1/2}(P_X gh + Q_X g' h')
\end{aligned}
\tag{26.47}
$$

The same quantities without suffix X are obtained by replacing $L_X, M_X, \ldots Q_X$ by $L_0, M_0, ..Q_0$. Addition of o to the subscript indicates that this quantity is to be evaluated at the object plane, $z = z_o$.

$$
\begin{aligned}
f_{mn1} &= M_0 g^2 + N_0 g'^2\\
f_{mn2} &= M_0 h^2 + N_0 h'^2\\
f_{mn3} &= M_0 gh + N_0 g' h'
\end{aligned}
$$

The quantities in ε denote

$$\varepsilon_{Xijk} = \int_{z_o}^{z} f_{Xijk} d\zeta \tag{26.48}$$

26.3.2 Third-Rank Aberrations

We are now concerned with aberrations that are linear in the position or gradient coordinates and quadratic in $\Delta\phi/\phi$ and $\Delta B/B$. They are defined by

$$\Delta \boldsymbol{r} = -\widehat{\Phi}\left(\frac{\Delta\phi}{\phi}\right)^2 - \widehat{B}\left(\frac{\Delta B}{B}\right)^2 - \widehat{\Psi}\left(\frac{\Delta\phi}{\phi}\right)\left(\frac{\Delta B}{B}\right) \tag{26.49}$$

As in the previous section, each coefficient is the sum of an 'intrinsic' part (label i) and a 'combined' part (label c). We have

$$\begin{aligned}
\widehat{\Phi} &= \left(C_{iC} + C_{cC}^{(\phi)}\right) r_o' + \left(C_{iM} + C_{cM}^{(\phi)}\right) r_o + \left(C_{iR} + C_{cR}^{(\phi)}\right) r_o^* + C_{cS}^{(\phi)} r_o'^* \\
\widehat{B} &= C_{cC}^{(B)} r_o' + C_{cM}^{(B)} r_o + C_{cR}^{(B)} r_o^* + C_{cS}^{(B)} r_o'^* \\
\widehat{\Psi} &= C_{cC}^{(\Psi)} r_o' + C_{cM}^{(\Psi)} r_o + C_{cR}^{(\Psi)} r_o^* + C_{cS}^{(\Psi)} r_o'^*
\end{aligned} \tag{26.50}$$

Intrinsic part

$$\begin{aligned}
C_{iC} &= \frac{2}{\phi_o^{1/2}} \int_{z_o}^{z_i} f_{\phi 2,010}\, dz \\
C_{iM} &= \frac{1}{\phi_o^{1/2}} \int_{z_o}^{z_i} f_{\phi 2,001}\, dz \\
C_{iR} &= \frac{1}{\phi_o^{1/2}} \int_{z_o}^{z_i} f_{\phi 2,000}\, dz
\end{aligned} \tag{26.51a}$$

in which

$$\begin{aligned}
f_{\phi 2,010} &= -\frac{\phi_o^2}{64\phi^{5/2}} \left\{ \left(3\phi'' + \eta^2 B^2\right) h^2 + 4\phi h'^2 \right\} \\
f_{\phi 2,001} &= -\frac{\phi_o^2}{32\phi^{5/2}} \left\{ \left(3\phi'' + \eta^2 B^2\right) gh + 4\phi g' h' \right\} \\
f_{\phi 2,000} &= -\frac{1}{16} \left(\frac{\phi_o}{\phi}\right)^{5/2} \eta B
\end{aligned} \tag{26.51b}$$

Combined part

$$C_{cC}^{(X)} = \frac{1}{\phi_o}\int_{z_o}^{z_i}\left(\Gamma_{C1}^{(X)}h^2 + \Gamma_{C2}^{(X)}gh + \Gamma_{C3}^{(X)}hh' + \Gamma_{C4}^{(X)}g'h + \Gamma_{C5}^{(X)}gh' + \Gamma_{C6}^{(X)}h'^2 + \Gamma_{C7}^{(X)}g'h'\right)dz$$

$$C_{cM}^{(X)} = \frac{1}{\phi_o}\int_{z_o}^{z_i}\left(\Gamma_{M1}^{(X)}h^2 + \Gamma_{M2}^{(X)}gh + \Gamma_{M3}^{(X)}hh' + \Gamma_{M4}^{(X)}g'h + \Gamma_{M5}^{(X)}gh' + \Gamma_{M6}^{(X)}h'^2 + \Gamma_{M7}^{(X)}g'h'\right)dz$$

$$C_{cR}^{(X)} = \frac{1}{\phi_o}\int_{z_o}^{z_i}\left(\Gamma_{R1}^{(X)}h^2 + \Gamma_{R2}^{(X)}gh + \Gamma_{R3}^{(X)}hh' + \Gamma_{R4}^{(X)}g'h + \Gamma_{R5}^{(X)}gh' + \Gamma_{R6}^{(X)}h'^2 + \Gamma_{R7}^{(X)}g'h'\right)dz$$

$$C_{cS}^{(X)} = \frac{1}{\phi_o}\int_{z_o}^{z_i}\left(\Gamma_{S1}^{(X)}h^2 + \Gamma_{S2}^{(X)}gh + \Gamma_{S3}^{(X)}hh' + \Gamma_{S4}^{(X)}g'h + \Gamma_{S5}^{(X)}gh' + \Gamma_{S6}^{(X)}h'^2 + \Gamma_{S7}^{(X)}g'h'\right)dz \tag{26.52}$$

in which, as before, X represents ϕ, B or Ψ. The quantities Γ are as follows:

Electrostatic group

$$\Gamma_{C1}^{(\phi)} = 2A_\phi\varepsilon_{\phi001} + 4A_\phi f_{\phi010,o} - C_\phi(f_{\phi000})^* \quad \Gamma_{C2}^{(\phi)} = -4A_\phi\varepsilon_{\phi010} \quad \Gamma_{C3}^{(\phi)} = 2B_\phi f_{\phi001}$$
$$\Gamma_{C4}^{(\phi)} = 0 \quad \Gamma_{C5}^{(\phi)} = -4B_\phi f_{\phi010} \quad \Gamma_{C6}^{(\phi)} = 2B_\phi\left(\varepsilon_{\phi001} + 2f_{\phi010,o}\right) \quad \Gamma_{C7}^{(\phi)} = -4B_\phi\varepsilon_{\phi010}$$
$$\Gamma_{M1}^{(\phi)} = 4A_\phi\varepsilon_{\phi100} + 2A_\phi f_{\phi001,o} \quad \Gamma_{M2}^{(\phi)} = -2A_\phi\varepsilon_{\phi001} - C_\phi(f_{\phi000})^* \quad \Gamma_{M3}^{(\phi)} = 4B_\phi f_{\phi100}$$
$$\Gamma_{M4}^{(\phi)} = -C_\phi(\varepsilon_{\phi000})^* \quad \Gamma_{M5}^{(\phi)} = -2B_\phi f_{\phi001} + C_\phi(\varepsilon_{\phi000})^* \; \Gamma_{M6}^{(\phi)} = 2B_\phi\left(2\varepsilon_{\phi100} + f_{\phi001,o}\right) \; \Gamma_{M7}^{(\phi)} = -2B_\phi\varepsilon_{\phi001}$$

$$\Gamma_{R1}^{(\phi)} = -2\left(C_\phi f_{\phi100} - A_\phi(f_{\phi000,o})^*\right) \quad \Gamma_{R2}^{(\phi)} = C_\phi f_{\phi001} - 2A_\phi(\varepsilon_{\phi000})^* \quad \Gamma_{R3}^{(\phi)} = 0 \quad \Gamma_{R4}^{(\phi)} = C_\phi\varepsilon_{\phi001}$$
$$\Gamma_{R5}^{(\phi)} = -C_\phi\varepsilon_{\phi001} - 2B_\phi(f_{\phi000})^* \quad \Gamma_{R6}^{(\phi)} = 2B_\phi(f_{\phi000,o})^* \quad \Gamma_{R7}^{(\phi)} = -2B_\phi(\varepsilon_{\phi000})^*$$
$$\Gamma_{S1}^{(\phi)} = -C_\phi f_{\phi001} - 2A_\phi(\varepsilon_{\phi000})^* \quad \Gamma_{S2}^{(\phi)} = 2C_\phi f_{\phi010} \quad \Gamma_{S3}^{(\phi)} = -2B_\phi(f_{\phi000})^*$$
$$\Gamma_{S4}^{(\phi)} = 2C_\phi\varepsilon_{\phi010} \quad \Gamma_{S5}^{(\phi)} = -2C_\phi\varepsilon_{\phi010} \quad \Gamma_{S6}^{(\phi)} = -2B_\phi(\varepsilon_{\phi000})^* \quad \Gamma_{S7}^{(\phi)} = 0 \tag{26.53}$$

Magnetic group

These terms are obtained from the electrostatic group above by replacing ϕ everywhere by B.

Mixed group

These terms are a modified version of the sums of the electrostatic and magnetic terms above:

$$\Gamma_{Lj}^{(\Psi)} \to \Gamma_{Lj}^{(\phi)} + \Gamma_{Lj}^{(B)} \quad j = 1, \ldots, 7$$

where L represents C, M, R or S. The modification requires the first appearance of ϕ in the electrostatic terms to be replaced by B and the first appearance of B in the magnetic terms

to be replaced by ϕ. This rule will be useful when programming the formulae. To prevent confusion, we now list the terms in $I^{(\psi)}$:

$$
\begin{aligned}
I_{C1}^{(\Psi)} &= 2A_B\varepsilon_{\phi 001} + 4A_B f_{\phi 010,o} - C_B(f_{\phi 000})^* + 2A_\phi\varepsilon_{B001} + 4A_\phi f_{B010,o} - C_\phi(f_{B000})^* \\
I_{C2}^{(\Psi)} &= -4(A_B\varepsilon_{\phi 010} + A_\phi\varepsilon_{B010}) \\
I_{C3}^{(\Psi)} &= 2(B_B f_{\phi 001} + B_\phi f_{B001}) \\
I_{C4}^{(\Psi)} &= 0 \\
I_{C5}^{(\Psi)} &= -4(B_B f_{\phi 010} + B_\phi f_{\phi B010}) \\
I_{C6}^{(\Psi)} &= 2B_\phi\left(\varepsilon_{B001} + 2f_{B010,o}\right) + 2B_B\left(\varepsilon_{\phi 001} + 2f_{\phi 010,o}\right) \\
I_{C7}^{(\Psi)} &= -4(B_B\varepsilon_{\phi 010} + B_\phi\varepsilon_{B010}) \\
I_{M1}^{(\Psi)} &= 2(2A_B\varepsilon_{\phi 100} + A_B f_{\phi 001,o} + 2A_\phi\varepsilon_{B100} + A_\phi f_{B001,o}) \\
I_{M2}^{(\Psi)} &= -\{2A_B\varepsilon_{\phi 001} + C_B(f_{\phi 000})^* + 2A_\phi\varepsilon_{B001} + C_\phi(f_{B000})^*\} \\
I_{M3}^{(\Psi)} &= 4(B_B f_{\phi 100} + B_\phi f_{B100}) \\
I_{M4}^{(\Psi)} &= -\{C_B(\varepsilon_{\phi 000})^* + C_\phi(\varepsilon_{B000})^*\} \\
I_{M5}^{(\Psi)} &= -2(B_B f_{\phi 001} + B_\phi f_{B001}) + C_B(\varepsilon_{\phi 000})^* + C_\phi(\varepsilon_{B000})^* \\
I_{M6}^{(\Psi)} &= 2B_B\left(2\varepsilon_{\phi 100} + f_{\phi 001,o}\right) + 2B_\phi\left(2\varepsilon_{B100} + f_{B001,o}\right) \\
I_{M7}^{(\Psi)} &= -2(B_B\varepsilon_{\phi 001} + B_\phi\varepsilon_{B001}) \\
I_{R1}^{(\Psi)} &= -2\left(C_B f_{\phi 100} - A_B(f_{\phi 000,o})^* + C_\phi f_{B100} - A_\phi(f_{B000,o})^*\right) \\
I_{R2}^{(\Psi)} &= C_B f_{\phi 001} + C_\phi f_{B001} - 2\{A_B(\varepsilon_{\phi 000})^* + A_\phi(\varepsilon_{B000})^*\} \\
I_{R3}^{(\Psi)} &= 0 \\
I_{R4}^{(\Psi)} &= C_B\varepsilon_{\phi 001} + C_\phi\varepsilon_{B001} \\
I_{R5}^{(\Psi)} &= -(C_B\varepsilon_{\phi 001} + C_\phi\varepsilon_{B001}) - 2\{B_B(f_{\phi 000})^* + B_\phi(f_{B000})^*\} \\
I_{R6}^{(\Psi)} &= 2\{B_B(f_{\phi 000,o})^* + B_\phi(f_{B000,o})^*\} \\
I_{R7}^{(\Psi)} &= -2\{B_B(\varepsilon_{\phi 000})^* + B_\phi(\varepsilon_{B000})^*\} \\
I_{S1}^{(\Psi)} &= -(C_B f_{\phi 001} + C_\phi f_{B001}) - 2\{A_B(\varepsilon_{\phi 000})^* + A_\phi(\varepsilon_{B000})^*\} \\
I_{S2}^{(\Psi)} &= 2(C_B f_{\phi 010} + C_\phi f_{B010}) \\
I_{S3}^{(\Psi)} &= -2\{B_B(f_{\phi 000})^* + B_\phi(f_{B000})^*\} \\
I_{S4}^{(\Psi)} &= 2(C_B\varepsilon_{\phi 010} + C_\phi\varepsilon_{B010}) \\
I_{S5}^{(\Psi)} &= -2(C_B\varepsilon_{\phi 010} + C_\phi\varepsilon_{B010}) \\
I_{S6}^{(\Psi)} &= -2\{B_B(\varepsilon_{\phi 000})^* + B_\phi(\varepsilon_{B000})^*\} \\
I_{S7}^{(\Psi)} &= 0
\end{aligned}
\tag{26.54}
$$

CHAPTER 27

Aberration Matrices and the Aberrations of Lens Combinations

We have seen in Section 16.2 that the cardinal elements of doublets and hence multiplets can be obtained straightforwardly by multiplying the appropriate transfer matrices. A similar procedure can be devised to express the aberration coefficients of a doublet (or multiplet) in terms of those of the individual components. The coefficients of the aberration polynomials of the multiplet can likewise be expressed in terms of those of the separate lenses.

For this, we introduce column vectors in an arbitrary pair of conjugate planes; we write

$$\boldsymbol{u}_o = \begin{pmatrix} \boldsymbol{u}_o \\ \boldsymbol{u}'_o \\ \boldsymbol{u}_o r_o^2 \\ \boldsymbol{u}'_o r_o^2 \\ \boldsymbol{u}_o V_o \\ \boldsymbol{u}'_o V_o \\ \boldsymbol{u}_o \theta_o^2 \\ \boldsymbol{u}'_o \theta_o^2 \\ \boldsymbol{u}_o v_o \\ \boldsymbol{u}'_o v_o \end{pmatrix} \tag{27.1}$$

in the object plane, with a similar expression for $\boldsymbol{u}_m$, the corresponding vector in the image plane, magnification M. Then we may write

$$\boldsymbol{u}_m = \boldsymbol{M}\boldsymbol{u}_o \tag{27.2}$$

where $\boldsymbol{M}$ is a 10×10 matrix, which divides naturally into four block matrices:

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{M}_3 & \boldsymbol{M}_4 \end{pmatrix} \tag{27.3}$$

$\boldsymbol{M}_1$ is the 2×2 paraxial matrix (16.20),

$$\boldsymbol{M}_1 = \begin{pmatrix} M & 0 \\ c & rm \end{pmatrix} \tag{27.4}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00027-4
© 2018 Elsevier Ltd. All rights reserved.

in which we have introduced the *convergence*, c and the relative refractive index r (25.27):

$$c := -\frac{1}{f_i} \quad r := \frac{f_o}{f_i} = \left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/2} \tag{27.5}$$

The matrix $\boldsymbol{M}_2$ has two rows and eight columns, the upper row containing the coefficients describing the aberrations of position and the lower row those of gradient:

$$\boldsymbol{M}_2 =: \begin{pmatrix} Mm_{11} & Mm_{12} & Mm_{13} & Mm_{14} & Mm_{15} & Mm_{16} & Mm_{17} & Mm_{18} \\ m_{21} & m_{22} & m_{23} & m_{24} & m_{25} & m_{26} & m_{27} & m_{28} \end{pmatrix} \tag{27.6}$$

and

$$\begin{aligned} m_{11} &= D + \mathrm{i}d & m_{15} &= K + \mathrm{i}k \\ m_{12} &= F - A & m_{16} &= C \\ m_{13} &= 2A + \mathrm{i}a & m_{17} &= a \\ m_{14} &= 2K & m_{18} &= 2k \end{aligned} \tag{27.7}$$

$$\begin{aligned} m_{21} &= c(D + \mathrm{i}d) - ri_1 m \\ m_{22} &= c(F - A) - (D - \mathrm{i}d)rm + \frac{1}{2}c^2 rm \\ m_{23} &= c(2A + \mathrm{i}a) - 2Drm + c^2 rm \\ m_{24} &= 2cK - (2A - \mathrm{i}a)rm + cr^2 m^2 \\ m_{25} &= c(K + \mathrm{i}k) - (F - A)rm + \frac{1}{2}cr^2 m^2 \\ m_{26} &= cC - (K - \mathrm{i}k)rm - \frac{1}{2}rm(1 - r^2 m^2) \\ m_{27} &= ca - 2drm \\ m_{28} &= 2ck - arm \end{aligned} \tag{27.8}$$

in which i_1 is defined in Eq. (25.31). The matrix $\boldsymbol{M}_3$ is null and $\boldsymbol{M}_4$, generated by $\boldsymbol{M}_1$, encodes the rules needed for adding the aberration coefficients. It is easily seen that

$$\boldsymbol{M}_4 = \begin{pmatrix} M^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ cM^2 & rM & 0 & 0 & 0 & 0 & 0 & 0 \\ cM^2 & 0 & rM & 0 & 0 & 0 & 0 & 0 \\ c^2M & rc & rc & r^2m & 0 & 0 & 0 & 0 \\ c^2M & 0 & 2rc & 0 & r^2m & 0 & 0 & 0 \\ c^3 & rc^2m & 2rc^2m & 2r^2cm^2 & r^2cm^2 & r^3m^3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & rM & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & rc & r^2m \end{pmatrix} \tag{27.9}$$

Suppose now that a second lens (or lens system) is characterized by a similar 10×10 matrix, $\boldsymbol{M}'$, and that just as $\boldsymbol{u}_m = \boldsymbol{M}\boldsymbol{u}_o$, so $\boldsymbol{u}'_m = \boldsymbol{M}'\boldsymbol{u}_m$ or

$$\boldsymbol{u}'_{\boldsymbol{m}} = \boldsymbol{M}'\boldsymbol{M}\boldsymbol{u}_o \tag{27.10}$$

It is convenient to write

$$\boldsymbol{P} := \boldsymbol{M}'M \quad P := M'M \quad p := P^{-1} \tag{27.11}$$

and to write $\boldsymbol{u}_p$ instead of $\boldsymbol{u}'_m$, so that (27.10) becomes

$$\boldsymbol{u}_p = \boldsymbol{P}\boldsymbol{u}_o \tag{27.12}$$

Clearly $\boldsymbol{P}$ must have the same block structure as $\boldsymbol{M}$ and $\boldsymbol{M}'$ and indeed $\boldsymbol{P}_1$ is just the paraxial matrix of the combination, which we have already met in Chapter 16, Gaussian Optics of Rotationally Symmetric Systems: Asymptotic Image Formation (Eqs 16.29−16.34):

$$\boldsymbol{P}_1 = \boldsymbol{M}'_1\boldsymbol{M}_1 = \begin{pmatrix} P & 0 \\ c_p & r_p p \end{pmatrix} = \begin{pmatrix} P & 0 \\ D_p cc' & r_p p \end{pmatrix} \tag{27.13}$$

with $D_p := z'_{Fo} - z_{Fi}$ and $r_p := rr'$; $\boldsymbol{P}_3$ is of course null and $\boldsymbol{P}_4$ has the same structure as $\boldsymbol{M}_4$ (27.9). It is $\boldsymbol{P}_2$ that is of interest, for it gives us the recipe for adding aberrations, long known but intermittently rediscovered (e.g. Orloff, 1983a). Explicitly, we find

$$p_{11} = m_{11} + M^2 m'_{11} + Mc(m'_{12} + m'_{13}) + c^2(m'_{14} + m'_{15}) + c^3 m m'_{16}$$

or

$$\begin{aligned} D_p + \mathrm{i}d_p &= D + \mathrm{i}d + M^2(D' + \mathrm{i}d') + Mc(F' + A' + \mathrm{i}a') + c^2(3K' + \mathrm{i}k') + c^3 C' \\ p_{12} &= m_{12} + r m'_{12} + rcm m'_{14} + rc^2 m m'_{16} \end{aligned}$$

or

$$\begin{aligned} F_p &= F + rF' + 4rcmK' + 2rc^2 mC' \\ p_{13} &= m_{13} + r m'_{13} + rcm(m'_{14} + 2m'_{15}) + 2rc^2 m m'_{16} \end{aligned}$$

or

$$\begin{aligned} 2A_p + \mathrm{i}a_p &= 2A + \mathrm{i}a + r(2A' + \mathrm{i}a') + rcm(4K' + 2\mathrm{i}k') + 2rc^2 mC' \\ p_{14} &= m_{14} + r^2 m^2 m'_{14} + 2r^2 cm^3 m'_{16} \end{aligned}$$

or

$$\begin{aligned} K_p &= K + r^2 m^2 K' + r^2 cm^3 C' \\ p_{15} &= m_{15} + r^2 m^2 m'_{15} + r^2 cm^3 m'_{16} \end{aligned}$$

or

$$K_p + \mathrm{i}k_p = K + \mathrm{i}k + r^2m^2(K' + \mathrm{i}k') + r^2cm^3C'$$
$$p_{16} = m_{16} + r^3m^4m'_{16}$$

or

$$C_p = C + r^3m^4C'$$
$$p_{17} = m_{17} + rm'_{17} + rcmm'_{18}$$

or

$$a_p = a + ra' + 2rcmk'$$
$$p_{18} = m_{18} + r^2m^2m'_{18}$$

or

$$k_p = k + r^2m^2k' \tag{27.14}$$

(The last two relations, for p_{17} and p_{18}, merely repeat information already provided by p_{13} and p_{15}.) The fact that, with the exception of spherical aberration, each of the aberration coefficients for a lens combination includes several of the coefficients of the individual members is exploited in the design of aberration correctors (Chapter 41 of Volume 2).

Just as $\boldsymbol{P}$ must have the same overall structure as $\boldsymbol{M}$, so must the elements of the aberration submatrix $\boldsymbol{P}_2$ have the polynomial dependence on p that the elements of $\boldsymbol{M}_2$ have on m. A somewhat lengthy calculation reveals that the relation between the polynomial coefficients appearing in $\boldsymbol{P}_2$ and those in $\boldsymbol{M}_2$ and $\boldsymbol{M}'_2$ is as follows, in which i_j corresponds to $\boldsymbol{M}$, i'_j to $\boldsymbol{M}'$ and $i_j^{(p)}$ to $\boldsymbol{P}$ ($j = 1-9$), cf. (25.29−25.30).

$$i_1^{(p)} = i_1 + \frac{c^4}{r}\left\{ i'_1D_p^4 - 4i'_2D_p^3\frac{r'}{c} + 2(i'_3 + i'_4)D_p^2\left(\frac{r'}{c'}\right)^2 - 4i'_5D_p\left(\frac{r'}{c'}\right)^3 + i'_6\left(\frac{r'}{c'}\right)^4 + \frac{1}{2}D_p^3\left(c'^2 + \frac{1}{D_p^2}\right)\right\}$$

$$i_2^{(p)} = \frac{c'}{r'}\left(i_2D_p - \frac{i_1}{c}\right) + \frac{c^3}{r}\left\{ i'_2D_p^3 - (i'_3 + i'_4)D_p^2\frac{r'}{c'} + 3i'_5D_p\left(\frac{r'}{c'}\right)^2 - i'_6\left(\frac{r'}{c'}\right)^3 - \frac{1}{2}D_p\frac{c'}{r'}\right\}$$

$$i_3^{(p)} = \left(\frac{c'}{r'}\right)^2\left(i_3D_p^2 - 2i_2\frac{D_p}{c} + \frac{i_1}{c^2}\right) + \frac{c^2}{r}\left\{ i'_3D_p^2 - 2i'_5D_p\frac{r'}{c'} + i'_6\left(\frac{r'}{c'}\right)^2 + \frac{1}{2}D_p\left(\frac{c'}{r'}\right)^2\right\}$$

$$i_4^{(p)} = \left(\frac{c'}{r'}\right)^2\left(i_4D_p^2 - 4i_2\frac{D_p}{c} + 2\frac{i_1}{c^2}\right) + \frac{c^2}{r}\left\{ i'_4D_p^2 - 4i'_5D_p\frac{r'}{c'} + 2i'_6\left(\frac{r'}{c'}\right)^2 + D_p\left(\frac{c'}{r'}\right)^2\right\}$$

$$
\begin{aligned}
i_5^{(p)} &= \left(\frac{c'}{r'}\right)^3 \left\{ i_5 D_p^3 - (i_3 + i_4)\frac{D_p^2}{c} + 3i_2\frac{D_p}{c^2}) - \frac{i_1}{c^3} \right\} + \frac{c}{r}\left\{ i_5' D_p - i_6'\frac{r'}{c'} - \frac{1}{2}D_p\left(\frac{c'}{r'}\right)^3 \right\} \\
i_6^{(p)} &= \left(\frac{c'}{r'}\right)^4 \left\{ i_6 D_p^4 - 4i_5\frac{D_p^3}{c} + 2(i_3 + i_4)\frac{D_p^2}{c^2} - 4i_2\frac{D_p}{c^3} + \frac{i_1}{c^4} \right\} + \frac{1}{r}\left\{ i_6' + \frac{1}{2}D_p\left(D_p^2\frac{c^2}{r^2} + 1\right)\left(\frac{c'}{r'}\right)^4 \right\} \\
i_7^{(p)} &= i_7 + c^2\left\{ i_7' D_p^2 - i_8' D_p\frac{r'}{c'} + i_9'\left(\frac{r'}{c'}\right)^2 \right\} \\
i_8^{(p)} &= \frac{c'}{r'}\left(-\frac{2i_7}{c} + i_8 D_p \right) + c\left(i_8' D_p - 2i_9'\frac{r'}{c'} \right) \\
i_9^{(p)} &= \left(\frac{c'}{r'}\right)\left(\frac{i_7}{c^2} - \frac{i_8 D_p}{c} + i_9 D_p^2 \right) + i_9'
\end{aligned}
\tag{27.15}
$$

Another problem that is more easily solved with the aid of these aberration matrices than in any other way is that of determining the aberration coefficients for some magnification $\overline{M}$ given those for another magnification M. We have provided a solution in Chapter 25, Asymptotic Aberration Coefficients (Eqs 25.58−25.60) but this required a detailed knowledge of the aberration structure. The same result can be obtained with somewhat less effort by straightforward matrix multiplication. Suppose that the elements of $\boldsymbol{M}_2$ are known and that

$$\boldsymbol{u}_i = \boldsymbol{M}\boldsymbol{u}_o \tag{27.16}$$

If the object and image are shifted so that the magnification is now $\overline{M}$, we seek shift matrices $\boldsymbol{S}^{(o)}$ and $\boldsymbol{S}^{(i)}$ that take us from $\boldsymbol{u}_i$ to $\overline{\boldsymbol{u}}_i$ and from $\boldsymbol{u}_o$ to $\overline{\boldsymbol{u}}_o$, where

$$\overline{\boldsymbol{u}}_i = \overline{\boldsymbol{M}}\overline{\boldsymbol{u}}_o \tag{27.17}$$

Explicitly, we seek matrices $\boldsymbol{S}^{(o)}$ and $\boldsymbol{S}^{(i)}$ such that

$$\overline{\boldsymbol{u}}_i =: \boldsymbol{S}^{(i)}\boldsymbol{u}_i \quad \text{and} \quad \boldsymbol{u}_o =: \boldsymbol{S}^{(o)}\overline{\boldsymbol{u}}_o \tag{27.18}$$

and hence

$$\overline{\boldsymbol{u}}_i = \boldsymbol{S}^{(i)}\boldsymbol{u}_i = \boldsymbol{S}^{(i)}\boldsymbol{M}\boldsymbol{u}_o = \boldsymbol{S}^{(i)}\boldsymbol{M}\boldsymbol{S}^{(o)}\overline{\boldsymbol{u}}_o \tag{27.19}$$

The unknown matrix $\overline{\boldsymbol{M}}$ will then be obtained from (27.17) and (27.19):

$$\overline{\boldsymbol{M}} = \boldsymbol{S}^{(i)}\boldsymbol{M}\boldsymbol{S}^{(o)} \tag{27.20}$$

Since u_o and $\overline{u}_o$ are connected by the incident asymptote and u_i and $\overline{u}_i$, by the emergent asymptote we have

$$\begin{aligned} \overline{u}_i &= u_i + (\overline{z}_i - z_i)u'_i \qquad \overline{u}'_i = u'_i \\ u_o &= \overline{u}_o - (\overline{z}_o - z_o)\overline{u}'_o \qquad u'_o = \overline{u}'_o \end{aligned} \tag{27.21}$$

and so $\boldsymbol{S}^{(i)}$ and $\boldsymbol{S}^{(o)}$ will have block structures similar to that of $\boldsymbol{M}$:

$$\boldsymbol{S}^{(o)} = \begin{pmatrix} \boldsymbol{S}_1^{(o)} & \boldsymbol{S}_2^{(o)} \\ \boldsymbol{S}_3^{(o)} & \boldsymbol{S}_4^{(o)} \end{pmatrix} \quad \boldsymbol{S}^{(i)} = \begin{pmatrix} \boldsymbol{S}_1^{(i)} & \boldsymbol{S}_2^{(i)} \\ \boldsymbol{S}_3^{(i)} & \boldsymbol{S}_4^{(i)} \end{pmatrix} \tag{27.22}$$

where now $\boldsymbol{S}_2^{(o)} = \boldsymbol{S}_3^{(o)} = \boldsymbol{S}_2^{(i)} = \boldsymbol{S}_3^{(i)} = 0$,

$$\boldsymbol{S}_1^{(o)} = \begin{pmatrix} 1 & -\zeta_o \\ 0 & 1 \end{pmatrix} \quad \boldsymbol{S}_1^{(i)} = \begin{pmatrix} 1 & \zeta_i \\ 0 & 1 \end{pmatrix} \tag{27.23}$$

and

$$\boldsymbol{S}_4^{(o)} = \begin{pmatrix} 1 & -\zeta_o & -2\zeta_o & 2\zeta_o^2 & \zeta_o^2 & -\zeta_o^3 & 0 & 0 \\ 0 & 1 & 0 & -2\zeta_o & 0 & \zeta_o^2 & 0 & 0 \\ 0 & 0 & 1 & -\zeta_o & -\zeta_o & \zeta_o^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -\zeta_o & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\zeta_o & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\zeta_o \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{27.24}$$

$\boldsymbol{S}_4^{(i)}$ has the same appearance as $\boldsymbol{S}_4^{(o)}$ but ζ_i replaces $-\zeta_o$ everywhere. We have written

$$\zeta_o := \overline{z}_o - z_o \quad \zeta_i := \overline{z}_i - z_i \tag{27.25}$$

We are interested only in $\overline{\boldsymbol{M}}_2$ and from the common block structure of $\boldsymbol{S}^{(o)}$, $\boldsymbol{S}^{(i)}$ and $\boldsymbol{M}$, it is obvious that

$$\overline{\boldsymbol{M}}_2 = \boldsymbol{S}_1^{(i)} \boldsymbol{M}_2 \boldsymbol{S}_4^{(o)} \tag{27.26}$$

Since $\boldsymbol{S}_1^{(i)}$ is only 2×2 and $\boldsymbol{S}_4^{(o)}$ is so sparse, the elements of $\overline{\boldsymbol{M}}_2$ may be read off immediately and of course prove to be identical with those already listed (25.39). The expressions (25.59) for the change in aberrations when object and aperture are shifted can likewise be obtained with less effort by deriving the appropriate shift matrices, as discussed in detail by Brouwer (1957; cf. Hawkes, 1985a). The appropriate starting point is the matrix equation $\boldsymbol{u_i} = \boldsymbol{M}\boldsymbol{u}_o$ analogous to Eq. (27.16) except that u'_o is replaced by u_{ao} in $\boldsymbol{u}_o$ and u'_i by u_{ai} in $\boldsymbol{u}_i$, where u_{ao} is the complex Cartesian coordinate in the entrance pupil and u_{ai} is

that in the exit pupil, between which the magnification is N. The paraxial block matrix $\boldsymbol{M}_1$ therefore has the form

$$\boldsymbol{M}_1 = \begin{pmatrix} M & 0 \\ 0 & N \end{pmatrix} \tag{27.27}$$

and $\boldsymbol{M}_4$ is diagonal. $\boldsymbol{M}_3$ is of course still null and $\boldsymbol{M}_2$ is the new aberration matrix. The upper row of $\boldsymbol{M}_2$ has the usual form but the lower row, representing the aberrations between the pupil planes, must be derived from expressions analogous to (25.7a) or from (25.57−25.60), setting $m_2 = n_1$ and $n_2 = m_1$. The results are as follows, in which the new coefficient $\tilde{C}$ is equal to I_{04} (25.53):

$$\begin{aligned}
m_{21} &= \tilde{C} \\
m_{22} &= -D + \mathrm{i}d + \frac{n^2-1}{2f^2(n-m)^2} \\
m_{23} &= -2D + \frac{n^2-1}{f^2(n-m)^2} \\
m_{24} &= -2A + \mathrm{i}a + \frac{1-mn}{f^2(n-m)^2} \\
m_{25} &= -(F-A) + \frac{1-mn}{2f^2(n-m)^2} \\
m_{26} &= -K + \mathrm{i}k + \frac{m^2-1}{2f^2(n-m)^2} \\
m_{27} &= -2d \\
m_{28} &= -a
\end{aligned} \tag{27.28}$$

The shift matrices may be established without difficulty and the shifted aberration coefficients are obtained as the elements of $\boldsymbol{S}_1^{(i)}\boldsymbol{M}_2\boldsymbol{S}_4^{(o)}$ (27.26). The results agree with those listed earlier.

CHAPTER 28

The Aberrations of Mirrors and Cathode Lenses

In this chapter, we extend the theories introduced in Chapter 18, Electron Mirrors, beyond the paraxial domain, using the notation introduced there. There are again two possible representations of the theory, the temporal form and the Cartesian form. We go into details only for the former.

28.1 The Modified Temporal Theory

In order to go beyond the paraxial theory of Section 18.2, we retain the terms F_u and F_h in Eqs (18.5, 18.6). We reproduce the main steps in the calculation of the geometrical aberrations; for the chromatic aberrations, we simply list the expressions derived by Preikszas. As in Chapter 18, Electron Mirrors, only the nonrelativistic approximation is given.

We impose the condition that the aberrated ray coincides with its paraxial approximation at τ_T thereby maintaining the symmetry in τ. An arbitrary paraxial ray may be written

$$\begin{aligned} u(\tau) &= \overline{p}p(\tau) + \overline{q}q(\tau) \\ h(\tau) &= \nu h_\nu(\tau) + \sigma h_\sigma(\tau) \end{aligned} \tag{28.1}$$

With $e_1 = \phi'/4\phi_0$ we have

$$\begin{aligned} &\overline{p} = u(\tau_T) \quad \overline{q} = \dot{u}(\tau_T) \\ &\sigma = -2e_1(\zeta_T)h(\tau_T) \quad \nu = \frac{\dot{h}(\tau_T)}{2e_1(\zeta_T)} \end{aligned} \tag{28.2}$$

It is convenient to choose the time of departure of the arbitrary ray in such a way that the reference and arbitrary electrons reach their turning points at the same time τ_T. For this, $\dot{h}(\tau_T) = 0$. We know that $\dot{h}_\sigma(\tau_T) = 0$ and so $\nu\dot{h}_\nu(\tau_T)$ will vanish only if $\nu = 0$. Thus u and h depend only on $\overline{p}$, $\overline{q}$, $\overline{p}^*$, $\overline{q}^*$ and σ and we expand them as series in these variables:

$$u = \sum_1^\infty u^{(r)} \quad h = \sum_1^\infty h^{(r)}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00028-6
© 2018 Elsevier Ltd. All rights reserved.

On substituting these into F_u and F_h, we find

$$F_u = \sum_2^\infty F_u^{(r)} \quad F_h = \sum_2^\infty F_h^{(r)}$$

and

$$\begin{aligned} u^{(r)} &= u^{(r-1)} + c_p^{(r)} p + c_q^{(r)} q \\ h^{(r)} &= h^{(r-1)} + c_\sigma^{(r)} h_\sigma + c_\nu^{(r)} h_\nu \end{aligned} \tag{28.3}$$

where

$$\begin{aligned} c_p^{(r)} &= \int_{\tau_T}^{\tau} F_u^{(r)} q \, d\tau \quad c_q^{(r)} = -\int_{\tau_T}^{\tau} F_u^{(r)} p \, d\tau \\ c_\sigma^{(r)} &= \int_{\tau_T}^{\tau} F_h^{(r)} h_\nu \, d\tau \quad c_\nu^{(r)} = -\int_{\tau_T}^{\tau} F_h^{(r)} h_\sigma \, d\tau \end{aligned} \tag{28.4}$$

Finally, we obtain

$$\begin{aligned} u(\tau) &= c_p(\tau) p(\tau) + c_q(\tau) q(\tau) \\ z(\tau) &= \zeta(\tau) + h(\tau) = \zeta(\tau) + c_\sigma(\tau) h_\sigma(\tau) + c_\nu(\tau) h_\nu(\tau) \end{aligned} \tag{28.5}$$

$$c_p = \overline{p} + \sum_{r=2}^\infty c_p^{(r)} \quad c_q = \overline{q} + \sum_{r=2}^\infty c_q^{(r)} \quad c_\sigma = \sigma + \sum_{r=2}^\infty c_\sigma^{(r)} \quad c_\nu = \sum_{r=2}^\infty c_\nu^{(r)} \tag{28.6}$$

In order to express the aberrations in the same language as that of electron lenses, the time-like argument τ must be replaced by the axial coordinate z. We distinguish functions with argument z from those in τ by adding a circumflex:

$$\hat{u}(z) = u(\tau(z)) \tag{28.7}$$

In the Gaussian approximation,

$$\hat{u}^{(1)}(z) = \omega \hat{u}_\omega(z) + \rho \hat{u}_\rho(z) \tag{28.8}$$

We impose the condition that the Wronskian is equal to one:

$$(\hat{u}'_\omega \hat{u}_\rho - \hat{u}_\omega \hat{u}'_\rho)\dot{\zeta} = 1 \tag{28.9}$$

As boundary condition, the ray $\hat{u}(z)$ is required to coincide with the paraxial ray $\hat{u}^{(1)}(z)$ before entering the field of the mirror. After reflection at the mirror, the ray $\hat{u}(z)$ tends to the asymptote

$$\begin{aligned} &\hat{u}(z) \to c_\omega \hat{u}_\omega(z) + c_\rho \hat{u}_\rho(z) \\ &c_\omega = \omega + \sum_{r=2}^\infty c_\omega^{(r)} \quad c_\rho = \rho + \sum_{r=2}^\infty c_\rho^{(r)} \end{aligned} \tag{28.10}$$

The calculation is simplified if we choose the initial time τ_1 and the final time τ_2 in such a way that the positions of the reference electron lie in the same plane outside the mirror field. We write

$$\zeta_2 = \zeta(\tau_2) \quad \zeta_1 = \zeta(\tau_1) \quad \tau_1 + \tau_2 = 2\tau_T \tag{28.11}$$

and so $\dot{\zeta}(\tau_1) = -1, \ \dot{\zeta}(\tau_2) = 1$ outside the mirror field. Any ray is rectilinear outside the mirror field and the position and gradient at τ_2 are given by

$$\begin{aligned} \hat{u}_2 + \frac{\dot{u}_2}{1+\dot{h}_2} h_2 &= u_2 \\ \hat{u}'_2 &= \frac{\dot{u}_2}{1+\dot{h}_2} \end{aligned} \tag{28.12}$$

From this we deduce that

$$\begin{aligned} c_\omega &= \frac{\dot{u}_2}{1+\dot{h}_2} \hat{u}_{\rho 2} - \left(u_2 - \frac{\dot{u}_2}{1+\dot{h}_2} h_2 \right) \hat{u}'_{\rho 2} \\ c_\rho &= \left(u_2 - \frac{\dot{u}_2}{1+\dot{h}_2} h_2 \right) \hat{u}'_{\omega 2} - \frac{\dot{u}_2}{1+\dot{h}_2} \hat{u}_{\omega 2} \end{aligned} \tag{28.13}$$

We shall derive aberration coefficients in the Gaussian image plane, where $\hat{u}_\omega = 0$ (like $h(z)$ in round lens calculations) and we need only study c_ρ therefore. Preikszas and Rose (1997) denote the resulting coefficients R with four or five suffixes. Here, we write

$$c_\rho = R_{\omega^a \omega^{*b} \rho^c \rho^{*d} \kappa^k} \omega^a \omega^{*b} \rho^c \rho^{*d} \kappa^k =: R(\omega^a \omega^{*b} \rho^c \rho^{*d} \kappa^k) \omega^a \omega^{*b} \rho^c \rho^{*d} \kappa^k \tag{28.14}$$

Hitherto we have been considering only geometrical aberrations but as a reminder that the theory can be straightforwardly generalized to include chromatic aberrations, we have retained the chromatic aberration terms here ($\kappa = \Delta E / e\phi_0$). Thus $a + b + c + d$ is the order of the aberration, k is its degree and $a + b + c + d + k$ the rank. Thus $R(\omega k)$ is the axial chromatic aberration and $R(\omega\omega\omega^*)$ is the spherical aberration. These and the other third-order aberration coefficients are listed below.[1] We shall adopt Preikszas' very convenient shorthand for the derivatives of potential and magnetic flux

$$e_n = \frac{\phi^{(n)}}{4\phi_0} \quad b_n = -\frac{\eta}{2\phi_0^{1/2}} B^{(n-1)}(z) = -\frac{1}{2} \hat{\eta} B_z^{(n-1)} \tag{28.15}$$

[1] Expressions for the chromatic and spherical aberration coefficients when both electrostatic and magnetic fields are present are given in Preikszas (1995) and Preikszas and Rose (1997); the other third-order coefficients have been derived by Preikszas (personal communication, 2016) for the electrostatic case for inclusion here. We thank him warmly for this participation.

and note that, with $D := e_2 + b_1^2$, we find $\ddot{p} = -Dp \quad \dot{D} = Lh_\nu$. We define

$$\begin{aligned}
L(\lambda_1\lambda_2\lambda_3) &:= \int_{\tau_T}^{\tau} U(\lambda_1\lambda_2\lambda_3)q \; d\tau \\
X(\lambda_1\lambda_2\lambda_3) &:= -\int_{\tau_T}^{\tau} U(\lambda_1\lambda_2\lambda_3)p \; d\tau \\
S(\lambda_1\lambda_2\lambda_3) &:= \int_{\tau_T}^{\tau} H(\lambda_1\lambda_2\lambda_3)h_\nu \; d\tau \\
N(\lambda_1\lambda_2\lambda_3) &:= -\int_{\tau_T}^{\tau} H(\lambda_1\lambda_2\lambda_3)h_\sigma \; d\tau \\
u(\lambda_1\lambda_2\lambda_3) &:= L(\lambda_1\lambda_2\lambda_3)p + X(\lambda_1\lambda_2\lambda_3)q \\
h(\lambda_1\lambda_2\lambda_3) &:= S(\lambda_1\lambda_2\lambda_3)h_\sigma + N(\lambda_1\lambda_2\lambda_3)h_\nu
\end{aligned} \tag{28.16}$$

in which the parameters λ_i ($i = 1, 2, 3$) represent $\overline{p}$, $\overline{q}$, σ. The individual terms are given by:

$$\begin{aligned}
U(\overline{p}\sigma) &:= -ph_\sigma e_3 \\
U(\overline{q}\sigma) &:= -qh_\sigma e_3 \\
U(\overline{ppp}^*) &:= \frac{1}{8}p^3 e_4 - pe_3 h(pp^*) \\
U(\overline{ppq}^*) &:= \frac{1}{8}p^2 q e_4 - pe_3 h(pq^*) \\
U(\overline{pqp}^*) &:= \frac{1}{4}p^2 q e_4 - pe_3 h(pq^*) - qe_3 h(pp^*) \\
U(\overline{pqq}^*) &:= \frac{1}{4}pq^2 e_4 - pe_3 h(qq^*) - qe_3 h(pq^*) \\
U(\overline{qqp}^*) &:= \frac{1}{8}pq^2 e_4 - qe_3 h(pq^*) \\
U(\overline{qqq}^*) &:= \frac{1}{8}q^3 e_4 - qe_3 h(qq^*) \\
U(\overline{p}\sigma\sigma) &:= -\frac{1}{2}ph_\sigma^2 e_4 - pe_3 h(\sigma\sigma) - h_\sigma e_3 u(p\sigma) \\
U(\overline{q}\sigma\sigma) &:= -\frac{1}{2}qh_\sigma^2 e_4 - qe_3 h(\sigma\sigma) - h_\sigma e_3 u(q\sigma) \\
H(\overline{pp}^*) &:= -\frac{1}{2}p^2 e_3 \quad H(\overline{pq}^*) = H(\overline{qp}^*) = -\frac{1}{2}pqe_3 \\
H(\overline{qq}^*) &:= -\frac{1}{2}q^2 e_3 \quad H(\sigma\sigma) = h_\sigma^2 e_3
\end{aligned} \tag{28.17}$$

In the second stage, in which we transfer from τ to z, as explained above, we write

$$w(\lambda_1\lambda_2\lambda_3) = B(\lambda_1\lambda_2\lambda_3)w_\omega + R(\lambda_1\lambda_2\lambda_3)w_\rho$$

In the following formulae, we have lightened the notation by omitting bars in the arguments of L, N, S and X. Thus $L(\overline{p},\ \overline{q},\ \sigma)$, for example, appears as $L(p,\ q,\ \sigma)$.

Spherical aberration

$$\begin{aligned}
R(\omega\omega\omega^*) &= \dot{p}_2^2\left(\dot{p}_2^2 h_{\sigma 2} - p_2\dot{p}_2 - 2N(pp^*)_2 - X(p\sigma)_2\right) - 2X(p\sigma)_2 S(pp^*)_2 + 2X(ppp^*)_2 \\
&= \dot{p}_2^4 h_{\sigma 2} - p_2\dot{p}_2^3 - 2\int_{\tau_T}^{\tau}\left[\frac{1}{8}(e_4 + 4b_1b_3)p^4 + \left\{\dot{p}^2 + (e_2 + b_1^2)p^2\right\}(e_3 + 2b_1b_2)p^2 h_\sigma\right] d\tau \\
&= -2\dot{p}^2 N(pp^*) - p\dot{p}^3 + \dot{p}^4 h_\sigma + 2X(ppp^*) - \dot{p}^2 X(p\sigma) - 2S(pp^*)X(p\sigma) \\
B(\omega\omega\omega^*) &= 0
\end{aligned} \tag{28.19}$$

Coma

$$\begin{aligned}
B(\omega\omega\rho^*) &= 2L(ppq^*) + 2\dot{p}\dot{q}N(pq^*) + 2\dot{p}\dot{q}h_\sigma S(pq^*) - 2\dot{p}qS(pq^*) \\
R(\omega\omega\rho^*) &= p\dot{p}^2\dot{q} - \dot{p}^3\dot{q}h_\sigma + 2p\dot{p}S(pq^*) - 2\dot{p}^2 h_\sigma S(pq^*) + \dot{p}\dot{q}X(p\sigma) + 2X(p\sigma)S(pq^*) \\
B(\omega\omega^*\rho) &= 2\dot{q}^2 N(pp^*) + \dot{p}^2 q\dot{q} - \dot{p}^2\dot{q}^2 h_\sigma + 2L(pqp^*) + 2\dot{p}\dot{q}N(pq^*) \\
&\quad + 2\dot{p}\dot{q}h_\sigma S(pq^*) - \dot{p}^2 L(q\sigma) - 2S(pp^*)L(q\sigma) - 2\dot{p}qS(pq^*) \\
R(\omega\omega^*\rho) &= p\dot{p}^2\dot{q} - \dot{p}^3\dot{q}h_\sigma + 2p\dot{p}S(pq^*) - 2\dot{p}^2 h_\sigma S(pq^*) + \dot{p}\dot{q}X(p\sigma) + 2X(p\sigma)S(pq^*)
\end{aligned} \tag{28.20}$$

Astigmatism and field curvature

$$\begin{aligned}
B(\omega\rho\rho^*) &= -\dot{p}q\dot{q}^2 + \dot{p}\dot{q}^3 h_\sigma - 2q\dot{q}S(pq^*) + 2\dot{q}^2 h_\sigma S(pq^*) + \dot{p}\dot{q}L(q\sigma) + 2L(q\sigma)S(pq^*) \\
R(\omega\rho\rho^*) &= -2\dot{p}^2 N(qq^*) - p\dot{p}\dot{q}^2 + \dot{p}^2\dot{q}^2 h_\sigma + 2X(pqq^*) - 2\dot{p}\dot{q}N(pq^*) \\
&\quad - 2\dot{p}\dot{q}h_\sigma S(pq^*) - \dot{q}^2 X(p\sigma) - 2S(qq^*)X(p\sigma) + 2p\dot{q}S(pq^*) \\
B(\omega^*\rho\rho) &= -\dot{p}q\dot{q}^2 + \dot{p}\dot{q}^3 h_\sigma - 2q\dot{q}S(pq^*) + 2\dot{q}^2 h_\sigma S(pq^*) + \dot{p}\dot{q}L(q\sigma) + 2L(q\sigma)S(pq^*) \\
R(\omega^*\rho\rho) &= 2X(qqp^*) - 2\dot{p}\dot{q}N(pq^*) - 2\dot{p}\dot{q}h_\sigma S(pq^*) + 2p\dot{q}S(pq^*)
\end{aligned} \tag{28.21}$$

Distortion

$$\begin{aligned}
B(\rho\rho\rho^*) &= 2\dot{q}^2 N(qq^*) + q\dot{q}^3 - \dot{q}^4 h_\sigma + 2L(qqq^*) - \dot{q}^2 L(q\sigma) - 2S(qq^*)L(q\sigma) \\
R(\rho\rho\rho^*) &= 0
\end{aligned} \tag{28.22}$$

Second-rank chromatic aberration

$$\begin{aligned} R(\omega k)/a^2 &= p_2\dot{p}_2 - \dot{p}_2^2(h_{\sigma 2} + 2h_{\kappa 2}) + 2X((p\kappa)_2 + X(p\sigma)_2) \\ &= p_2\dot{p}_2 - \dot{p}_2^2 h_{\sigma 2} + \int_{\tau_T}^{\tau} Lp^2 h_\sigma \, d\tau \end{aligned}$$

or

$$\begin{aligned} R(\omega k) &= p\dot{p} - \dot{p}^2 h_\sigma + X(p\sigma) \\ B(\rho k) &= -q\dot{q} + \dot{q}^2 h_\sigma + L(q\sigma) \\ B(\omega k) &= R(\rho k) = 0 \end{aligned} \tag{28.23}$$

Third-rank chromatic aberration

$$\begin{aligned} B(\omega kk) &= \dot{p}^2 q\dot{q}h_\sigma - (1/2)\{\dot{p}^2\dot{q}^2 h_\sigma^2 + \dot{p}\dot{q}h_\sigma + p\dot{p}q\dot{q} + q\dot{q}X(p\sigma) \\ &\quad - \dot{q}^2 h_\sigma X(p\sigma) - p\dot{p}L(q\sigma) + \dot{p}^2 h_\sigma L(q\sigma) - X(p\sigma)L(q\sigma)\} \\ R(\omega kk) &= -(1/4)p\dot{p} + (1/4)\dot{p}^2 h_\sigma - (1/2)\dot{p}^2 N(\sigma\sigma) - p\dot{p}^2\dot{q}h_\sigma + (1/2)\dot{p}^3\dot{q}h_\sigma^2 \\ &\quad + (1/2)p^2\dot{p}\dot{q} + (1/4)X(p\sigma) + (1/2)X(p\sigma\sigma) - \dot{p}\dot{q}h_\sigma X(p\sigma) - (1/2)S(\sigma\sigma)X(p\sigma) \\ &\quad + p\dot{q}X(p\sigma) - (1/2)L(p\sigma)X(p\sigma) \\ B(\rho kk) &= (1/4)q\dot{q} - (1/4)\dot{q}^2 h_\sigma + (1/2)\dot{q}^2 N(\sigma\sigma) - \dot{p}q\dot{q}^2 h_\sigma + (1/2)\dot{p}\dot{q}^3 h_\sigma^2 \\ &\quad + (1/2)\dot{p}q^2\dot{q} + (1/4)L(q\sigma) + (1/2)L(q\sigma\sigma) + \dot{p}\dot{q}h_\sigma L(q\sigma) - (1/2)S(\sigma\sigma)L(q\sigma) \\ &\quad - \dot{p}qL(q\sigma) - (1/2)L(q\sigma)X(q\sigma) \\ R(\rho kk) &= p\dot{p}\dot{q}^2 h_\sigma - (1/2)\{\dot{p}^2\dot{q}^2 h_\sigma^2 - \dot{p}\dot{q}h_\sigma + p\dot{p}q\dot{q} - p\dot{p}L(q\sigma) \\ &\quad + \dot{p}^2 h_\sigma L(q\sigma) + q\dot{q}X(p\sigma) - \dot{q}^2 h_\sigma X(p\sigma) - X(p\sigma)L(q\sigma)\} \end{aligned} \tag{28.24}$$

Preikszas and Rose observe that the choice of unit magnification, for which

$$\hat{u}_{\omega 2} = ap_2 \quad \hat{u}'_{\omega 2} = a\dot{p}_2 \ \ (a \text{ arbitrary})$$

is particularly suitable for aberration correction.

One last point needs to be considered: the spatial and temporal Wronskians may not be the same. If they are not, then the coefficients need to be scaled appropriately. We can also allow for a change of reference potential from ϕ_0 to $\tilde{\phi}_0$ in the definition of κ.

Axial scaling:

$$\begin{aligned} A &:= w'_\omega/\dot{p} \quad \text{for} \quad \dot{p} \neq 0 \\ &:= w_\omega/p \quad \text{for} \quad \dot{p} = 0 \end{aligned} \tag{28.25a}$$

Field scaling

$$\begin{aligned} F &:= w'_\rho/\dot{q} \quad \text{for} \quad \dot{q} \neq 0 \\ &:= w_\rho/q \quad \text{for} \quad \dot{q} = 0 \end{aligned} \tag{28.25b}$$

Wronskian = 1/AF

$$W = \frac{1}{w'_{\omega}.w_{\rho} - w_{\omega}.w'_{\rho}} \tag{28.25c}$$

Energy scaling

$$K = \frac{\tilde{\phi}_0}{\phi_0} \tag{28.25d}$$

The coefficients then become

$$KB(\omega k),\ WKA^2R(\omega k),\ WKF^2B(\rho k),\ KR(\rho k),\ A^2B(\omega\omega\omega^*),\ WA^4R(\omega\omega\omega^*),\ AFB(\omega\omega\rho^*),\ A^2R(\omega\omega\rho^*),\ AFB(\omega\omega^*\rho),\ A^2R(\omega\omega^*\rho),\ F^2B(\omega\rho\rho^*),\ AFR(\omega\rho\rho^*),\ F^2B(\omega^*\rho\rho),\ AFR(\omega^*\rho\rho),\ WF^4B(\rho\rho\rho^*),\ F^2R(\rho\rho\rho^*),\ K^2B(\omega kk),\ WK^2A^2R(\omega kk),\ WK^2F^2B(\rho kk),\ K^2R(\rho kk). \tag{28.26}$$

Note that the upper limit of the integrals is τ_2.

As well as the cited publications, see Preikszas and Rose (1996a,b) and Rose and Preikszas (1995).

28.2 The Cartesian Theory

Cartesian theories of the aberrations of cathode lenses and, by extension, of electron mirrors have been presented intermittently for several decades and their legitimacy and accuracy have been explored in some detail. These theories fall into two families, those that remain very close to the familiar aberration theory of round lenses and those that introduce a new independent variable as explained in Chapter 18, Electron Mirrors. The first group begins with the paper of Ximen (1957), to be followed by Bonshtedt (1964) (who does not cite Ximen). The Chinese work was repeated and completed by Zhou et al. (1983) and by Ximen et al. (1983) but, meanwhile, numerous Russian publications had been published on the subject, in particular Kulikov (1971, 1972, 1973, 1975), Monastyrskii and Kulikov (1976, 1978), Monastyrskii (1978a,b, 1980), Kulikov et al. (1978) and Smirnov et al. (1979).

The second group is associated initially with the work of Kel'man et al. (1971a–c, 1972, 1973) who examined rotationally and cylindrically symmetric systems; their analysis was recast into a more general form by Daumenov et al. (1978).

It was clearly important to establish whether one of these approaches was superior and Dodin and Nesvizhskii (1981) in particular have examined the approximations involved. They come to the conclusion that the easier method associated with Bonshtedt, Kulikov et al. and Ximen is reliable.

The most detailed analyses of the aberrations by the earlier method are to be found in the papers of Kulikov et al. (1978), Monastyrskii and Kulikov (1978) and Ximen et al. (1983),

in which aberration integrals for the usual geometrical aberration coefficients and for chromatic aberration coefficients defined in terms of the axial component of the initial electron velocity are listed. The behaviour of these expressions close to the cathode is analysed by Monastyrskii (1978b). We draw particular attention to a more recent paper by Nesvizhskii (1986), who shows how the aberration coefficients of complex cathode systems can be derived with the aid of the traditional methods of aberration calculation by an ingenious change of variable (see also the references for Chapter 38 of Volume 2).

Similar lists of coefficients are to be found in the papers of Kel'man et al. already cited but the amount of computation needed seems very heavy when the chromatic aberrations are being calculated. This can be seen by returning to Eq. (18.23), which describes any polyenergetic paraxial beam, for which the energy $\hat{\varepsilon}$ will have a spectrum of values. For each of these, the corresponding turning point has to be found. Series expansions about each of these points must now be made: the origin is not the appropriate singularity when $\hat{\varepsilon} \neq 0$, as is shown in Fig. 28.1.

After a considerable amount of algebra, given in full by Yakushev and Sekunova (1986), integrals for the aberration coefficients are found. Here, we merely list the results including the improved form of the spherical aberration coefficient given by Bimurzaev et al. (2004). In the image plane, we have as usual

$$\begin{aligned}
\Delta u_i = {} & C(x_o'^2 + y_o'^2)u_o' && (\textit{spherical aberration}) \\
& + 2(K + \mathrm{i}k)(x_o'^2 + y_o'^2)u_o + (K - \mathrm{i}k)(x_o'^2 - y_o'^2 + 2\mathrm{i}x_o'y_o')u_o^* && (\textit{coma}) \\
& + (A + \mathrm{i}a)r_o^2(x_o' - \mathrm{i}y_o') && (\textit{astigmatism}) \\
& + Fr_o^2u_o' && (\textit{field curvature}) \\
& + (D + \mathrm{i}d)r_o^2u_o && (\textit{distortion}) \\
& + \frac{\varepsilon}{\phi_o}\left\{u_0'K_{1r} + u_o(K_{2r} + \mathrm{i}k_r)\right\}
\end{aligned} \tag{28.27}$$

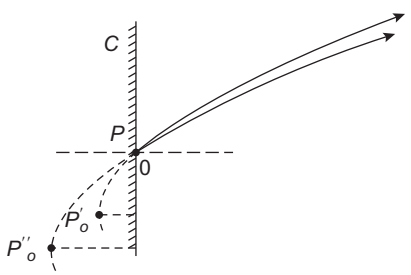


Figure 28.1
Points of inflexion P_o' and P_o'' of two rays starting from a common point P of a cathode surface with different initial velocities.

in which

$$
\begin{aligned}
C &= \frac{1}{R^3}\left\{J_1 Z_{ov}^4 + 2(2J_2 + J_3)Z_{ov}^2 Z_{oc}^2 + J_4 Z_{oc}^4 + 3Z_{ov}^2 + Z_{oc}^2\right\}\\
K &= -\frac{1}{R^3}\left\{J_1 Z_{ov}^3 + 2(2J_2 + J_3)Z_{ov}Z_{oc}Z_{oF} + J_4 Z_{oc}^3 + Z_{ov}Z_{oc}(2Z_{ov} + Z_{oc})\right\}\\
A &= \frac{1}{R^3}\left\{J_1 Z_{ov}^2 + 4J_2 Z_{oF}^2 + 2J_3 Z_{ov}Z_{oc} + J_4 Z_{oc}^2 + + J_7 R^2 + Z_{ov}(Z_{ov} + 2Z_{oc})\right\}\\
F &= \frac{1}{R^3}\left\{2J_1 Z_{ov}^2 + 4J_2(Z_{oF}^2 + Z_{ov}Z_{oc}) + J_3(Z_{ov}+Z_{oc})^2 + 2J_4 Z_{oc}^2 - J_7 R^2 + 2Z_{ov}(Z_{ov} + 2Z_{oc})\right\}\\
D &= -\frac{1}{R^3}\left\{J_1 Z_{ov} + 2(2J_2 + J_3)Z_{oF} + J_4 Z_{oc} + 3Z_{ov}\right\}\\
K_{1r} &= \frac{\phi_o}{R}(J_8 Z_{ov}^2 + J_9 Z_{oc}^2)\\
K_{2r} &= -\frac{\phi_o}{R}(J_8 Z_{ov} + J_9 Z_{oc})\\
k &= \frac{1}{R^2}(J_5 Z_{ov}^2 + J_6 Z_{oc}^2)\\
a &= -\frac{2}{R^2}(J_5 Z_{ov} + J_6 Z_{oc})\\
d &= \frac{1}{R^2}(J_5 + J_6)\\
k_r &= \phi_o J_{10}
\end{aligned}
\tag{28.28}
$$

and J_1–J_{10} denote

$$
\begin{aligned}
J_1 &= \frac{g'_o}{8\phi'_k p'^3_o}\int_{z_k}^{z_o}\frac{p}{\phi^{1/2}}\left\{R_4 p^3 + 4R'_2 p^2 p' + 32p''(\phi p'^2 + R_2 p^2/4)\right\}dz\\
J_2 &= \frac{1}{8\phi'_k p'_o g'_o}\int_{z_k}^{z_o}\frac{g}{\phi^{1/2}}\left\{R_4 p^2 g + 4R'_2 pp' g + 32p''(\phi p' g' + R_2 pg/4)\right\}dz\\
J_3 &= \frac{1}{8\phi'_k p'_o g'_o}\int_{z_k}^{z_o}\frac{q}{\phi^{1/2}}\left\{\phi R_4 p^2 q + 2R'_2 p^2(2\phi q' - \phi' q) + 32\phi q''(\phi p'^2 + R_2 p^2/4)\right\}dz
\end{aligned}
$$

$$
\begin{aligned}
J_4 &= \frac{p'_o}{8\phi'_k g'^3_o}\int_{z_k}^{z_o}\phi^{1/2}q\left\{\phi R_4 q^3 + 2R'_2 q^2(2\phi q' - \phi' q) + 32q''(\phi g'^2 + R_2 g^2/4)\right\}dz \\
J_5 &= \frac{\sigma\eta}{8p'^2_o}\int_{z_k}^{z_o}\frac{p}{\phi^{1/2}}\left\{B''p + 4B'p' + 8Bp''\right\}dz \\
J_6 &= -\frac{\sigma\eta}{16p'^2_o}\int_{z_k}^{z_o}\frac{q}{\phi^{1/2}}\left\{5\phi' B' q - 4\phi(B'q' + 4Bq'')\right\}dz \\
J_7 &= \frac{\eta^2}{8p'_o g'_o}\int_{z_k}^{z_o}\frac{1}{\phi^{3/2}}\left\{2B_k\phi B' + B(\phi'_k B - B_k\phi')\right\}dz \\
J_8 &= -\frac{4g'_o}{\phi'_k p'_o}\int_{z_k}^{z_o}\frac{pp''}{\phi^{1/2}}dz \\
J_9 &= -\frac{4p'_o}{\phi'_k g'_o}\int_{z_k}^{z_o}\phi^{1/2}qq''dz \\
J_{10} &= \frac{\sigma\eta}{2\phi'_k}\int_{z_k}^{z_o}\frac{1}{\phi^{3/2}}\left\{2\phi B' + B(\phi'_k - \phi')\right\}dz
\end{aligned}
\tag{28.29}
$$

The distances Z are as follows:

$$
\begin{aligned}
Z_{ov} &= z_o - z_v \quad Z_{oc} = z_o - z_c \quad Z_{oF} = z_o - z_F \\
Z_v &= z - z_v \quad Z_c = z - z_c \quad Z_F = z - z_F \quad Z_o = z - z_o
\end{aligned}
\tag{28.30}
$$

In the special case of an electrostatic mirror operating at high magnification, for which the object plane and the focus coincide, the expression for C_s collapses to

$$C_s = \tilde{J}_1 + \tilde{J}_2 + \tilde{J}_3 + \delta \tag{28.31}$$

in which convenient forms of the four contributions are (Bimurzaev et al., 2004):

$$
\begin{aligned}
\tilde{J}_1 &= \frac{1}{256p'^4_o\phi^{1/2}_o}\int_{z_k}^{z_o}\int_{z_k}^{z_o}\left[\phi^{-1/2}\left\{\frac{\phi'''_k\phi'}{2\phi}(p^4-1) + 32pp''f_1\right\} - \frac{1}{2}\left\{\phi'' - \phi''_k - \phi'''_k(z-z_k)\right\}\left(\frac{p^4\phi'}{\phi^{3/2}}\right)'\right]dz \\
\tilde{J}_2 &= \frac{1}{(16p'_o g'_o)^2\phi^{1/2}_o}\int_{z_k}^{z_o}\left[(\phi''_k - \phi'')\left\{\frac{2pq}{\phi^{1/2}}\left(\tfrac{5}{2}\phi' pq + \phi'_k\right)\right\}' + 64q(p''f_2 + \phi^{1/2}q''f_1\right]dz \\
\tilde{J}_3 &= \frac{1}{256g'^4_o\phi^{1/2}_o}dz\int_{z_k}^{z_o}\left\{\frac{7}{2}(\phi'' - \phi''_k)(\phi^{1/2}q^4\phi')' + 32\phi^{1/2}qq''f_3\right\}dz \\
\delta &= \frac{\phi'_k}{16\phi^{1/2}_o p'_o g'_o}\left(\frac{3}{2} + \frac{\phi''\phi'^2_k}{(16p'_o g'_o)^2\phi^2_o}\right) - \frac{\phi'''_k}{256p'^4_o\phi_o}
\end{aligned}
\tag{28.32}
$$

If a magnetic field is also present, Eq. (28.31) must be replaced by

$$C_s = \tilde{J}_1 + \tilde{J}_2 + \tilde{J}_3 + \frac{\eta^2}{64 p_o'^4 \phi_o^{1/2}} (\hat{J}_1 + \hat{J}_2 + \hat{J}_3) + \delta \tag{28.33}$$

in which (Bimurzaev, 2006, personal communication)

$$\hat{J}_1 = \int_{z_k}^{z_o} \frac{p^2 B}{\phi^{1/2}} (p^2 B'' + 2pp'B') dz = \int_{z_k}^{z_o} \left(\frac{p^2 B}{\phi^{1/2}} \right)' (B_k' - p^2 B') dz$$

$$\hat{J}_2 = 2 \int_{z_k}^{z_o} \phi^{1/2} pqB (3pqB'' + 4p'qB' + 2pq'B') dz = - \int_{z_k}^{z_o} pqB' \left\{ 6\phi^{1/2} (pqB)' + B\phi^{-1/2} (2pq\phi' + \phi_k') \right\} dz$$

$$\hat{J}_3 = \int_{z_k}^{z_o} \phi^{1/2} q^2 B \left\{ \phi (q^2 B'' + 2qq'B') - \phi' q^2 B' \right\} dz = - \int_{z_k}^{z_o} \phi^{1/2} q^2 B' \left\{ \phi (q^2 B)' + 5\phi' q^2 B/2 \right\} dz \tag{28.34}$$

For more details, we refer to the papers cited above, and in particular to those of Yakushev and Sekunova (1986), Takaoka (1995), Bimurzaev and Yakushev (2004) and Zhukov and Zav'yalova (2006), who examined a combined electrostatic−magnetic mirror. For earlier work on these topics, or other approaches, see Artsimovich (1944), Ximen (1957), Ximen et al. (1983), Ioanoviciu et al. (1989), Lenc and Müllerová (1992a,b) and Flory et al. (1996) and for a fresh investigation, Wang et al. (2008a,b) and Yakushev (2013).

28.3 Devices with Curved Cathodes

So far we have tacitly assumed the object to be planar but this excludes cathode lenses with curved cathode surfaces. Concave spherical cathodes are, however, quite common in image converters. Usually, the radius R of curvature is fairly large, so that $2z = r^2/R$ is a sufficiently good approximation for the equation of the cathode surface (Fig. 28.2).

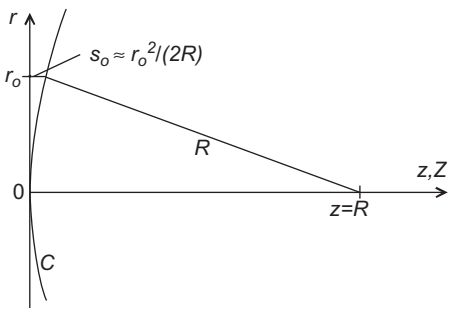


Figure 28.2
Parameters characterizing a spherical cathode surface *C*.

The theory can easily be generalized to include this case. The cathode surface must be an equipotential $\Phi(\boldsymbol{r}) = 0$. From the series expansion (7.37), we see that

$$\Phi(z, r) = z\phi'_o + \frac{z^2}{2}\phi''_o - \frac{r^2}{4}\phi''_o + \ldots \tag{28.35}$$

is valid in the vicinity of the cathode. Introducing $2z = r^2/R$ for the surface $\Phi = 0$ and retaining only terms of order r^2, we find that

$$\phi''_o = 2\phi'_o/R \tag{28.36}$$

It only remains to define the initial conditions for the longitudinal aberration $h(\tau)$. We now have

$$\tau_o = 0, \quad \zeta_o = 0, \quad z_o \equiv h_o = r_o^2/2R \tag{28.37}$$

Recalling that $\Phi(\boldsymbol{r}_o)$ is zero on the cathode surface, we obtain the condition

$$\dot{z}_o^2 \equiv \dot{h}_o^2 = \Phi_T/U, \quad \dot{h}_o \geq 0 \tag{28.38}$$

Apart from these special initial conditions, no other aspects of systems with a curved cathode need be considered here. Note that although we have suggested how systems with a curved cathode surface may be included, the present theory must not be applied to electron guns with sharply pointed cathodes. In the latter, the curvature of the cathode is so great and the associated aberrations are so large that a series expansion in the form (28.27) makes no sense. A completely different approach is then necessary, which is the subject of Part IX.

28.4 Practical Mirror Studies

The expressions for the aberration coefficients of electron lenses, deflectors and quadrupoles are simple enough for us to deduce general properties; the sign of the spherical aberration coefficient is a good example. The same is not true of the aberration coefficients of mirrors listed in the previous sections, with the result that practical studies of mirrors have relied on simple models or numerical results. Thus the many publications on mirrors as correctors for the photoemission electron microscope from the Portland State University group use a "hyperbolic" model, inspired by a suggestion of Rüdenberg (1948, cf. Glaser and Robl, 1951). The simplest version, a two-electrode or diode mirror in which the anode has the form of a hyperbola of revolution with a small hole on the axis, was studied at length by Rempfer (1990a), following preliminary work by Rempfer and Mauck (1985, 1986). The potential is assumed to be given by

$$\Phi(r, z) = \Phi_M + k\left(z^2 - r^2/2\right) \tag{28.39}$$

where Φ_M is the potential on the cone $z = r/\sqrt{2}$ and $k = (\Phi_A - \Phi_M)/L^2$; Φ_A denotes the anode potential and L is defined in Fig. 28.3. The potential of such a mirror for aberration

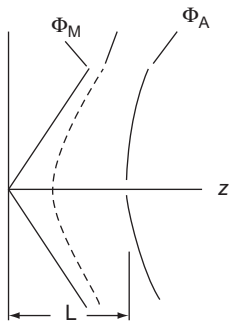


Figure 28.3
Rempfer's hyperbolic mirror model.

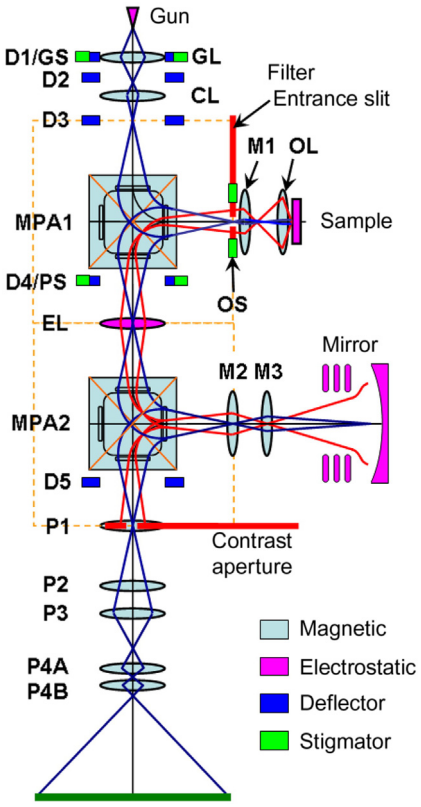


Figure 28.4
A LEEM/PEEM instrument with a mirror aberration corrector. After Tromp et al. (2013). *Courtesy:* Elsevier.

correction is discussed by Shao and Wu (1989, 1990a,b), whose work was criticized by Rempfer (1990b), and by Rempfer (1990a) and Rempfer and Mauck (1992) and again by Rempfer et al. (1997). It was revived by Könenkamp et al. (2008). The diode mirror has the serious defect that the spherical and chromatic aberration coefficients cannot be varied independently for a given geometry and a three-electrode mirror was therefore investigated by Fitzgerald et al. (2011, 2012a). This too has limitations and four-electrode mirrors were explored for the SpectroMicroscope for All Relevant Techniques (SMART) and PEEM3 instruments (see Chapters 37 and 41 of Volume 2). A less complicated aberration-corrected LEEM−PEEM instrument incorporating an energy filter (Fig. 28.4) has been constructed by Tromp et al. (2010, 2013). Further work on this theme is mentioned in Chapters 37 and 41 (Tromp, 2015a,b; Geelen et al., 2015; Mankos, 2011; Mankos and Shadman, 2013; Mankos et al., 2008, 2010, 2014).

CHAPTER 29

The Aberrations of Quadrupole Lenses and Octopoles

29.1 Introduction

The aberration coefficients of systems characterized by a pair of symmetry planes are of less interest at moderate energies than those of rotationally symmetric components but they are needed in two practical situations: instruments in which a line focus is required and quadrupole–octopole systems intended to reduce or cancel the spherical and chromatic aberration of round lenses. In the latter case, a sequence of quadrupole lenses, typically four, may be suitably combined with octopoles either as a corrector, the combination providing little or no paraxial focusing, or as a corrected lens, in which case both focusing and correction are required. Spherical aberration compensation by means of quadrupoles and octopoles was one of the ways of circumventing Scherzer's theorem proposed by Scherzer himself (1947); the idea attracted much attention over the years, both experimental and theoretical, and was finally successful in the 1990s, as we shall see in Chapter 41 of Volume 2.

29.2 Geometrical Aberration Coefficients

Although it is in theory necessary to distinguish between real and asymptotic coefficients, we concentrate on the latter for it is unlikely that a real specimen or target would be immersed within the field of a quadrupole. Real aberration coefficients will therefore be mentioned only in passing; they are accorded much more space in earlier surveys (Hawkes, 1966, 1970a) and in the textbooks of Strashkevich (1959, 1966), Yavor (1968) and Baranova and Yavor (1986, 1989).

We set out from the perturbation characteristic function denoted by S^I_{12} in Chapter 22, Perturbation Theory: General Formalism (Eq. 22.21); for two planes $z = z_o$ and $z = z_c$, we may write

$$S^I_{oc} = \begin{pmatrix} x_o^2 \\ y_o^2 \\ x_o'^2 \\ y_o'^2 \\ x_o x_o' \end{pmatrix}^T \begin{pmatrix} (\overline{4000}) & (\overline{2200}) & (\overline{2020}) & (\overline{2002}) & (\overline{3010}) & (\overline{2101}) \\ 0 & (\overline{0400}) & (\overline{0220}) & (\overline{0202}) & (\overline{1210}) & (\overline{0301}) \\ 0 & 0 & (\overline{0040}) & (\overline{0022}) & (\overline{1030}) & (\overline{0121}) \\ 0 & 0 & 0 & (\overline{0004}) & (\overline{1012}) & (\overline{0103}) \\ 0 & 0 & 0 & 0 & 0 & (\overline{1111}) \end{pmatrix} \begin{pmatrix} x_o^2 \\ y_o^2 \\ x_o'^2 \\ y_o'^2 \\ x_o x_o' \\ y_o y_o' \end{pmatrix} \tag{29.1}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00029-8
© 2018 Elsevier Ltd. All rights reserved.

The notation for the matrix elements has been chosen so that

$$S^I_{oc} = \sum_{0 \le p,q,r,s \le 4} (\overline{pqrs})\, x_o^p y_o^q x_o'^r y_o'^s, \quad p+q+r+s = 4 \tag{29.2}$$

Introducing the dimensionless coordinates

$$\xi := \frac{x_o}{f_{xi}}, \quad \eta := \frac{y_o}{f_{yi}} \tag{29.3}$$

S^I_{oc} becomes

$$S^I_{oc} = \begin{pmatrix} \xi_o^2 \\ \eta_o^2 \\ x_o'^2 \\ y_o'^2 \\ \xi_o x_o' \end{pmatrix}^T \begin{pmatrix} (4000) & (2200) & (2020) & (2002) & (3010) & (2101) \\ 0 & (0400) & (0220) & (0202) & (1210) & (0301) \\ 0 & 0 & (0040) & (0022) & (1030) & (0121) \\ 0 & 0 & 0 & (0004) & (1012) & (0103) \\ 0 & 0 & 0 & 0 & 0 & (1111) \end{pmatrix} \begin{pmatrix} \xi_o^2 \\ \eta_o^2 \\ x_o'^2 \\ y_o'^2 \\ \xi_o x_o' \\ \eta_o y_o' \end{pmatrix} \tag{29.4}$$

with

$$\begin{array}{lll} (4000) = (\overline{4000})\ f_{xi}^4 & (0400) = (\overline{0400})\ f_{yi}^4 & (0022) = (\overline{0022}) \\ (2200) = (\overline{2200})\ f_{xi}^2 f_{yi}^2 & (0220) = (\overline{0220})\ f_{yi}^2 & (1030) = (\overline{1030})\ f_{xi} \\ (2020) = (\overline{2020})\ f_{xi}^2 & (0202) = (\overline{0202})\ f_{yi}^2 & (0121) = (\overline{0121})\ f_{yi} \\ (2002) = (\overline{2002})\ f_{xi}^2 & (1210) = (\overline{1210})\ f_{xi} f_{yi}^2 & (0004) = (\overline{0004}) \\ (3010) = (\overline{3010})\ f_{xi}^3 & (0301) = (\overline{0301})\ f_{yi}^3 & (1012) = (\overline{1012})\ f_{xi} \\ (2101) = (\overline{2101})\ f_{xi}^2 f_{yi} & (0040) = (\overline{0040}) & (0103) = (\overline{0103})\ f_{yi} \\ & & (1111) = (\overline{1111})\ f_{xi} f_{yi} \end{array} \tag{29.5}$$

or

$$(pqrs) = f_{xi}^p f_{yi}^q (\overline{pqrs})$$

If the lens or system of lenses is *symmetric*, in the sense that all the field or potential functions are symmetric about some midplane, the object and image focal lengths are equal and several of the $(pqrs)$ are likewise equal. To see this, we write $x_o \to -f_x x'_c$, $y_o \to -f_y y'_c$, $x'_o \to x_c/f_x =: \zeta_c$ and $y'_o \to y_c/f_y =: \eta_c$ in (29.4), giving

$$S^I_{oc} = \begin{pmatrix} \xi_c^2 \\ \eta_c^2 \\ x_c'^2 \\ y_c'^2 \\ \xi_c x_c' \end{pmatrix}^T \begin{pmatrix} (0040) & (0022) & (2020) & (0220) & (1030) & (0121) \\ 0 & (0004) & (2002) & (0202) & (1012) & (0103) \\ 0 & 0 & (4000) & (2200) & (3010) & (2101) \\ 0 & 0 & 0 & (0400) & (1210) & (0301) \\ 0 & 0 & 0 & 0 & 0 & (1111) \end{pmatrix} \begin{pmatrix} \xi_c^2 \\ \eta_c^2 \\ x_c'^2 \\ y_c'^2 \\ \xi_c x_c' \\ \eta_c y_c' \end{pmatrix} \tag{29.6}$$

so that

$$\begin{array}{ll} (4000) = (0040) & (1030) = (3010) \\ (0400) = (0004) & (0103) = (0301) \quad \text{(symmetric system)} \\ (2200) = (0022) & (0121) = (2101) \\ (2002) = (0220) & (1210) = (1012) \end{array} \tag{29.7}$$

or

$$(pqrs) = (rspq)$$

In another common combination, the system is *antisymmetric*: the centre plane is a plane of geometrical symmetry and electrical antisymmetry (Fig. 29.1). Writing Eq. (29.4) in the form

$$S^I_{oc} = \begin{pmatrix} \eta_c^2 \\ \xi_c^2 \\ y_c'^2 \\ x_c'^2 \\ \eta_c y_c' \end{pmatrix}^T \begin{pmatrix} (0004) & (0022) & (0202) & (2002) & (0103) & (1012) \\ 0 & (0040) & (0220) & (2020) & (0121) & (1030) \\ 0 & 0 & (0400) & (2200) & (0301) & (1210) \\ 0 & 0 & 0 & (4000) & (2101) & (3010) \\ 0 & 0 & 0 & 0 & 0 & (1111) \end{pmatrix} \begin{pmatrix} \eta_c^2 \\ \xi_c^2 \\ y_c'^2 \\ x_c'^2 \\ \eta_c y_c' \\ \xi_c x_c' \end{pmatrix} \tag{29.8}$$

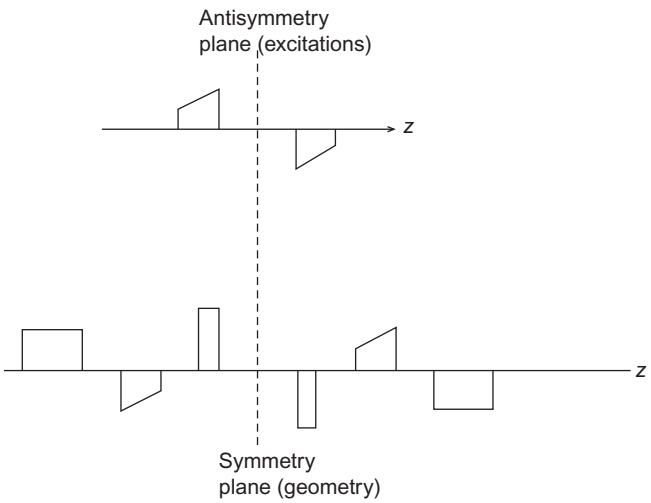


Figure 29.1
Antisymmetric multiplets. The simplest case is the doublet (above) but the quadruplet and the sextuplet (below), having more variable parameters, are more useful.

we see that

$$\begin{array}{ll} (4000) = (0004) & (2101) = (1012) \\ (0022) = (2200) & (0400) = (0040) \\ (2020) = (0202) & (1210) = (0121) \\ (3010) = (0103) & (0301) = (1030) \end{array} \tag{29.9}$$

or

$$(pqrs) = (srqp)$$

Proceeding as in Chapters 22 and 24, Perturbation Theory: General Formalism and The Geometrical Aberrations of Round Lenses, we obtain expressions for the aberrations but with an additional degree of complexity, already mentioned in Chapter 19, Quadrupole Lenses. A quadrupole lens focuses a point object into a pair of real or virtual line foci, and since quadrupoles are mostly used in combination, the 'object' for an intermediate member of a multiplet will almost always be astigmatic. We therefore list the aberrations for the general case of astigmatic object and image. If z_{ox} and z_{ix} are conjugate in the plane $x-z$ and z_{oy} and z_{iy} in the plane $y-z$, with magnifications M_x and M_y respectively, we write

$$\begin{aligned} \Delta x_i &\coloneqq \frac{x(x_{ix}) - M_x x_o}{M_x} \\ \Delta y_i &\coloneqq \frac{y(z_{iy}) - M_y y_o}{M_y} \end{aligned} \tag{29.10}$$

and

$$\begin{aligned} \Delta x_i =: {} & \left\{(3000)\xi_o^2 + (1200)\eta_o^2\right\}\xi_o && \text{distortions} \\ & + \left\{(2010)\xi_o^2 + (0210)\eta_o^2\right\}x'_o + (1101)\xi_o\eta_o y'_o && \left\{\begin{array}{l}\text{field curvature}\\ \text{and astigmatism}\end{array}\right. \\ & + \left\{(1020)x_o'^2 + (1002)y_o'^2\right\}\xi_o + (0111)\eta x'_o y'_o && \text{comas} \\ & + \left\{(0030)x_o'^2 + (0012)y_o'^2\right\}x'_o && \left\{\begin{array}{l}\text{aperture}\\ \text{aberrations}\end{array}\right. \end{aligned} \tag{29.11}$$

$$\begin{aligned} \Delta y_i =: {} & \left\{(0300)\eta_o^2 + (2100)\xi_o^2\right\}\eta_o && \text{distortions} \\ & + \left\{(2001)\xi_o^2 + (0201)\eta_o^2\right\}y'_o + (1110)\xi_o\eta_o x'_o && \left\{\begin{array}{l}\text{field curvature}\\ \text{and astigamtism}\end{array}\right. \\ & + \left\{(0120)x_o'^2 + (0102)y_o'^2\right\}\eta_o + (1011)\xi_o x'_o y'_o && \text{comas} \\ & + \left\{(0003)y_o'^2 + (0021)x_o'^2\right\}y'_o && \left\{\begin{array}{l}\text{aperture}\\ \text{aberrations}\end{array}\right. \end{aligned}$$

In the case of real aberrations and a stigmatic object, $x_c^{(1)}$ and $y_c^{(1)}$ are given by Eq. (22.30), into which we substitute

$$\begin{aligned} x(z) &= x_o s_x(z) + x_a t_x(z) \\ y(z) &= y_o s_y(z) + y_a t_y(z) \end{aligned} \tag{29.12}$$

giving

$$\begin{aligned} x_c^{(1)}(z) &= \frac{t_x}{W_x}\frac{\partial S_{ac}^I}{\partial x_o} - \frac{g_x}{W_x}\frac{\partial S_{oc}^I}{\partial x_a} \\ y_c^{(1)}(z) &= \frac{t_y}{W_y}\frac{\partial S_{ac}^I}{\partial y_o} - \frac{g_y}{W_y}\frac{\partial S_{oc}^I}{\partial y_a} \end{aligned} \tag{29.13}$$

where W_x, W_y are the appropriate Wronskians. In the line-image plane where $t_x = 0$, $s_x = M_x$ for example, we have

$$\Delta x_i = -\frac{1}{W_x}\frac{\partial S_{oc}^I}{\partial x_a} \tag{29.14a}$$

while in the other line-image plane, where $t_y = 0$ and $g_y = M_y$,

$$\Delta y_i = -\frac{1}{W_y}\frac{\partial S_{oc}^I}{\partial y_a} \tag{29.14b}$$

In the case of asymptotic aberrations, we generalize Eq. (25.9) to include astigmatic imagery:

$$\begin{aligned} S_x^I &= S^I(z_{ox}, -\infty) + S^I(-\infty, \infty) + S^I(\infty, z_{ix}) \\ S_y^I &= S^I(z_{oy}, -\infty) + S^I(-\infty, \infty) + S^I(\infty, z_{iy}) \end{aligned} \tag{29.15}$$

and the aberration coefficients are then obtained from equations analogous to Eq. (25.7):

$$\begin{aligned} x^{(1)}(z_c) &= \frac{1}{\hat{\phi}_o^{1/2}}\left(\tilde{H}_{xc}\frac{\partial S_x^I}{\partial x_o} - \tilde{G}_{xc}\frac{\partial S_x^I}{\partial x_o'}\right) - \frac{1}{2}\tilde{H}_{xc}x_o'(x_o'^2 + y_o'^2) \\ y^{(1)}(z_c) &= \frac{1}{\hat{\phi}_o^{1/2}}\left(\tilde{H}_{yc}\frac{\partial S_y^I}{\partial y_o} - \tilde{G}_{yc}\frac{\partial S_y^I}{\partial y_o'}\right) - \frac{1}{2}\tilde{H}_{yc}y_o'(x_o'^2 + y_o'^2) \end{aligned} \tag{29.16}$$

in which the tilde indicates that the asymptote to $G_x(z)$, $G_y(z)$, $H_x(z)$, $H_y(z)$ is to be understood.

Each of the numerous coefficients appearing in Eq. (29.11) has a polynomial structure in inverse magnification, but now the two inverse magnifications, $m_x \coloneqq 1/M_x$ and $m_y \coloneqq 1/M_y$, appear. Explicitly, we find

Aperture aberrations

$$\begin{aligned}(0030) &= \sum_{i=0}^{4}(0030)_i m_x^i \quad (0012) = \sum_{i,j=0}^{2}(0012)_{ij} m_x^i m_y^j \\ (0003) &= \sum_{i=0}^{4}(0003)_i m_y^i \quad (0021) = \sum_{i,j=0}^{2}(0021)_{ij} m_x^i m_y^j\end{aligned} \tag{29.17a}$$

Comas

$$\begin{aligned}(1020) &= \sum_{i=0}^{3}(1020)_i m_x^i \qquad (0102) = \sum_{i=0}^{3}(0102)_i m_y^i \\ (1002) &= \sum_{i=0}^{1}\sum_{j=0}^{2}(1002)_{ij} m_x^i m_y^j \quad (0120) = \sum_{i=0}^{2}\sum_{j=0}^{1}(0120)_{ij} m_x^i m_y^j \\ (0111) &= \sum_{i=0}^{2}\sum_{j=0}^{1}(0111)_{ij} m_x^i m_y^j \quad (1011) = \sum_{i=0}^{1}\sum_{j=0}^{2}(1011)_{ij} m_x^i m_y^j\end{aligned} \tag{29.17b}$$

Astigmatisms

$$\begin{aligned}(2010) &= \sum_{i=0}^{2}(2010)_i m_x^i \qquad (0201) = \sum_{i=0}^{2}(0201)_i m_y^i \\ (0210) &= \sum_{i=0}^{2}(0210)_i m_x^i \qquad (2001) = \sum_{i=0}^{2}(2001)_i m_y^i \\ (1101) &= \sum_{i,j=0}^{1}(1101)_{ij} m_x^i m_y^j \quad (1110) = \sum_{i,j=0}^{1}(1110)_{ij} m_x^i m_y^j\end{aligned} \tag{29.17c}$$

Distortions

$$\begin{aligned}(3000) &= \sum_{i=0}^{1}(3000)_i m_x^i \quad (1200) = \sum_{i=0}^{1}(1200)_i m_x^i \\ (0300) &= \sum_{i=0}^{1}(0300)_i m_y^i \quad (2100) = \sum_{i=0}^{1}(2100)_i m_y^i\end{aligned} \tag{29.17d}$$

The coefficients $(pqrs)_i$ and $(pqrs)_{ij}$ are themselves simply related to a set of integrals. We write (see Eq. 25.27)

$$r = \frac{f_{zo}}{f_{xi}} = \frac{f_{yo}}{f_{yi}} = \left(\frac{\hat{\phi}_o}{\hat{\phi}_i}\right)^{1/2} \tag{29.18}$$

Aperture aberrations

$$
\begin{aligned}
(0030)_4 &= -(4000)r^3 & (0003)_4 &= -4(0400)r^3\\
(0030)_3 &= \left\{4(3010) - \frac{f_{xo}}{2}\right\}r^2 & (0003)_3 &= \left\{4(0301) - \frac{f_{yo}}{2}\right\}r^2\\
(0030)_2 &= -4(2020)r & (0003)_2 &= -4(0202)r\\
(0030)_1 &= 4(1030) - \frac{f_{xo}}{2} & (0003)_1 &= 4(0103) - \frac{f_{yo}}{2}\\
(0030)_0 &= -4(0040)r^{-1} & (0003)_0 &= -4(0004)r^{-1}\\
\\
(0012)_{22} &= -2(2200)_F r^3 & (0021)_{22} &= -2(2200)_G r^3\\
(0012)_{21} &= 2(2101)r^2 & (0021)_{21} &= \left\{2(2101) - \frac{f_{yo}}{2}\right\}r^2\\
(0012)_{12} &= \left\{(2110) - \frac{f_{xo}}{2}\right\}r^2 & (0021)_{12} &= 2(1210)r^2\\
(0012)_{20} &= -2(2002)r & (0021)_{20} &= -2(2002)r\\
(0012)_{11} &= -2(1111)r & (0021)_{11} &= -2(1111)r\\
(0012)_{02} &= -2(0220)r & (0021)_{02} &= -2(0220)r\\
(0012)_{10} &= 2(1012) - \frac{f_{xo}}{2} & (0021)_{10} &= 2(1012)\\
\\
(0012)_{01} &= 2(0121) & (0021)_{01} &= 2(0121) - \frac{f_{yo}}{2}\\
(0012)_{00} &= -2(0022)_F r^{-1} & (0021)_{00} &= -2(0022)_G r^{-1}
\end{aligned}
\tag{29.19a}
$$

Comas

$$
\begin{aligned}
(1020)_3 &= 12(4000)r^2 & (0102)_3 &= 12(0400)r^2\\
(1020)_2 &= \left\{-9(3010) + 3\frac{f_{xo}}{2}\right\}r & (0102)_2 &\left\{-9(0301) + 3\frac{f_{yo}}{2}\right\}r\\
(1020)_1 &= 6(2020) & (0102)_1 &= 6(0202)\\
(1020)_0 &= -3(1030)r^{-1} & (0102)_0 &= -3(0103)r^{-1}\\
(1002)_{12} &= 2(2200)_F r^2 & (0120)_{21} &= 2(2200)_G r^2\\
(1002)_{11} &= -2(2101)r & (0120)_{11} &= -2(1210)r\\
(1002)_{02} &= \left\{-(1210) + \frac{f_{xo}}{2}\right\}r & (0120)_{20} &= \left\{-(2101) + \frac{f_{yo}}{2}\right\}r\\
(1002)_{10} &= 2(2002) & (0120)_{10} &= (1111)\\
(1002)_{01} &= (1111) & (0120)_{01} &= 2(0220)\\
(1002)_{00} &= -(1012)r^{-1} & (0120)_{00} &= -(0121)r^{-1}
\end{aligned}
$$

$$
\begin{aligned}
(0111)_{21} &= 4(2200)_F r^2 & (1011)_{12} &= 4(2200)_G r^2 \\
(0111)_{20} &= -2(2101)r & (1011)_{02} &= -2(1210)r \\
(0111)_{11} &= \{-4(1210) + f_{xo}\}r & (1011)_{11} &= \{-4(2101) + f_{yo}\}r \\
(0111)_{10} &= 2(1111) & (1011)_{10} &= 4(2002) \\
(0111)_{01} &= 4(0220) & (1011)_{01} &= 2(1111) \\
(0111)_{00} &= -2(0121)r^{-1} & (1011)_{00} &= -2(1012)r^{-1}
\end{aligned}
\tag{29.19b}
$$

Astigmatism

$$
\begin{aligned}
(2010)_2 &= -12(4000)r & (0201)_2 &= -12(0400)r \\
(2010)_1 &= 6(3010) - 3\frac{f_{xo}}{2} & (0201)_1 &= 6(0301) - 3\frac{f_{yo}}{2} \\
(2010)_0 &= -2(2020)r^{-1} & (0201)_0 &= -2(0202)r^{-1} \\
(1101)_{11} &= -4(2200)_F r & (1110)_{11} &= -4(2200)_G r \\
(1101)_{10} &= 2(2101) & (1110)_{10} &= 2(2101) - f_{yo} \\
(1101)_{01} &= 2(1210) - f_{xo} & (1110)_{01} &= 2(1210) \\
(1101)_{00} &= -(1111)r^{-1} & (1110)_{00} &= -(1111)r^{-1} \\
(0210)_2 &= -2(2200)_F r & (2001)_2 &= -2(2200)_G r \\
(0210)_1 &= 2(1210) - \frac{f_{xo}}{2} & (2001)_1 &= 2(2101) - \frac{f_{yo}}{2} \\
(0210)_0 &= -2(0220)r^{-1} & (2001)_0 &= -2(2002)r^{-1}
\end{aligned}
\tag{29.19c}
$$

Distortions

$$
\begin{aligned}
(3000)_1 &= 4(4000) & (0300)_1 &= 4(0400) \\
(3000)_0 &= \left\{-(3010) + \frac{f_{xo}}{2}\right\}r^{-1} & (0300)_0 &= \left\{-(0301) + \frac{f_{yo}}{2}\right\}r^{-1} \\
(1200)_1 &= 2(2200)_F & (2100)_1 &= 2(2200)_G \\
(1200)_0 &= \left\{-(1210) + \frac{f_{xo}}{2}\right\}r^{-1} & (2100)_0 &= \left\{-(2101) + \frac{f_{yo}}{2}\right\}r^{-1}
\end{aligned}
\tag{29.19d}
$$

The coefficients characterizing the aberrations of gradient, defined by

$$
\begin{aligned}
\Delta x_i' &= f_x\left\{x'(z_{ix}) + x_o - f_x\frac{x_o'}{M_x}\right\} \\
&= \xi\{\,[3000]\xi^2 + [1200]\eta^2\} + x_o'\{\,[0030]x_o'^2 + [0012]y_o'^2\} \\
&+ x_o'\{\,(2010)\xi^2 + [0210]\eta^2\} + \xi\{\,[1020]x_o'^2 + [1002]y_o'^2\} \\
&+ \eta y_o'\{\,[1101]\xi + [0111]x_o'\}
\end{aligned}
$$

$$\begin{aligned}\Delta y_i' &= f_y\left\{y'(z_{iy}) + y_o - f_y\frac{y_o'}{M_y}\right\}\\ &= \eta\{\,[0300]\eta^2 + [2100]\xi^2\} + y_o'\{\,[0003]y_o'^2 + [0021]x_o'^2\}\\ &+ y_o'\{\,[2001]\xi^2 + [0201]\eta^2\} + \eta\{\,[0120]x_o'^2 + [0102]y_o'^2\}\\ &+ \xi x_o'\{\,[1110]\eta + [1011]y_o'\}\end{aligned} \tag{29.20}$$

are also polynomials in inverse magnification. These have the following form:

$$[3000] = (3010) - \frac{1}{2}f_{xo}$$

$$[1200] = (1210) - \frac{1}{2}f_{xo}$$

$$[2010] = -3\left\{(3010) - \frac{1}{2}f_{xo}\right\}rm_x + 2(2020)$$

$$[0210] = -\left\{(1210) - \frac{1}{2}f_{zo}\right\}rm_x + 2(0220)$$

$$[1101] = -2\left\{(1210) - \frac{1}{2}f_{xo}\right\}rm_y + (1111)$$

$$[1020] = 3\left\{(3010) - \frac{1}{2}f_{xo}\right\}r^2m_x - 4(2020)rm_x + 3(1030)$$

$$[1002] = \left\{(1210) - \frac{1}{2}f_{xo}\right\}r^2m_y^2 - (1111)rm_y + (1012)$$

$$\begin{aligned}[0111] &= 2\left\{(1210) - \frac{1}{2}f_{xo}\right\}r^2m_xm_y - 4(0220)rm_y\\ &\quad - (1111)m_x + 2(0121)\end{aligned}$$

$$\begin{aligned}[0030] &= -\left\{(3010) - \frac{1}{2}f_{xo}\right\}r^3m_x^3 + 2(2020)r^2m_x^2\\ &\quad - 3(1030)rm_x + 4(0040)\end{aligned}$$

$$\begin{aligned}[0012] &= -\left\{(1210) - \frac{1}{2}f_{xo}\right\}r^3m_xm_y^2 + 2(0220)r^2m_y^2\\ &\quad + (1111)r^2m_xm_y - (1012)rm_x - 2(0121)rm_y\\ &\quad + 2(0022)_F\end{aligned}$$

$$
\begin{aligned}
[0300] &= (0301) - \frac{1}{2} f_{yo} \\
[2100] &= (2101) - \frac{1}{2} f_{yo} \\
[2001] &= -\left\{(2101) - \frac{1}{2} f_{yo}\right\} r m_y + 2(2002) \\
[0201] &= -3\left\{(0301) - \frac{1}{2} f_{yo}\right\} r m_y + 2(0202) \\
[1110] &= -2\left\{(2101) - \frac{1}{2} f_{yo}\right\} r m_x + (1111) \\
[0120] &= \left\{(2101) - \frac{1}{2} f_{yo}\right\} r^2 m_x^2 - (1111) r m_x + (0121) \\
[0102] &= 3\left\{(0301) - \frac{1}{2} f_{yo}\right\} r^2 m_y^2 - (0202) r m_y + 3(0103) \\
[1011] &= 2\left\{(2101) - \frac{1}{2} f_{yo}\right\} r^2 m_x m_y - 4(2002) r m_x - (1111) r m_y + 2(1012) \\
[0003] &= -\left\{(0301) - \frac{1}{2} f_{yo}\right\} r^3 m_y^3 + 2(0202) r^2 m_y^2 - 3(0103) r m_y + 4(0004) \\
[0021] &= -\left\{(2101) - \frac{1}{2} f_{yo}\right\} r^3 m_x^2 m_y + 2(2002) r^2 m_x^2 \\
&\quad + (1111) r^2 m_x m_y - 2(0102) r m_x - (0121) r m_y + 2(0022)_G
\end{aligned}
\tag{29.21}
$$

The coefficients $(pqrs)$, $p + q + r + s = 4$, denote the following integrals:

$$
\begin{aligned}
(4000) &= f_{xi}^4 \int_{-\infty}^{\infty} (A_x G_x^4 + B_x G_x^2 G_x'^2 + N G_x'^4) dz + \frac{1}{8} \lim_{z \to \infty} (z - z_{Fi}^{(x)}) \\
(0400) &= f_{yi}^4 \int_{-\infty}^{\infty} (A_y G_y^4 + B_y G_y^2 G_y'^2 + N G_y'^4) dz + \frac{1}{8} \lim_{z \to \infty} (z - z_{Fi}^{(y)}) \\
(0040) &= \int_{-\infty}^{\infty} (A_x \overline{I}_x^4 + B_x \overline{I}_x^2 \overline{I}_x'^2 + N \overline{I}_x'^4) dz - \frac{1}{8} r \lim_{z \to -\infty} (z - z_{Fo}^{(x)}) \\
(0004) &= \int_{-\infty}^{\infty} (A_y \overline{I}_y^4 + B_y \overline{I}_y^2 \overline{I}_y'^2 + N \overline{I}_y'^4) dz - \frac{1}{8} r \lim_{z \to -\infty} (z - z_{Fo}^{(y)})
\end{aligned}
$$

$$(3010) = 4f_{xi}^3 \int_{-\infty}^{\infty} \left\{ A_x G_x^3 \overline{\Gamma}_x + \frac{1}{4} B_x (G_x^2)' (G_x \overline{\Gamma}_x)' + N G_x'^3 \overline{\Gamma}_x' \right\} dz$$

$$(0301) = 4f_{yi}^3 \int_{-\infty}^{\infty} \left\{ A_y G_y^3 \overline{\Gamma}_y + \frac{1}{4} B_y (G_y^2)' (G_y \overline{\Gamma}_y)' + N G_y'^3 \overline{\Gamma}_y' \right\} dz$$

$$(1030) = 4f_{xi} \int_{-\infty}^{\infty} \left\{ A_x G_x \overline{\Gamma}_x^3 + \frac{1}{4} B_x (G_x \overline{\Gamma}_x)' (\overline{\Gamma}_x^2)' + N G_x' \overline{\Gamma}_x'^3 \right\} dz$$

$$(0103) = 4f_{yi} \int_{-\infty}^{\infty} \left\{ A_y G_y \overline{\Gamma}_y^3 + \frac{1}{4} B_y (G_y \overline{\Gamma}_y)' (\overline{\Gamma}_y^2)' + N G_y' \overline{\Gamma}_y'^3 \right\} dz$$

$$(2020) = 6f_{xi}^2 \int_{-\infty}^{\infty} \left\{ A_x G_x^2 \overline{\Gamma}_x^2 + \frac{1}{6} B_x (G_x^2 \overline{\Gamma}_x'^2 + G_x'^2 \overline{\Gamma}_x^2 + (G_x^2)' (\overline{\Gamma}_x^2)') + N G_x'^2 \overline{\Gamma}_x'^2 \right\} dz$$

$$(0202) = 6f_{yi}^2 \int_{-\infty}^{\infty} \left\{ A_y G_y^2 \overline{\Gamma}_y^2 + \frac{1}{6} B_y (G_y^2 \overline{\Gamma}_y'^2 + G_y'^2 \overline{\Gamma}_y^2 + (G_y^2)' (\overline{\Gamma}_y^2)') + N G_y'^2 \overline{\Gamma}_y'^2 \right\} dz$$

$$(2002) = f_{xi}^2 \int_{-\infty}^{\infty} \left\{ C G_y^2 \overline{\Gamma}_y^2 + B_x G_x^2 \overline{\Gamma}_y'^2 + B_y G_x'^2 \overline{\Gamma}_y^2 + 2N G_x'^2 \overline{\Gamma}_y'^2 + R G_x \overline{\Gamma}_y (G_x \overline{\Gamma}_y' - G_x' \overline{\Gamma}_y) \right\} dz$$

$$(0220) = f_{yi}^2 \int_{-\infty}^{\infty} \left\{ C \overline{\Gamma}_x^2 G_y^2 + B_x \overline{\Gamma}_x^2 G_y'^2 + B_y \overline{\Gamma}_x'^2 G_y^2 + 2N \overline{\Gamma}_x'^2 G_y'^2 + R \overline{\Gamma}_x G_y (\overline{\Gamma}_x G_y' - \overline{\Gamma}_x' G_y) \right\} dz$$

$$(2200) = f_{xi}^2 f_{yi}^2 \int_{-\infty}^{\infty} \left\{ C G_x^2 G_y^2 + B_x G_x^2 G_y'^2 + B_y G_x'^2 G_y^2 + 2N G_x'^2 G_y'^2 + R G_x G_y (G_x G_y' - G_x' G_y) \right\} dz$$

$$(2200)_F = (2200) + \frac{1}{4} \lim_{z \to \infty} (z - z_{Fi}^{(x)})$$

$$(2200)_G = (2200) + \frac{1}{4} \lim_{z \to \infty} (z - z_{Fi}^{(y)})$$

$$(0022) = \int_{-\infty}^{\infty} \left\{ C \overline{\Gamma}_x^2 \overline{\Gamma}_y^2 + B_x \overline{\Gamma}_x^2 \overline{\Gamma}_y'^2 + B_y \overline{\Gamma}_x'^2 \overline{\Gamma}_y^2 + 2N \overline{\Gamma}_x'^2 \overline{\Gamma}_y'^2 + R \overline{\Gamma}_x \overline{\Gamma}_y (\overline{\Gamma}_x \overline{\Gamma}_y' - \overline{\Gamma}_x' \overline{\Gamma}_y) \right\} dz$$

$$(0022)_F = (0022) - \frac{1}{4} r \lim_{z \to -\infty} (z - z_{Fo}^{(x)})$$

$$(0022)_G = (0022) - \frac{1}{4} r \lim_{z \to -\infty} (z - z_{Fo}^{(y)})$$

$$(2101) = 2f_{xi}^2 f_{yi} \int_{-\infty}^{\infty} \left[CG_x^2 G_y \overline{\Gamma}_y + B_x G_x^2 G_y' \overline{\Gamma}_y' + B_y G_x'^2 G_y \overline{\Gamma}_y + 2NG_x'^2 G_y' \overline{\Gamma}_y' + \frac{1}{2} R \left\{ G_x^2 (G_y \overline{\Gamma}_y)' - (G_x^2)' G_y \overline{\Gamma}_y \right\} \right] dz$$

$$(1210) = 2f_{xi} f_{yi}^2 \int_{-\infty}^{\infty} \left[CG_x \overline{\Gamma}_x G_y^2 + B_x G_x \overline{\Gamma}_x G_y'^2 + B_y G_x' \overline{\Gamma}_x' G_y^2 + 2NG_x' \overline{\Gamma}_x' G_y'^2 + \frac{1}{2} R \left\{ G_x \overline{\Gamma}_x (G_y^2)' - (G_x \overline{\Gamma}_x)' G_y^2 \right\} \right] dz$$

$$(1012) = 2f_{xi} \int_{-\infty}^{\infty} \left[CG_x \overline{\Gamma}_x \overline{\Gamma}_y^2 + B_x G_x \overline{\Gamma}_x \overline{\Gamma}_y'^2 + B_y G_x' \overline{\Gamma}_x' \overline{\Gamma}_y^2 + 2NG_x' \overline{\Gamma}_x' \overline{\Gamma}_y'^2 + \frac{1}{2} R \left\{ G_x \overline{\Gamma}_x (\overline{\Gamma}_y^2)' - (G_x \overline{\Gamma}_x)' \overline{\Gamma}_y^2 \right\} \right] dz \tag{29.22}$$

$$(0121) = 2f_{yi} \int_{-\infty}^{\infty} \left[C\overline{\Gamma}_x^2 G_y \overline{\Gamma}_y + B_x \overline{\Gamma}_x^2 G_y' \overline{\Gamma}_y' + B_y \overline{\Gamma}_x'^2 G_y \overline{\Gamma}_y + 2N\overline{\Gamma}_x'^2 G_y' \overline{\Gamma}_y' + \frac{1}{2} R \left\{ \overline{\Gamma}_x^2 (G_y \overline{\Gamma}_y)' - (\overline{\Gamma}_x^2)' G_y \overline{\Gamma}_y \right\} \right] dz$$

$$(1111) = 4f_{xi} f_{yi} \int_{-\infty}^{\infty} \left[CG_x \overline{\Gamma}_x G_y \overline{\Gamma}_y + B_x G_x \overline{\Gamma}_x G_y' \overline{\Gamma}_y' + B_y G_x' \overline{\Gamma}_x' G_y \overline{\Gamma}_y + 2NG_x' \overline{\Gamma}_x' G_y' \overline{\Gamma}_y' + \frac{1}{2} R \left\{ G_x \overline{\Gamma}_x (G_y \overline{\Gamma}_y)' - (G_x \overline{\Gamma}_x)' G_y \overline{\Gamma}_y \right\} \right] dz$$

in which

$$A_x = \frac{1}{128} \left(\frac{\hat{\phi}}{\hat{\phi}_i} \right)^{1/2} \left(\frac{\gamma \phi^{(4)}}{\hat{\phi}} - \frac{\phi''^2 + 4p_2^2}{\hat{\phi}^2} - \frac{8\gamma p_2''}{3\hat{\phi}} + \frac{4p_2 \phi''}{\hat{\phi}^2} + \frac{8\eta Q_2''}{3\hat{\phi}^{1/2}} \right) + \Xi$$

$$A_y = \frac{1}{128} \left(\frac{\hat{\phi}}{\hat{\phi}_i} \right)^{1/2} \left(\frac{\gamma \phi^{(4)}}{\hat{\phi}} - \frac{\phi''^2 + 4p_2^2}{\hat{\phi}^2} + \frac{8\gamma p_2''}{3\hat{\phi}} - \frac{4p_2 \phi''}{\hat{\phi}^2} - \frac{8\eta Q_2''}{3\hat{\phi}^{1/2}} \right) + \Xi$$

$$
\begin{aligned}
B_x &= -\frac{\gamma}{16}\left(\frac{\hat{\phi}}{\hat{\phi}_i}\right)^{1/2}\frac{\phi'' - 2p_2}{\hat{\phi}} \\
B_y &= -\frac{\gamma}{16}\left(\frac{\hat{\phi}}{\hat{\phi}_i}\right)^{1/2}\frac{\phi'' + 2p_2}{\hat{\phi}} \\
C &= \frac{1}{64}\left(\frac{\hat{\phi}}{\hat{\phi}_i}\right)^{1/2}\left(\frac{\gamma\phi^{(4)}}{\hat{\phi}} - \frac{\phi''^2 - 4p_2^2}{\hat{\phi}^2}\right) - 6\Xi \\
N &= -\frac{1}{8}\left(\frac{\hat{\phi}}{\hat{\phi}_i}\right)^{1/2} \\
R &= \frac{\eta Q_2'}{4\hat{\phi}_i^{1/2}}
\end{aligned}
\tag{29.23}
$$

and $\Xi(z)$ characterizes the octopole distribution:

$$
\Xi(z) \coloneqq \frac{1}{48}\left(\frac{\hat{\phi}}{\hat{\phi}_i}\right)^{1/2}\left(\frac{\gamma p_4}{\hat{\phi}} - \frac{2\eta Q_4}{\hat{\phi}^{1/2}}\right) \tag{29.24}
$$

In the foregoing formulae, we have imposed no restrictions on $\phi(z)$, $p_2(z)$ and $Q_2(z)$, which may all be present simultaneously. Commonly, however, ϕ is constant and any electrostatic and magnetic quadrupole fields do not overlap. In these conditions, the formulae can be recast in a much simpler form by writing

$$
\begin{aligned}
p_2(z) &=: p_{20}\, q(z) \quad & Q_2(z) &=: Q_{20}\, q(z) \\
p_4(z) &=: p_{40}\, \omega(z) \quad & Q4(z) &=: Q_{40}\, \omega(z)
\end{aligned}
\tag{29.25}
$$

where $q(z)$ and $\omega(z)$ are functions that reach a maximum value of unity. We define an excitation parameter β^2 and a label n as follows:

$$
\frac{\gamma p_{20}}{2\hat{\phi}} =: \beta_E^2 \quad \frac{\eta Q_{20}}{\hat{\phi}^{1/2}} =: \beta_M^2 \quad \beta^2 \coloneqq \beta_E^2 - \beta_E^2 \quad n \coloneqq \frac{\beta_E^2}{\beta_M^2 - \beta_E^2} \tag{29.26}
$$

Thus for purely electrostatic quadrupoles, $n = -1$, for magnetic quadrupoles, $n = 0$, and as we shall see in Section 29.4, for achromatic quadrupoles, $n = \gamma^2 \approx 1$. From Eqs (29.25–29.26), we see that

$$
\frac{\gamma p_2(z)}{\hat{\phi}} = 2n\beta^2 q(z) \text{ and } \frac{\eta Q_2(z)}{\hat{\phi}^{1/2}} = (n+1)\beta^2 q(z) \tag{29.27}
$$

For octopoles, we write

$$\frac{\gamma p_{40}}{48\hat{\phi}} =: -\tau_E \quad \frac{\eta Q_{40}}{24\hat{\phi}^{1/2}} =: -\tau_M \quad \tau_M - \tau_E := \tau \tag{29.28}$$

so that

$$\Xi(z) = \tau\omega(z) \tag{29.29}$$

The functions A_x, A_y. . .R now become

$$\begin{aligned}
A_x &= -\frac{1}{8}\beta^2\left(\beta^2 q^2\overline{n}^2 + \frac{n-1}{6}q''\right) + \tau\omega \\
A_y &= -\frac{1}{8}\beta^2\left(\beta^2 q^2\overline{n}^2 - \frac{n-1}{6}q''\right) + \tau\omega \\
B_x &= -B_y = \frac{1}{4}\beta^2 nq \\
C &= \frac{1}{4}\beta^4\overline{n}^2 q^2 - 6\tau\omega \\
N &= -\frac{1}{8} \\
R &= \frac{1}{4}\beta^2(n+1)q'
\end{aligned} \tag{29.30}$$

where $\overline{n} := n/\gamma$. The following set of coefficients ($pqrs$) is now more convenient:

$$\begin{aligned}
(4000) &= \int_{-\infty}^{\infty} a_x \Gamma_x^4\, dz \\
&= \int_{-\infty}^{\infty} b\Gamma_x^4\, dz - \frac{1}{24}\int_{-\infty}^{\infty} \Gamma_x'^4\, dz + \frac{1}{24}\lim_{z\to\infty}(z - z_{Fi}^{(x)}) \\
(0400) &= \int_{-\infty}^{\infty} a_y \Gamma_y^4\, dz \\
&= \int_{-\infty}^{\infty} b\Gamma_y^4\, dz - \frac{1}{24}\int_{-\infty}^{\infty} \Gamma_y'^4\, dz + \frac{1}{24}\lim_{z\to\infty}(z - z_{Fi}^{(y)}) \\
(0040) &= \int_{-\infty}^{\infty} a_x \overline{\Gamma}_x^4\, dz \\
&= \int_{-\infty}^{\infty} b\overline{\Gamma}_x^4\, dz - \frac{1}{24}\int_{-\infty}^{\infty} \overline{\Gamma}_x'^4\, dz - \frac{1}{24}\lim_{z\to-\infty}(z - z_{Fo}^{(x)})
\end{aligned}$$

$$(0004) = \int_{-\infty}^{\infty} a_y \overline{\Gamma}_y^4 \, dz$$

$$= \int_{-\infty}^{\infty} b\overline{\Gamma}_y^4 \, dz - \frac{1}{24} \int_{-\infty}^{\infty} \overline{\Gamma}_y'^4 \, dz - \frac{1}{24} \lim_{z \to -\infty} (z - z_{Fo}^{(y)})$$

$$(3010) = 4 \int_{-\infty}^{\infty} a_x \Gamma_x^3 \overline{\Gamma}_x \, dz + \frac{f_x}{8}$$

$$= 4 \int_{-\infty}^{\infty} b\Gamma_x^3 \overline{\Gamma}_x \, dz - \frac{1}{6} \int_{-\infty}^{\infty} \Gamma_x'^3 \overline{\Gamma}_x' \, dz + \frac{f_x}{12}$$

$$(0301) = 4 \int_{-\infty}^{\infty} a_y \Gamma_y^3 \overline{\Gamma}_y \, dz + \frac{f_y}{8}$$

$$= 4 \int_{-\infty}^{\infty} b\Gamma_y^3 \overline{\Gamma}_y \, dz - \frac{1}{6} \int_{-\infty}^{\infty} \Gamma_y'^3 \overline{\Gamma}_y' \, dz + \frac{f_y}{12}$$

$$(1030) = 4 \int_{-\infty}^{\infty} a_x \Gamma_x \overline{\Gamma}_x^3 \, dz + \frac{f_x}{8}$$

$$= 4 \int_{-\infty}^{\infty} b\Gamma_x \overline{\Gamma}_x^3 \, dz - \frac{1}{6} \int_{-\infty}^{\infty} \Gamma_x' \overline{\Gamma}_x'^3 \, dz + \frac{f_x}{12}$$

$$(0103) = 4 \int_{-\infty}^{\infty} a_x \Gamma_y \overline{\Gamma}_y^3 \, dz + \frac{f_y}{8}$$

$$= 4 \int_{-\infty}^{\infty} b\Gamma_y \overline{\Gamma}_y^3 \, dz - \frac{1}{6} \int_{-\infty}^{\infty} \Gamma_y' \overline{\Gamma}_y'^3 \, dz + \frac{f_y}{12}$$

$$(2020) = 6 \int_{-\infty}^{\infty} a_x \Gamma_x^2 \overline{\Gamma}_x^2 \, dz$$

$$= 6 \int_{-\infty}^{\infty} b\Gamma_x^2 \overline{\Gamma}_x^2 \, dz - \frac{1}{4} \int_{-\infty}^{\infty} \Gamma_x'^2 \overline{\Gamma}_x'^2 \, dz$$

$$(0202) = 6\int_{-\infty}^{\infty} a_y \Gamma_y^2 \overline{\Gamma}_y^2 \, dz$$

$$= 6\int_{-\infty}^{\infty} b\Gamma_y^2 \overline{\Gamma}_y^2 \, dz - \frac{1}{4}\int_{-\infty}^{\infty} \Gamma_y'^2 \overline{\Gamma}_y'^2 \, dz$$

$$(2002) = \int_{-\infty}^{\infty} \left\{ c\Gamma_x^2 \overline{\Gamma}_y^2 + \frac{3}{8}\beta^2 q' \Gamma_x \overline{\Gamma}_y (\Gamma_x \overline{\Gamma}_y' - \overline{\Gamma}_x' \overline{\Gamma}_y) \right\} dz$$

$$= \int_{-\infty}^{\infty} \left\{ \overline{c}\Gamma_x^2 \overline{\Gamma}_y^2 - \frac{3}{8}\beta^2 q (\Gamma_x^2 \overline{\Gamma}_y'^2 - \Gamma_x'^2 \overline{\Gamma}_y^2) \right\} dz$$

$$= \int_{-\infty}^{\infty} \overline{c}\Gamma_x^2 \overline{\Gamma}_y^2 \, dz - \frac{3}{4}\int_{-\infty}^{\infty} \Gamma_x'^2 \overline{\Gamma}_y'^2 \, dz$$

$$(0220) = \int_{-\infty}^{\infty} \left\{ c\overline{\Gamma}_x^2 \Gamma_y^2 + \frac{3}{8}\beta^2 q' \overline{\Gamma}_x \Gamma_y (\overline{\Gamma}_x \Gamma_y' - \overline{\Gamma}_x' \Gamma_y) \right\} dz$$

$$= \int_{-\infty}^{\infty} \left\{ \overline{c}\overline{\Gamma}_x^2 \Gamma_y^2 - \frac{3}{8}\beta^2 q (\overline{\Gamma}_x^2 \Gamma_y'^2 - \overline{\Gamma}_x'^2 \Gamma_y^2) \right\} dz$$

$$= \int_{-\infty}^{\infty} \overline{c}\overline{\Gamma}_x^2 \Gamma_y^2 \, dz - \frac{3}{4}\int_{-\infty}^{\infty} \overline{\Gamma}_x'^2 \Gamma_y'^2 \, dz$$

$$(2200) = \int_{-\infty}^{\infty} \left\{ c\Gamma_x^2 \Gamma_y^2 + \frac{3}{8}\beta^2 q' \Gamma_x \Gamma_y (\Gamma_x \Gamma_y' - \Gamma_x' \Gamma_y) \right\} dz$$

$$-\frac{1}{8}\lim_{z\to\infty}(2z - z_{Fi}^{(x)} - z_{Fi}^{(y)})$$

$$= \int_{-\infty}^{\infty} \left\{ \overline{c}\Gamma_x^2 \Gamma_y^2 - \frac{3}{8}\beta^2 q (\Gamma_x^2 \Gamma_y'^2 - \Gamma_x'^2 \Gamma_y^2) \right\} dz - \frac{1}{8}\lim_{z\to\infty}(2z - z_{Fi}^{(x)} - z_{Fi}^{(y)})$$

$$= \int_{-\infty}^{\infty} \overline{c}\Gamma_x^2 \Gamma_y^2 \, dz - \frac{3}{4}\int_{-\infty}^{\infty} \Gamma_x'^2 \Gamma_y'^2 dz + \frac{1}{4}\lim_{z\to\infty}(2z - z_{Fi}^{(x)} - z_{Fi}^{(y)})$$

$$(2200)_F = \int_{-\infty}^{\infty} \left\{ c\Gamma_x^2\Gamma_y^2 + \frac{3}{8}\beta^2 q'\Gamma_x\Gamma_y(\Gamma_x\Gamma_y' - \Gamma_x'\Gamma_y) \right\} dz - \frac{z_{Fi}^{(x)} - z_{Fi}^{(y)}}{8}$$

$$= \int_{-\infty}^{\infty} \left\{ \bar{c}\Gamma_x^2\Gamma_y^2 - \frac{3}{8}\beta^2 q(\Gamma_x^2\Gamma_y'^2 - \Gamma_x'^2\Gamma_y^2) \right\} dz - \frac{z_{Fi}^{(x)} - z_{Fi}^{(y)}}{8} \Big)$$

$$= \int_{-\infty}^{\infty} \bar{c}\Gamma_x^2\Gamma_y^2\, dz - \frac{3}{4}\int_{-\infty}^{\infty} \Gamma_x'^2\Gamma_y'^2\, dz + \frac{1}{4}\lim_{z\to\infty}(3z - 2z_{Fi}^{(x)} - z_{Fi}^{(y)})$$

$$(2200)_G = \int_{-\infty}^{\infty} \left\{ c\Gamma_x^2\Gamma_y^2 - \frac{3}{8}\beta^2 q'\Gamma_x\Gamma_y\Gamma(\Gamma_x\Gamma_y' - \Gamma_x'\Gamma_y) \right\} dz + \frac{z_{Fi}^{(x)} - z_{Fi}^{(y)}}{8}$$

$$= \int_{-\infty}^{\infty} \left\{ \bar{c}\Gamma_x^2\Gamma_y^2 - \frac{3}{8}\beta^2 q(\Gamma_x^2\Gamma_y'^2 - \Gamma_x'^2\Gamma_y^2) \right\} dz + \frac{z_{Fi}^{(x)} - z_{Fi}^{(y)}}{8}$$

$$= \int_{-\infty}^{\infty} \Gamma_x'^2\Gamma_y'^2\, dz - \frac{3}{4}\int_{-\infty}^{\infty} \Gamma_x'^2\Gamma_y'^2\, dz + \frac{1}{4}\lim_{z\to\infty}(3z - 2z_{Fi}^{(y)} - z_{Fi}^{(x)})$$

$$(0022) = \int_{-\infty}^{\infty} \left\{ c\overline{\Gamma}_x^2\overline{\Gamma}_y^2 + \frac{3}{8}\beta^2 q'\overline{\Gamma}_x\overline{\Gamma}_y(\overline{\Gamma}_x\overline{\Gamma}'_y - \overline{\Gamma}'_x\overline{\Gamma}_y) \right\} dz + \frac{1}{8}\lim_{z\to-\infty}(2z - z_{Fo}^{(x)} - z_{Fo}^{(y)})$$

$$= \int_{-\infty}^{\infty} \left\{ \bar{c}\overline{\Gamma}_x^2\overline{\Gamma}_y^2 - \frac{3}{2}\beta^2 q(\overline{\Gamma}_x^2\overline{\Gamma}_v'^2 - \overline{\Gamma}_x'^2\overline{\Gamma}_y^2) \right\} dz + \frac{1}{8}\lim_{z\to-\infty}(2z - z_{Fo}^{(x)} - z_{Fo}^{(y)})$$

$$= \int_{-\infty}^{\infty} \bar{c}\overline{\Gamma}_x^2\overline{\Gamma}_y^2\, dz - \frac{3}{4}\int_{-\infty}^{\infty} \overline{\Gamma}_x'^2\overline{\Gamma}_y'^2\, dz - \frac{1}{4}\lim_{z\to-\infty}(2z - z_{Fo}^{(x)} - z_{Fo}^{(y)})$$

$$(0022)_F = \int_{-\infty}^{\infty} \left\{ c\overline{\Gamma}_x^2\overline{\Gamma}_y^2 + \frac{3}{8}\beta^2 q'\overline{\Gamma}_x\overline{\Gamma}_y(\overline{\Gamma}_x\overline{\Gamma}'_y - \overline{\Gamma}'_x\overline{\Gamma}_y) \right\} dz + \frac{z_{Fo}^{(x)} - z_{Fo}^{(y)}}{8}$$

$$= \int_{-\infty}^{\infty} \left\{ \bar{c}\overline{\Gamma}_x^2\overline{\Gamma}_y^2 - \frac{3}{8}\beta^2 q(\overline{\Gamma}_x^2\overline{\Gamma}_y'^2 - \overline{\Gamma}_x'^2\overline{\Gamma}_y^2) \right\} dz + \frac{z_{Fo}^{(x)} - z_{Fo}^{(y)}}{8}$$

$$= \int_{-\infty}^{\infty} \bar{c}\overline{\Gamma}_x^2\overline{\Gamma}_y^2\, dz - \frac{3}{4}\int_{-\infty}^{\infty} \overline{\Gamma}_{\mathrm{x}}'^2\overline{\Gamma}_{\mathrm{y}}'^2\, dz - \frac{1}{4}\lim_{z\to-\infty}\left(3z - 2z_{Fo}^{(x)} - z_{Fo}^{(y)}\right)$$

$$(0022)_G = \int_{-\infty}^{\infty} \left\{ c\overline{\Gamma}_x^2\overline{\Gamma}_y^2 + \frac{3}{8}\beta^2 q'\overline{\Gamma}_x\overline{\Gamma}_y\left(\overline{\Gamma}_x\overline{\Gamma}_y' - \overline{\Gamma}_x'\overline{\Gamma}_y\right)\right\} dz - \frac{z_{Fo}^{(x)} - z_{Fo}^{(y)}}{8}$$

$$= \int_{-\infty}^{\infty} \left\{ \overline{c}\overline{\Gamma}_x^2\overline{\Gamma}_y^2 - \frac{3}{8}\beta^2 q(\overline{\Gamma}_x^2\overline{\Gamma}_{\mathrm{y}}'^2 - \overline{\Gamma}_{\mathrm{x}}'^2\overline{\Gamma}_y^2)\right\} dz - \frac{z_{Fo}^{(x)} - z_{Fo}^{(y)}}{8}$$

$$= \int_{-\infty}^{\infty} \overline{c}\overline{\Gamma}_x^2\overline{\Gamma}_y^2\, dz - \frac{3}{4}\int_{-\infty}^{\infty} \overline{\Gamma}_{\mathrm{x}}'^2\overline{\Gamma}_{\mathrm{y}}'^2\, dz - \frac{1}{4}\lim_{z\to-\infty}(3z - 2z_{Fo}^{(y)} - z_{Fo}^{(x)})$$

$$(2101) = \int_{-\infty}^{\infty} \left[2c\Gamma_x^2\Gamma_y\overline{\Gamma}_y + \frac{3}{8}\beta^2 q'\left\{ \Gamma_x^2(\Gamma_y\overline{\Gamma}_y)' - (\Gamma_x^2)'\Gamma_y\overline{\Gamma}_y\right\}\right] dz + \frac{f_y}{8}$$

$$= 2\int_{-\infty}^{\infty} \left\{ \overline{c}\Gamma_x^2\Gamma_y\overline{\Gamma}_y - \frac{3}{8}\beta^2 q(\Gamma_x^2\Gamma_y'\overline{\Gamma}_y' - \Gamma_{\mathrm{x}}'^2\Gamma_y\overline{\Gamma}_y)\right\} dz + \frac{f_y}{8}$$

$$= 2\int_{-\infty}^{\infty} \overline{c}\Gamma_x^2\Gamma_y\overline{\Gamma}_y\, dz - \frac{3}{2}\int_{-\infty}^{\infty} \Gamma_{\mathrm{x}}'^2\Gamma_y'\overline{\Gamma}_{\mathrm{y}}'\, dz - \frac{f_y}{4}$$

$$(1210) = \int_{-\infty}^{\infty} \left[2c\Gamma_x\overline{\Gamma}_x\Gamma_y^2 + \frac{3}{8}\beta^2 q'\left\{ \Gamma_x\overline{\Gamma}_x(\overline{\Gamma}_y^2)' - (\Gamma_x\overline{\Gamma}_x)'\Gamma_y^2\right\}\right] dz + \frac{f_x}{8}$$

$$= 2\int_{-\infty}^{\infty} \left\{ \overline{c}\Gamma_x\overline{\Gamma}_x\Gamma_y^2 - \frac{3}{8}\beta^2 q(\Gamma_x\overline{\Gamma}_x\Gamma_{\mathrm{y}}'^2 - \Gamma_{\mathrm{x}}'\overline{\Gamma}_x'\Gamma_{\mathrm{y}}^2)\right\} dz + \frac{f_x}{8}$$

$$= 2\int_{-\infty}^{\infty} \overline{c}\Gamma_x\overline{\Gamma}_x\Gamma_y^2 dz - \frac{3}{2}\int_{-\infty}^{\infty} \Gamma_{\mathrm{x}}'\overline{\Gamma}_x'\Gamma_{\mathrm{y}}'^2\, dz - \frac{f_x}{4}$$

$$(1012) = \int_{-\infty}^{\infty} \left[2c\Gamma_x\overline{\Gamma}_x\overline{\Gamma}_y^2 + \frac{3}{2}\beta^2 q'\left\{ \Gamma_x\overline{\Gamma}_x(\overline{\Gamma}_y^2)' - (\Gamma_x\overline{\Gamma}_x)'\overline{\Gamma}_y^2\right\} dz + \frac{f_x}{8}\right]$$

$$= 2\int_{-\infty}^{\infty} \left\{ \overline{c}\Gamma_x\overline{\Gamma}_x\overline{\Gamma}_y^2 - \frac{3}{8}\beta^2 q(\Gamma_x\overline{\Gamma}_x\overline{\Gamma}_{\mathrm{y}}'^2 - \Gamma_{\mathrm{x}}'\overline{\Gamma}_x'\overline{\Gamma}_y^2)\right\} dz + \frac{f_x}{8}$$

$$= 2\int_{-\infty}^{\infty} \overline{c}\Gamma_x\overline{\Gamma}_x\overline{\Gamma}_y^2\, dz - \frac{3}{2}\int_{-\infty}^{\infty} \Gamma_{\mathrm{x}}'\overline{\Gamma}_x'\overline{\Gamma}_{\mathrm{y}}'^2\, dz - \frac{f_x}{4}$$

$$
\begin{aligned}
(0121) &= \int_{-\infty}^{\infty} \left[2c\overline{\Gamma}_x^2 \Gamma_y \overline{\Gamma}_y + \frac{3}{8}\beta^2 q' \left\{ \overline{\Gamma}_x^2 (\Gamma_y \overline{\Gamma}_y)' - (\overline{\Gamma}_x^2)' \Gamma_y \overline{\Gamma}_y \right\} \right] dz + \frac{f_y}{8} \\
&= 2 \int_{-\infty}^{\infty} \left\{ \overline{c}\overline{\Gamma}_x^2 \Gamma_y \overline{\Gamma}_y - \frac{3}{8}\beta^2 q (\Gamma_x^2 \Gamma_y' \overline{\Gamma}_y - \overline{\Gamma}_x'^2 \Gamma_y \overline{\Gamma}_y) \right\} dz + \frac{f_y}{8} \\
&= 2 \int_{-\infty}^{\infty} \overline{c}\overline{\Gamma}_x^2 \Gamma_y \overline{\Gamma}_y \, dz - \frac{3}{2} \int_{-\infty}^{\infty} \overline{\Gamma}_x'^2 \Gamma_y' \overline{\Gamma}_y' \, dz - \frac{f_y}{4} \\
(1111) &= 2 \int_{-\infty}^{\infty} \left[2c\Gamma_x \overline{\Gamma}_x \Gamma_y \overline{\Gamma}_y + \frac{3}{8}\beta^2 q' \left\{ \Gamma_x \overline{\Gamma}_x (\Gamma_y \overline{\Gamma}_y)' - (\Gamma_x \overline{\Gamma}_x)' \Gamma_y \overline{\Gamma}_y \right\} \right] dz \\
&\quad 4 \int_{-\infty}^{\infty} \left[\overline{c}\Gamma_x \overline{\Gamma}_x \overline{\Gamma}_y - \frac{3}{8}\beta^2 q \left\{ \Gamma_x \overline{\Gamma}_x \Gamma_y' \overline{\Gamma}_y' - \Gamma_x' \overline{\Gamma}_x' \Gamma_y \overline{\Gamma}_y \right\} \right] dz \qquad (29.31) \\
&= 4 \int_{-\infty}^{\infty} \overline{c}\Gamma_x \overline{\Gamma}_x \Gamma_y \overline{\Gamma}_y \, dz - 3 \int_{-\infty}^{\infty} \Gamma_x' \overline{\Gamma}_x' \Gamma_y' \overline{\Gamma}_y' \, dz
\end{aligned}
$$

In these expressions, the functions a_x, a_y, b, c and $\overline{c}$ are as follows:

$$
\begin{aligned}
a_x &:= -\frac{1}{24}\beta^4 (3\overline{n}^2 - 2n + 3) q^2 - \frac{1}{96}\beta^2 q'' + \tau\omega \\
a_y &:= -\frac{1}{24}\beta^4 (3\overline{n}^2 - 2n + 3) q^2 + \frac{1}{96}\beta^2 q'' + \tau\omega \\
b &:= -\frac{1}{24}\beta^4 (3\overline{n}^2 - 2n + 2) q^2 + \tau\omega \\
c &:= \frac{1}{4}\beta^4 (\overline{n}^2 - 2n + 1) q^2 - 6\tau\omega \\
\overline{c} &:= \frac{1}{4}\beta^4 (\overline{n}^2 - 2n - 2) q^2 - 6\tau\omega
\end{aligned} \qquad (29.32)
$$

We have lightened the notation by setting

$$
\Gamma_x = f_x G_x \quad \overline{\Gamma}_x = f_x \overline{G}_x \quad \Gamma_y = f_y G_y \quad \overline{\Gamma}_y = f_y \overline{G}_y \qquad (29.33)
$$

so that Γ_x and Γ_y are the rays incident from object space parallel to the axis at heights f_x and f_y in the two principal sections. If the function $q(z)$ is even in z, $\Gamma_x(z) = \overline{\Gamma}_x(-z)$ and $\Gamma_y(z) = \overline{\Gamma}_y(-z)$; if $q(z)$ is odd, then $\Gamma_x(z) = \overline{\Gamma}_y(-z)$ and $\Gamma_y(z) = \overline{\Gamma}_x(-z)$. The interrelations between the coefficients, derived earlier by symmetry arguments (29.7, 29.9), become obvious.

29.3 Aperture Aberrations

Each of the aberration coefficients of quadrupoles is associated with an aberration figure. We shall examine only the aperture aberrations, however, and we likewise confine the detailed analysis of the aberration formulae to this family of defects.

In a general plane z_c, the real aperture aberrations expressed in terms of x_o, y_o, x_a and y_a are given by

$$\begin{aligned} x_c - s_{xc}x_o &= x_a\{\, t_{xc} + (30)x_a^2 + (12)y_a^2\} \\ y_c - s_{yc}y_o &= y_a\{\, t_{yc} + (03)y_z^2 + (21)x_a^2\} \end{aligned} \tag{29.34}$$

in which we have shortened $(00rs)$ to (rs). As in Chapter 24, The Geometrical Aberrations of Round Lenses, we consider a family of rays intersecting the aperture plane around a circle, $x_a = r_a \cos\theta$, $y_a = r_a \sin\theta$. We find

$$\begin{aligned} x_c - s_{xc}x_o &= r_a\{\, t_{xc} + (30)r_a^2\} \cos\theta + r_a^3\{(12) - (30)\} \sin^2\theta \cos\theta \\ y_c - s_{yc}y_o &= r_a\{\, t_{yc} + (03)r_a^2\} \sin\theta + r_a^3\{(21) - (03)\} \cos^2\theta \sin\theta \end{aligned} \tag{29.35}$$

or writing

$$\begin{aligned} \delta x &:= x_c - s_{xc}x_o =: (\kappa + \lambda \sin^2\theta) \cos\theta \\ \delta y &:= y_c - s_{yc}y_o =: (\mu + \nu \cos^2\theta) \sin\theta \end{aligned} \tag{29.36}$$

we obtain

$$\begin{aligned} \delta X &= (1 + \varepsilon + \alpha \sin^2\theta) \cos\theta \\ \delta Y &= (1 - \varepsilon + \alpha \cos^2\theta) \sin\theta \end{aligned} \tag{29.37}$$

with

$$\begin{aligned} \delta X &= \frac{2\nu}{\kappa\nu + \lambda\mu}\delta x \qquad \delta Y = \frac{2\lambda}{\kappa\nu + \lambda\mu}\delta y \\ \varepsilon &= \frac{\kappa\nu - \lambda\mu}{\kappa\nu + \lambda\mu} \qquad \alpha = \frac{2\lambda\nu}{\kappa\nu + \lambda\mu} \end{aligned} \tag{29.38}$$

Eqs (29.37) are the parametric representation of the aberration figure in a general plane. The resulting curve is always symmetric about the planes $\theta = 0$, $\theta = \pi/2$ and has an oval, star- or rosette-shaped appearance according to the relative values of α and ε. Fig. 29.2 shows a division of the $\alpha - \varepsilon$ plane into regions associated with various aberration figures, proposed by Meads (1963) (and extensively analysed in Hawkes (1966)). A classification of

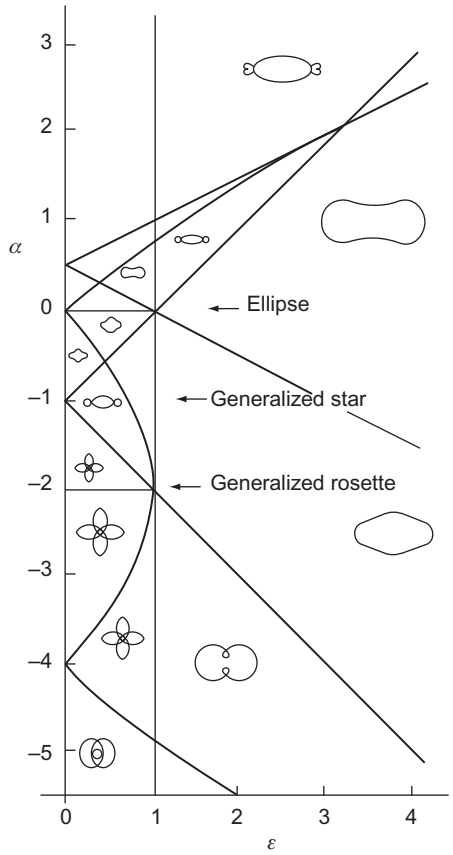


Figure 29.2
The various forms of the aberration figure associated with the aperture aberrations.

the aberrations for quadrupole systems producing a stigmatic but not necessarily orthomorphic image has also been proposed by Burfoot (1954a,b), who writes

$$\begin{aligned} x_i - s_{xi}s_o &= x_a\{\alpha_x r_a^2 + \beta x_a^2 + \gamma(x_a^2 - y_a^2)\} \\ y_i - s_{yi}y_o &= y_a\{\alpha_y r_a^2 - \beta y_a^2 - \gamma(x_a^2 - y_a^2)\} \end{aligned} \tag{29.39}$$

so that

$$\begin{aligned} \alpha_x &= \frac{1}{4}\{(30) + (03) + 3(12) - (21)\} \\ \alpha_y &= \frac{1}{4}\{(30) + (03) - (12) + 3(21)\} \\ \beta &= \frac{1}{2}\{(30) - (03) - (12) + (21)\} \\ \gamma &= \frac{1}{4}\{(30) + (03) - (12) - (21)\} \end{aligned} \tag{29.40}$$

The pair of coefficients α_x and α_y convert an image point into an ellipse which reduces to a circle, the familiar spherical aberration disc of round lenses, if $M_x = M_y$, so that (12) = (21). The coefficient β measures the so-called 'star' aberration, and γ the 'rosette'.

The aperture aberration coefficients can be written in many different ways. We now list a few of the formulae that have been found convenient in practice.

i. *Real aberrations*, $\phi(z)$, $p_2(z)$ and $Q_2(z)$ may all be present. For this case only, we list distortions as well as aperture aberrations, to show that they can be derived from one another by symmetry arguments.

$$\begin{aligned}
W_x x^{(1)}(z_c) = &\left(t_{xc}\int_a^c \sigma_{x4}\,dz - s_{xc}\int_o^c \sigma_{x3}\,dz\right)x_o^3 \\
&+ \left(t_{xc}\int_a^c \sigma_{xy}\,dz - s_{xc}\int_o^c \sigma_{12}\,dz\right)x_o y_o^2 \\
&+ \left(t_{xc}\int_a^c \tau_{x3}\,dz - s_{xc}\int_o^c \tau_{x4}\,dz\right)x_a^3 \\
&+ \left(t_{xc}\int_a^c \tau_{12}\,dz - s_{xc}\int_o^c \tau_{xy}\,dz\right)x_a y_a^2 \\
W_y y^{(1)}(z_c) = &\left(t_{yc}\int_a^c \sigma_{y4}\,dz - s_{yc}\int_o^c \sigma_{y3}\,dz\right)y_o^3 \\
&+ \left(t_{yc}\int_a^c \sigma_{xy}\,dz - s_{yc}\int_o^c \sigma_{21}\,dz\right)x_o^2 y_o \\
&+ \left(t_{yc}\int_a^c \tau_{y3}dz - s_{yc}\int_o^c \tau_{y4}\,dz\right)y_a^3 \\
&+ \left(t_{yc}\int_a^c \tau_{21}\,dz - s_{yc}\int_o^c \tau_{xy}\,dz\right)x_a^2 y_a
\end{aligned} \tag{29.41}$$

with $c \coloneqq z_c$ and

$$\begin{aligned}
\sigma_{x4} &= 4(A_x s_x^4 + B_x s_x^2 s_x'^2 + N s_x'^4)\hat{\phi}_i^{1/2} \\
\sigma_{y4} &= 4(A_y s_y^4 + B_y s_y^2 s_y'^2 + N s_y'^4)\hat{\phi}_i^{1/2} \\
\tau_{x4} &= 4(A_x t_x^4 + B_x t_x^2 t_x'^2 + N t_x'^4)\hat{\phi}_i^{1/2} \\
\tau_{y4} &= 4(A_y t_y^4 + B_y t_y^2 t_y'^2 + N t_y'^4)\hat{\phi}_i^{1/2}
\end{aligned}$$

$$
\begin{aligned}
\sigma_{x3} &= 2\hat{\phi}_i^{1/2}(2A_x s_x^3 t_x + B_x s_x s_x'(s_x t_x)' + 2N s_x'^3 t_x') \\
\sigma_{y3} &= 2\hat{\phi}_i^{1/2}(2A_y s_y^3 t_y + B_y s_y s_y'(s_y t_y)' + 2N s_y'^3 t_y') \\
\tau_{x3} &= 2\hat{\phi}_i^{1/2}(2A_x s_x t_x^3 + B_x (s_x t_x)' t_x t_x' + 2N s_x' t_x'^3) \\
\tau_{y3} &= 2\hat{\phi}_i^{1/2}(2A_y s_y t_y^3 + B_y (s_y t_y)' t_y t_y' + 2N s_y' t_y'^3) \\
\sigma_{xy} &= \hat{\phi}_i^{1/2}\Big[2C s_x^2 s_y^2 + R\{s_x^2 (s_y^2)' - (s_x^2)' s_y^2\} \\
&\qquad + 2B_x s_x^2 s_y'^2 + 2B_y s_x'^2 s_y^2 + 4N s_x'^2 s_y'^2\Big] \\
\tau_{xy} &= \hat{\phi}_i^{1/2}\Big[2C t_x^2 t_y^2 + R\{t_x^2 (t_y^2)' - (t_x^2)' t_y^2\} \\
&\qquad + 2B_x t_x^2 t_y'^2 + 2B_y t_x'^2 t_y^2 + 4N t_x'^2 t_y'^2\Big] \\
\sigma_{21} &= \hat{\phi}_i^{1/2}\Big[2C s_x^2 s_y t_y + R\{s_x^2 (s_y t_y)' - (s_x^2)' s_y t_y\} \\
&\qquad + 2B_x s_x^2 s_y' t_y' + 2B_y s_x'^2 s_y t_y + 4N s_x'^2 s_y' t_y'\Big] \\
\tau_{21} &= \hat{\phi}_i^{1/2}\Big[2C t_x^2 s_y t_y + R\{t_x^2 (s_y t_y)' - (t_x^2)' s_y t_y\} \\
&\qquad + 2B_x t_x^2 s_y' t_y' + 2B_y t_x'^2 s_y t_y + 4N t_x'^2 s_y' t_y'\Big] \\
\sigma_{12} &= \hat{\phi}_i^{1/2}\Big[2C s_x t_x s_y^2 + R\{s_x t_x (s_y^2)' - (s_x t_x)' s_y^2\} \\
&\qquad + 2B_x s_x t_x s_y'^2 + 2B_y s_x' t_x' s_y^2 + 4N s_x' t_x' s_y'^2\Big] \\
\tau_{12} &= \hat{\phi}_i^{1/2}\Big[2C s_x t_x s_y^2 + R\{s_x t_x (t_y^2)' - (s_x t_x)' t_y^2\} \\
&\qquad + 2B_x s_x t_x t_y'^2 + 2B_y s_x' t_x' t_y^2 + 4N s_x' t_x' t_y'^2\Big]
\end{aligned}
\tag{29.42}
$$

ii. *Real aberrations, quadrupoles only, formulae indicating the sign of the various coefficients.*

At the real line image at which t_x vanishes, $z = z_i =: i$, we have

$$
\begin{aligned}
(0030) &= \frac{1}{6}\int_o^i t_x'^4\, dz + \frac{1}{6}\int_o^i \left\{4\left(\frac{3\gamma p_2}{\hat{\phi}} - \frac{4\eta Q_2}{\hat{\phi}^{1/2}}\right)^2 + \frac{4p_2^2}{\hat{\varphi}^2}(6-\gamma^2)\right\} t_x^4\, dz \\
&= \frac{1}{6}\int_o^i t_x'^4\, dz + \frac{1}{6}(3\overline{n}^2 - 2n + 2)\beta^4 \int_o^i q^2 t_x^4\, dz
\end{aligned}
$$

$$
\begin{aligned}
(0012) &= -\frac{1}{2}t'^2_{xl}t_{yi}t'_{yi} + \frac{3}{2}\int_o^i t'^2_x t'^2_y\, dz \\
&\quad -\frac{1}{8}\int_o^i \left(\frac{p_2^2}{\hat{\phi}} + \frac{4\eta\gamma p_2 Q_2}{\hat{\phi}^{3/2}} - \frac{8\eta^2 Q_2^2}{\hat{\phi}}\right) t_x^2 t_y^2\, dz \\
&= -\frac{1}{2}t'^2_{xl}t_{yi}t'_{yi} + \frac{3}{2}\int_o^i t'^2_x t'^2_y\, dz - \frac{1}{2}(\overline{n}^2 - 2n - 2)\beta^2 \int_o^i q^2 t_x^2 t_y^2\, dz
\end{aligned}
\tag{29.43}
$$

At the real line image at which t_y vanishes, $z = z_j$, (0003) and (0021) are obtained from (29.43) by interchanging t_x and t_y. If the image is stigmatic, $z_i = z_j$, then (0012) = (0021). As before, $\overline{n} = n/\gamma$.

It is clear from Eq. (29.43) that (0030) and (0003) cannot change sign in the nonrelativistic approximation ($\gamma = 1$) and that the mixed coefficients cannot change sign for stigmatic magnetic quadrupole systems (Ovsyannikova and Yavor, 1965; Moses, 1966; Hawkes, 1966/67b). Still in the nonrelativistic case, (0012) = (0021) is always positive in stigmatic systems provided that $1-\sqrt{3} \le n \le 1 + \sqrt{3}$; this range includes magnetic ($n = 0$) and achromatic ($n = 1$) systems. When the relativistic correction becomes large, however, (0030) and (0003) can vanish as a negative term appears in the integrand when $6 - \gamma^2 \le 0$, that is, when the accelerating voltage exceeds about 0.7 MV (Rose, 1967).

iii. *Quadrupoles and octopoles only* ($\phi = const$). Asymptotic aberrations expressed in terms of x_o and x'_o in $z = z_{ox}$, y_o and y'_o in $z = z_{oy}$ and evaluated at the asymptotic image planes $z = z_{ix}$ and $z = z_{iy}$ conjugate to $z = z_{ox}$ and $z = z_{oy}$, respectively.

$$
\begin{aligned}
(0030) &= \int_{-\infty}^{\infty} (F - \overline{F}) H_x^4\, dz \\
(0003) &= \int_{-\infty}^{\infty} (F + \overline{F}) H_y^4\, dz \\
(0012) &= \int_{-\infty}^{\infty} \left\{ K H_x^2 H_y^2 + L H_x H_y (H_x H'_y - H'_x H_y) \right\} dz + \frac{1}{4}(d_i m_x^2 m_y^2 - d_o) \\
(0021) &= \int_{-\infty}^{\infty} \left\{ K H_x^2 H_y^2 + L H_x H_y (H_x H'_y - H'_x H_y) \right\} dz - \frac{1}{4}(d_i m_x^2 m_y^2 - d_o)
\end{aligned}
\tag{29.44}
$$

where

$$
\begin{aligned}
F &= \frac{1}{6\hat{\phi}^2}\left(2p_2^2 - 4\eta\gamma p_2 Q_2 \hat{\phi}^{1/2} + 3\eta^2 Q_2^2 \hat{\phi}\right) + \frac{5\varepsilon p_2^2}{6\hat{\phi}} - 2\Xi \\
&= \frac{\beta^4 q^2}{6}(3\overline{n}^2 - 2n + 3) - 2\tau\omega \\
\overline{F} &= \frac{1}{48\hat{\phi}}\left(\gamma p''_2 - 2\eta Q''_2 \hat{\phi}^{1/2}\right) = -\frac{1}{24}\beta^2 q'' \\
K &= -\frac{1}{2\hat{\phi}^2}\left(\gamma p_2 - \eta Q_2 \hat{\phi}^{1/2}\right)^2 + \frac{\varepsilon p_2^2}{2\hat{\phi}} + 6\Xi \\
&= -\frac{1}{2}\beta^4 q^2(\overline{n}^2 - 2n + 1) + 6\tau\omega \\
L &= \frac{3}{8}\frac{\gamma p'_2 - 2\eta Q'_2 \hat{\phi}^{1/2}}{\hat{\phi}} = -\frac{3}{4}\beta^2 q'_2 \\
d_i &= z_{ix} - z_{iy} \\
d_o &= z_{ox} - z_{oy}
\end{aligned}
\tag{29.45}
$$

iv. *Quadrupoles and octopoles only* ($\phi = const$). As (iii) but free of derivatives of p_2 and Q_2.

$$
\begin{aligned}
(0030) = \int_{-\infty}^{\infty} \Bigg\{ & \frac{1}{24}\left(\frac{7 + 16\varepsilon\hat{\phi}}{\hat{\phi}} p_2^2 - 12\eta\gamma \frac{p_2 Q_2}{\hat{\phi}^{3/2}} + \frac{8\eta^2 Q_2^2}{\hat{\phi}}\right) H_x^4 \\
& - \frac{1}{4}\left(\frac{\gamma p_2}{\hat{\phi}} - \frac{2\eta Q_2}{\hat{\phi}^{1/2}}\right) H_x^2 H_x'^2 \Bigg\} dz - 2\int_{-\infty}^{\infty} \Xi H_x^4 \, dz \\
= \int_{-\infty}^{\infty} & \left\{ \frac{1}{6}\left(3\overline{n}^2 - 2n + 2\right)\beta^4 q^2 H_x^4 + \frac{1}{2}\beta^2 q H_x^2 H_x'^2 \right\} dz \\
& - 2\tau \int_{-\infty}^{\infty} \omega H_x^4 \, dz
\end{aligned}
$$

$$
\begin{aligned}
(0003) &= \int_{-\infty}^{\infty} \left\{ \frac{1}{24}\left(\frac{7+16\varepsilon\hat{\phi}}{\hat{\phi}} p_2^2 - 12\eta\gamma \frac{p_2 Q_2}{\hat{\phi}^{3/2}} + \frac{8\eta^2 Q_2^2}{\hat{\phi}} \right) H_y^4 \right. \\
&\quad \left. + \frac{1}{4}\left(\frac{\gamma p_2}{\hat{\phi}} - \frac{2\eta Q_2}{\hat{\phi}^{1/2}} \right) H_y^2 H_y'^2 \right\} dz - 2 \int_{-\infty}^{\infty} \Xi H_y^4 \, dz \\
&= \int_{-\infty}^{\infty} \left\{ \frac{1}{6}\left(3\overline{n}^2 - 2n + 2 \right) \beta^4 q^2 H_y^4 - \frac{1}{2}\beta^2 q H_y^2 H_y'^2 \right\} dz \\
&\quad - 2\tau \int_{-\infty}^{\infty} \omega H_y^4 \, dz \\
\begin{pmatrix} 0012 \\ 0021 \end{pmatrix} &= -\frac{1}{8} \int_{-\infty}^{\infty} \left\{ \left(\frac{p_2^2}{\hat{\phi}^2} + \frac{4\eta\gamma p_2 Q_2}{\hat{\phi}^{3/2}} - \frac{8\eta^2 Q_2^2}{\hat{\phi}} \right) H_x^2 H_y^2 \right. \\
&\quad \left. + 3\left(\frac{\gamma p_2}{\hat{\phi}} - \frac{2\eta Q_2}{\hat{\phi}^{1/2}} \right) (H_x^2 H_y'^2 - H_x'^2 H_y^2) \right\} dz + 6 \int_{-\infty}^{\infty} \Xi H_x^2 H_y^2 \, dz \\
&\quad \pm \frac{1}{4}(d_i m_x^2 m_y^2 - d_o) \\
&= -\int_{-\infty}^{\infty} \left\{ \left(\frac{\overline{n}}{2} - n - 1 \right) \beta^4 q^2 H_x^2 H_y^2 + \frac{3}{4}\beta^2 q (H_x^2 H_y'^2 - H_x'^2 H_y^2) \right\} dz \\
&\quad + 6\tau \int_{-\infty}^{\infty} \omega H_x^2 H_y^2 \, dz \pm \frac{1}{4}(d_i m_x^2 m_y^2 - d_o)
\end{aligned}
\tag{29.46}
$$

29.4 Chromatic Aberrations

In the general case, in which $\phi(z)$, $p_2(z)$ and $Q_2(z)$ may all be present, the asymptotic chromatic aberration is given in the planes z_{ix} and z_{iy} conjugate to z_{ox} and z_{oy} respectively by the following expressions:

$$
\begin{aligned}
\Delta x_i &= (C_{cx} x_o' + C_{Mx} x_o) \frac{\Delta\phi}{\hat{\phi}_o} \\
\Delta y_i &= (C_{cy} y_o' + C_{My} y_o) \frac{\Delta\phi}{\hat{\phi}_o}
\end{aligned}
\tag{29.47}
$$

where

$$\begin{aligned} C_{cx} &= \int_{-\infty}^{\infty} (k_\phi + k_q) H_x^2 \, dz \\ C_{cy} &= \int_{-\infty}^{\infty} (k_\phi - k_q) H_y^2 \, dz \\ C_{Mx} &= \int_{-\infty}^{\infty} (k_\phi + k_q) G_x H_x \, dz \\ C_{My} &= \int_{-\infty}^{\infty} (k_\phi - k_q) G_y H_y \, dz \end{aligned} \tag{29.48}$$

and

$$\begin{aligned} k_\phi &= -\left(\frac{\hat{\phi}_0}{\hat{\phi}}\right)^{1/2} \frac{\gamma(5+\gamma^2)\phi'^2}{16\hat{\phi}^2} \\ k_q &= \left(\frac{\hat{\phi}_0}{\hat{\phi}}\right)^{1/2} \frac{(1+\gamma^2)p_2 - 2\eta\gamma Q_2 \hat{\phi}^{1/2}}{4\hat{\phi}} \end{aligned} \tag{29.49}$$

With the notation of Eqs (29.25−29.29), the coefficients (29.48) take the following form in the pure quadrupole case:

$$\begin{aligned} C_{cx} &= \frac{1}{2}\nu\beta^2 \int_{-\infty}^{\infty} q(z) H_x^2 \, dz; \qquad C_{cy} = -\frac{1}{2}\nu\beta^2 \int_{-\infty}^{\infty} q(z) H_y^2 \, dz; \\ C_{Mx} &= \frac{1}{2}\nu\beta^2 \int_{-\infty}^{\infty} q(z) G_x H_x dz; \quad C_{My} = -\frac{1}{2}\nu\beta^2 \int_{-\infty}^{\infty} q(z) G_y H_y \, dz; \end{aligned} \tag{29.50}$$

in which

$$\nu := \frac{n - \gamma^2}{\gamma} \approx n - 1 \tag{29.51}$$

Thus if $n = \gamma^2 \approx 1$, which implies $\beta_M^2 = 2\beta_E^2$, a combined electrostatic and magnetic quadrupole will be *achromatic*, provided that the field functions $p_2(z)$ and $Q_2(z)$ are identical. Fig. 29.3 shows a quadrupole in which this condition is satisfied to a good approximation.

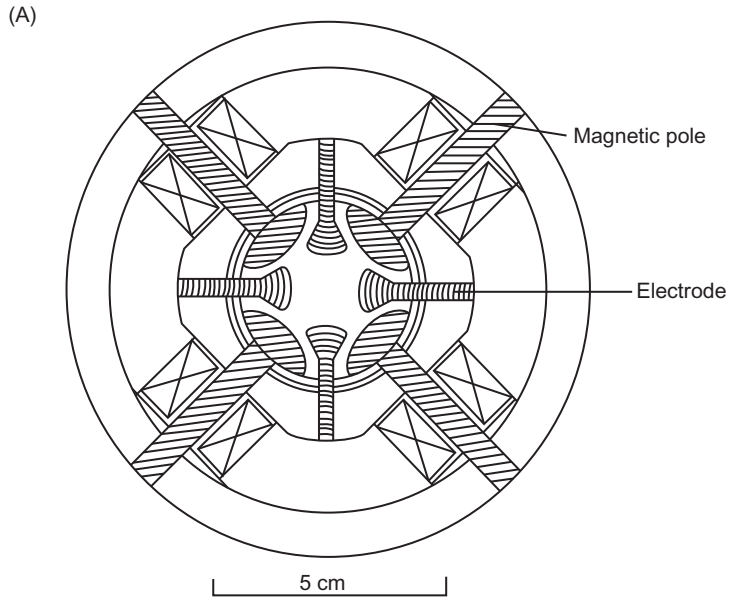


Figure 29.3
A combined electrostatic and magnetic quadrupole, with which the achromatic condition can be satisfied to a good approximation. (A) Cross-section and (B) view of such a lens. *Courtesy D.F. Hardy.*

This achromatic condition was first derived by Kel'man and Yavor (1961) and rediscovered independently by Septier (1963). The relativistic formulae were first given by Hawkes (1964, 1965c). Extensive experimental work on mixed magnetic and electromagnetic quadrupoles was subsequently performed by Hardy (1967). The achromatic condition was originally established for real aberration coefficients but these are identical with the asymptotic ones apart from the presence of different pairs of paraxial solutions. The condition is thus applicable in both cases.

In a practical design, it is more than likely that $p_2(z)$ and $Q_2(z)$ will not have quite the same shape. The effect of small differences has been analysed with the aid of a particular field model, the rectangular distribution (see Chapter 39 of Volume 2) by Shpak and Yavor (1964, 1965) (cf. Yavor et al., 1964).

29.5 Quadrupole Multiplets

Since quadrupole lenses have a diverging action in one plane and a converging effect in the other, they are commonly combined into multiplets. Formulae from which the total aberration coefficients can be calculated, given those of the individual components, are therefore required. These have a strong family resemblance to their counterparts for round lenses, listed in Chapter 27, Aberration Matrices and the Aberrations of Lens Combinations; we give only the principal expressions here, referring to Hawkes (1970a) for further details.

As in Chapter 27, Aberration Matrices and the Aberrations of Lens Combinations, we introduce vectors,

$$\begin{aligned}\boldsymbol{x} &= (x\ \ x'\ \ x^3\ \ xy^2\ \ x^2x'\ \ y^2x'\ \ xyy'\ \ xx'^2\ \ xy'^2\ \ yx'y'\ \ x'^3\ \ x'y'^2)^T \\ \boldsymbol{y} &= (y\ \ y'\ \ y^3\ \ x^2y\ \ y^2y'\ \ x^2y'\ \ xyx'\ \ yy'^2\ \ yx'^2\ \ xx'\ \ y'\ \ y'^3\ \ x'^2y')^T\end{aligned} \tag{29.52}$$

(in which T denotes transpose as usual) with the convention that suffix o attached to x, x' indicates quantities in the object plane z_{ox}, but attached to y, y', quantities in $z = z_{oy}$ are meant. We write

$$\boldsymbol{x}_m = \boldsymbol{M}\boldsymbol{x}_o \quad \boldsymbol{y}_n = \boldsymbol{N}\boldsymbol{y}_o \tag{29.53}$$

in which $\boldsymbol{M}_{\boldsymbol{x}}$ is now denoted by $\boldsymbol{M}$ and $\boldsymbol{M}_{\boldsymbol{y}}$ by $\boldsymbol{N}$, to prevent undue proliferation of the suffixes. The transfer matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ relating z_{ix} and z_{ox}, z_{iy} and z_{oy} respectively, have the block structure of Eq. (27.3):

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{M}_3 & \boldsymbol{M}_4 \end{pmatrix} \quad \boldsymbol{N} = \begin{pmatrix} \boldsymbol{N}_1 & \boldsymbol{N}_2 \\ \boldsymbol{N}_3 & \boldsymbol{N}_4 \end{pmatrix} \tag{29.54}$$

Here $\boldsymbol{M}_1$ and $\boldsymbol{N}_1$ are the 2×2 paraxial transfer matrices in the $x-z$ and $y-z$ planes; $\boldsymbol{M}_3$ and $\boldsymbol{N}_3$ are null; $\boldsymbol{M}_4$ and $\boldsymbol{N}_4$ encode the addition rules for the aberration coefficients; and $\boldsymbol{M}_2$, $\boldsymbol{N}_2$ are the 2×10 aberration matrices similar to Eq. (27.6). We write

$$\boldsymbol{M}_2 = \begin{pmatrix} Mm_{11} & Mm_{12} & \cdots & Mm_{19} & Mm_{1,10} \\ m_{21} & m_{22} & \cdots & m_{29} & m_{2,10} \end{pmatrix}$$
$$\boldsymbol{N}_2 = \begin{pmatrix} Nn_{11} & Nn_{12} & \cdots & Nn_{19} & Nn_{1,10} \\ n_{21} & n_{22} & \cdots & n_{29} & n_{2,10} \end{pmatrix} \tag{29.55}$$

Suppose now that we have a second quadrupole characterized by a similar pair of matrices, $\boldsymbol{M}'$ and $\boldsymbol{N}'$, and that the total magnifications in the two planes are P, Q:

$$P = M'M \quad Q = \boldsymbol{N}'N \tag{29.56}$$

Then

$$\begin{aligned} \boldsymbol{x_p} &= \boldsymbol{M}'\boldsymbol{x_m} = \boldsymbol{M}'\boldsymbol{M}\boldsymbol{x_o} =: \boldsymbol{P}\boldsymbol{x_o} \\ \boldsymbol{y_p} &= \boldsymbol{N}'\boldsymbol{y_n} = \boldsymbol{N}'\boldsymbol{N}\boldsymbol{y_o} =: \boldsymbol{Q}\boldsymbol{y_o} \end{aligned} \tag{29.57}$$

where $\boldsymbol{P}$ and $\boldsymbol{Q}$ of course have the same block structure as $\boldsymbol{M}$, $\boldsymbol{M}'$, $\boldsymbol{N}$ and $\boldsymbol{N}'$. The paraxial submatrices $\boldsymbol{P}_1$ and $\boldsymbol{Q}_1$ have already been discussed in Section 19.1. It is the elements of the upper row of $\boldsymbol{P}_2$ and $\boldsymbol{Q}_2$, denoted by Pp_{1j} and Qq_{1j} ($j = 1-10$), that are of interest. It can be shown that

$$\begin{aligned}
p_{11} &= m_{11} + M^2 m'_{11} + c_m M m'_{13} + c_m^2 m'_{16} + c_m^3 m m'_{19} \\
p_{12} &= m_{12} + N^2 m'_{12} + c_m m N^2 m'_{14} + c_n N m'_{15} + c_n^2 m'_{17} + c_m c_n m N m'_{18} \\
&\quad + c_m c_n^2 m m'_{1,10} \\
p_{13} &= m_{13} + r m'_{13} + 2 r c_m m m'_{16} + 3 r c_m^2 m^2 m'_{19} \\
p_{14} &= m_{14} + r m^2 N^2 m'_{14} + r c_n m^2 N m'_{18} + r c_n^2 m^2 m'_{1,10} \\
p_{15} &= m_{15} + r m'_{15} + 2 r c_n n m'_{17} + r c_m m m'_{18} + 2 r c_m c_n m n m'_{1,10} \\
p_{16} &= m_{16} + r^2 m^2 m'_{16} + 3 r^2 c_m m^3 m'_{19} \\
p_{17} &= m_{17} + r^2 n^2 m'_{17} + r^2 c_m m n^2 m'_{1,10} \\
p_{18} &= m_{18} + r^2 m^2 m'_{18} + 2 r^2 c_n m^2 n m'_{1,10} \\
p_{19} &= m_{19} + r^3 m^4 m'_{19} \\
p_{1,10} &= m_{1,10} + r^3 m^2 n^2 m'_{1,10} \\
q_{11} &= n_{11} + N^2 n'_{11} + c_n N n'_{13} + c_n^2 n'_{16} + c_n^3 n n'_{19} \\
q_{12} &= n_{12} + M^2 n'_{12} + c_n M^2 n n'_{14} + c_m M n'_{15} + c_m^2 n'_{17} + c_m c_n M n n'_{18} \\
&\quad + c_n c_m^2 n n'_{1,10} \\
q_{13} &= n_{13} + r n'_{13} + 2 r c_n n n'_{16} + 3 r c_n^2 n^2 n'_{19} \\
q_{14} &= n_{14} + r M^2 n^2 n'_{14} + r c_m M n^2 n'_{18} + r c_m^2 n^2 n'_{1,10}
\end{aligned}$$

$$
\begin{aligned}
q_{15} &= n_{15} + rn'_{15} + 2rc_m mn'_{17} + rc_n nn'_{18} + 2rc_m c_n mnn'_{1,10} \\
q_{16} &= n_{16} + r^2 n^2 n'_{16} + 3r^2 c_n n^3 n'_{19} \\
q_{17} &= n_{17} + r^2 m^2 n'_{17} + r^2 c_n m^2 nn'_{1,10} \\
q_{18} &= n_{18} + r^2 n^2 n'_{18} + 2r^2 c_m n^2 mn'_{1,10} \\
q_{19} &= n_{19} + r^3 n^4 n'_{19} \\
q_{1,10} &= n_{1,10} + r^3 m^2 n^2 n'_{1,10}
\end{aligned}
\tag{29.58}
$$

in which $c_m \coloneqq -1/f_{xi}$, and $c_n = -1/f_{yi}$ as in Eq. (27.5).

The coefficients m_{1j}, n_{1j} m'_{1j}, n'_{1j}, and hence p_{1j} and q_{1j} also, may be written as polynomials in the corresponding inverse magnification m and n (for m_{1j} and n_{1j}), m' and n' (for m'_{1j} and n'_{1j}) and p and q (for p_{1j} and q_{1j}). The formulae giving the numerous coefficients that occur in these polynomials for the complete system in terms of those of the two individual lenses (or partial systems) are set out explicitly in Hawkes (1970e) and are not reproduced here.

CHAPTER 30

The Aberrations of Cylindrical Lenses

The aberrations of cylindrical lenses, the paraxial properties of which were outlined in Chapter 20, Cylindrical Lenses, have been analysed by several authors. For electrostatic lenses, these were first calculated by Leitner (1942), who used the method of variation of parameters to obtain formulae in several different forms for all the aberration coefficients. Strashkevich (1965) likewise listed formulae for the latter, some of which had appeared earlier in Strashkevich and Pilat (1951, 1952) and Strashkevich and Gluzman (1954). Full expressions for the primary aberrations of combined electrostatic and magnetic lenses were derived by Shtepa (1952), Laudet (1955a,b) and Rheinfurth (1955). Shi (1956) used Seman's formulation of the method of characteristic functions to analyse electrostatic lenses, employing the reduced equations of motion and eliminating derivatives of the Gaussian solutions. The coefficients, recalculated using characteristic functions, are listed in Hawkes (1966/67a), where a comparison with all the earlier formulae is to be found. For details of the calculation, the reader is referred to that paper; here we merely indicate the form of the results in the general case and list the full formulae only for electrostatic cylindrical lenses. We express the aberrations in terms of object and aperture coordinates; for details of the alternative choice, again see the paper cited above.

The paraxial solution expressed in terms of position coordinates in the object and aperture planes was not derived explicitly in Chapter 20, Cylindrical Lenses. With $s(z_o) = t(z_a) = 1$ and $t(z_o) = s(z_a) = 0$, a lengthy but straightforward calculation yields

$$\begin{aligned} x(z) &= -r(z)(y_a - y_o) + x_o\left\{ s(z) + \eta r(z)\int_o^a \frac{Bs}{\phi^{1/2}}\,dz \right\} \\ &\quad + x_a\left\{ t(z) + \eta r(z)\int_o^a \frac{Bt}{\phi^{1/2}}\,dz \right\} \\ y(z) &= \{1 - R(z)\}y_o + Ry_a + sx_o + Tx_a \end{aligned} \tag{30.1}$$

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00030-4
© 2018 Elsevier Ltd. All rights reserved.

with

$$
\begin{aligned}
r(z) &:= -\frac{\eta}{FW}\left(s\int_o^z \frac{Bt}{\phi^{1/2}}d\zeta - t\int_o^z \frac{Bs}{\phi^{1/2}}d\zeta\right)\\
R(z) &:= \frac{1}{F}\int_o^z \frac{d\zeta}{\phi^{1/2}}\\
S(z) &:= \eta\int_o^z \frac{Bs}{\phi^{1/2}}\,d\zeta - \frac{\eta}{F}\int_o^a \frac{Bs}{\phi^{1/2}}dz\int_o^z \frac{d\zeta}{\phi^{1/2}}\\
&\quad + \eta^2\int_o^a \frac{Bs}{\phi^{1/2}}\,dz\int_o^z \frac{Br}{\phi^{1/2}}d\zeta\\
T(z) &:= \eta\int_o^z \frac{Bt}{\phi^{1/2}}d\zeta - \frac{\eta}{F}\int_o^a \frac{Bt}{\phi^{1/2}}dz\int_o^z \frac{d\zeta}{\phi^{1/2}}\\
&\quad + \eta^2\int_o^a \frac{Bt}{\phi^{1/2}}dz\int_o^z \frac{Br}{\phi^{1/2}}d\zeta\\
F &:= \int_o^a \frac{dz}{\phi^{1/2}} + \frac{\eta^2}{W}\int_o^a \frac{Bs}{\phi^{1/2}}\left(\int_o^z \frac{Bt}{\phi^{1/2}}d\zeta\right)dz\\
&\quad - \frac{\eta^2}{W}\int_o^a \frac{Bt}{\phi^{1/2}}\left(\int_o^z \frac{Bs}{\phi^{1/2}}d\zeta\right)dz\\
W &:= \phi^{1/2}(st' - s't)
\end{aligned}
\tag{30.2}
$$

The expressions (30.1) are substituted into the perturbation $M^{(4)}$ (M^I in Chapter 22, Perturbation Theory: General Formalism), and aberration coefficients are extracted by partial differentiation (22.30). In the general case, the geometrical aberrations in an arbitrary plane take the form

$$
\begin{aligned}
x^{(1)}(z) = {} & (300)_x x_o^3 + (210)_x x_o^2 x_a + (201)_x x_o^2 Y\\
& + (120)_x x_o x_a^2 + (102)_x x_o Y^2 + (111)_x x_o x_a Y\\
& + (030)_x x_a^3 + (021)_x x_a^2 Y + (012)_x x_a Y^2 + (003)_x Y^3
\end{aligned}
\tag{30.3}
$$

with a similar expression for $y^{(1)}(z)$. We have introduced $Y := y_a - y_o$. The formulae for the coefficients are extremely complicated in the mixed case, in which electrostatic and magnetic fields are both present; they are listed in full in Hawkes (1966/67a) and are not

reproduced here. In the simpler case of electrostatic cylindrical lenses, there are fewer terms. The results are then as follows.

The function $M^{(4)}$ is given by

$$M^{(4)} = \left(\frac{\phi^{(4)}}{48\phi^{1/2}} - \frac{\phi''^2}{32\phi^{3/2}}\right)x^4 - \frac{\phi''}{8\phi^{1/2}}x^2\left(x'^2 + y'^2\right) - \frac{1}{8}\phi^{1/2}\left(x'^2 + y'^2\right)^2 \tag{30.4}$$

in the nonrelativistic approximation. Into this, we substitute

$$\begin{aligned} x &= sx_o + tx_a \\ y &= (1-R)y_o + Ry_a \end{aligned} \tag{30.5}$$

where

$$R = \frac{R_z}{F}, \quad R_z \coloneqq \int_{z_o}^{z} \frac{d\zeta}{\phi^{1/2}}, \quad F \coloneqq \int_{z_o}^{z_a} \frac{dz}{\phi^{1/2}} \tag{30.6}$$

giving

$$\begin{aligned} x^{(1)}(z_c) &= (300)x_o^3 + (030)x_a^3 + (120)x_o x_a^2 + (210)x_o^2 x_a \\ &\quad + (102)x_o Y^2 + (012)x_a Y^2 \\ y^{(1)}(z_c) &= (201)x_o^2 Y + (111)x_o x_a Y + (021)x_a^2 Y + (003)Y^3 \end{aligned} \tag{30.7}$$

in some arbitrary plane $z = z_c$. The coefficients are given by

$$\begin{aligned} (300) = \frac{t}{2W}\int_a^c \left\{ \left(\frac{\phi^{(4)}}{6\phi^{1/2}} - \frac{\phi''^2}{4\phi^{3/2}}\right)s^4 - \frac{\phi''}{\phi^{1/2}}s^2 s'^2 - \phi^{1/2}s'^4 \right\} dz \\ - \frac{s}{2W}\int_o^c \left\{ \left(\frac{\phi^{(4)}}{6\phi^{1/2}} - \frac{\phi''^2}{4\phi^{3/2}}\right)s^3 t - \frac{\phi''}{\phi^{1/2}}(s^2)'(st)' - \phi^{1/2}s'^3 t' \right\} dz \end{aligned}$$

and $(300) \rightarrow (030)$ when $s \rightarrow t$, $t \rightarrow -s$, $o \leftrightarrow a$;

$$\begin{aligned} (102) = -\frac{t}{4F^2}\int_a^c \left(\frac{\phi''}{\phi^{3/2}}s^2 + 2\phi^{-1/2}s'^2\right) dz \\ + \frac{s}{4F^2}\int_o^c \left(\frac{\phi''}{\varphi^{3/2}}st + 2\phi^{-1/2}s't'\right) dz \end{aligned}$$

and (102) → (012) when $s \to t$, $t \to -s$, $o \leftrightarrow a$;

$$(120) = \frac{3t}{4W}\int_a^c \left\{ \left(\frac{\phi^{(4)}}{3\phi^{1/2}} - \frac{\phi''^2}{2\phi^{3/2}} \right) s^2 t^2 - \frac{\phi''}{3\phi^{1/2}} \left(s^2 t'^2 + 4ss'tt' + s'^2 t^2 \right) \right\} dz$$

$$- \frac{3s}{4W}\int_o^c \left\{ \left(\frac{\phi^{(4)}}{3\phi^{1/2}} - \frac{\phi''^2}{2\phi^{3/2}} \right) st^3 - \frac{\phi''}{\phi^{1/2}} (st)'tt' - 2\phi^{1/2} s't'^3 \right\} dz$$

and (120) → (210) when $s \to t$, $t \to -s$, $o \leftrightarrow a$

$$(003) = \frac{1}{2F^4} \left(F\int_o^c \frac{dz}{\phi^{3/2}} - R_z \int_o^c \frac{dz}{\phi^{3/2}} \right)$$

$$(201) = \frac{1}{4F}\int_o^c \left(\frac{\phi''}{\phi^{3/2}} s^2 + 2\phi^{-1/2} s'^2 \right) dz - \frac{R_z}{4F}\int_o^c \left(\frac{\phi''}{\phi^{3/2}} s^2 + 2\phi^{-1/2} s'^2 \right) dz$$

and (201) → (111) when $s^2 \to 2st$, $s'^2 \to 2s't'$ while (201) → (021) when $s^2 \to t^2$ and $s'^2 \to t'^2$.

CHAPTER 31

Parasitic Aberrations

We have so far been assuming that the electron optical system under consideration is perfect, in the sense that the various symmetries are exactly respected. This is certainly unrealistic, for every real system is imperfect: there is a limit to the precision attainable during the manufacture and assembly of its components. The electrodes and polepieces of a lens system designed to have rotational symmetry are never truly round but slightly elliptical, and after assembly their individual axes may be shifted and tilted with respect to one another. Moreover, in magnetic lenses, the polepiece material will be slightly inhomogeneous, and in electrostatic lenses, an asymmetric layer of insulating but charged contamination may be deposited on the electrode surfaces, again impairing the original symmetry. It is clear that the number of degrees of freedom in the imperfections increases drastically with the number of individual components in the device, so that complex systems are particularly prone to problems of this kind.

The electron optical defects caused by the various forms of imperfection are known as *parasitic aberrations.* Unlike the systematic aberrations of perfect systems, they cannot be determined accurately since the mechanical defects that cause them are never known exactly and may not be stable in time. The aim of the theory is to furnish an understanding of the possible types of parasitic aberrations and to establish tolerance limits on the precision of machining and alignment of the various components of a system.

This chapter is organized in two parts. The first part, Sections 31.1–31.4, follows the text of the first edition closely, while the second part (Section 31.6) introduces the deeper analysis that has become necessary with the proliferation of aberration correctors. The discussion of stigmators, formerly in Chapter 32, Paraxial Properties of Deflection Systems, has been placed at the end of the first part (Section 31.5) and the section on numerical methods has been extended. A final short section (31.7) describes the 'Uhlemann effect'. This too is a parasitic feature of magnetic lenses, though not in the same sense as the rest of this chapter, since no mechanical defect is involved. The first part has been retained in the belief that newcomers to the topic will appreciate this simpler approach.

31.1 Small Deviations from Rotational Symmetry; Axial Astigmatism

This situation is the most important, for in the vast majority of practical devices, with the exception of aberration correctors and deflectors, the lens systems are intended to be round.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00031-6
© 2018 Elsevier Ltd. All rights reserved.

Any deviations of the boundaries from the ideal rotational symmetry generate weak multipole fields of various orders, as outlined in Section 9.4.6 where magnetic lenses were considered. From the properties of multipole fields presented in Chapter 7, Series Expansions, it is obvious that the fields with the lowest multipole orders, $m = 1$, 2 and 3, will be the most important in the paraxial domain, as their potentials increase as r^m with the radial coordinate r.

Eccentricities, in the sense of shifts and tilts of the local optic axes, produce very weak dipole fields ($m = 1$), which merely cause a very small lateral deflection of the whole electron beam. The image is hence displaced bodily but remains stigmatic, with the result that these defects are not serious as long as they remain small.

Quadrupole fields ($m = 2$) are caused by ellipticity of the electrodes or polepieces and produce an astigmatism; the latter does not vanish even for the axial object point and is therefore often called the *axial* astigmatism. Owing to its great practical importance, this aberration was studied comprehensively in earlier decades. Substantial contributions were made by Glaser (1942/3), Bertein (1947e, 1948a), Bertein et al. (1947), Hillier and Ramberg (1947), Rang (1949a,b), Sturrock (1951b,1955), which should be read in conjunction with Archard (1953), Glaser (1952), Glaser and Schiske (1953), Der-Shvarts (1954) and Stoyanov (1955a,b). With the advance of modern computational methods, these early investigations have lost much of their significance, with the exception of Sturrock's work, and we discuss them only briefly. It should, however, not be forgotten that the idea of classifying departures from rotational symmetry according to their Fourier components, thereby revealing the character of the aberrations to be expected, first appeared in these studies, notably those of Bertein and Sturrock; this in turn led to the idea of the *stigmator* and permitted Archard to relate individual types of imperfection of magnetic lenses (corrugation of the polepieces, eccentricity and misorientation of the axes of the latter) to particular image defects.

It is customary to introduce dimensionless ellipticity parameters $\varepsilon_1(z)$, $\varepsilon_2(z)$; the electrostatic potential Φ in the paraxial domain then takes the form

$$\Phi(z, r, \varphi) = \phi(z) - \frac{r^2}{4}\phi''(z)\{1 + \varepsilon_1(z)\cos 2\varphi + \varepsilon_2(z)\sin 2\varphi\} + O(r^4) \tag{31.1}$$

This represents the superposition of a dominant round focusing field and two weak quadrupole fields with variable strengths and different orientations; we must assume that $|\varepsilon_1| \ll 1$, $|\varepsilon_2| \ll 1$. In the case of *constant* ellipticity, $\varepsilon_1 = \text{const}$, $\varepsilon_2 = \text{const}'$, it is possible to relate these parameters to the semiaxes $a > b$ of the bores in the electrodes. Without loss of generality, we introduce a constant rotation of the coordinate system about the optic axis, so that $\varepsilon_1 =: \bar{\varepsilon} > 0$, $\varepsilon_2 = 0$. We then have

$$\overline{\varepsilon} = \frac{a-b}{a+b} \tag{31.2}$$

The motions of the electrons in the two symmetry planes, $X_1 = 0$ and $X_2 = 0$, are decoupled and each coordinate satisfies a paraxial ray equation of the type associated with quadrupoles:

$$\phi X''_{1,2} + \frac{1}{2}\phi' X'_{1,2} + \frac{1}{4}\phi''(1 \pm \overline{\varepsilon})X_{1,2} = 0 \tag{31.3}$$

With the Picht transformation (15.40), $X_j = \upsilon_j \phi^{-1/4}$, this becomes

$$\upsilon''_{1,2} + \left\{ \frac{3}{16}\left(\frac{\phi'}{\phi}\right)^2 \pm \frac{\overline{\varepsilon}\phi''}{4\phi} \right\} \upsilon_{1,2} = 0 \tag{31.4}$$

In the Busch approximation for weak lenses, we first find the average focal length $\overline{f}$,

$$1/\overline{f} = (f_i f_o)^{-1/2} = \frac{3}{16} \int_{-\infty}^{\infty} \left(\frac{\phi'}{\phi}\right)^2 dz \tag{31.5}$$

The ellipticity causes small deviations from $\overline{f}$, given by

$$\Delta\left(1/\overline{f}\right) = \pm \frac{\overline{\varepsilon}}{4} \int_{-\infty}^{\infty} \frac{\phi''}{\phi} dz$$

Partial integration yields an expression proportional to formula (31.5) for $\overline{f}^{-1}$ and as this deviation is very small, we can simplify the relation to

$$\Delta\overline{f} = \mp \frac{4\overline{\varepsilon}}{3}\overline{f} \tag{31.6}$$

This expression is certainly an oversimplification since in most practical applications, the lenses are not weak, but it gives an estimate of the axial astigmatism and this is sufficient for rough calculations.

For the scalar potential in magnetic lenses an expansion similar to Eq. (31.1) holds and, in the case of constant ellipticity of the bores a relation of the form (31.2) can be derived, but thereafter the situation becomes far more complicated. Owing to the Larmor rotation of the electron rays in the lens, there is no coordinate system in which the ray equations decouple like Eq. (31.1). We now have to employ one of the general methods for calculating aberrations. This theory is comprehensively presented by Glaser (1952); we shall not repeat it here, since better techniques are now available.

31.2 Classification of the Parasitic Aberrations

Quite generally, there must exist an eikonal or characteristic function, from which the lateral aberrations can be obtained by partial differentiation. For rotationally symmetric systems these relations can be cast into the convenient form

$$\Delta x_o = \frac{\partial}{\partial x'_o} S\left(x_o, y_o, x'_o y'_o, z_o, z_i\right), \quad \Delta y_o = \frac{\partial S}{\partial y'_o} \tag{31.7}$$

where S is a characteristic function, also called the 'wave aberration', and Δx_o, Δy_o are the lateral aberrations referred back to the object plane in the usual way. These relations remain valid even if only the paraxial unperturbed contributions are rotationally symmetric; small asymmetric aberrations are then superimposed on the intrinsic aberrations of the round lens.

The assumption that the asymmetries are small allows us to neglect their fourth-order contributions to S in comparison with those from the rotationally symmetric part. We can therefore terminate the series expansion of the eikonal after the *third* order if only the asymmetric aberrations are being studied. Even so, the possible aberrations are quite numerous since S is now a polynomial consisting of all terms in the variables x_o, y_o, x'_o and y'_o up to third order. We first examine the *second-order* terms. In the expansion, we omit any terms that depend on x_o and y_o only, since they do not contribute to the derivatives (31.7); we then find

$$\begin{aligned} S = {} & x'_o(c_1 + a_{11}x_o + a_{12}y_o) + y'_o(c_2 + a_{21}x_o + a_{22}y_o) \\ & + \frac{1}{2}\left(b_{11}x'^2_o + 2b_{12}x'_o y'_o + b_{22}y'^2_o\right) \end{aligned} \tag{31.8}$$

Here a_{12} is not necessarily the same as a_{21}. It is now convenient to introduce new coefficients, writing

$$\begin{aligned} a_{11} &=: a'_1 + a_1, \quad a_{22} =: a'_1 - a_1, \quad a_{21} =: a_2 + a'_2, \quad a_{12} =: a_2 - a'_2, \\ b_{11} &=: b'_1 + b_1, \quad b_{22} =: b'_1 - b_1, \quad b_{12} =: b_2 \end{aligned}$$

We can then rewrite Eq. (31.8) as the sum of a rotationally symmetric term

$$S_s = a'_1\left(x_o x'_o + y_o y'_o\right) + a'_2\left(x_o y'_o - x'_o y_o\right) + \frac{1}{2}b'^2_1\left(x'^2_o + y'^2_o\right)$$

and an essentially asymmetric contribution

$$\begin{aligned} S_a = {} & a_1\left(x_o x'_o - y_o y'_o\right) + a_2\left(x_o y'_o + y_o x'_o\right) \\ & + \frac{1}{2}b_1\left(x'^2_o - y'^2_o\right) + b_2 x'_o y'_o + c_1 x'_o + c_2 y'_o \end{aligned} \tag{31.9}$$

with $S = S_a + S_s$. The expression for S_s, being composed of rotation-invariants, yields the aberrations to be expected in perfectly round lenses. These are an isotropic and an anisotropic alteration of the magnification (a_1' and a_2', respectively) and a defocus (b_1'). Since these are of no interest in the present context, we disregard them here.

The lateral aberrations, arising from Eq. (31.9), are now given by

$$\Delta x_o = c_1 + a_1 x_o + a_2 y_o + b_1 x_o' + b_2 y_o' \tag{31.10a}$$

$$\Delta y_o = c_2 + a_2 x_o - a_1 y_o + b_2 x_o' - b_1 y_o' \tag{31.10b}$$

The constants c_1 and c_2 describe a *lateral shift* of the image: $\Delta x_i = Mc_1$, $\Delta y_i = Mc_2$, M being the magnification. Such a shift results from tilts or lateral displacements of the lens parts relative to each other, which produce a weak deflection. Since the image remains stigmatic, this is not serious. The terms that are proportional to x_o and y_o describe a *paraxial distortion.* From the affine imaging relation,

$$\begin{aligned} x_i &= M(x_o + \Delta x_o) = M\{(1 + a_1)x_o + a_2 y_o\} \\ y_i &= M(y_o + \Delta y_o) = M\{a_2 x_o + (1 - a_1)y_o\} \end{aligned} \tag{31.11}$$

it can be seen that there are two principal directions in the image, which are orthogonal to each other and in which the magnification takes the extreme values

$$M_{1,2} = M\left\{1 \pm \left(a_1^2 + a_2^2\right)^{1/2}\right\} \tag{31.12}$$

Since $|M_1 - M_2| \ll |M|$, the deviations from uniform magnification are so small that this error is usually of very little practical concern.

The terms in x_o' and y_o' in Eq. (31.10) are the most important. By an appropriate rotation of the coordinate system about the optic axis, we can bring them into an uncoupled form:

$$\Delta\tilde{x}_o = b\tilde{x}_o' \quad \Delta\tilde{y}_o = -b\tilde{y}_o, \quad b = \sqrt{b_1^2 + b_2^2} \tag{31.13}$$

These describe the *axial astigmatism*, which is presented schematically in Fig. 31.1. Eq. (31.13) represent the lateral aberration in the circle of least confusion: with $\tilde{x}_o' = \alpha \cos\varphi$

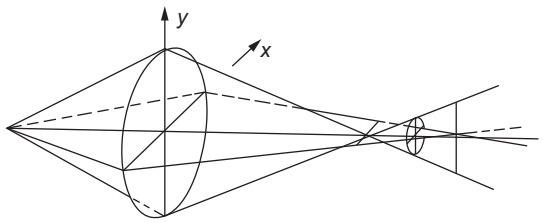


Figure 31.1
Axial astigmatism.

and $\tilde{y}'_o = \alpha \sin\varphi$, α being the semiaperture angle, we find $\Delta\tilde{x}_o = b\alpha\cos\varphi$, $\Delta\tilde{y}_o = -b\alpha\sin\varphi$. Obviously the sense of rotation on the circle of least confusion is opposite to that on the aperture cone. Referred back to object space, the distance between the two line foci is $2b$.

Among the lateral aberrations of *second* order, the corresponding pure aperture term is the most important. This is described by the wave aberration term

$$S_3 = A_{3,0}x_o'^3 + A_{2,1}x_o'^2 y_o' + A_{1,2}x_o' y_o'^2 + A_{0,3}y_o'^3 \tag{31.14}$$

with the corresponding lateral aberrations

$$\Delta x_o = 3A_{3,0}x_o'^2 + 2A_{2,1}x_o' y_o' + A_{1,2}y_o'^2 \tag{31.15a}$$

$$\Delta y_o = A_{2,1}x_o'^2 + 2A_{1,2}x_o' y_o' + 3A_{0,3}y_o'^2 \tag{31.15b}$$

This aberration consists of two components with different azimuthal symmetry, as can be seen by writing Eq. (31.14) in polar coordinates,

$$x_o' = r_o'\cos\theta, \quad y_o' = r_o'\sin\theta$$

Expressing all powers of x_o' and y_o' in terms of trigonometric functions of multiple arguments, we find $S_3 = S_3^{(1)} + S_3^{(3)}$ with

$$S_3^{(1)} = \frac{1}{4}r_o'^3\left\{\left(3A_{3,0} + A_{1,2}\right)\cos\theta + \left(A_{2,1} + 3A_{0,3}\right)\sin\theta\right\} \tag{31.16a}$$

$$S_3^{(3)} = \frac{1}{4}r_o'^3\left\{\left(A_{3,0} - A_{1,2}\right)\cos 3\theta + \left(A_{2,1} - A_{0,3}\right)\sin 3\theta\right\} \tag{31.16b}$$

The different multiplicities with respect to the azimuth θ show that the aberrations that occur in Eq. (31.16a) are caused by weak *deflection* fields, produced by misalignments, while the aberrations corresponding to Eq. (31.16b) are caused by *threefold* deviations from the rotational symmetry. The latter are produced by corresponding threefold corrugations of the electrodes or polepieces. We shall call these two different aberrations the deflection astigmatism and the threefold astigmatism, respectively. They are hardly ever seen in their pure forms, since they are practically always superimposed on the numerous other lens aberrations.

We return to this topic in Section 31.6.

31.3 Numerical Determination of Parasitic Aberrations

31.3.1 Introduction

Determination of tolerance limits for the machining and alignment of electron optical systems requires calculation of the parasitic aberrations. Apart from a few oversimplified cases, this can

only be accomplished numerically. The main problem is the computation of the electromagnetic fields in the imperfect system. A rigorous calculation would proceed as follows:

1. First, the fields in the perfect system are calculated by means of the methods outlined in Part II.
2. From the solution of the appropriate boundary-value problem, the boundary values of the field strengths $\boldsymbol{E}(\boldsymbol{r})$ and $\boldsymbol{H}(\boldsymbol{r})$ are determined. This is often a very tedious task, since the numerical differentiation of potentials in the vicinity of their surface sources is rarely straightforward.
3. It is now necessary to adopt realistic values of the shifts, tilts, ellipticities and any other deformations and to determine from these a surface deformation $\boldsymbol{s}(\boldsymbol{r})$, defined as the shift from a point $\boldsymbol{r}$ at the ideal surface to the point $\boldsymbol{r}' = \boldsymbol{r} + \boldsymbol{s}$ at the real surface. With the necessary assumption that $|\boldsymbol{s}|$ is very small, such a deformation $\boldsymbol{s}$ corresponds to an alteration of the potential on the *ideal* boundary.

$$\delta\Phi = \boldsymbol{E}\cdot\boldsymbol{s} \quad \text{or} \quad \delta\chi = \boldsymbol{H}\cdot\boldsymbol{s} \tag{31.17}$$

4. With the boundary values $\delta\Phi(r)$ or $\delta\chi(\boldsymbol{r})$, the boundary-value problem in the *undeformed* domain is solved and the resulting solution is added to the unperturbed one. This procedure is usually much easier than a direct field calculation in the deformed domain, since the explicit treatment of irregularly deformed surfaces is extremely tedious.
5. Several electron trajectories with appropriate initial conditions are then traced through the total field. From the sets of final coordinates and slopes in a terminal (image) plane $z = z_i$, the aberration figures and − if needed − the aberration coefficients are determined. The corresponding methods are outlined in Part V. Very often the least-squares-fit methods explained there are much simpler than the evaluation of aberration integrals.

This whole procedure requires a major effort and is therefore often much simplified. The greatest simplification is obtained if the axial harmonics, the axial amplitudes of the various multipole components in the field, are determined not from the solution of a boundary-value problem but from simple analytic models. An approximate field calculation is then possible by evaluation of the corresponding radial series expansions given in Chapter 7, Series Expansions. This is certainly advisable whenever some parts of the boundaries are not round even in the ideal case, as in electric deflectors or in quadrupole systems.

For systems with circular boundaries, such as those composed of ordinary lenses and magnetic deflectors with a round ferrite core, the more rigorous method is not too complicated. The Fourier-series method, outlined in Chapter 7, Series Expansions, can then be applied to the potential variations $\delta\Phi(r)$ and $\delta\chi(r)$, and there remains the task of repeatedly solving the boundary-value problems for differential equations having the structure of (7.10). Calculations of this kind were first carried out, with a view to determining tolerance limits for the shift $\boldsymbol{s}(\boldsymbol{r})$, by Janse (1971), who demonstrated their

practical applicability. A numerical study starting from Sturrock's theory, has been made by Munro (1988), using the finite-difference method. We recapitulate this briefly and then summarize recent work using other methods.

31.3.2 Use of the Finite-Difference Method

A set of finite-difference programs for computing the parasitic aberrations of electrostatic lenses was developed by Munro (1988). These predict the beam shift and paraxial coma caused by misalignment and tilt and the astigmatism associated with ellipticity. For misalignment of an electrode by a distance $\boldsymbol{s}$, which for convenience we assume to be in the x-direction, $\boldsymbol{s} = (s_x, 0, 0)$, we have

$$\delta\Phi = -\nabla\Phi\cdot\boldsymbol{s} = -\frac{\partial\Phi}{\partial r}\cos\theta\cdot s_x$$

For tilt through a small angle $\delta\alpha$ in the x–z plane about a pivot point $(0, 0, \bar{z})$, we find

$$\delta\Phi = \left\{ r\frac{\partial\Phi}{\partial z} - (z-\bar{z})\frac{\partial\Phi}{\partial r} \right\}\cos\theta\cdot\delta\alpha$$

When an electrode is elliptical instead of circular,

$$\delta\Phi = -\varepsilon\frac{\partial\Phi}{\partial r}\cos 2\theta$$

Laplace's equation is then solved for $\delta\Phi$ and, since $\delta\Phi$ is proportional to cos θ (tilt and misalignment) or cos2θ (ellipticity), the equation involves only r and z. In fact, it is still simpler to write

$$\delta\Phi = F_1(r,z)r\cos\theta$$

and F_1 then satisfies

$$\frac{\partial^2 F_1}{\partial r^2} + \frac{3}{r}\frac{\partial F_1}{\partial r} + \frac{\partial^2 F_1}{\partial z^2} = 0$$

and on the electrode,

$$F_1 = \frac{1}{r}\frac{\partial\Phi}{\partial r}s_x$$

On the axis, $\partial F_1/\partial r = 0$. The function F_1 is then obtained in the usual way (Chapter 11, The Finite-Difference Method (FDM)). For tilt, F_1 satisfies the boundary condition

$$F_1 = \left(\frac{\partial\Phi}{\partial z} - \frac{z-\bar{z}}{r}\frac{\partial\Phi}{\partial r}\right)\delta\alpha$$

For ellipticity, a term in r^2 is introduced:

$$\delta\Phi(r,\theta,z) = F_2(r,z)r^2\cos 2\theta$$

and F_2 satisfies

$$\frac{\partial^2 F_2}{\partial r^2} + \frac{5}{r}\frac{\partial F_2}{\partial r} + \frac{\partial^2 F_2}{\partial z^2} = 0$$

On the perturbed electrode, on the axis and on the outer boundary $F_2 = -(\varepsilon/r^2)\partial\Phi/\partial r$ (on the unperturbed electrode, $F_2 = 0$). On the axis, these functions are readily identified with $F_1(z)$ and $p_2(z)$ (7.36). Munro lists integrals for the resulting aberration coefficients, not reproduced here. This work can be generalized to misalignment and tilts in any direction and to magnetic lenses.

31.3.3 Use of the Finite-Element Method

These parasitic aberrations are calculated with the aid of the finite-element method by Zlámal and Lencová (2011); the corresponding program is included in their program suite EOD, described in Chapter 34, Numerical Calculation of Trajectories, Paraxial Properties and Aberrations. Their results agree well with those of Munro (1988) and they are able to calculate such aberrations as third-order field curvature and astigmatism that are also liable to occur. With their approach, saturation in magnetic lenses can also be included.

31.4 The Isoplanatic Approximation

The isoplanatic approximation is invoked whenever we study the aberrations that affect the imaging of very small objects. This situation arises in practice in electron microscopes operating at very high magnification, say $|M| > 10^5$, where only a very small part of the object is visible on the viewing screen. In a perfectly round system, the best image quality would be obtained for an object in the vicinity of the straight optic axis, but since no unique axis exists in a real and imperfect instrument, the question of the best alignment is not at all trivial.

We confine the following discussion to the parasitic aberrations of first order, together with the intrinsic aberrations of a round system. We express all lateral aberrations in complex form in terms of a complex object coordinate $u_o \coloneqq x_o + \mathrm{i}y_o$ and a complex aperture coordinate $u_A \coloneqq x_A + \mathrm{i}y_A$. The resulting lateral image aberration Δu_i can then be written

$$\begin{aligned}\Delta u_i = {} & \sigma + Pu_o + P'u_o^* + Qu_A + Q'u_A^* \\ & + Cu_A^2u_A^* + 2(K+\mathrm{i}k)u_ou_Au_A^* + (K-\mathrm{i}k)u_o^*u_A^2 \\ & + (A+\mathrm{i}a)u_o^2u_A^* + Fu_ou_o^*u_A + (D+\mathrm{i}d)u_o^2u_o^*\end{aligned} \tag{31.18}$$

where the aberrations of third order are those dealt with comprehensively in Chapter 24, The Geometrical Aberrations of Round Lenses. The constant σ is a small shift due to weak parasitic deflection fields. The coefficient P is the deviation of the magnification from its nominal value arising from a defocus, an error in the image rotation, and to lateral chromatic aberrations

$$P = \delta M + \mathrm{i}M\delta\theta + (C_M + \mathrm{i}C_\theta)\frac{\Delta\hat{\phi}_i}{\hat{\phi}_i} \tag{31.19}$$

The term $P'u_o^*$ represents the paraxial distortion. The factor Q expresses the effect of the defocus Δ and the axial chromatic aberration and is always real:

$$Q = \Delta + \frac{\Delta\hat{\phi}_i}{\hat{\phi}_i}C_c \tag{31.20}$$

Finally the term $Q'u_A^*$ represents the axial astigmatism caused by ellipticities.

It is a crucial feature of imperfect systems that there is no way of determining whether the aperture is aligned perfectly relative to the optic axis because there is no unique axis. It hence makes sense to assume that the centre of the aperture is *shifted laterally* relative to the z-axis of the coordinate system, $w^{(0)} \coloneqq x_A^{(0)} + \mathrm{i}y_A^{(0)}$, and to express the aberrations in terms of

$$w_A \coloneqq u_A - w^{(0)} = x_A - x_A^{(0)} + \mathrm{i}\left(y_A - y_A^{(0)}\right) \tag{31.21}$$

We then notice that there is one particular shift

$$w^{(0)} = -(K + \mathrm{i}k)\frac{u_o}{C} \tag{31.22}$$

for which the coma vanishes. Experimentally, this value is always chosen approximately, since one always tries to get the centre of the image as sharp as possible by an adjustment of the aperture. With Eqs. (31.21) and (31.22) we now establish a series expansion in terms of u_o and w_A instead of Eq. (31.17), the new coefficients being

$$\begin{aligned}
&C_0 = C, \quad K_o + \mathrm{i}k_o = 0\\
&A_0 + \mathrm{i}a_0 = A + \mathrm{i}a - (K+\mathrm{i}k)^2/C\\
&F_0 = F - 2\left(K^2 - k^2\right)/C\\
&D_0 + \mathrm{i}d_0 = D + \mathrm{i}d - F(K + \mathrm{i}k)/C + (A + \mathrm{i}a)(K - \mathrm{i}k)/C\\
&\qquad + 2(K+\mathrm{i}k)^2(K - \mathrm{i}k)/C^2\\
&S_0 = S, \quad Q_0 = Q, \quad Q_0' = Q'\\
&P_0 = P - (K + \mathrm{i}k)Q/C\\
&P_0' = P' - (K - \mathrm{i}k)Q'/C
\end{aligned} \tag{31.23}$$

In the isoplanatic approximation the object field in question is assumed to be so narrow that the variable u_o can be replaced by the complex coordinate of the *centre* of the observed field. This replacement is to be made in all the aberration terms. We then arrive at the representation:

$$\Delta u_i = s + q w_A + q' w_A^* + C w_A^2 w_A^* \tag{31.24}$$

with the coefficients

$$\begin{aligned} s &= S_0 + P_0 u_o + P_0' u_o^* + (D_0 + \mathrm{i}d_0) u_o^2 u_o^* \\ q &= Q_o + F_0 u_o u_o^*, \quad q' = Q_0' + (A_0 + \mathrm{i}a_0) u_o^2 \end{aligned} \tag{31.25}$$

The resulting lateral shift s is of very little practical importance, since the microscope is always adjusted in such a way that the centre of the object field of interest is imaged in the centre of the screen; this is achieved by means of weak deflection fields. We can thus ignore the shifts. The remaining aberrations are now the defocus, the axial chromatic aberration, (both combined in the term qw_A), the astigmatism $q'w_A^*$ and the spherical aberration $Cw_A^2 w_A^*$ The astigmatism, here the combined effect of the axial and off-axis contributions, can be compensated with a stigmator, as shown below; the resolution of the instrument is thus finally determined by the spherical and the axial chromatic aberration.

Elimination of the coma by a shift of the aperture is important; it is shown schematically in Fig. 31.2. It can be understood as a straightening of the beam axis passing simultaneously through the centres of the object, the lens, the aperture and the image. Apart from the axial aberrations of perfectly round lenses there remains only an *astigmatism*, generated by the ellipticity and the inclination of the beam axis relative to the lens.

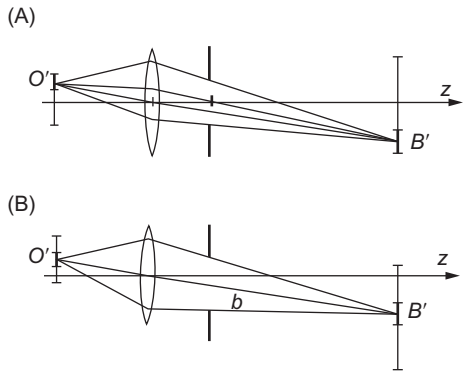


Figure 31.2
Lateral shift of an aperture. (A) Aperture centred on the lens axis leading to asymmetric confinement of the beam with coma. (B) Aperture shifted in such a way that the beam is confined symmetrically; the beam axis b is now straight and the coma vanishes.

31.5 Stigmators

Stigmators are systems of weak electric or magnetic multipole fields, suitably located within an electron optical device, the role of which is to provide minor correction of the beam. Such corrections are necessary to eliminate some of the aberrations caused by lenses and deflectors and, still more important, those caused by misalignments and similar imperfections. The corrections may be both static and dynamic. Static correction of the geometric imperfections is always necessary; in this case the electrode voltages or coil currents are constant. Dynamic correction is mainly employed in electron lithography devices; the correction currents are then functions of the time-dependent deflection currents.

31.5.1 Necessary Simplifications

In this section we shall outline the operating principles of a stigmator. Since an exact treatment requires extensive numerical calculations, we make a number of simplifying assumptions that permit analytic relations to be found. These assumptions are approximately valid in practice. First of all, we assume that the necessary correction fields are *weak* and not superimposed on other strong fields. This means that the stigmator has to be located outside the lens or deflector field region. In the lowest order approximation, the trajectories are then straight lines within the stigmator, and a first-order perturbation yields the small deviations from these straight lines.

The second assumption is illustrated in Fig. 31.3. The stigmator fields must be essentially confined between two planes $z = -a$ and $z = +a$ and symmetric with respect to the midplane $z = 0$. For asymmetric fields, the theory will become more complicated but the conclusions made below will be essentially the same. Moreover, there is no need to produce asymmetric correction fields, which would complicate the construction with no real gain; the limitation to symmetric fields is thus not a serious restriction. This assumption implies that the functions Φ, E_x, E_y, A_z, B_x, B_y are positively symmetric in z, while E_z, A_x, A_y, B_z are

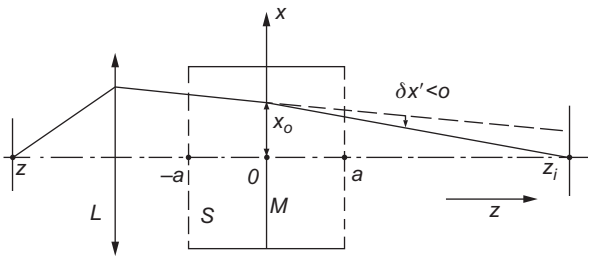


Figure 31.3
Action of a stigmator. *L*: lens, *S*: stigmator, *M*: midplane. The stigmator field is confined to the domain $-a \leq z \leq a$.

antisymmetric. These symmetry requirements can still be satisfied by certain round lens fields, but we shall exclude these, since it is not the purpose of a stigmator to alter the focusing. This is better done by alteration of the lens excitations.

Finally we assume that $x'^2 + y'^2 \ll 1$ for the slopes of the trajectories. Since some correction will result even for $x' = 0$ or $y' = 0$, it is permissible to neglect all terms of second and higher order in x' and y'. The latter would already represent aberrations caused by the stigmator itself. Since the stigmator is assumed to be weak, such aberrations will be extremely small and can hence be ignored. With all these simplifying assumptions, the theory of the stigmator can be expressed in closed form.

31.5.2 The Wave Aberration

We start from Eqs. (4.34) and (4.35), and with $x'^2 + y'^2 \ll 1$, $Q = -e$, $g = \sqrt{2m_0 e\hat{\Phi}}$, $z_o = -a$, $z_1 = a$ we obtain

$$S \approx \int_{-a}^{a} \sqrt{2m_0 e}\left\{\hat{\Phi}^{1/2}(\boldsymbol{r}) - \eta\left(x'A_x + y'A_y + A_z\right)\right\}dz$$

The electric potential has the form

$$\Phi(\boldsymbol{r}) = \phi + \Phi_s(\boldsymbol{r}), \quad |\Phi_s| \ll \phi$$

ϕ being the accelerating voltage and Φ_s the very small stigmator potential.
The approximation

$$\sqrt{\hat{\Phi}} = \sqrt{\hat{\phi}} + \frac{\gamma}{2\hat{\phi}^{1/2}}\Phi_s$$

is then justified. We now drop the constant term in S and normalize with respect to the factor $(2m_0 e\hat{\phi})^{1/2}$; this yields a convenient form of the characteristic function, conventionally called the *wave aberration* even in geometrical optics:

$$W \coloneqq 2a - \frac{S}{\left(2m_0 e\hat{\phi}\right)^{1/2}} = \int_{-a}^{a}\left\{-\frac{\gamma}{2\hat{\phi}}\Phi_s(\boldsymbol{r}) + \frac{\eta}{\hat{\phi}^{1/2}}\left(x'A_x + y'A_y + A_z\right)\right\}dz \tag{31.26}$$

This integral is to be evaluated for an arbitrary *straight* trajectory with small slope. The corresponding Cartesian representation is

$$x(z) = x_0 + zx', \quad y(z) = y_0 + zy' \tag{31.27}$$

x_0 and y_0 being the coordinates of the point of intersection with the symmetry plane. Introducing Eq. (31.27) into (31.26) and recalling the assumptions made above, we find that the terms in A_x and A_y cancel out. In order to evaluate the remaining terms, we introduce the Taylor series expansions of Φ_s and A_z in x' and y' and retain only linear terms:

$$\Phi_s(\boldsymbol{r}) = \Phi_s(z, x_0, y_0) + z\left(x'\Phi_{s|x} + y'\Phi_{s|y}\right)_{(z,x_0,y_0)}$$
$$A_z(\boldsymbol{r}) = A_z(z, x_0, y_0) + z\left(x'A_{z|x} + y'A_{z|y}\right)_{(z,x_0,y_0)}$$

Introducing this into Eq. (31.26) and bearing in mind the symmetry properties, we see that the terms in x' and y' again cancel out. We hence obtain a very simple formula for W:

$$W(x_0, y_0) = \int_{-a}^{a} \left(-\frac{\gamma \Phi_s(z, x_0, y_0)}{2\hat{\phi}} + \frac{\eta}{\hat{\phi}^{1/2}} A_z(z, x_0, y_0) \right) dz \tag{31.28}$$

The corrections introduced by the stigmator are therefore determined by the coordinates (x_0, y_0) of the trajectory in the midplane.

An interesting rule is obtained by forming the two-dimensional Laplacian of W. Differentiation under the integral results in

$$\nabla_2^2 W(x_0, y_0) \equiv \frac{\partial^2 W}{\partial x_0^2} + \frac{\partial^2 W}{\partial y_0^2}$$
$$= -\frac{\gamma}{2\hat{\phi}} \int_{-a}^{a} \left(\frac{\partial^2 \Phi_s}{\partial x_0^2} + \frac{\partial^2 \Phi_s}{\partial y_0^2} \right) dz + \frac{\eta}{\hat{\phi}^{1/2}} \int_{-a}^{a} \left(\frac{\partial^2 A_z}{\partial x_0^2} + \frac{\partial^2 A_z}{\partial y_0^2} \right) dz$$

Now $\Phi_s(\boldsymbol{r})$ and $A_z(\boldsymbol{r})$ must satisfy the Laplace equation, since the space charge of the beam can be neglected in the vast majority of practical applications and other sources are always located outside the beam. For Φ_s,

$$\int_{-a}^{a} \left(\frac{\partial^2 \Phi_s}{\partial x_0^2} + \frac{\partial^2 \Phi_s}{\partial y_0^2} \right) dz = -\int_{-a}^{a} \frac{\partial^2 \Phi}{\partial z^2} dz = \left[-\Phi_{s|z}(z, x_0, y_0) \right]_{-a}^{a} = 0$$

since at and beyond the boundaries $z = \pm a$, all the corrector functions considered vanish in view of our assumptions about the confinement. A similar argument shows that the magnetic term also vanishes and hence

$$\nabla_2^2 W \equiv \frac{\partial^2 W}{\partial x_0^2} + \frac{\partial^2 W}{\partial y_0^2} = 0 \tag{31.29}$$

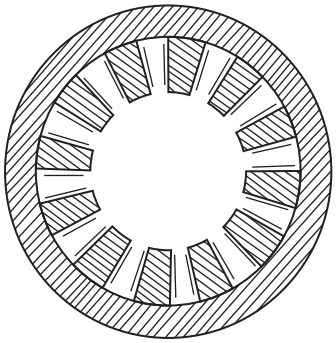


Figure 31.4
Cross-section through a magnetic 12-pole stigmator.

The operation of a stigmator can now be understood in the following way. In the midplane, the system of lenses and deflectors located upstream produces a wave aberration $W_L(x_0, y_0)$. The stigmator adds the contribution $W(x_0, y_0)$ so that for the part of the beam downstream the wave aberration is $W_c = W_L + W$. The average value of $|W_c|$ over the cross-section of the beam at $z = 0$ should be made as small as possible. It is obvious that only geometric aberrations whose wave aberrations already satisfy $\nabla_2^2 W_L = 0$ can be eliminated completely. These are paraxial deflections and *astigmatisms* of various multiplicity. Their wave aberrations are best given in polar form

$$W_L(r_0, \varphi_0) = \sum_{n=1}^{N} r_0^n \left(a_n \cos n\varphi_0 + b_n \sin n\varphi_0\right) \tag{31.30}$$

$4N$ being the number of electrodes or poles of the stigmator. Fig. 31.4 shows a magnetic 12-pole stigmator which is suitable for the correction of astigmatisms up to $N = 3$. It is, however, never possible to correct spherical aberration, since the latter produces a wave aberration.

$$W_L = \text{const}\left(x_0^2 + y_0^2\right)^2 = \text{const} \times r_0^4, \quad \nabla_2^2 W_L \neq 0$$

31.5.3 The Deflection of Trajectories

The influence of a stigmator on the electron beam can also be understood in another useful manner. The trajectories are weakly bent by the stigmator fields, the alteration of direction being simply given by

$$\Delta x' = -\frac{\partial W}{\partial x_0}, \quad \Delta y' = -\frac{\partial W}{\partial y_0} \tag{31.31}$$

Carrying out this differentiation in the integrand of Eq. (31.28) and recalling that $\boldsymbol{E} = -\nabla\Phi$, we obtain

$$\Delta x'_j = -\frac{\gamma}{2\hat{\phi}}\int_{-a}^{a} E_j(z, x_0, y_0)\, dz - \frac{\eta}{\hat{\phi}^{1/2}}\int_{-a}^{a} A_{z|j}\, dz, \quad j = 1, 2$$

In the second term we substitute $A_{z|x} = A_{x|z} - B_y$ and $A_{z|y} = A_{y|z} + B_x$. The derivatives with respect to z do not contribute to the integral, since $\int_{-a}^{a} A_{x|z} dz = [A_x]_{-a}^{a} = 0$, for example. Finally, then, we obtain

$$\begin{aligned}
\Delta x' &= -\frac{\gamma}{2\hat{\phi}}\int_{-a}^{a} E_x(z, x_0, y_0)\, dz + \frac{\eta}{\hat{\phi}^{1/2}}\int_{-a}^{a} B_y(z, x_0, y_0)\, dz \\
\Delta y' &= -\frac{\gamma}{2\hat{\phi}}\int_{-a}^{a} E_y(z, x_0, y_0)\, dz - \frac{\eta}{\hat{\phi}^{1/2}}\int_{-a}^{a} B_x(z, x_0, y_0)\, dz
\end{aligned} \tag{31.32}$$

This is a Busch approximation of first order, since the true trajectory over which the integration should strictly be made is replaced by a simple straight line parallel to the optic axis. Apart from correction terms of *second* order in x' and y', this gives essentially the same results as integration over the straight line given by Eq. (31.27).

The stigmator can now be considered as an optically thin element, which bends the trajectories sharply in the midplane, the change of gradient being given by Eq. (31.32). This result is very useful for practical applications.

A very detailed account of the practical design of stigmators is given by Riecke (1982). For an advanced design, see Bärtle and Plies (2000). The early work of Kanaya is also very relevant (Kanaya and Kawakatsu, 1961; Kawakatsu and Kanaya, 1961).

31.6 Advanced Theory

31.6.1 Introduction

It is only with the advent of aberration correctors that any but the lowest order of the parasitic aberrations, the (first order) axial astigmatism, have been studied. In the 1990s, Krivanek drew attention to the need to eliminate any (second order) threefold astigmatism (Krivanek and Leber, 1993; Krivanek, 1994; Krivanek and Stadelmann, 1995; see also Ishizuka, 1994; Saxton, 1995; Overwijk et al. 1996, 1997; Stenkamp, 1998) and subsequently the whole family of parasitic aberrations has received attention (Batson, 2008, 2009). Since these aberrations become harmful only when the highest resolution is required, attention is usually confined to the axial aberrations, those that are independent of the position coordinates in the object plane. For devices with a curved axis, such as

monochromators, off-axis parasitic aberrations may also need to be analysed. We shall examine the various parasitic aberrations of increasing order and indicate their origin. First however, we summarize the principal notations in use. Throughout the rest of this book, we have denoted spherical aberrations by C (with a suffix s or 3 or 5), comas by K, astigmatisms by A, field curvature by F and distortion by D. This convention is not adopted by Haider, Krivanek and Sawada and in the remainder of this chapter we have therefore preferred to use their notations.

31.6.2 Notation

In Chapter 41 (Volume 2) on aberration correction, we shall meet many more aberrations than the third and higher order geometrical aberrations and the second rank chromatic aberrations studied in Part IV. In particular, the large family of parasitic aberrations have to be cancelled or rendered harmless in systems designed for the highest resolving power. Three notations for all these aberrations are in regular use. The first, introduced by Uhlemann and Haider (1998), extended in Haider et al. (2008) and inspired by earlier publications of Plies and Rose (1971) and Pöhner and Rose (1974), employs subscripted letters to identify astigmatisms, comas, 'spherical' aberrations and other families of aberrations. Another, used by Krivanek and colleagues and fully explained in Krivanek et al. (2009), employs multiple subscripts to characterize order, multiplicity and any further distinctions. Krivanek argues that his notation makes it easier to understand the interrelations between aberrations of different order. A third notation is used by Sawada (Sawada et al., 2008, 2011; Sawada, 2015). This is not very different from that of Uhlemann and Haider but the choice of letters is not the same and the numerical subscripts are one greater, thus corresponding to the path difference (S' in 22.21 and the end of Section 24.2) or wave aberration (Volume 3, Eq. 65.30) and not to the order of the resulting aberrations.

Recent publications on the classification of aberrations usually set out from the wave aberration, which is based on the phase shift χ,

$$\chi \coloneqq \frac{2\pi}{\lambda} W \tag{31.33}$$

Here, W is a measure of the distance between the reference sphere S_0 and the spherical surface to which electron trajectories are normal when aberrations are considered. In (24.3) this distance was written S'/W, where W denoted the Wronskian of the paraxial solutions. Now, however, for consistency with Chapter 65 and much published work, we shall denote the distance between the reference sphere and the true aspheric surface by W, as in (65.30 in Volume 3) and Eq. (31.33) above. Thus

$$W \text{ (this chapter)} \leftrightarrow S'/W \text{ (Chapter 24)} \tag{31.34}$$

Table 31.1: Notations employed for axial aberration coefficients.

Aberration	Haider	Krivanek	Sawada
Shift	A_0	$C_{0,1}$	
Defocus	C_1	$C_{1,0}$	O_2
Twofold astigmatism	A_1	$C_{1,2}$	A_2
Second-order axial coma	B_2	$(1/3)C_{2,1}$	P_3
Threefold astigmatism	A_2	$C_{2,3}$	A_3
Third-order spherical aberration	C_3	$C_{3,0}$	O_4
Third-order star aberration	S_3	$(1/4)C_{3,2}$	Q_4
Fourfold astigmatism	A_3	$C_{3,4}$	A_4
Fourth-order axial coma	B_4	$(1/4)C_{4,1}$	P_5
Fourth-order 3-lobe aberration	D_4	$(1/4)\ C_{4,3}$	R_5
Fivefold astigmatism	A_4	$C_{4,5}$	A_5
Fifth-order spherical aberration	C_5	$C_{5,0}$	O_6
Fifth-order star aberration	S_5	$(1/6)\ C_{5,2}$	Q_6
Fifth-order rosette aberration	R_5	$(1/6)\ C_{5,4}$	
Sixfold astigmatism	A_5	$C_{5,6}$	A_6
Sixth-order axial coma	B_6	$(1/7)\ C_{6,1}$	P_7
Sixth-order 3-lobe aberration	D_6	$(1/7)\ C_{6,3}$	R_7
Sixth-order pentacle aberration	F_6	$(1/7)\ C_{6,5}$	
Sevenfold astigmatism	A_6	$C_{6,7}$	A_7
Seventh-order spherical aberration	C_7	$C_{7,0}$	O_8
Seventh-order star aberration	S_7	$(1/8)C_{7,2}$	Q_8
Seventh-order rosette aberration	R_7	$(1/8)C_{7,4}$	
Seventh-order chaplet aberration	G_7	$(1/8)C_{7,6}$	
Eightfold astigmatism	A_7	$C_{7,8}$	A_8

An expression for the terms of low order in the isoplanatic approximation (terms independent of off-axial distance in the object plane) is given in Chapter 65 (65.29) but, with the advent of aberration correction, the notation used there has become too restrictive. The notations adopted by Uhlemann and Haider, Krivanek et al. and Sawada or the four-digit notation of Chapter 23 are better suited to this situation. Table 31.1 shows the relations between the various symbols employed, which can be understood from the following expressions for W. In these, u_o denotes $x_o + \mathrm{i}y_o$ and θ denotes $x'_o + iy'_o := \alpha + \mathrm{i}\beta$.

Uhlemann and Haider.

$$W = \Re\left\{A_0\theta^* + \frac{1}{2}C_1\theta\theta^* + \frac{1}{2}A_1\theta^{*2} + B_2\theta^2\theta^* + \frac{1}{3}A_2\theta^{*3} + \frac{1}{4}C_3(\theta\theta^*)^2 + S_3\theta^3\theta^* + \frac{1}{4}A_3\theta^{*4}\right.$$

$$
\begin{aligned}
&+B_4\theta^3\theta^{*2}+D_4\theta^4\theta^*+\frac{1}{5}A_4\theta^{*5}\\
&+\frac{1}{6}C_5(\theta\theta^*)^3+S_5\theta^4\theta^{*2}+R_5\theta^5\theta^*+\frac{1}{6}A_5\theta^{*6}\\
&+B_6\theta^4\theta^{*3}+D_6\theta^5\theta^{*2}+F_6\theta^6\theta^*+\frac{1}{7}A_6\theta^{*7}\\
&+\frac{1}{8}C_7(\theta\theta^*)^4+S_7\theta^5\theta^{*3}+R_7\theta^6\theta^{*2}+G_7\theta^7\theta^*+\frac{1}{8}A_7\theta^{*8}\Big\}
\end{aligned}
\tag{31.35a}
$$

or

$$
\begin{aligned}
W=&A_0\alpha+a_0\beta+\frac{1}{2}C_1\left(\alpha^2+\beta^2\right)+\frac{1}{2}A_1\left(\alpha^2-\beta^2\right)+a_1\alpha\beta\\
&+B_2\alpha\left(\alpha^2+\beta^2\right)-b_2\beta\left(\alpha^2+\beta^2\right)+\frac{1}{2}A_2\alpha\left(\alpha^2-3\beta^2\right)+\frac{1}{3}a_2\beta\left(3\alpha^2-\beta^2\right)\\
&+\frac{1}{4}C_3\left(\alpha^2+\beta^2\right)^2+S_3\left(\alpha^4-\beta^4\right)-2s_3\alpha\beta\left(\alpha^2+\beta^2\right)+\frac{1}{4}A_3\left(\alpha^4-6\alpha^2\beta^2+\beta^4\right)+a_3\alpha\beta\left(\alpha^2-\beta^2\right)\\
&+B_4\alpha\left(\alpha^2+\beta^2\right)^2-b_4\beta\left(\alpha^2+\beta^2\right)^2+D_4\alpha\left(\alpha^2+\beta^2\right)\left(\alpha^2-3\beta^2\right)-d_4\beta\left(\alpha^2+\beta^2\right)\left(3\alpha^2-\beta^2\right)\\
&+\frac{1}{5}A_4\left(\alpha^5-10\alpha^3\beta+5\alpha\beta^4\right)+\frac{1}{5}a_4\left(5\alpha^4\beta-10\alpha^2\beta^3+\beta^5\right)\\
&+\frac{1}{6}C_5\left(\alpha^2+\beta^2\right)^3+S_5\left(\alpha^2+\beta^2\right)^2\left(\alpha^2-\beta^2\right)-2s_5\alpha\beta\left(\alpha^2+\beta^2\right)^2+R_5\left(\alpha^2+\beta^2\right)\left(\alpha^4-6\alpha^2\beta^2+\beta^4\right)\\
&-4r_5\alpha\beta\left(\alpha^4-\beta^4\right)+\frac{1}{6}A_5\left(\alpha^6-15\alpha^4\beta^2+15\alpha^2\beta^4-\beta^6\right)+\frac{1}{3}a_5\left(3\alpha^5\beta-10\alpha^3\beta^3+3\alpha\beta^5\right)\\
&+B_6\alpha\left(\alpha^2+\beta^2\right)^3-b_6\beta\left(\alpha^2+\beta^2\right)^3+D_6\alpha\left(\alpha^2+\beta^2\right)^2\left(\alpha^2-3\beta^2\right)-d_6\beta\left(\alpha^2+\beta^2\right)^2\left(3\alpha^2-\beta^2\right)\\
&+F_6\left(\alpha^2+\beta^2\right)\left(\alpha^5-10\alpha^3\beta^2+5\alpha\beta^4\right)-f_6\left(\alpha^2+\beta^2\right)\left(5\alpha^4\beta-10\alpha^2\beta^3+\beta^5\right)\\
&+\frac{1}{7}A_6\left(\alpha^7-21\alpha^5\beta^2+35\alpha^3\beta^4+7\alpha\beta^6\right)+\frac{1}{7}a_6\left(7\alpha^6\beta-35\alpha^4\beta^3+21\alpha^2\beta^5-\beta^7\right)\\
&+\frac{1}{8}C_7\left(\alpha^2+\beta^2\right)^4+S_7\left(\alpha^2+\beta^2\right)^3\left(\alpha^2-\beta^2\right)-2s_7\alpha\beta\left(\alpha^2+\beta^2\right)^3\\
&+R_7\left(\alpha^2+\beta^2\right)^2\left(\alpha^4-6\alpha^2\beta^2+\beta^4\right)-4r_7\alpha\beta\left(\alpha^2-\beta^2\right)\\
&+G_7\left(\alpha^2+\beta^2\right)\left(\alpha^6-15\alpha^4\beta^2+15\alpha^2\beta^4-\beta^6\right)-2g_7\left(\alpha^2+\beta^2\right)\left(3\alpha^5\beta-10\alpha^3\beta^2+3\alpha\beta^5\right)\\
&+\frac{1}{8}A_7\left(\alpha^8-28\alpha^6\beta^2+70\alpha^4\beta^4-28\alpha^2\beta^6+\beta^8\right)+a_7\alpha\beta\left(\alpha^6-7\alpha^4\beta^2+7\alpha^2\beta^4-\beta^6\right)
\end{aligned}
\tag{31.35b}
$$

Krivanek et al. The corresponding expansions in the notation of Krivanek are as follows:

$$
\begin{aligned}
W &= \frac{1}{n+1}\sum_n\sum_m \Re\{C_{n,m}\theta^{n+1}\exp(-\mathrm{i}m\varphi)\} \\
&= \frac{1}{n+1}\sum_n\sum_m \left(C_{n,m,\,a}\theta^{n+1}\cos m\varphi + C_{n,m,b}\theta^{n+1}\sin m\varphi\right) \\
&= \frac{1}{n+1}\sum_n\sum_m \left(\alpha^2+\beta^2\right)^{(n-m+1)/2}\Re\{C_{n,m,a}(\alpha-\mathrm{i}\beta)^m + \mathrm{i}C_{n,m,b}(\alpha-\mathrm{i}\beta)^m\} \\
&= C_{0,1,a}\alpha + C_{0.2.b}\beta + \frac{1}{2}C_{1,0}\left(\alpha^2+\beta^2\right) \\
&\quad + C_{1,2,a}\left(\alpha^2-\beta^2\right)/2 + C_{1,2,b}\alpha\beta \\
&\quad + C_{2,1,a}\alpha\left(\alpha^2+\beta^2\right)/3 + C_{2,1,b}\beta\left(\alpha^2+\beta^2\right)/3 \\
&\quad + C_{2,3,a}\alpha\left(\alpha^2-3\beta^2\right)/3 + C_{2,3,b}\beta\left(3\alpha^2-\beta^2\right) \\
&\quad + C_{3,0}\left(\alpha^2+\beta^2\right)^2/4 + C_{3,2,a}\left(\alpha^4-\beta^4\right)/4 + C_{3,2,b}\alpha\beta\left(\alpha^2+\beta^2\right)/2 \\
&\quad + C_{3,4,a}\left(\alpha^4-6\alpha^2\beta^2+\beta^4\right)/4 + C_{3,4,b}\alpha\beta\left(\alpha^2-\beta^2\right) \\
&\quad + \left(C_{4,1,a}\alpha + C_{4,1,b}\beta\right)\left(\alpha^2+\beta^2\right)^2/5 \\
&\quad + C_{4,3,a}\alpha\left(\alpha^4-2\alpha^2\beta^2-3\beta^4\right)/5 + C_{4,3,b}\beta\left(3\alpha^4+2\alpha^2\beta^2-\beta^4\right)/5 \\
&\quad + C_{4,5,a}\alpha\left(\alpha^4-10\alpha^2\beta^2+5\beta^4\right)/5 + C_{4,5,b}\beta\left(5\alpha^4-10\alpha^2\beta^2+\beta^4\right)/5 \\
&\quad + C_{5,0}\left(\alpha^2+\beta^2\right)^3/6 + C_{5,2,a}\left(\alpha^2+\beta^2\right)\left(\alpha^4-\beta^4\right)/6 + C_{5,2,b}\alpha\beta\left(\alpha^2+\beta^2\right)^2/3 \\
&\quad + C_{5,4,a}\left(\alpha^6-5\alpha^4\beta^2-5\alpha^2\beta^4+\beta^6\right)/6 + 2C_{5,4,b}\alpha\beta\left(\alpha^4-\beta^4\right)/3 \\
&\quad + C_{5,6,a}\left(\alpha^6-15\alpha^4\beta^2+15\alpha^2\beta^4-\beta^6\right)/6 + C_{5,6,b}\alpha\beta\left(3\alpha^4-10\alpha^2\beta^2+3\beta^4\right)/3
\end{aligned}
\tag{31.36}
$$

Sawada. The notation employed by Sawada is similar to that of Uhlemann and Haider but not identical with it. The starting point is again

$$
W = \Re\left(\sum_{m,\,n}\frac{1}{m+n}C_{mn}\theta^n\theta^{*m}\right) \tag{31.37}
$$

Aberrations that contain only θ^{*m} and not θ are named m-fold astigmatisms, A_2, A_3, A_4,..., generated by $\Re\left(\frac{1}{2}A_2\theta^{*2}\right)$, $\Re\left(\frac{1}{3}A_3\theta^{*3}\right)$, $\Re\left(\frac{1}{4}A_4\theta^{*4}\right)$, ... Aberrations in $(\theta\theta^*)^m$ are denoted by O_m: thus O_2 represents defocus, O_4 is third-order spherical aberration (otherwise C_s or C_3) and O_6 is fifth-order spherical aberration. Comas are labelled P_m: P_3 is axial coma, arising from $\Re\left(P_3\theta\theta^{*2}/3\right)$ and P_5 is fourth-order axial coma, $\Re\left(P_5\theta^2\theta^{*3}/5\right)$.

The aberrations denoted by Q_m correspond to twofold symmetry and arise from $\Re\left(Q_4\theta\theta^{*3}/4\right)$, $\Re\left(Q_6\theta^2\theta^{*4}\right), \ldots$. Terms in R_m correspond to threefold symmetry: $\Re\left(R_5\theta\theta^{*4}/5\right), \ldots$ and terms in S_m to fourfold symmetry: $\Re\left(S_6\theta\theta^{*5}/6\right), \ldots$

These long lists include only the 'axial' or isoplanatic aberrations, those that are independent of the off-axis coordinates in the object plane. Krivanek et al. mention that their coefficients could be extended to include the other aberrations by adding exponents:

$$W = \frac{1}{n+1}\sum_p\sum_q\sum_n\sum_m r^p\left(\alpha^2+\beta^2\right)^{(n+1)/2}\left\{C^{p,\,q,\,a}_{n,\,m,\,a}\cos q\psi\cos m\varphi + C^{p,\,q,\,b}_{n,\,m,\,a}\sin q\psi\cos m\varphi + C^{p,\,q,\,a}_{n,\,m,\,b}\cos q\psi\cos m\varphi + C^{p,\,q,\,b}_{n,\,m,\,b}\sin q\psi\cos m\varphi\right\} \tag{31.38}$$

Just as $\theta_x = \alpha = \cos\varphi$ and $\theta_y = \beta = \sin\varphi$, so $x_0 = r_o\cos\psi$ and $y_0 = r_o\sin\psi$. The label p indicates the dependence on the off-axis distance in the object plane; The 'multiplicity index' q is an even number between 0 and p when the latter is even and an odd number between 1 and p when the latter is odd. Thus field curvature, for example, corresponds to $C^{2,0}_{1,0}$ (quadratic dependence on x_o, y_0), astigmatism to $C^{2,0}_{1,2}$ and $C^{2,2}_{1,2}$; coma corresponds to $C^{1,1}_{2,1}$(linear dependence on x_o, y_0) and $C^{1,1,a}_{2,1,a} = C^{1,1,b}_{2,1,a}$, $C^{1,1,a}_{2,1,b} = C^{1,1,b}_{2,1,b} = 0$. This has not, however, been much used in practice. Erni (2015) suggests that these aberrations could be identified by adding a second subscript to the axial aberration coefficients of Uhlemann and Haider. The (third order) coma of round lenses would thus be characterized by B_{31}, astigmatism by A_{32}, field curvature by F_{32} and distortion by D_{33}; the second subscript indicates the dependence on off-axial distance in the object plane.

31.6.3 Further Analysis of the Aberrations, Classified by Order

Second-order aberrations. These are essentially comas (B_2 or $C_{2,1}$) and a threefold astigmatism (A_2 or $C_{2,3}$). A sextupole is capable of correcting the latter. The coma can be corrected either by introducing a weak deflection field (a dipole field) or by contriving that the paraxial beam is astigmatic when it passes through a sextupole.

Third-order aberrations. Two terms may occur here. S_3 or $C_{3,2}$ is known as (axial) star aberration and A_3 or $C_{3,4}$ as fourfold (axial) astigmatism. They need attention in systems that include a quadrupole−octopole corrector since they are the result of azimuthal misalignment of the many multipole elements present. C_3 or $C_{3,0}$ is the spherical aberration of round lenses.

Fourth-order aberrations. Three terms occur here. Coma (B_4 or $C_{4,1}$), a three-lobed aberration (D_4 or $C_{4.3}$) and fivefold (axial) astigmatism (A_4 or $C_{4,5}$).

Fifth-order aberrations. Again we have three terms: star (S_5 or $C_{5,2}$), rosette (R_5 or $C_{5,4}$) and an astigmatism (A_5 or $C_{5,4}$). C_5 or $C_{5,0}$ is the fifth-order spherical aberration of round lenses.

Sixth-order aberrations. There are now four terms: coma (B_6 or $C_{6,1}$), three-lobed aberration (D_6 or $C_{6,3}$), pentacle (F_6 or $C_{6,5}$) and astigmatism (A_6 or $C_{6,7}$).

Seventh-order aberrations: Star (S_7 or $C_{7,2}$), rosette (R_7 or $C_{7,4}$), chaplet (G_7 or $C_{7,6}$) and astigmatism (A_7 or $C_{7,8}$). C_7 or $C_{7,0}$ is the seventh-order spherical aberration of round lenses.

We refer to specialized texts for description of the procedures employed to restrict the adverse effects of all these aberrations. The subject is dealt with at length by Erni (2015, Chapter 8) and more briefly by Rose (2012). For more details, the publications of Krivanek and colleagues at Nion and Haider and co-workers at CEOS must be consulted (see the bibliography in Hawkes, 2015).

It is not always possible, or indeed advantageous, to cancel all the aberrations of a given system up to a particular order. It is frequently possible to gain some improvement by balancing terms of different order. The principle is easily understood when we recall that the aberration disc arising from third-order spherical aberration is smallest in the plane of least confusion situated between the Gaussian image plane and the marginal focus (24.50). If we think of defocus as a first-order aberration (C_1 or $C_{1,0}$), then we may say that we have improved the system by balancing a first-order aberration against a third-order one (C_3 or $C_{3,0}$). An example in which third- and fifth-order spherical aberrations of a probe-forming lens are balanced is described by Dellby et al. (2001).

General rules for these *combination aberrations* – aberration terms that arise when the fields of two elements interact – have been established (see Krivanek et al., 2009, recapitulated in Erni, 2015, Section 8.4.6). An aberration of element 1 is characterized by its order n_1 and its symmetry or multiplicity (defined below) N_1. If a second element characterized by n_2 and N_2 interacts with the first element, then combination aberrations of order $n_1 + n_2 - 1$ and symmetry $|N_2 - N_1|$ or symmetry $N_1 + N_2$ can occur. From this we see that two elements with third-order spherical aberration ($n_1 = n_2 = 3$, $N_1 = N_2 = 0$) can generate spherical aberration of fifth order, since $n_1 + n_2 - 1 = 5$.

Care is needed when applying these rules as some combinations are not permitted. If, for a given order, we consider the aberrations of highest symmetry, N_1 and N_2, combination aberrations of symmetry $N_1 + N_2$ do not occur. Thus a combination of twofold astigmatism (A_1, a quadrupole effect, for which $n = 1$ and $N = 2$) and fourfold astigmatism (A_3, an octopole effect, for which $n = 3$ and $N = 4$) generates a third-order combination aberration ($n_1 + n_2 - 1 = 3$, $|N_2 - N_1| = 2$, which we recognize as a star aberration (S_2); there is no third-order combination aberration with symmetry $N_1 + N_2 = 6$.

The symmetry or multiplicity of an aberration is most easily understood by returning to the wave aberrations in Section 31.6.2. An aberration for which the term in the wave aberration satisfies

$$W(\theta) = W\left(\theta \exp(2\pi \mathrm{i}/N)\right) \tag{31.39}$$

is of symmetry or multiplicity N, N being the largest integer for which this relation is satisfied.

31.7 The Uhlemann Effect

The most sophisticated type of aberration corrector is designed to correct the chromatic aberration as well as many of the geometrical aberrations. When the first microscope equipped with such a corrector (TEAM 1) was tested, however, its performance was not improved as had been anticipated on theoretical grounds. The explanation was found by Uhlemann, who showed that the resolution was adversely affected by magnetic field noise caused by thermally driven currents in the conductive material of the lenses and multipoles. The transverse components of these noise fields caused random deflection of the beam, large enough to counteract the benefits of the chromatic aberration corrector. These currents are temperature-dependent and the effect can be attenuated by cooling. For details, see Uhlemann et al. (2013a,b, 2014, 2015).

PART V

Deflection Systems

CHAPTER 32

Paraxial Properties of Deflection Systems

32.1 Introduction

In virtually every kind of electron beam device, some provision is made for deflection, that is, for lateral shift of the electrons with as little disturbance of the beam structure as possible. In fixed-beam instruments, essentially in conventional transmission electron microscopes, deflection plays a minor role and is provided only to permit nonmechanical alignment of the column. In scanning devices, however, the role of the optics of the deflection system is to move a focused spot (often called the electron probe) in a raster pattern over a prescribed area of a specimen or a viewing screen, or to sweep a two-dimensional image over a small detector, and its design is at least as important as that of the lenses. The obvious examples here are cathode-ray oscillographs and television tubes, together with the various types of scanning electron and ion microscopes. In the more recent generations of electron microscopes, provision is made for both fixed-beam (conventional) transmission imaging and scanning transmission microscopy, so that the deflection system becomes just one member of the complex sequence of optical components that make up these hybrid instruments. It is the properties of deflection systems, mainly of the magnetic type, that determine the optical performance of electron lithography devices and they have therefore attracted considerable attention. Here, not only is the raster technique employed but – far more advantageously – the shaped-beam technique: instead of a narrow beam, focused into a small probe, a comparatively broad beam, shaped by masks, is deflected. The accuracy required in such devices is very high. More details will be found in Chapter 40 of Volume 2.

In scanning electron microscopes, the magnification is altered by varying the area scanned with the result that high resolution is associated with small deflection angles. The probe size is then determined by the spherical and chromatic aberrations of the probe-forming lens and it is important to place the specimen as close as possible to the lens to keep the corresponding coefficients small. The scan coils are hence placed upstream from the lens. In lithography devices, on the other hand, the area scanned is large and the deflection aberrations become important; there is therefore no need to keep the working distance (the distance between the lens and the target) small and the scan coils can be inserted into this space, which has long been known to be advantageous (e.g., Owen and Nixon, 1973). In the most recent designs, however, the fields of the probe-forming lens and of the scan coils overlap.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00032-8
© 2018 Elsevier Ltd. All rights reserved.

Deflection systems may be either magnetic or electric and each type finds practical application. In old-fashioned television tubes, for example, the power consumption is higher for magnetic deflectors, but these are nevertheless preferred, for it is easier to provide the required deflection current than the voltage needed in an electric system; moreover, the distortions are lower. In the range of frequencies used in television systems, inductance effects do not yet play a role.

In oscilloscopes, on the other hand, electric fields have been widely used since the deflection plates draw little or no current and only at very high frequency does the displacement cease to follow the signal. In magnetic systems, particularly in those with ferrite shielding, the inductance of the device begins to cause problems at much lower frequencies. In spite of this, magnetic systems are almost exclusively used in scanning microscopes since their deflection aberrations can be kept small; electric deflection is often included for beam blanking, that is, switching off the illuminating beam by abruptly deflecting it sideways so that it no longer passes through some centred aperture. Magnetic systems are also usually adopted in lithography devices, though electrostatic and hybrid arrangements are of interest for particular purposes.

In the electrostatic case, deflectors usually consist of pairs of plates, symmetrical about a plane through the optical axis. They may have a wide variety of shapes, ranging from simple rectangles parallel to each other to tilted or curved surfaces. Some typical designs are illustrated in Fig. 32.1. A more advanced design consists of a family of 8 or 12 plates,

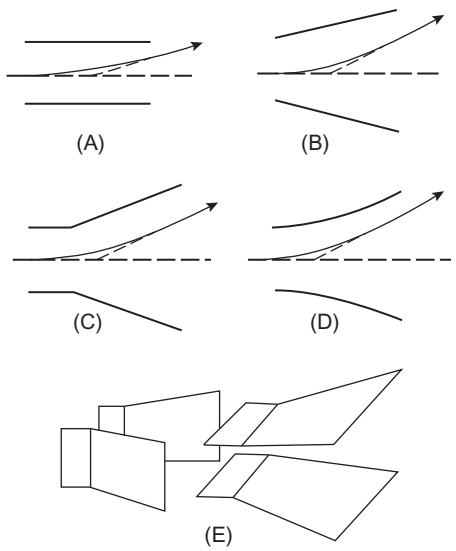


Figure 32.1
(A)–(D) Cross-sections through electric deflectors of various kinds; the flared design (C) is frequently employed. (E) Perspective view of an electric deflector with flared plates.

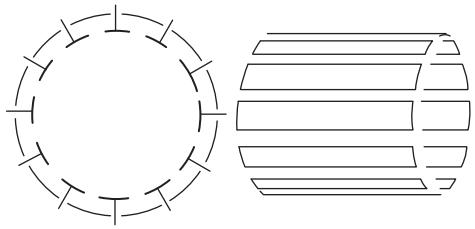


Figure 32.2
Simplified drawing of an electric 12-pole deflector. The outer screening cylinder is omitted in the perspective view (right).

arranged symmetrically around the axis and forming sectors of a cylindrical surface (Fig. 32.2). By choosing the electrode potentials appropriately, not only can deflection at an arbitrary azimuth be achieved but the aberrations can also be partially corrected.

In the magnetic case, two geometries are common: saddle coils and toroidal coils. *Saddle coils* (Fig. 32.3A) are usually enclosed in a ferrite sheath, thereby reducing the wastage of flux. The shielding is omitted only in devices designed to function at high deflection frequencies in order to decrease the inductance. In *toroidal structures* (Fig. 32.3B and C) the turns of the two individual coils are wound meridionally, or in other words, lengthwise around a ferrite yoke, which may have a more complicated shape than a simple cylinder or cone (Fig. 32.3D). This was especially true of television tubes, where distortions had to be kept very small. In these, hybrid schemes were often used; Fig. 32.4 illustrates a deflection system in which the vertical deflection is achieved by means of toroidal coils while the horizontal deflection is produced by the shielded saddle coils. The former have smaller aberrations, the latter lower inductance.

These introductory remarks can give only a superficial impression of the host of deflection systems in practical use. More technical details will be found in Chapter 40 of Volume 2. We now develop the corresponding theory of paraxial properties; aberrations are studied in Chapter 33, The Aberrations of Deflection Systems.

32.2 The Paraxial Optics of Deflection Systems

The study of the optics of deflection systems passes through essentially the same stages as those already encountered for round lenses or quadrupoles; the novel aspects arise from the new symmetry conditions. Hitherto, we have considered systems in which an electron travelling along the straight optic axis – the axis of rotational symmetry of round lenses and the line of intersection of the symmetry or antisymmetry planes in quadrupole systems – experienced no transverse force. Terms of odd degree in x and y were thus

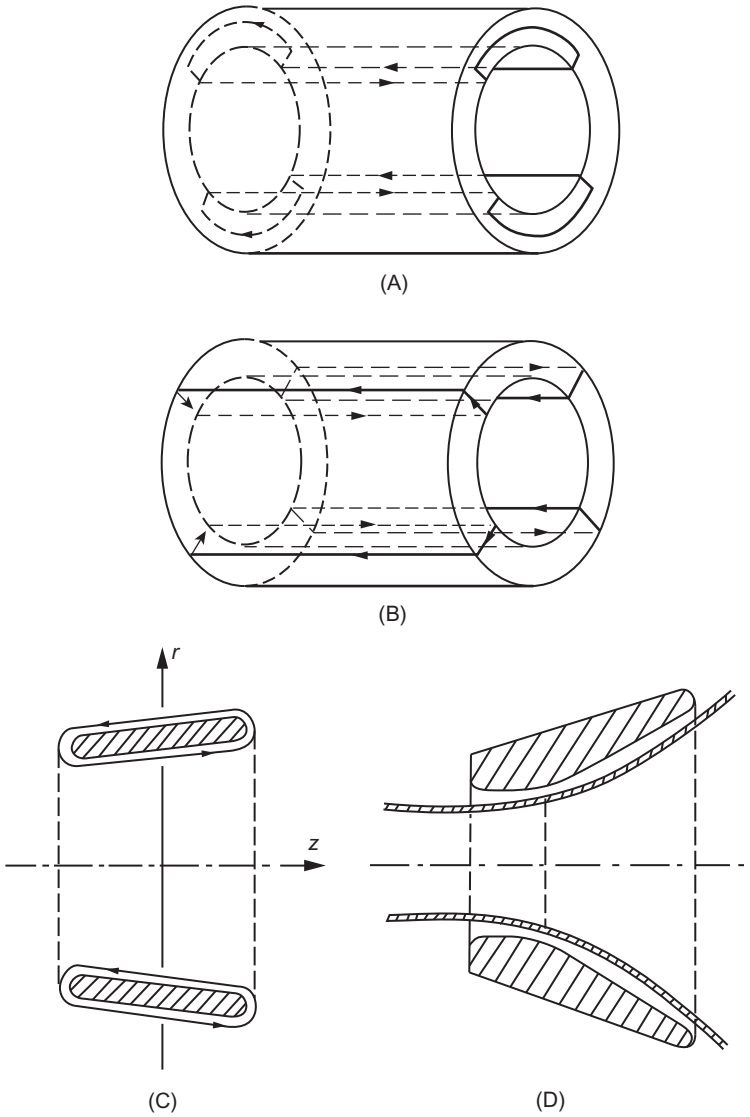


Figure 32.3
Different types of magnetic deflector. (A) Windings in a pair of saddle coils. (B) Windings in a pair of toroidal coils. (C) Axial section through a pair of toroidal coils wound round a conical yoke. (D) Axial section through a ferrite core in a television tube; part of the tube wall is indicated.

excluded from the expansions (7.36) for $\Phi(\boldsymbol{r})$ and (7.45) for A_z, as were terms of even degree from A_x, A_y in Eqs. (7.43−7.44).

In this section, we first derive the most general form of the paraxial equations and then specialize these to the cases of practical importance. Just as for round lenses and

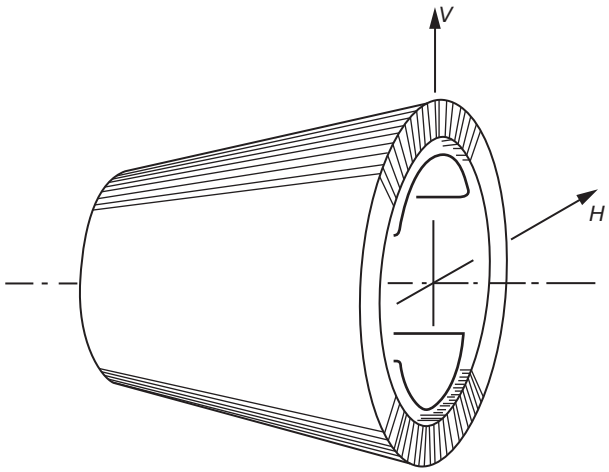


Figure 32.4
Hybrid system consisting of saddle coils for the horizontal deflection (H) and toroidal coils for the vertical deflection (V). On the optic axis, the magnetic field generated by the saddle coils is in the vertical direction while that created by the toroidal coils is horizontal.

quadrupole systems, these equations can be obtained in two equivalent ways, namely, by using the trajectory method or the eikonal method. Here we use the trajectory method, which is convenient for the paraxial theory; the eikonal method is presented in detail by Glaser (1952, 1956), Kaashoek (1968), Ximen (1986) and Rose (2012).

32.2.1 The General Paraxial Equations

The paraxial equations for an arbitrary system with a straight optic axis are easily obtained by linearization of Eq. (3.22). Reduction to linear form implies that $\rho = 1$ from Eq. (3.16), and $B_t = B_z$ from Eq. (3.19). A nontrivial simplification arises from the fact that there is a linear relation between the transverse deflection field strengths at the optic axis and the coordinates x_i, y_i of the trajectory in a given recording plane $z = z_i$, as will become obvious later. The axial values of $E_x = -\Phi_{|x}$, $E_y = -\Phi_{|y}$, B_x and B_y are thus to be considered as *linear* quantities, though this is not explicitly obvious from the corresponding series expansions. We may therefore replace $\hat{\Phi}(\boldsymbol{r})$ in the denominators of Eq. (3.22) by its axial value $\hat{\phi}(z)$ and $\nabla\hat{\Phi}$ by $(1 + 2\varepsilon\phi)\nabla\Phi(\boldsymbol{r})$. Bringing all this together, we obtain

$$\left.\begin{aligned} x'' &= \frac{\gamma}{2\hat{\phi}}(\Phi_{|x} - x'\Phi_{|z}) + \eta\hat{\phi}^{-1/2}(B_y - y'B_z) \\ y'' &= \frac{\gamma}{2\hat{\phi}}(\Phi_{|y} - y'\Phi_{|z}) + \eta\hat{\phi}^{-1/2}(-B_x + x'B_z) \end{aligned}\right\} \tag{32.1}$$

Differentiating Eq. (7.36) and retaining only terms that are of linear order when substituted into Eq. (32.1), we find $\Phi_{|z} = \phi'(z)$ and

$$\begin{aligned} \Phi_{|x} &= -F_1 - \frac{\phi''}{2}x + xp_2 + yq_2 \\ \Phi_{|y} &= -F_2 - \frac{\phi''}{2}y + xq_2 - yp_2 \end{aligned} \tag{32.2a}$$

The linear terms of $\boldsymbol{B}$ (Eqs. 7.46–7.48) are given by $B_z = B$ and

$$\begin{aligned} B_x &= B_1 - \frac{x}{2}B' - P_2 x - Q_2 y \\ B_y &= B_2 - \frac{y}{2}B' - Q_2 x + P_2 y \end{aligned} \tag{32.2b}$$

All this is now to be introduced into Eq. (32.1). The result of these elementary calculations takes a very concise form if the following complex quantities are introduced:

$$w \coloneqq x + \mathrm{i}y \tag{32.3a}$$

$$F_T(z) \coloneqq F_1 + \mathrm{i}F_2 \tag{32.3b}$$

$$B_T(z) \coloneqq B_1 + \mathrm{i}B_2 \tag{32.3c}$$

$$Q_E(z) \coloneqq p_2 + \mathrm{i}q_2 \tag{32.3d}$$

$$Q_M(z) \coloneqq P_2 + \mathrm{i}Q_2 \tag{32.3e}$$

These have the following physical meaning: w is the transverse coordinate, F_T and B_T are the complex transverse axial field strengths, while Q_E and Q_M are quadrupole coefficients. Quadrupole terms are not introduced deliberately in deflection systems but may arise from constructional imperfections. Using Eq. (32.3), an elementary calculation yields

$$w''(z) = A_1 w^* + A_2 w + A_3 w' + A_4 \tag{32.4}$$

the coefficients being given by

$$A_1 = -\frac{\gamma}{2}\frac{Q_E}{\hat{\phi}} + \mathrm{i}\eta\hat{\phi}^{-1/2}Q_M \tag{32.5a}$$

$$A_2 = -\frac{\gamma}{4}\frac{\phi''}{\hat{\phi}} + \frac{\mathrm{i}}{2}\eta\hat{\phi}^{-1/2}B' \tag{32.5b}$$

$$A_3 = -\frac{\gamma}{2}\frac{\phi'}{\hat{\phi}} + \mathrm{i}\eta\hat{\phi}^{-1/2}B \tag{32.5c}$$

$$A_4 = -\frac{\gamma}{2}\frac{F_T}{\hat{\phi}} - \mathrm{i}\eta\hat{\phi}^{-1/2}B_T \tag{32.5d}$$

The complex linear differential equation (32.4) is seldom solved in full generality, though its numerical solution provides absolutely no problem. The reason is that an electron optical system of so general a type is very rarely encountered in practice. We now turn to the important special cases.

First of all, a round electrostatic lens is never combined with electrostatic elements of lower symmetry in such a way that the fields overlap, since the boundary conditions governing the electric field on the electrode surfaces do not allow this (although round lens components can of course be generated by elements such as quadrupoles, suitably excited). Round electrostatic lenses well separated from any other elements are better treated independently by the methods outlined earlier and we hence ignore them here.

A second simplification concerns the term $A_1 w^*$ in Eq. (32.4), which may be neglected. Such a term would cause axial astigmatism, which can be compensated by means of stigmators; in practice, therefore, this term is always negligible. Stigmators are dealt with in Section 31.5; we say no more about them here and assume forthwith that $A_1(z) \equiv 0$.

The combination of round magnetic lenses and deflection units is common in electron lithography systems, but advantageous only under very special conditions, as will become obvious later. If these conditions are not satisfied, it is better to remove the round lens from the deflection field region since its presence destroys the electron optical symmetry properties of the deflection unit. In order to establish these properties, we omit the magnetic lens in the first step.

32.2.2 Ideal Deflection

With all these simplifying assumptions, Eqs. (32.4) and (32.5) reduce to

$$w'' = D(z) \tag{32.6}$$

where

$$D(z) \equiv A_4(z) = -\frac{\gamma F_T(z)}{2\hat{\phi}} - \mathrm{i}\eta\hat{\phi}^{-1/2} B_T(z) \tag{32.7}$$

in which $\hat{\phi}$ is now a constant. Eq. (32.6) is readily integrated, the solution with the general initial conditions

$$w(z_o) = w_o, \quad w'(z_o) = w'_o$$

being given by

$$w(z) = w_o + (z - z_o) w'_o + w_d(z) \tag{32.8}$$

with the particular solution

$$w_d(z) = \int_{z_o}^{z} (z - \zeta) D(\zeta) d\zeta \tag{32.9}$$

This is easily verified. The solution (32.9) is of particular interest since it describes the trajectory of a particle incident along the optic axis; $w_d(z)$ is thus the equation of a curved axis representing the central ray of a deflected beam. The most important property of $w_d(z)$ is its asymptotic behaviour for large values of z beyond the domain of the deflecting field. This is given by

$$w_d(z) = z \int_{-\infty}^{\infty} D(\zeta) d\zeta - \int_{-\infty}^{\infty} \zeta D(\zeta) d\zeta \tag{32.10}$$

The limits of integration have been extended to infinity, which is justified if $D(\zeta)$ has appreciable values only in a finite interval and the coordinates z_o and z are located outside the latter. In practice this simplification is always permissible.

Eq. (32.10) describes a straight line, the asymptote of the deflected principal ray. In general this does *not* intersect the optic axis: it is a skew ray. This is easily seen by writing Eq. (32.10) in real form:

$$\left.\begin{aligned} x_d(z) &= z \int_{-\infty}^{\infty} D_x(\zeta) d\zeta - \int_{-\infty}^{\infty} \zeta D_x(\zeta) d\zeta \\ y_d(z) &= z \int_{-\infty}^{\infty} D_y(\zeta) d\zeta - \int_{-\infty}^{\infty} \zeta D_y(\zeta) d\zeta \end{aligned}\right\} \tag{32.11}$$

This can be cast into the form

$$\left.\begin{aligned} x_d(z) &= x'_d(\infty)(z - z_1) \\ y_d(z) &= y'_d(\infty)(z - z_2) \end{aligned}\right\} \tag{32.12}$$

with the asymptotic slopes

$$x'_d(\infty) = \int_{-\infty}^{\infty} D_x(z) dz, \quad y'_d(\infty) = \int_{-\infty}^{\infty} D_y(z) dz \tag{32.13}$$

and pivot-point coordinates

$$z_1 = \frac{\int_{-\infty}^{\infty} z D_x(z) dz}{\int_{-\infty}^{\infty} D_x(z) dz}, \quad z_2 = \frac{\int_{-\infty}^{\infty} z D_y(z) dz}{\int_{-\infty}^{\infty} D_y(z) dz} \tag{32.14}$$

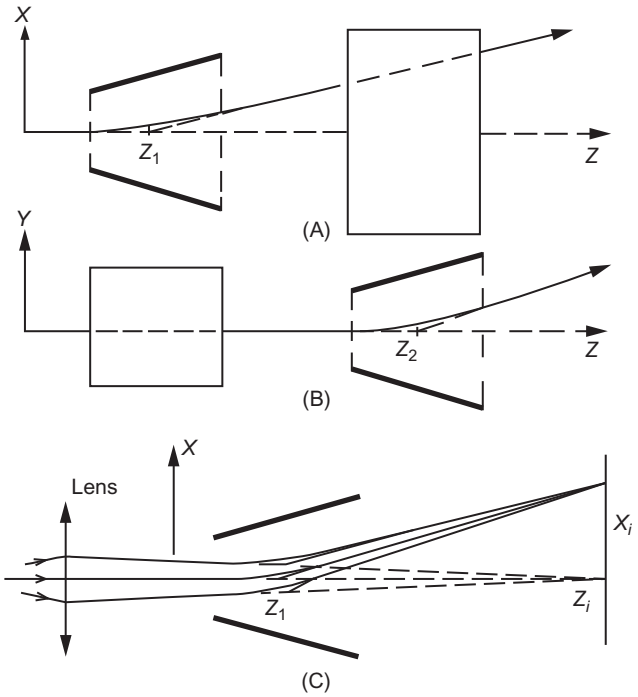


Figure 32.5
(A)–(B) Principal sections through an electric deflection system. (A) x–z plane; (B) y–z plane. (C) Schematic description of the 'ideal' deflection in the x–z plane; the situation in the y–z plane is analogous.

The latter depend only on the geometric forms of the function $D_x(z)$ and $D_y(z)$, not on their absolute values. In the general case, these forms are different and, hence $z_1 \neq z_2$. Figs 32.5A and B illustrate this for an electrostatic deflection system, of the type employed in oscilloscopes.

The general solution (32.8) represents the linear superposition of the principal ray solution $w_d(z)$ and a free-space motion $w_o + (z - z_o)w'_o$. A consequence of Eq. (32.8) is that if there is stigmatic focusing in the absence of deflection ($w_d \equiv 0$), this remains true with deflection; there is only a lateral shift of the focus. Let us assume that with no deflection, the image is a point $(\overline{x}_i, \overline{y}_i)$ in the plane $z = z_i$: $w_d(z) \equiv 0$, $w(z_i) = \overline{w}_i$ irrespective of w'_o, as is sketched in Fig. 32.5C. The initial values in the plane $z = z_o$ in front of the deflection unit must satisfy the condition

$$w_o + (z_i - z_o)w'_o = \overline{w}_i$$

Substituting this into Eq. (32.8) we find

$$w(z) = (z - z_i)w'_o + w_d(z) + \overline{w}_i \tag{32.15}$$

In the plane of the focus, $z = z_i$, we then have

$$w_i := w(z_i) = w_d(z_i) + \overline{w}_i \tag{32.16}$$

or in real form, with the abbreviations $x'_i = x'_d(\infty)$, $y'_i = y'_d(\infty)$,

$$\left.\begin{aligned} x_i &= x'_i(z_i - z_1) + \overline{x}_i \\ y_i &= y'_i(z_i - z_2) + \overline{y}_i \end{aligned}\right\} \tag{32.17}$$

irrespective of the initial slopes x'_o, y'_o. This is characteristic of ideal deflection and illustrated in Fig. 32.5C.

32.2.3 The Dependence on the Electrical Input Signals

The input signals for deflection systems are time-dependent capacitor voltages or coil currents, which generate the necessary deflection fields. Production of the appropriate time dependence of the signals may be a major technical problem, but this does not concern us here. The inertia of the electron beam and the contributions of the Maxwell-displacement terms to the field equations can always be neglected. The time is thus a mere parameter, and a quasi-stationary approximation is quite sufficient. There is no need to study any time-dependent effects here.

The axial electric field strengths F_1, F_2 are proportional to the corresponding deflection voltages U_1, U_2 at the capacitor plates:

$$F_j(z) = U_j a_j(z), \quad j = 1, 2 \tag{32.18}$$

The functions $a_j(z)$ depend only on the geometric forms of the axial field distributions and have the dimension of a reciprocal length. The axial magnetic field strengths B_1, B_2 are likewise proportional to the corresponding coil currents I_1, I_2:

$$B_j(z) = \mu_0 I_j b_j(z), \quad j = 1, 2 \tag{32.19}$$

The permeability μ_0 has been introduced for dimensional reasons: b_1 and b_2 also have the dimensions of a reciprocal length.

Let us first consider systems with no round magnetic lens term. We rewrite Eq. (32.7) in real form and introduce Eqs. (32.18) and (32.19), giving

$$\left.\begin{aligned} D_x &= -\frac{\gamma a_1(z)}{2\hat{\phi}} U_1 + \frac{\mu_0 \eta b_2(z)}{\hat{\phi}^{1/2}} I_2 \\ D_y &= -\frac{\gamma a_2(z)}{2\hat{\phi}} U_2 - \frac{\mu_0 \eta b_1(z)}{\hat{\phi}^{1/2}} I_1 \end{aligned}\right\} \tag{32.20}$$

We see that there is indeed a linear relation between the input signals and the deflection functions D_x, D_y, but not a simple one. If we introduce Eqs. (32.20) into (32.13) and (32.14), and then substitute the resulting expressions into Eq. (32.17), we certainly obtain linear relations between the input signals U_1, U_2, I_1, I_2 and the corresponding deflections x_i, y_i, as we assumed at the outset, but these linear relations are often too complicated for practical applications. We recall that the functions $a_1(z)$, $a_2(z)$, $b_1(z)$, $b_2(z)$ are in general different from one other, and the same may also be true of integrals over these functions.

In practice, apart from very special applications like Wien filters, combined electric-magnetic deflection systems are of rare occurrence though the possibility of using magnetic deflection for large shifts and electrostatic fields for small deflections has been explored in lithography. We first specialize to the pure cases.

In purely *electric* deflection systems, Eq. (32.14) become

$$z_j = \frac{\int_{-\infty}^{\infty} z a_j(z)dz}{\int_{-\infty}^{\infty} a_j(z)dz}, \quad j = 1, 2 \tag{32.21}$$

and Eq. (32.17) then take the form

$$x_i = d_1^E U_1, \qquad y_i = d_2^E U_2 \tag{32.22}$$

with the proportionality factors

$$d_j^E = -\frac{\gamma}{2\hat{\phi}}(z_i - z_j)\int_{-\infty}^{\infty} a_j(z)dz, \quad j = 1, 2 \tag{32.23}$$

These factors are called the *sensitivities* and have the dimensions of a reciprocal field strength. Even though $z_1 \neq z_2$ (see Fig. 32.5), the condition $d_1^E = d_2^E$ can be satisfied by appropriate choice of the functions $a_j(z)$, and in practice this can be achieved by suitable design of the deflection plates. The response is then *isotropic*. Whether this is necessary depends on the particular purpose of the device in question.

In the case of purely *magnetic* deflection systems, the formulae are analogous; it is only necessary to modify certain factors, signs and subscripts. Eq. (32.14) may be written

$$z_j = \frac{\int_{-\infty}^{\infty} z b_j(z)dz}{\int_{-\infty}^{\infty} b_j dz}, \quad j = 1, 2 \tag{32.24}$$

The change of sign in Eq. (32.20) here cancels out. The response relations take the form

$$x_i = d_2^M I_2, \quad y_i = -\, d_1^M I_1 \tag{32.25}$$

with the sensitivities

$$d_j^M = \mu_0 \eta \hat{\phi}^{-1/2} (z_i - z_j) \int_{-\infty}^{\infty} b_j(z) dz, \quad j = 1, 2 \tag{32.26}$$

Once again, the sensitivities can be made equal by appropriate design of the deflection coils, even if the functions $b_1(z)$ and $b_2(z)$ are different. The case $b_1(z) \neq b_2(z)$ certainly occurred in older television tubes equipped with hybrid deflection systems (Fig. 32.4).

32.2.4 Rotation-Invariant Systems

We now examine the effect on the electron optical properties of superimposing a round magnetic field on a purely magnetic deflection system, a situation that occurs in lithography devices. In this situation, Eqs. (32.4) and (32.5), in combination with (32.19), become

$$w'' = \mathrm{i}\eta \hat{\phi}^{-1/2} \left(\frac{1}{2} B' w + B w' \right) + D(z) \tag{32.27}$$

in which the inhomogeneous term $D(z)$ takes the form

$$D(z) = -\, \mathrm{i}\mu_0 \eta \hat{\phi}^{-1/2} \left\{ I_1 b_1(z) + \mathrm{i} I_2 b_2(z) \right\} \tag{32.28}$$

In the theory of round magnetic lenses, it is advantageous to introduce the familiar coordinate frame rotating with the Larmor frequency. In the present case, however, it is no longer obvious that this is advantageous since the natural symmetries of the series expansion for the magnetic deflection field would be destroyed. For this reason, we shall continue to use the fixed (Cartesian) frame (X, Y, z), which we here denote (x, y, z). This causes no particular complication. We note that this choice is not mandatory, and some authors have pursued the calculation in the rotating frame.

We must now solve Eq. (32.27) for arbitrary initial conditions $w(z_o) = w_o$, $w'(z_o) = w'_o$, in some given object plane. The fundamental solutions $\sigma(z)$ and $\tau(z)$ of the homogeneous differential equation with the initial conditions

$$\sigma(z_o) = \tau'(z_o) = 1, \quad \sigma'(z_o) = \tau(z_o) = 0 \tag{32.29a}$$

and the Wronskian

$$W_0 = \mathrm{e}^{-2\mathrm{i}\theta(z)} (\tau' \sigma - \sigma' \tau) =: \tau' \tilde{\sigma} - \sigma' \tilde{\tau} \equiv 1 \tag{32.29b}$$

are now appropriate, $\theta(z)$ being the angle of Larmor rotation (2.39, 15.9); the functions $\sigma(z)$, $\tau(z)$ are now complex. Solution of (32.27) by variation of parameters is straightforward and we find

$$w(z) = w_o\sigma(z) + w_o'\tau(z) + w_d(z) \tag{32.30}$$

with the particular integral

$$\begin{aligned} w_d(z) &= \int_{z_o}^{z} \mathrm{e}^{-2\mathrm{i}\theta(\zeta)} D(\zeta)\big\{\sigma(\zeta)\tau(z) - \tau(\zeta)\sigma(z)\big\}d\zeta \\ &= \int_{z_o}^{z} D(\zeta)\big\{\tilde{\sigma}(\zeta)\tau(z) - \tilde{\tau}(\zeta)\sigma(z)\big\}d\zeta \end{aligned} \tag{32.31}$$

The value of this solution in the image plane $z = z_i$ is the most important. There we have $\tau(z_i) = 0$ and $\sigma(z_i) = M\exp(\mathrm{i}\theta_i)$, $\theta_i = \theta(z_i)$ being the angle of image rotation relative to the object. Hence,

$$w_d(z_i) = -M\mathrm{e}^{\mathrm{i}\theta_i}\int_{z_o}^{z_i} D(z)\tilde{\tau}(z)dz =: w_{di}$$

It will prove convenient to introduce the real fundamental solution $h(z)$ in the rotating frame, with $h(z_o) = 0$, $h'(z_o) = 1$ as usual. From Eq. (32.29) with $\theta(z_o) = 0$, we see that

$$\tau(z) = h(z)\mathrm{e}^{\mathrm{i}\theta(z)}$$

and so

$$w_{di} = -M\mathrm{e}^{\mathrm{i}\theta_i}\int_{z_o}^{z_i} \mathrm{e}^{-\mathrm{i}\theta(z)} D(z)h(z)dz \tag{32.32}$$

We notice that, since $\tau(z_i) = 0$, the paraxial deflection is again *ideal*, in the sense that the general solution $w(z_i) = w_o\sigma(z_i) + w_{di}$ is independent of w_o'. Eq. (32.28) shows that it also depends linearly on the currents I_1 and I_2. In the general case, however, it is *not rotation-invariant.* This means that if we apply the coil currents

$$I_1 = I_0\cos\alpha, \quad I_2 = I_0\sin\alpha \tag{32.33}$$

with a time-dependent phase α, intending to trace out a circle, the deflection signal (with $w_o = 0$) does not in fact respect this but instead describes an ellipse. This effect is called the *paraxial distortion.* In order to obtain a circle, the deflection signal w_{di}, given by Eq. (32.32), must be proportional to the complex current variable I_c,

$$I_c := I_1 + \mathrm{i}I_2 = I_0\mathrm{e}^{\mathrm{i}\alpha} \tag{32.34}$$

This proportionality can be achieved only if

$$\int_{z_o}^{z_i} \mathrm{e}^{-\mathrm{i}\theta} b_1(z) h(z)\, dz = \int_{z_o}^{z_i} \mathrm{e}^{-\mathrm{i}\theta} b_2(z) h(z) dz \tag{32.35}$$

When different types of coil systems are combined, in hybrid toroidal- and saddle-coil systems for instance, this condition is in general not satisfied. Even if Eq. (32.35) is true for a particular lens excitation, a slight variation of the latter to adjust the focusing renders the two integrals unequal. The only reasonable way of satisfying Eq. (32.35) in general is hence to impose the condition $b_1(z) \equiv b_2(z)$, which means that the coils, rotated at 90° to each other, must have the same geometry. Systems of this type are *rotation-invariant.* They are the only reasonable choice when a deflection field is to be combined with a magnetic round lens field and, in our discussion of aberration theory, we shall consider only this case.

We can then drop the suffices of $b_1(z)$ and $b_2(z)$; combining Eqs. (32.32) with (32.28), we find

$$w_{di} = d_c I_c \tag{32.36}$$

in which the *complex sensitivity* is given by

$$d_c = \mathrm{i}\mu_0 \eta \hat{\phi}^{-1/2} M \mathrm{e}^{\mathrm{i}\theta_i} \int_{z_o}^{z_i} \mathrm{e}^{-i\theta(z)} b(z) h(z) dz \tag{32.37}$$

The modulus of this complex quantity is of great technical importance, whereas the phase is not important since it merely describes a *constant* rotation between the input and output signals.

CHAPTER 33

The Aberrations of Deflection Systems

In this chapter, we again distinguish between systems with pure deflection fields and hybrid systems containing a round magnetic lens because the symmetry properties in these two situations are essentially different. We concentrate on the nature of the aberrations, referring to the original publications for the full lists of aberration integrals for the following reasons: in the more general case, the terms are extremely numerous whereas in the simpler cases of pure deflection fields, only a few of the aberration integrals have found practical use. We do, however, reproduce one set of aberration integrals, which is particularly convenient in that the expressions required in many practical situations can be obtained as special cases.

33.1 Pure Deflection Systems

The most important technical devices employing this kind of deflector are oscilloscopes and the older types of television tubes. These do of course also contain electrostatic lenses to focus the beam on the viewing screen but these lenses are well separated from the deflectors, and have little effect so far as the deflection aberrations are concerned. The situation in scanning electron microscopes is more complicated. For medium and high resolution operation, the deflection structure most commonly used, the double deflector, can be situated far enough before the lens for the field overlap to be negligible. If, however, the microscope is designed to produce a good image at very low magnification where the angles involved become large, it will be necessary to place the deflector within the lens field and the situation is that discussed in Section 33.2, always recalling that the deflector precedes the probe-forming lens. The so-called Grigson coils that are used to scan a diffraction pattern over a small detector do, on the other hand, belong to the present section.

For simplicity, we shall assume that the undeflected beam is directed along the optic axis and produces a round spot, blurred by spherical and chromatic aberrations. When the beam is deflected, the aberration coordinates Δx_i, Δy_i of a particular ray in the undeflected beam can simply be added to the corresponding deflection aberration components.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00033-X
© 2018 Elsevier Ltd. All rights reserved.

33.1.1 Two Different Symmetries

Devices that fall into this category are always constructed so as to have two orthogonal planes of symmetry, intersecting along the optic axis within the limits of experimental accuracy. We have already tacitly assumed this in Sections 32.2.2 and 32.2.3, where these symmetry planes were the coordinate planes $x = 0$ and $y = 0$.

In magnifying systems, it is usually the object plane that is of interest and the aberrations are therefore most usefully expressed in terms of position and gradient in that plane or in terms of position in the object and entrance pupil planes. In probe-forming and other demagnifying systems, on the other hand, the plane of interest is the target or image plane and the most suitable variables for characterizing deflection aberrations are therefore the position coordinates x_i and y_i corresponding to ideal deflection, introduced earlier, and the paraxial difference of gradient between that of the ray in question (gradient x_i', y_i') and the deflected central ray (gradient $x_i'(0),\ \ y_i'(0)$); we write $\alpha \coloneqq x_i' - x_i'(0), \quad \beta \coloneqq y_i' - y_i'(0)$.

The lateral aberrations Δx_i and Δy_i can be derived from a perturbation eikonal, which we denote as in Part IV by S. The symmetry properties impose the following conditions on $S(x_i,\ y_i,\ \alpha,\ \beta)$:

$$\begin{aligned} S(-x_i, y_i, -\alpha, \beta) &= S(x_i, y_i, \alpha, \beta) \\ S(x_i, -y_i, \alpha, -\beta) &= S(x_i, y_i, \alpha, \beta) \end{aligned} \tag{33.1}$$

These are the same symmetry properties as those we have encountered for well-aligned systems of quadrupole lenses. At first sight, this is surprising, since in quadrupoles the potentials are positively symmetric with respect to the symmetry planes, while in deflection systems they are antisymmetric. We must, however, remember that the deflection input signals are not fixed constants, but proportional to x_i and y_i respectively.

From Eq. (33.1) it can be concluded that S will contain aberration terms of *even* order only, the lowest order being characterized by terms of fourth degree. We can simply adopt the classification employed in Chapter 29, The Aberrations of Quadrupole Lenses and Octopoles, without further ado; the cartesian representation of the transverse aberrations is given by Eq. (29.12) if we exchange the variables as follows:

$$\xi_o \to x_i, \quad \eta_o \to y_i, \quad x_o' \to \alpha, \quad y_o' \to \beta$$

One additional simplification arises from the condition that, when the deflection fields are switched off ($x_i = y_i = 0$), the aberration must collapse to the spherical aberration of a round lens and hence $C \coloneqq (0030) = (0012) = (0003) = (0021)$ in Eq. (29.12).

We now introduce a simplified notation, in terms of which the perturbation eikonal has the basic form

$$
\begin{aligned}
S = {} & \frac{1}{4}(E_{11}x_i^4 + 2E_{12}x_i^2y_i^2 + E_{22}y_i^4) \\
& + \alpha x_i(D_{11}x_i^2 + D_{12}y_i^2) + \beta y_i(D_{21}x_i^2 + D_{22}y_i^2) \\
& + \frac{\alpha^2}{2}(A_{11}x_i^2 + A_{12}y_i^2) + \frac{\beta^2}{2}(A_{21}x_i^2 + A_{22}y_i^2) \\
& + A_3\alpha\beta x_i y_i + \frac{C}{4}(\alpha^2+\beta^2)^2 \\
& + \frac{x_i\alpha}{3}(K_{11}\alpha^2 + 3K_{12}\beta^2) + \frac{y_i\beta}{3}(3K_{21}\alpha^2 + K_{22}\beta^2)
\end{aligned}
\tag{33.2}
$$

There are three phase shifts (E_{jk}), four distortions (D_{jk}), five astigmatisms (A_{jk}), disregarding differences between these and the field curvature, four comas (K_{jk}) and the spherical aberration (C). Since the phase shifts (E_{jk}) do not contribute to the lateral aberrations, 14 relevant real independent terms remain.

The function S can be normalized in such a way that the lateral aberrations are simply given by

$$
\Delta x_i = \frac{\partial S}{\partial \alpha}, \quad \Delta y_i = \frac{\partial S}{\partial \beta} \tag{33.3}
$$

Differentiation of Eq. (33.2) then gives

$$
\begin{aligned}
\Delta x_i = {} & x_i(D_{11}x_i^2 + D_{12}y_i^2) \\
& + \alpha(A_{11}x_i^2 + A_{12}y_i^2) + A_3\beta x_i y_i \\
& + x_i(K_{11}\alpha^2 + K_{12}\beta^2) + 2K_{21}y_i\alpha\beta \\
& + C\alpha(\alpha^2 + \beta^2)
\end{aligned}
\tag{33.4a}
$$

$$
\begin{aligned}
\Delta y_i = {} & y_i(D_{21}x_i^2 + D_{22}y_i^2) \\
& + \beta(A_{21}x_i^2 + A_{22}y_i^2) + A_3\alpha x_i y_i \\
& + y_i(K_{21}\alpha^2 + K_{22}\beta^2) + 2K_{12}x_i\alpha\beta \\
& + C\beta(\alpha^2 + \beta^2)
\end{aligned}
\tag{33.4b}
$$

33.1.2 Fourfold Symmetry

A further simplification is obtained if the two symmetry planes $x = 0$ and $y = 0$ are equivalent, in the sense that the system is invariant under rotation through 90° about the

optic axis. This means that interchanging x and y ($x \leftrightarrow y$) and α and β ($\alpha \leftrightarrow \beta$) must leave the formulae unaffected, which in turn implies that for all sets of coefficients,

$$\lambda_{11} = \lambda_{22}, \quad \lambda_{12} = \lambda_{21} \tag{33.5}$$

where λ stands for each of E, D, A and K.

In order to reveal the physical meaning of the surviving coefficients, it is helpful to introduce complex and polar position and angular coordinates thus:

$$w_i \coloneqq x_i + \mathrm{i}y_i = r_i \mathrm{e}^{\mathrm{i}\varphi_i} \tag{33.6a}$$

$$s \coloneqq \alpha + \mathrm{i}\beta = \sigma \mathrm{e}^{\mathrm{i}\varphi_a} \tag{33.6b}$$

On expanding Eq. (33.2) as a Fourier series with respect to the azimuths φ_i and φ_a, we find

$$\begin{aligned} S = {} & \frac{1}{4} r_i^4 (E_0 + E_4 \cos 4\varphi_i) \\ & + \sigma r_i^3 \left\{ D_0 \cos(\varphi_a - \varphi_i) + D_4 \cos(3\varphi_i + \varphi_a) \right\} \\ & + \frac{1}{2} \sigma^2 r_i^2 \left\{ A_0 \cos(2\varphi_a - 2\varphi_i) + A_4 \cos(2\varphi_i + 2\varphi_a) \right\} \\ & + \sigma^3 r_i \left\{ K_0 \cos(\varphi_a - \varphi_i) + \frac{1}{3} K_4 \cos(\varphi_i + 3\varphi_a) \right\} \\ & + \frac{1}{2} F_0 \sigma^2 r_i^2 + \frac{1}{4} C_0 \alpha^4 \end{aligned} \tag{33.7}$$

in which

$$\begin{aligned} E_0 &= \frac{1}{4}(3E_{11} + E_{12}), & E_4 &= \frac{1}{4}(E_{11} - E_{12}) \\ D_0 &= \frac{1}{4}(3D_{11} + D_{12}), & D_4 &= \frac{1}{4}(D_{11} - D_{12}) \\ A_0 &= \frac{1}{4}(A_{11} - A_{12} + A_3), & A_4 &= \frac{1}{4}(A_{11} - A_{12} - A_3) \\ K_0 &= \frac{1}{4}(K_{11} + K_{12}), & K_4 &= \frac{1}{4}(K_{11} - 3K_{12}) \\ F_0 &= \frac{1}{2}(A_{11} + A_{12}), & C_0 &\equiv C \end{aligned} \tag{33.8}$$

The transverse aberrations are now obtained by forming the gradient in complex polar coordinates:

$$\Delta w_i = \frac{\partial S}{\partial \alpha} + \mathrm{i}\frac{\partial S}{\partial \beta} = \mathrm{e}^{\mathrm{i}\varphi_a}\left(\frac{\partial}{\partial \sigma} + \frac{\mathrm{i}}{\sigma}\frac{\partial}{\partial \varphi_a}\right)S \tag{33.9}$$

the result being

$$\begin{aligned}\Delta w_i = {} & r_i^3(D_0\mathrm{e}^{\mathrm{i}\varphi_i} + D_4\mathrm{e}^{-3\mathrm{i}\varphi_i}) \\ & + r_i^2\sigma(A_0\mathrm{e}^{\mathrm{i}(2\varphi_i-\varphi_a)} + A_4\mathrm{e}^{-\mathrm{i}(2\varphi_i+\varphi_a)}) \\ & + r_i^2\sigma F_0\mathrm{e}^{\mathrm{i}\varphi_a} + 2\sigma^2 r_i K_0\mathrm{e}^{\mathrm{i}\varphi_i} \\ & + r_i\sigma^2(K_0\mathrm{e}^{\mathrm{i}(2\varphi_a-\varphi_i)} + K_4\mathrm{e}^{-\mathrm{i}(2\varphi_a+\varphi_i)}) \\ & + \sigma^3 C_0\mathrm{e}^{\mathrm{i}\varphi_a}\end{aligned} \tag{33.10}$$

or in the general complex form with Eq. (33.8):

$$\begin{aligned}\Delta w_i = {} & D_0 w_i^2 w_i^* + D_4 w_i^{*3} \\ & + A_0 w_i^2 s^* + A_4 w_i^{*3} s^* \\ & + F_0 w_i w_i^* s + 2K_0 s s^* w_i \\ & + K_0 s^2 w_i^* + K_4 s^{*2} w_i^* \\ & + C_0 s^2 s^*\end{aligned} \tag{33.11}$$

Apart from the choice of notation and the limitation to systems without superimposed magnetic lens fields, this is the representation which has long been known and has been explored in great detail in connection with electron lithography devices (Munro, 1974; Goto and Soma, 1977). In comparison with the aberrations of round lenses (see 24.36–24.37), there are three new types of error, those indicated by the subscript 4; these are known as *fourfold aberrations* and in particular as fourfold distortion (D_4), astigmatism (A_4) and coma (K_4). This nomenclature reflects the behaviour of these aberrations when the entire system is rotated bodily about the optic axis. Let us substitute $\varphi_i = \varphi'_i + \psi$ and $\varphi_a = \varphi'_a + \psi$ in Eq. (33.10). It quickly becomes clear that all the *isotropic* aberration terms, labelled with the subscript zero and familiar from the theory of round lenses, are modified by a common factor $\exp(\mathrm{i}\psi)$, which means that these lateral aberrations rotate in synchronism with the system, just as in round systems. The *fourfold* error terms, on the other hand, are modified by a common factor $\exp(-3\mathrm{i}\psi)$, and their rotation relative to the system is therefore described by a phase factor $\exp(-4\mathrm{i}\psi)$.

The pure fourfold errors have fourfold symmetry; some typical examples are shown in Fig. 33.1. In practice, however, such symmetric error figures are never obtained. The complex superposition of all the terms appearing in Eq. (33.10) or (33.11) can result in

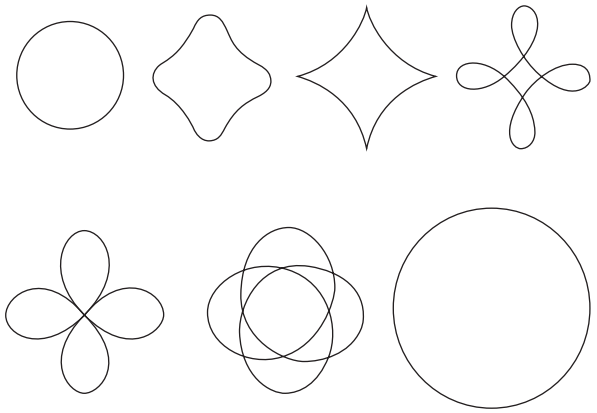


Figure 33.1
Various combinations of fourfold aberrations and defocus values.

quite complicated asymmetric patterns. The possibility of designing electrostatic deflection electrodes free of fourfold aberrations was examined by Dodin (1983).

33.1.3 General Considerations

Although the classification of the possible error types in deflection systems is an intellectually pleasing and somehow beautiful task, the determination of the corresponding aberration coefficients is very laborious. In principle, the calculation is straightforward, whether the trajectory method or the eikonal method is employed, and formulae for the various coefficients have been derived by Glaser (1938, 1941, 1949, see 1952), Wendt (1939, 1942a,b, 1947, 1953), Picht and Himpan (1941a–c), Picht (1943) and Hutter (1947, 1948). They were subsequently reconsidered by Haantjes and Lubben (1957, 1959), by Kanaya (Kanaya and Kawakatsu, 1961a,b, 1962; Kawakatsu and Kanaya, 1961; Kanaya et al., 1961, 1963, 1964) and especially by Kaashoek (1968) and reexamined by Ding (1982). We shall not, however list them here as they are not much used in practice. For numerical values, see Wang (1966, 1967a–c, 1971) and Amboss and Wolf (1971) as well as the papers cited above.

In older television devices and oscilloscopes the deflection angles can become quite large, since the viewing screen must have a given minimum size and the tube-length must be kept reasonably small. For such large angles a perturbation theory of third order is clearly insufficient, but the practical evaluation of error terms of higher than third order by means of the standard perturbation calculus becomes extremely laborious. The fifth-order aberrations of magnetic deflectors were examined by Kaashoek (1968) and Rao and Nixon (1981) and a thorough study of these aberrations for general systems has been made by Uno et al. (1995); we return to this at the end of this chapter.

In electron lithography devices, the off-axis distances and deflection angles remain small enough for the third-order approximation to be adequate but, in such systems, a round magnetic lens is usually strongly coupled to the deflection. The classification given above is then incomplete. This is the subject of the next section.

Before leaving the present topic, we briefly discuss how systems with large deflection angles can be investigated. It is unreasonable to analyse all the aberrations of fifth and even higher order, since the number of terms becomes unmanageable. The discussion must be confined to the technically most important defects, which are clearly the distortions and, with lower priority, the astigmatisms (see Figs 33.2 and 33.3). Since fourfold symmetry is not always present, a more general form of Eq. (33.4) is required. For the distortions of third and fifth order, we may write

$$\begin{aligned}\Delta x_i &= x_i(D_{11}x_i^2 + D_{12}y_i^2 + D_{13}x_i^4 + 2D_{14}x_i^2y_i^2 + D_{15}y_i^4)\\ \Delta y_i &= y_i(D_{21}x_i^2 + D_{22}y_i^2 + D_{23}x_i^4 + 2D_{24}x_i^2y_i^2 + D_{25}y_i^4)\end{aligned} \tag{33.12}$$

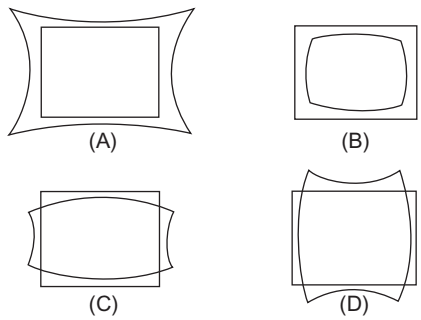


Figure 33.2
Deflection distortions. (A) Cushion; (B) barrel; (C)–(D) hammock.

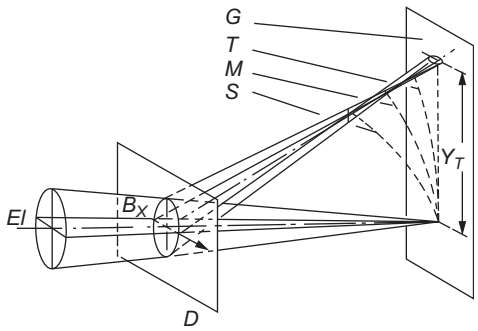


Figure 33.3
Deflection astigmatism (the deflection is shown in only one symmetry plane); *G*: Gaussian image plane; *T*, *M*, *S*: tangential, mean and sagittal image curvature; *D*: principal plane of deflection.

The coefficients in these expansions can be determined by exact numerical calculation of a sufficiently large number of electron trajectories followed by a least-squares fit of Eq. (33.12) to the set of numerically determined aberrations (Δx_i, Δy_i).

33.2 Deflection Systems with Magnetic Lenses

We now study the influence of a round magnetic lens field on the deflection aberrations. As we have already mentioned, this is of great importance in the theory of electron lithography devices. The practical necessity for superimposing deflection and lens fields will be discussed in Chapter 40 of Volume 2.

33.2.1 Geometric Aberrations

We again begin by classifying the permitted types of aberrations. As we have seen in Section 32.2.4, only systems with a fourfold geometric symmetry of the coil system are advantageous, since the device will otherwise exhibit a paraxial distortion. In contrast to the case treated in the preceding section, the planes $x = 0$ and $y = 0$ are no longer planes of mirror symmetry, since the round magnetic lens forces the beam to rotate about the optic axis. There thus remains only an invariance of the eikonal with respect to rotations through 90° about the optic axis. Instead of Eq. (33.1) we have

$$\begin{aligned} S(y_i, -x_i, \beta, -\alpha) &= S(x_i, y_i, \alpha, \beta) \\ S(-x_i, -y_i, -\alpha, -\beta) &= S(x_i, y_i, \alpha, \beta) \end{aligned} \tag{33.13}$$

From the second property it can be concluded that the power series expansion of S can have only terms of *even* total order, the interesting ones being again those of fourth order. The first condition expresses the fourfold symmetry.

The mathematical form of the perturbation eikonal S is most easily obtained by a Fourier series expansion similar to Eq. (33.7). Only terms that are invariant with respect to the simultaneous transforms $\varphi_i = \varphi'_i + \pi/2$ and $\varphi_a = \varphi'_a + \pi/2$ can occur in this expression. Furthermore, terms higher than fourth order need not be retained. The only permissible trigonometric dependences are then

$$\cos n(\varphi_a - \varphi_i), \quad \sin n(\varphi_a - \varphi_i), \quad n = 0, 1, 2$$

together with the cosine and sine of the arguments

$$4\varphi_i, \quad 3\varphi_i + \varphi_a, \quad 2\varphi_i + 2\varphi_a, \quad \varphi_i + 3\varphi_a, \quad 4\varphi_a$$

the sum of the factors of φ_i and φ_a being always 4. In Eq. (33.7) only the cosine terms were obtained, since only these are compatible with Eqs (33.1) and (33.13). If Eq. (33.1) is given up, the corresponding sine terms are also permitted. A fourfold spherical aberration,

given by the terms in $4\varphi_a$, is compatible with the symmetries, but is excluded for technical reasons, since the ordinary spherical aberration of the round lens must be obtained when the deflection currents are switched off.

We now have to complete the Fourier series expansion (33.7) by including the appropriate sine terms. The resulting expression can be cast into a more concise form by introducing a complex notation. For this purpose we define the following complex aberration coefficients:

$$\begin{aligned} & & \tilde{E}_4 &= E_4 + \mathrm{i}e_4 \\ \tilde{D}_0 &= D_0 + \mathrm{i}d_0, & \tilde{D}_4 &= D_4 + \mathrm{i}d_4 \\ \tilde{A}_0 &= A_0 + \mathrm{i}a_0, & \tilde{A}_4 &= A_4 + \mathrm{i}a_4 \\ \tilde{K}_0 &= K_0 + \mathrm{i}k_0, & \tilde{K}_4 &= K_4 + \mathrm{i}k_4 \end{aligned} \tag{33.14}$$

The real parts denote isotropic coefficients, while the imaginary parts represent anisotropic ones. Then, instead of Eq. (33.7), we have

$$\begin{aligned} S = \Re\Big[& \frac{1}{4} r_i^4 \left(E_0 + \tilde{E}_4 \exp\left(-4\mathrm{i}\varphi_i\right)\right. \\ & + \sigma r_i^3 \left\{ \tilde{D}_0 \exp \mathrm{i}(\varphi_i - \varphi_a) + \tilde{D}_4 \exp \mathrm{i}(-3\varphi_i - \varphi_a) \right\} \\ & + \frac{1}{2}\sigma^2 r_i^2 \left\{ \tilde{A}_0 \exp 2\mathrm{i}(\varphi_i - \varphi_a) + \tilde{A}_4 \exp 2\mathrm{i}(-\varphi_i - \varphi_a) \right\} \\ & + \sigma^3 r_i \left\{ \tilde{K}_0 \exp \mathrm{i}(\varphi_i - \varphi_a) + \frac{1}{3}\tilde{K}_4 \exp \mathrm{i}(-\varphi_i - 3\varphi_a) \right\} \Big] \\ & + \frac{1}{2} F_0 \sigma^2 r_i^2 + \frac{1}{4} C_0 \sigma^4 \end{aligned} \tag{33.15}$$

After some elementary calculations, differentiation of Eq. (33.15) using (33.9) gives the geometric aberrations (suffix g)

$$\begin{aligned} \Delta w_g = & C_0 s^2 s^* + \tilde{D}_0 w_i^2 w_i^* + \tilde{D}_4 w_i^{*3} \\ & + F_0 w_i w_i^* s + \tilde{A}_0 w_i^2 s^* + \tilde{A}_4 w_i^{*2} s^* \\ & + 2\tilde{K}_0 s s^* w_i + \tilde{K}_0^* s^2 w_i^* + \tilde{K}_4 s^{*2} w_i^* \end{aligned}$$

$$
\begin{aligned}
&= (w_i^* s^*) \begin{pmatrix} \tilde{D}_0 & F_0 & \tilde{K}_0^* & \tilde{D}_4 & \tilde{K}_4 \\ \tilde{A}_0 & 2\tilde{K}_0 & C_0 & \tilde{A}_4 & 0 \end{pmatrix} \begin{pmatrix} w_i^2 \\ w_i s \\ s^2 \\ w_i^{*2} \\ s^{*2} \end{pmatrix} \\
&= (w_i^* s^*) \begin{pmatrix} \tilde{D}_0 & F_0 & \tilde{K}_0^* \\ \tilde{A}_0 & 2\tilde{K}_0 & C_0 \end{pmatrix} \begin{pmatrix} w_i^2 \\ w_i s \\ s^2 \end{pmatrix} \\
&\quad + (w_i^* s^*) \begin{pmatrix} \tilde{D}_4 & \tilde{K}_4 \\ \tilde{A}_4 & 0 \end{pmatrix} \begin{pmatrix} w_i^{*2} \\ s^{*2} \end{pmatrix}
\end{aligned}
\tag{33.16}
$$

Apart from the novel fact that, with the exception of F_0 and C_0, all the coefficients have become complex, Eq. (33.16) has the same structure as (33.11). The formula clearly exhibits an interesting pattern: fourfold error terms contain only complex conjugate variables, w_i^* and s^*; moreover, the field curvature and the astigmatisms have the same functional structure as the comas, if we exchange the variables $w_i \leftrightarrow s$, $w_i^* \leftrightarrow s^*$.

33.2.2 Chromatic Aberrations

The accuracy requirements in electron lithography devices are so demanding that chromatic aberrations must also be taken into account. As in the corresponding theory of round lenses, a *linear* approximation is sufficient. The chromatic effects can be easily understood in the following way. In Eq. (32.30) the principal solutions $\sigma(z)$, $\tau(z)$, and $w_d(z)$ depend on the acceleration voltage $\hat{\phi}$ and the lens excitation current I_L as parameters, which may fluctuate around their mean values. For sufficiently small fluctuations, $\Delta\hat{\phi}$ and ΔI_L, an expansion of the form

$$
\begin{aligned}
\Delta w_c = w_o &\left(\frac{\partial \sigma}{\partial \hat{\phi}} \Delta\hat{\phi} + \frac{\partial \sigma}{\partial I_L} \Delta I_L \right) + w_o' \left(\frac{\partial \tau}{\partial \hat{\phi}} \Delta\hat{\phi} + \frac{\partial \tau}{\partial I_L} \Delta I_L \right) \\
&+ \frac{\partial w_d}{\partial \hat{\phi}} \Delta\hat{\phi} + \frac{\partial w_d}{\partial I_L} \Delta I_L
\end{aligned}
$$

is justified. Usually, $|w_o|$ is so small that the corresponding term can be neglected; w_o' is proportional to the final slope s. Owing to the rotation invariance in the paraxial domain, the whole solution $w_d(z)$ is proportional to the complex deflection current I_c and also, from Eq. (32.36), to w_{di}. Since this is true for all values of $\hat{\phi}$ and I_L, this proportionality must also hold for the derivatives $\partial w_d/\partial\hat{\phi}$ and $\partial w_d/\partial I_L$. Moreover, $\tau(z)$, being a paraxial solution

for an ordinary magnetic lens, depends only on $I_L^2/\hat{\phi}$ and hence $\hat{\phi}\partial\tau/\partial\hat{\phi} = -2I_L\partial\tau/\partial I_c$. Putting all this together and referring to the image plane $z = z_i$, we obtain the general form

$$\Delta w_c = C_c s\left(\frac{\Delta\hat{\phi}}{\hat{\phi}} - 2\frac{\Delta I_L}{I_L}\right) + \left(C_{T1}\frac{\Delta\hat{\phi}}{\hat{\phi}} + C_{T2}\frac{\Delta I_L}{I_L}\right)w_{di} \tag{33.17}$$

with $\Delta w_{di} = \Delta w_c + \Delta w_g$ as the total lateral aberration. The first term in Eq. (33.17) is the familiar axial chromatic aberration of a round magnetic lens, referred to the image plane. The second term, being proportional to the deflection w_i, is called the *transverse* chromatic error and represents a new type of aberration. Between the corresponding coefficients C_{T1} and C_{T2} no simple relation can be expected, since any variation of $\theta(z)$ in Eq. (32.32) causes complications.

33.3 Detailed Aberration Analyses

Apart from the choice of notation, the classifications (33.16) and (33.17) are the same as those first derived by Ohiwa (1970) and Munro (1974), who used the eikonal and trajectory methods respectively. Munro lists the geometric and chromatic aberration integrals for purely magnetic deflection and round lens systems. Goto and Soma (1977) pointed out that Eq. (33.16) is too limited for practical purposes, since dynamic correction of the aberrations is not represented whereas such correction has long been an important feature of deflection units. Additional weak multipole fields are applied, generated by currents that are usually nonlinear functions of the principal currents. Owen and Nixon (1973) had already pointed out that deflection field curvature and both isotropic and anisotropic distortion can be eliminated in this way, and Goto and Soma demonstrated that this is also true of both types of astigmatism. Moreover, linear correction currents are capable of eliminating all the pure deflection aberrations. Their paper concludes with a list of aberration integrals, which are a generalization of those to be found in Munro (1974). Later, Soma (1977) generalized the formulae still further to include all possible focusing and deflection effects of both magnetic and electrostatic type for arbitrary superposition of the fields; even relativistic effects were taken into account. A correction to these expressions was made by Li (1983).

Meanwhile, the possibility of eliminating or compensating the aberrations in demagnifying systems had been examined by Koops (1972, 1973; Koops and Bernhard, 1975) and ways of correcting all the aberrations in scanning microscopes by Crewe and Parker (1976). A systematic analysis of aberration minimization for post-lens deflection, double deflection before the lens and the 'moving objective lens' (MOL) of Ohiwa (1970, 1978, 1979; Ohiwa et al., 1971), is to be found in Kern (1979). We shall return to the MOL in Chapter 40 of Volume 2 but we mention that Ohiwa (1979) relates this concept to the coma-free condition and states explicitly that the introduction of pre-lens deflection is equivalent to replacement of the position and angle variables by linear combinations of these and the deflection

current (in the magnetic case). A systematic procedure for eliminating aberrations from magnetic scanning and focusing combinations is set out in Hosokawa (1980), and the aberration coefficients of a double-deflection unit for which the second deflector coincides with a round magnetic lens field are listed in Kuroda (1980). More general formulae for adding deflection aberrations are to be found in Lencová (1981); these are applied very tellingly to several practical situations by Lencová (1988), using a model field (Lenc and Lencová, 1988). Expressions for the asymptotic aberration coefficients of combined focusing–deflection units have been used by Li (1994) to derive the addition formulae for the coefficients of two (or more) subsystems.

We postpone discussion of a series of papers by Munro and Chu while we examine the work of Ximen and Li. In the first contribution (Ximen, 1981), which extends this author's earlier work (Ximen, 1977, 1978), expressions for the terms of second and fourth order in the eikonal function are given explicitly in the rotating coordinate system; the fact that deflection has the effect of replacing position and angle (in the object or target plane) by a linear combination of these quantities and the deflecting current or voltage, mentioned above, is used to introduce a generating function, from which the various aberration coefficients are extracted by differentiation. This ingenious technique enables Ximen to formulate the aberration theory very compactly. Explicit aberration integrals are given in later papers by Li and Ximen (1982a,b), once again in the rotating coordinate system, together with formulae for adding the aberration coefficients of sets of deflectors. Finally, Ximen and Li (1982) show how dynamic correction can be used to eliminate various aberrations. Further details are given in Li (1981). The asymptotic aberration coefficients of both position and gradient are listed in Hawkes (1989), where the coefficients of the corresponding polynomials are also to be found. A less full list is given by Tang (1986).

One important family of aberrations is not considered in any of these papers, namely, the parasitic aberrations; this gap has been filled by Plies (1982a,b), who examined the aberrations that arise from misalignments of these complex hybrid systems. The purpose of his studies was to find practical tolerance limits by exploring the influence of various kinds of misalignment on the aberrations.

We now return to the series of papers by Munro and Chu (1982a,b; Chu and Munro, 1982a,b) devoted to the numerical analysis of electron beam lithography systems. The first two papers are concerned with field calculation and the fourth with computer optimization of complex systems. It is the third part that particularly concerns us here, for in it Chu and Munro give a list of aberration integrals, which can be used to study systems consisting of any combination of magnetic and electrostatic lenses and deflectors; these expressions are of immediate practical use and we reproduce them here, in the notation defined earlier.

The paraxial equation for a system consisting of any combination of round lenses and deflectors takes the (nonrelativistic) form

$$w'' + \frac{\phi'}{2\phi}w' + \frac{\phi''}{4\phi}w - \frac{\mathrm{i}\eta}{\phi^{1/2}}\left(Bw' + \frac{1}{2}B'w\right) = -\frac{F_T}{2\phi} - \mathrm{i}\frac{\eta}{\phi^{1/2}}B_T = -\frac{U_c}{2\phi}a(z) - \frac{\mathrm{i}\eta\mu_0 I_c}{\phi^{1/2}}b(z) \tag{33.18}$$

which is a generalization of Eq. (32.27) using the notation of (32.3b,c) and (32.18−32.29). The solution of Eq. (33.18) is written

$$w(z) = w_o\sigma(z) + w_o'\tau(z) + I_c m(z) + U_c e(z) \tag{33.19}$$

which replaces Eq. (32.30), I_c being the complex current of (32.34) and U_c its electrical analogue; $m(z)$ and $e(z)$ are the magnetic and electrical contributions to $w_d(z)$ (32.31), after separating the current I_c and the voltage U_c as factors:

$$\begin{aligned} m(z) &= -\frac{\mathrm{i}\eta\mu_0}{\phi_o^{1/2}}\int_{z_o}^{z} b(\zeta)\left\{\tau(z)\sigma^*(\zeta) - \sigma(z)\tau^*(\zeta)\right\}d\zeta \\ e(z) &= -\frac{1}{2\phi_o^{1/2}}\int_{z_o}^{z} \frac{a(\zeta)}{\phi^{1/2}(\zeta)}\left\{\tau(z)\sigma^*(\zeta) - \sigma(z)\tau^*(\zeta)\right\}d\zeta \end{aligned} \tag{33.20}$$

These expressions are obtained by means of the method of variation of parameters, applied to Eq. (33.18), as in (32.31); examination of the Wronskian now tells us that

$$\phi^{1/2}(\sigma\tau' - \sigma'\tau) = \mathrm{e}^{2\mathrm{i}\theta}\phi_o^{1/2} \tag{33.21a}$$

or

$$\phi^{1/2}(\tilde{\sigma}\tau' - \sigma'\tilde{\tau}) = \phi_o^{1/2} \tag{33.21b}$$

(cf. 32.29b).

The aberrations are obtained by retaining higher order terms in the potential and field expansions and again using the method of variation of parameters—the standard trajectory method. Into the inhomogeneous terms is substituted a slightly modified form of Eq. (33.19),

$$w(z) = w_i\frac{\sigma(z)}{\sigma_i} + s_i\frac{\tau(z)}{\tau_i'} + w_i^{(m)}\frac{m(z)}{m_i} + w_i^{(e)}\frac{e(z)}{e_i} \tag{33.22}$$

in which the image parameters in terms of which the aberrations are expressed have been introduced; $\sigma_i \coloneqq \sigma(z_i)$ and likewise for τ_i, m_i and e_i. The shifts $w_i^{(m)}$ and $w_i^{(e)}$ are given by equations analogous to (32.36): $w_i^{(m)} = d_c^{(m)}I_c$ and $w_i^{(e)} = d_c^{(e)}U_c$.

In the worst case, in which no term is negligible, 56 complex geometric aberration coefficients emerge from this calculation. It is fortunately extremely rare that all these are required and indeed, we return below to the utility of an analysis such as this when so large

a number of coefficients are relevant. At the other extreme, for a system in which the Gaussian spot size is negligible compared with the magnetic deflection $w_i^{(m)}$ and there is no electrostatic contribution, only nine coefficients survive (Munro, 1974, for example). For the intermediate cases, the numbers are as follows:

Gaussian round beam and dual-channel deflection ($U_c \neq 0$, $I_c \neq 0$), 27 coefficients.
Shaped beam and single-channel deflecction (either $I_c = 0$, $U_c \neq 0$ or $I_c \neq 0$, $U_c = 0$). 24 coefficients.

In practice, the most general situation that is liable to be encountered is the *dual-channel* system, in which the Gaussian spot size is typically negligible, and both magnetic and electric deflection are employed (Fig. 40.20): the magnetic deflection provides coverage of a large field while the electrostatic deflection offers fast response.

The nature and origins of all these aberrations have been examined very carefully by Plies (1982a,b) and in particular, by Plies and Elstner (1989a,b). We reproduce their tables, showing the types of aberration that occur and the field components that give rise to them. They note that Munro found different numbers of aberrations for the general case and for the shaped-beam one-channel case and explain why certain coefficients present in Munro's list vanish. Table 33.1 shows the aberrations in the absence of sextupole components of the field, some of which are not independent. The 40 complex coefficients are therefore reduced to 64 real coefficients. Note that in Plies and Elstner's papers, the aberrations are expressed in terms of quantities which are not quite the same as those employed here but the general patterns are unaffected. (The quantity denoted by γ by Plies and Elstner corresponds to w_o but α corresponds to $w_o' - w_o w_\gamma'(z_o)$, where w_γ is such that $w_\gamma(z_o) = 1$ but now $w_\gamma(z_a) = 0$ where z_a is an aperture plane.) Table 33.2 shows the number of complex third-order aberration coefficients for round lenses, deflectors and combinations of the two and their dependence on α, γ, U and I.

Substituting (33.22) into the aberration formula, 27 aberration coefficients are obtained, as shown in the Table 33.3. The coefficients are as follows:

Table 33.1: Nature of the Aberrations of Round Lenses and Deflectors

$\alpha^2\alpha^*$			Spherical aberration
$\alpha\alpha^*\gamma$			Coma
$\alpha^2\gamma^*$			Coma
$\alpha\gamma\gamma^*$	$\alpha\gamma U^*$, $\alpha\gamma^* U$, $\alpha\gamma I^*$, $\alpha\gamma^* I$	αUU^*, αUI^*, $\alpha U^* I$, αII^*,	Field curvature
$\alpha^*\gamma^2$	$\alpha^*\gamma U$, $\alpha^*\gamma I$	$\alpha^* U^2$, $\alpha^* UI$, $\alpha^* I^2$	Astigmatism
$\gamma^2\gamma^*$	$\gamma^2 U^*$, $\gamma^2 I^*$, $\gamma\gamma^* U$, $\gamma\gamma^* I$	$U^2 U^*$, $U^2 I^*$, $UU^* I$	Distortion
	γUU^*, γUI^*, $\gamma U^* I$, γII^*	UII^*, $U^* I^2$, $I^2 I^*$,	Distortion
	$\gamma^* U^2$, $\gamma^* UI$, $\gamma^* I^2$		Distortion
Round lens	Round lens and deflector	Round lens and deflector	

Table 33.2: Number and Types of Aberrations in Various Combinations of Round Lens and Multipole

1					α
			4	2	α, U
			4	2	α, I
			3	1	α, U, I
			1	1	U
			1	1	I
			4	2	U, I
4		0			α, γ
	3	1			α, γ, U
	3	1			α, γ, I
1		0			γ
	4	2			γ, U
	4	2			γ, I
	3	1			γ, U, I
Round Lens	Dipole	Additional sextupole	Dipole	Additional sextupole	
	Round lens and Deflector		Deflector		

Table 33.3: Third-order Geometrical Aberrations of a Combined Focusing and Dual-Channel Deflection System With Magnetic Main-Field Deflection and Electrostatic Secondary Deflection

	Axial Aberration	Magnetic Deflection Aberration	Electrostatic Deflection Aberration	Mixed Deflection Aberration
Spherical aberration	$C_0 s^2 s^*$			
Coma	—	$\tilde{K}_0^{(m)*} s^2 w^{(m)}*$	$\tilde{K}_0^{(e)*} s^2 w^{(e)}*$	
		$2\tilde{K}_0^{(m)} ss^* w^{(m)}$	$2\tilde{K}_0^{(e)} ss^* w^{(e)}$	
Astigmatism and field curvature	—	$\tilde{F}_0^{(m)} sw^{(m)} w^{(m)}*$	$\tilde{F}_0^{(e)} sw^{(e)} w^{(e)}*$	$\tilde{F}_0^{(em)} sw^{(m)}* w^{(e)}$
				$\tilde{F}_0^{(me)} sw^{(m)} w^{(e)}*$
		$\tilde{A}_0^{(m)} s^* w^{(m)2}$	$\tilde{A}_0^{(e)} s^* w^{(e)2}$	$\tilde{A}_0^{(me)} s^* w^{(m)} w^{(e)}$
Distortion	—	$\tilde{D}_0^{(m)} w^{(m)2} w^{(m)}*$	$\tilde{D}_0^{(e)} w^{(e)2} w^{(e)}*$	$\tilde{D}_0^{(mme*)} w^{(m)2} w^{(e)}*$
	—			$\tilde{D}_0^{(mme*)} w^{(m)} w^{(m)}* w^{(e)}$
	—			$\tilde{D}_0^{(mee*)} w^{(m)} w^{(e)} w^{(e)}*$
	—			$\tilde{D}_0^{(m*ee)} w^{(m)}* w^{(e)2}$
Fourfold coma	—	$\tilde{K}_4^{(m)} s^{*2} w^{(m)}*$	$\tilde{K}_4^{(e)} s^{*2} w^{(e)}*$	— — —
Fourfold astigmatism	—	$\tilde{A}_4^{(m)} s^* w^{(m)*2}$	$\tilde{A}_4^{(e)} s^* w^{(e)*2}$	$\tilde{A}_4^{(me)} s^* w^{(m)*} w^{(e)*}$
Fourfold distortion	—	$\tilde{D}_4^{(m)} w^{(m)*3}$	$\tilde{D}_4^{(e)} w^{(e)*3}$	$\tilde{D}_4^{(me)} w^{(m)*2} w^{(e)*}$
	—			$\tilde{D}_4^{(em)} w^{(m)*} w^{(e)*2}$

$$
\begin{aligned}
C_0 &= F(\tau,\tau,\tau^*)\\
\tilde{K}_0^{*(m)} &= F(\tau,\tau,m^*), \quad \tilde{K}_0^{*(e)} = F(\tau,\tau,e^*)\\
2\tilde{K}_0^{(m)} &= F(\tau,m,\tau^*) + F(m,\tau,\tau^*)\\
2\tilde{K}_0^{(e)} &= F(\tau,e,\tau^*) + F(e,\tau,\tau^*)\\
\tilde{A}_0^{(m)} &= F(m,m,\tau^*), \quad \tilde{A}_0^{(e)} = F(e,e,\tau^*)\\
\tilde{A}_0^{(me)} &= F(m,e,\tau^*) + F(e,m,\tau^*)\\
F_0^{(m)} &= F(\tau,m,m^*) + F(m,\tau,m^*)\\
F_0^{(e)} &= F(\tau,e,e^*) + F(e,\tau,e^*)\\
F_0^{(me)} &= F(\tau,m,e^*) + F(m,\tau,e^*)\\
F_0^{(em)} &= F(\tau,e,m^*) + F(e,\tau,m^*)\\
\tilde{D}_0^{(m)} &= F(m,m,m^*), \quad \tilde{D}_0^{(e)} = F(e,e,e^*)\\
\tilde{D}_0^{(mme^*)} &= F(m,m,e^*), \quad \tilde{D}_0^{(m^*ee)} = F(e,e,m^*)\\
\tilde{D}_0^{(mm^*e)} &= F(m,e,m^*) + F(e,m,m^*)\\
\tilde{D}_0^{(mee^*)} &= F(e,m,e^*) + F(m,e,e^*)\\
\tilde{K}_4^{(m)} &= G(\tau^*,\tau^*,m^*), \quad \tilde{K}_4^{(e)} = G(\tau^*,\tau^*,e^*)\\
\tilde{A}_4^{(m)} &= 2G(\tau^*,m^*,m^*), \quad \tilde{A}_4^{(e)} = 2G(\tau^*,e^*,e^*)\\
\tilde{A}_4^{(me)} &= 2G(\tau^*,m^*,e^*) + 2G(\tau^*,e^*,m^*)\\
\tilde{D}_4^{(m)} &= G(m^*,m^*,m^*), \quad \tilde{D}_4^{(e)} = G(e^*,e^*,e^*)\\
\tilde{D}_4^{(me)} &= G(m^*,m^*,e^*) + 2G(m^*,e^*,m^*)\\
\tilde{D}_4^{(em)} &= G(e^*,e^*,m^*) + 2G(e^*,m^*,e^*)
\end{aligned}
\tag{33.23}
$$

(These are given by Chu and Munro and are a striking example of the advantage of the eikonal method over the trajectory method: the fact that the formulae for $\tilde{K}_0^{(e)}$ and $\tilde{K}_0^{(m)}$ and their complex conjugates, given separately above, are the same is far from obvious.) See Li (1986, 1992) or Hosokawa (2002) for a list of all the relations between the coefficients.

The functions F and G are as follows:

$$
F(x_1,x_2,x_3) = F_1 + \frac{1}{\phi_i^{1/2}\tau_i'^*\upsilon_{1i}\upsilon_{2i}\upsilon_{3i}^*}\int_{z_o}^{z_i}\phi^{1/2}(F_2+F_3)dz \tag{33.24a}
$$

where

$$F_1 = \frac{n_1 n_2 n_3}{\phi_i^{1/2}\tau_i'^* v_{1i} v_{2i} v_{3i}^*}\left[\phi^{1/2}\tau'^* v_2\left\{v_1 v_3^*\left(\frac{\phi''}{32\phi} - \frac{\mathrm{i}\eta B'}{16\phi^{1/2}}\right) + v_1\left(\frac{s_3 a}{16\phi} + \frac{\mathrm{i}\eta m_3 b}{8\phi^{1/2}}\right) + v_3^*\left(\frac{s_1 a}{8\phi} + \frac{\mathrm{i}\eta m_1 b}{4\phi^{1/2}}\right)\right\}\right]_{z_o}^{z_i} \tag{33.24b}$$

$$\begin{aligned} F_2 = {} & \frac{1}{2}\tau'^* x_1' x_2' x_3'^* + \frac{3}{64}\left(\frac{\phi''}{\phi}\right)^2 \tau^* x_1 x_2 x_3''^* \\ & - \frac{\phi''}{32\phi}(\tau''^* x_1 x_2 x_3^* + 2\tau^* x''_1 x_2 x_3^* + \tau^* x_1 x_2 x_3''^* \\ & + 2\tau'^* x_1 x_2 x_3'^* + 2\tau^* x_1' x_2' x_3^*) \\ & + \frac{\phi''\phi'}{32\phi^2}(\tau'^* x_1 x_2 x_3^* + 2\tau x_1' x_2 x_3^* + \tau x_1 x_2 x_3'^*) \\ & - \frac{3\phi''\phi'^2}{128\phi^3}\tau^* x_1 x_2 x_3^* \\ & + \frac{\mathrm{i}\eta B'}{16\phi^{1/2}}(\tau''^* x_1 x_2 x_3^* - 2\tau^* x_1'' x_2 x_3^* + \tau^* x_1 x_2 x_3''^* \\ & + 2\tau'^* x_1 x_2 x_3'^* - 2\tau^* x_1' x_2' x_3^*) \end{aligned} \tag{33.24c}$$

$$\begin{aligned} F_3 = {} & \frac{s_1 a}{8\phi}\left\{\tau^*\left(\frac{\phi''}{\phi}x_2 x_3^* + \frac{s_2 a}{\phi}x_3^* + \frac{s_3 a}{\phi}x_2\right)\right. \\ & - (\tau''^* x_2 x_3^* + \tau^* x_2'' x_3^* + \tau^* x_2 x_3''^* + 2\tau'^*(x_2 x_3^*)') \\ & \left. + \frac{\phi'}{\phi}(\tau^* x_2 x_3^*)' - \frac{3\phi'^2}{4\phi^2}\tau^* x_2 x_3^*\right\} + \frac{s_2 a}{16\phi}x_3^*\left(4\tau'^* x_1' + \frac{\phi''}{\phi}\tau^* x_1\right) \\ & + \frac{s_3 a}{16\phi}\left\{\frac{3\phi''}{2\phi}\tau^* x_1 x_2 - \tau''^* x_1 x_2 - 2\tau x''_1 x_2 - 2\tau^* x_1' x_2'\right. \\ & \left. + \frac{\phi'}{\phi}x_2(\tau'^* x_1 + 2\tau^* x_1') - \frac{3\phi'^2}{4\phi^2}\tau^* x_1 x_2\right\} \\ & - \mathrm{i}\frac{\eta b m_1}{4\phi^{1/2}}(\tau''^* x_1 x_3^{\,*} - \tau x''_2 x_3^* + \tau x_2 x_3''^* + 2\tau'^* x_2 x_3'^*) \\ & - \mathrm{i}\frac{\eta b m_3}{8\phi^{1/2}}(\tau''^* x_1 x_2 - 2\tau^* x''_1 x_2 - 2\tau^* x_1' x_2') \end{aligned} \tag{33.24d}$$

and

$$G(x_1^*, x_2^*, x_3^*) \coloneqq \frac{1}{2\phi_i^{1/2}\tau'^* x_{1i}^* x_{2i}^* x_{3i}^*} \int_{z_o}^{z_i} \phi^{1/2} \left(\frac{s_3 p_e}{2\phi} + \frac{\eta m_3 P_m}{\phi^{1/2}} \right) \tau^* x_1^* x_2^* dz \tag{33.25}$$

The quantities v_j, n_j, m_j and s_j that figure in these definitions take the following values:

w_j	v_j	n_j	m_j	s_j
τ	τ'	0	0	0
σ	σ	1	0	0
m	m	1	1	0
e	e	1	0	1

(33.26)

The field functions p_e and P_m that appear in $G(x_1^*, x_2^*, x_3^*)$ denote the contributions from the field components with threefold symmetry:

$$p_e \coloneqq \frac{p_3 - \mathrm{i}q_3}{U_c} \quad P_m \coloneqq -\mathrm{i}\frac{P_3 - \mathrm{i}Q_3}{I_c} \tag{33.27}$$

(*Note*: the above formulae are slightly less general than those of Chu and Munro, who retain the possibility that the different deflection fields are not aligned; we have set the angles between them equal to zero.)

These expressions can be generalized to include the other principal situations, namely, dual-channel systems in which each channel can be electrostatic or magnetic and shaped-beam systems (Gaussian spot large) with electrostatic or magnetic deflection (but not both). For such shaped-beam systems, with magnetic deflection say, Eq. (33.22) becomes

$$w(z) = w_i \frac{\sigma(z)}{\sigma_i} + s_i \frac{\tau(z)}{\tau_i'} + w_i^{(m)} \frac{m(z)}{m_i} \tag{33.28}$$

so that $w_i^{(e)}$ must be replaced by w_i throughout Table 33.1 and column four now contains 'shaped-beam aberrations'; (33.23–33.25) give the aberration integrands on substituting $\sigma(z)$ for $e(z)$. A similar set of substitutions yields the corresponding expressions for a shaped-beam system with electrostatic deflection.

The general dual-channel system coefficients are again given by Eq. (33.23) but the functions F_1, F_3 and G are more complicated:

$$\begin{aligned} F_1(x_1, x_2, x_3) = \frac{n_1 n_2 n_3}{\phi^{1/2}\tau'^* v_{1i} v_{2i} v_{3i}^*} \Bigg[\phi^{1/2}\tau'^* v_2 \Bigg\{ v_1 v_3^* \left(\frac{\phi''}{32\phi} - \frac{\mathrm{i}\eta B'}{16\phi^{1/2}} \right) \\ + v_1 \left(\frac{m_3 a_1 + s_3 a_2}{16\phi} - \frac{\mathrm{i}\eta}{8\phi^{1/2}} (m_3 b_1 + s_3 b_2) \right) \\ + v_3^* \left(\frac{m_1 a_1 + s_1 a_2}{8\phi} + \frac{\mathrm{i}\eta}{4\phi^{1/2}} (m_1 b_1 + s_1 b_2) \right) \Bigg\} \Bigg]_{z_o}^{z_i} \end{aligned}$$

$$
\begin{aligned}
F_3(x_1,x_2,x_3) = &-\frac{m_1a_1+s_1a_2}{8\phi}\left\{\tau^*\left(\frac{\phi''}{\phi}x_2x_3^*\right.\right.\\
&+\left.\frac{m_2a_1+s_2a_2}{\phi}x_3^*+\frac{m_3a_1+s_3a_2}{\phi}x_2\right)\\
&-(\tau''^*x_2x_3^*+\tau^*x''_2x_3^*+\tau^*x_2x''^*_3+2\tau'^*(x_2x_3^*)')\\
&+\left.\frac{\phi'}{\phi}(\tau^*x_2x_3^*)'-\frac{3\phi'^2}{4\phi^2}\tau^*x_2x_3^*\right\}\\
&+\frac{m_2a_1+s_2a_2}{16\phi}x_3^*\left(4\tau'^*x_1'-\frac{\phi''}{\phi}\tau^*x_1\right)\\
&+\frac{m_2a_1+s_3a_2}{16\phi}\left\{\frac{3\phi''}{2\phi}\tau x_1x_2-\tau''^*x_1x_2-2\tau^*x_1''x_2-2\tau^*x_1'x_2'\right.\\
&+\left.\frac{\phi'}{\phi}x_2(\tau'^*x_1+2\tau^*x_1')-\frac{3\phi'^2}{4\phi^2}\tau x_1x_2\right\}\\
&-\frac{\mathrm{i}\eta}{4\phi^{1/2}}(m_1b_1+s_1b_2)(\tau''^*x_2x_3^*-\tau^*x_2''x_3^*+\tau^*x_2x_3''^*+2\tau^*x_2x_3'^*)\\
&-\frac{\mathrm{i}\eta}{8\phi^{1/2}}(m_3b_1+s_3b_2)(\tau''^*x_1x_2-2\tau^*x_1''x_2-2\tau^*x_1'x_2')
\end{aligned}
\tag{33.29}
$$

$$
\begin{aligned}
G(x_1^*,x_2^*,x_3^*) = &\frac{1}{2\phi_i^{1/2}\tau_i'^*v_{1i}^*v_{2i}^*v_{3i}^*}\int_{z_o}^{z_i}\phi^{1/2}\left\{\frac{m_3p_{e1}+s_3p_{e2}}{2\phi}\right.\\
&+\left.\frac{\eta}{\phi^{1/2}}(m_3P_{m1}+s_3P_{m2})\right\}\tau^*x_1^*x_2^*dz
\end{aligned}
\tag{33.30}
$$

The suffixes 1 and 2 added to $a(z)$, $b(z)$ and $p_e(z)$ and $P_m(z)$ indicate the role of the deflection field in the dual-channel system; thus a_1 and p_{e1} are the field functions for the principal electrostatic deflection field while b_1 and P_{m1} are those for the principal magnetic deflection field; these are of course mutually exclusive. The functions a_2, p_{e2}, b_2 and P_{m2} are the corresponding functions for the subfield.

Chu and Munro also give compact formulae for the chromatic aberration coefficients denoted by C_c and C_{T1} in Eq. (33.17). For the most important situation, magnetic main-field deflection and electrostatic subfield deflection, C_T is divided into two parts, contributing $C_T^{(m)}w_i^{(m)}\Delta\phi/\phi_i$ and $C_T^{(e)}w_i^{(e)}\Delta\phi/\phi_i$ to Δw_c. The coefficients are given by

$$
C_c = H(\tau) \quad C_T^{(m)} = H(m) \quad C_T^{(e)} = H(e) \tag{33.31}
$$

in which

$$H(x) = \frac{\phi_i^{1/2}}{\tau'^* v_{1i}} \int_{z_o}^{z_i} \phi^{-1/2} \left\{ -\frac{1}{2}\tau'^* x + \frac{\phi'}{\phi}(\tau'^* x + \tau^* x') - \frac{3}{16}\left(\frac{\phi'}{\phi}\right)^2 \tau^* x - \frac{s_1 a}{4\phi}\tau^* \right\} dz \tag{33.32}$$

A shaped-beam system may be incorporated as explained above. In the more general case of the dual-channel system, the function $H(x)$ becomes

$$H(x) = \frac{\phi_i^{1/2}}{\tau'^* v_{1i}} \int_{z_o}^{z_i} \phi^{-1/2} \left\{ -\frac{1}{2}\tau'^* x_1' + \frac{\phi'}{8\phi}(\tau'^* x_1 + \tau^* x_1') - \frac{3}{16}\left(\frac{\phi'}{\phi}\right)^2 \tau^* x_1 - \frac{\tau^*}{4\phi}(m_1 a_1 + s_1 a_2) \right\} dz \tag{33.33}$$

For a further generalization, see Smith and Munro (1986, 1987).

The formulae for the various deflection aberration coefficients, of which these are a representative sample, are manifestly extremely complicated in appearance and, in cases that are becoming common in practice, very numerous. Programming them with no errors is not a light task though it is possible to output the results of a computer algebra calculation (Chapter 34) in Fortran directly, which eliminates the chance of human error. Moreover, when the list of aberration coefficients is long, it is difficult to estimate their relative importance and one is driven to wonder whether this kind of aberration theory is the most appropriate. An alternative approach may prove to be more favourable.

Fifth-order aberrations of combined focusing and deflection fields have nevertheless been studied by Li (1986), Li et al. (1993) and Uno et al. (1995). The steps in the calculation are set out by these authors in full and are not reproduced here, we merely single out the more important features. An integral expression for the total aberration is obtained, into which the paraxial solutions are substituted. The resulting terms are then classified according to the powers of the quantities analogous to α, γ, U and I (above), here denoted α (angle), γ (position), γ_e and γ_m at the image. The 380 coefficients thus obtained contain many high-order derivatives of the field and potential functions; these are removed by partial integration using a computer algebra language. The coefficients are distributed as follows:

Pure image, rotationally symmetric: AAAA*A* (12 terms)
Pure deflection, rotationally symmetric (59 terms)
Mixed, rotationally symmetric (129 terms)

Pure deflection, fourfold symmetric: $AA^*A^*A^*A^*$ (44 terms)
Pure deflection, fourfold symmetric: AAAAA (20 terms)
Mixed, fourfold symmetric: $AA^*A^*A^*A^*$ (86 terms)
Mixed, fourfold symmetric: AAAAA (30 terms)

in which A may be α, β, γ_e or γ_m. Uno et al. use these calculations to study several practical devices. Not all these coefficients are independent and many relations between them are known. These can most easily be established by representing the aberrations in terms of position and momentum since the symplectic condition can then be applied (Sivkov, 1971; Wollnik and Berz, 1985). The resulting relations have been published by Hosokawa (2002) who retrieves the known relations between third-order coefficients and establishes 21 new relations between the fifth-order aberration ocoefficients. These are then used to confirm the correctness of calculations of these coefficient by the differential algebraic method.

We note that the fifth-order aberrations of purely magnetic deflectors had been examined by Kaashoek (1968) and Rao and Nixon (1981).

PART VI

Computer-Aided Electron Optics

CHAPTER 34

Numerical Calculation of Trajectories, Paraxial Properties and Aberrations

34.1 Introduction

In the preceding chapters, and also in many later paragraphs, various forms of trajectory equations, paraxial quantities and aberration coefficients are presented but their practical evaluation has not yet been discussed, important though this topic is. Attitudes to this question have changed dramatically with the rapid advance of computer technology in the past few decades. When the basic trajectory equations were derived (Störmer, 1907; Busch, 1927) and the aberration theory developed (Glaser, 1933, 1935, 1952; Scherzer, 1936), the resulting differential equations and aberration integrals had to be solved by laborious hand calculations. It is clear that under these circumstances the amount of numerical calculation necessary had to be reduced to the absolute minimum and that analytical calculations were preferred, however sophisticated, whenever this was possible. There is no doubt that these constraints strongly influenced the early development of electron optics.

Fortunately, the accurate solution of systems of ordinary differential equations is nowadays no longer an obstacle, even for comparatively complicated mathematical structures; in Section 34.2 we shall present a numerical procedure that has proved very useful. In combination with advanced techniques for the differentiation of axial fields (see Chapter 13, Field-Interpolation Techniques) and for integration, all the paraxial properties and aberration coefficients of electron optical systems can be calculated with great accuracy.

For highly complex systems like the arrangements of lenses and deflectors in lithography devices or the sequences of multipole lenses needed for aberration correction, the derivation of aberration coefficients and the writing of a computer program to evaluate them are themselves major tasks. The labour involved can be alleviated with the aid of computer algebra systems but without such facilities, when only hand calculation remains, the organization of the corresponding programs really does seem hopeless. A brief account of the types of task that can be performed by computer algebra languages is given in Section 34.9.

Principles of Electron Optics: Basic Geometrical Optics.
DOI: http://dx.doi.org/10.1016/B978-0-08-102256-6.00034-1
© 2018 Elsevier Ltd. All rights reserved.

Another way of determining the effects of aberrations is to plot the endpoints of a large number of accurately calculated electron trajectories, the spot diagrams of light optics. This will be discussed in some detail in Sections 34.5 and 34.6. Although this is not yet very common in electron optics, it does offer a means of assessing the performance of a complex system.

34.2 Numerical Solution of Ordinary Differential Equations

We have seen that the equations describing the motion of electrons can be cast into various different forms. We shall not discuss these again, but simply state that finally, after certain mathematical transformations, they can always be cast into the general form

$$y_i'(x) = f_i(x, y_1(x), \ldots y_N(x)) \quad i = 1 \ldots N \tag{34.1}$$

or in more concise vector notation

$$\boldsymbol{y}'(x) = \boldsymbol{f}(x, \boldsymbol{y}) \tag{34.2}$$

We assume that this transformation has been made before embarking on any attempt to obtain a numerical solution. The variables x and $\boldsymbol{y}$ may have various physical meanings or even none. For instance, the variable x may be the time, the arc-length, the axial coordinate or, indeed, none of these. In the following discussion, the particular meaning is unimportant. For physical reasons, the vector function $\boldsymbol{f}(x, \boldsymbol{y})$ must be smooth in all its arguments and we exclude forthwith any cases in which exceptions such as singularities or discontinuities appear.

The solution of Eq. (34.2) for given initial values

$$\boldsymbol{y}(x_0) = \boldsymbol{y}_0(x_0) \Rightarrow \boldsymbol{y}'(x_0) = \boldsymbol{f}(x_0, \boldsymbol{y}_0) =: \boldsymbol{y}'_0 \tag{34.3}$$

is a standard problem of numerical analysis. In practically every major textbook on this subject, it is dealt with in detail and a subprogram for solving it numerically is available in every computer routine library. The methods employed are, however, of unequal suitability for electron optical applications. We therefore outline the various standard methods very briefly with a few remarks concerning their advantages and disadvantages. Good general references are the collection of surveys edited by Jacobs (1977) and textbooks such as Stoer (1979), Hairer et al. (1993) and Hairer and Wanner (1996).

34.2.1 The Fox–Goodwin–Numerov Method

This method, proposed by Numerov (1923), Manning and Millman (1938) and Fox and Goodwin (1949) and introduced into electron optics by Burfoot (1952) and Jennings and Pratt (1955), is designed for the solution of differential equations of the form

$$y''(x) + f(x)y(x) = g(x)$$

which can be solved directly without conversion into the standard form Eq. (34.2). The method is a two-step procedure, which requires *constant* step-width h in x. Its only advantage is its simplicity for it does have serious drawbacks: lack of flexibility and the restriction to constant step-width together with the lack of any control over the accuracy. Moreover, the procedure is sensitive to rounding errors when a very small step-width is chosen. In view of all these disadvantages, this method has fallen into disuse.

34.2.2 The Runge–Kutta Method

This method, which is free of most of the disadvantages of the Fox–Goodwin–Numerov method, is always needed, even if it is used merely to launch other procedures such as the predictor–corrector method (Section 34.2.3). Here we shall describe it only in its general form, since details are given in any comprehensive work on numerical analysis. In order to simplify the notation we introduce the abbreviations

$$x_n = x_0 + nh, \quad \boldsymbol{y}_n = \boldsymbol{y}(x_n), \quad \boldsymbol{y}'_n = \boldsymbol{y}'(x_n) = \boldsymbol{f}(x_n, \boldsymbol{y}_n) \quad (n \geq 0) \tag{34.4}$$

The method can have different orders m of approximation, all of which take the general form

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \sum_{i=1}^{m} A_i \boldsymbol{K}_i \tag{34.5a}$$

with

$$\boldsymbol{K}_i = h\boldsymbol{f}\left(x_n + c_i h, \ \boldsymbol{y}_n + \sum_{j=1}^{i-1} B_{ij}\boldsymbol{K}_j\right) \quad (i = 1 \ldots m) \tag{34.5b}$$

This is an iterative procedure, in which the preceding evaluations of the function $\boldsymbol{f}$ are used to improve the later ones. The different methods are distinguished by the choice of the parameters. The matrix B is of rank m with vanishing upper tridiagonal part, including the diagonal elements. The scheme of highest order $m = 8$ was published by Fehlberg (1961). Its error is of order h^7 with a very small absolute factor. It is hence suitable to be used as the starting point of other methods. A later paper presents new high-order Runge–Kutta formulae with arbitrarily small truncation errors (Fehlberg, 1966). See also Fehlberg (1960, 1969).

All Runge–Kutta methods have the advantage of full flexibility, which means that the step-size h can be chosen arbitrarily, independent of the preceding steps. This should be exploited to keep the discretization error sufficiently small. There is, however, no reliable intrinsic criterion to achieve this. A straightforward method is the *repetition* of each step after halving the step-size and comparison of the results. The difference between them is a

reliable measure of the error and can be used to control and adjust the step-size h. Moreover the error can be reduced significantly by *extrapolation* of the final vector $\boldsymbol{y_n}$.

This scheme takes then the following form. Let $F(x_a, \boldsymbol{z_a}; x_b, \boldsymbol{z}_b)$ be the function that governs the Runge−Kutta integration from x_a to x_b. The following quantities are then computed:

$$F(x_n, \mathbf{y_n};\ \ x_n + h/2, \boldsymbol{z}_n), \quad F(x_n + h/2, \boldsymbol{z}_n;\ \ x_n + h, \boldsymbol{z}_{n+1}) \tag{34.6a}$$

$$F(x_n, \mathbf{y_n};\ \ x_n + h, \boldsymbol{y}_{n+1}) \tag{34.6b}$$

The difference vector $\boldsymbol{d_{n+1}} = \boldsymbol{z}_{n+1} - \boldsymbol{y}_{n+1}$ provides reliable information about the error of integration and can be used to minimize it by setting

$$\boldsymbol{y_{n+1}} \coloneqq \boldsymbol{z}_{n+1} + \boldsymbol{d}_{n+1}/(2^{m-2} - 1), \quad \boldsymbol{z}_n \coloneqq \boldsymbol{z}_n + \boldsymbol{d}_{n+1}/(2^{m-1} - 1) \tag{34.6c}$$

Here $\coloneqq$ means replacement with the expression on the right-hand side of the equation, as is familiar in computing. For Fehlberg's method with $m = 8$ the corresponding denominators are 63 and 127, respectively.

The accuracy thus achieved is very high. This is important, as unduly large initial errors cause a systematic drift, which can never be compensated later. The method could be used for complete ray tracing, but this would be very laborious since the triple step in Eq. (34.6) requires 3 m evaluations of the function $\boldsymbol{f}(x, \boldsymbol{y})$. If a single evaluation of this function is already very lengthy and time-consuming, the Runge−Kutta method becomes unfavourable. It is therefore wise to consider other integration methods that are more economic. These are the *predictor−corrector methods*, dealt with in the next section.

34.2.3 The Predictor−Corrector Method

The predictor−corrector (PC) method is a multistep procedure; this means that the current vector $\boldsymbol{y}_{n+1}$ is calculated by forming an appropriate linear combination of the preceding ones. Since this can be done in different ways, it automatically provides an accuracy control. We have to distinguish between predictor and corrector formulae. The former have the general structure

$$\boldsymbol{y}_{n+1} = \boldsymbol{P}_{n+1} = \sum_{i=0}^{m} a_i \boldsymbol{y}_{n-i} + h \sum_{k=0}^{m'} a'_k \boldsymbol{y}'_{n-k} \tag{34.7a}$$

The corresponding abscissae must be *equidistant* and $\boldsymbol{y}'_{n+1}$ must not appear on the right-hand side. By means of Eq. (34.7a) an approximate vector $\boldsymbol{y}'_{n+1} = \boldsymbol{f}(x_{n+1}, \boldsymbol{P}_{n+1})$ is obtained. This is then substituted into the corrector formula

$$\boldsymbol{y}_{n+1} = \boldsymbol{C}_{n+1} = \sum_{i=0}^{n} b_i \boldsymbol{y}_{n-i} + h \sum_{k=-1}^{n'} b'_k \boldsymbol{y}'_{n-k} \tag{34.7b}$$

which generally gives a better approximation. The error vector

$$\boldsymbol{D}_{n+1} = \boldsymbol{P}_{n+1} - \boldsymbol{C}_{n+1} \tag{34.7c}$$

represents the discretization error. The magnitude $\Delta_{n+1} = |\boldsymbol{D}_{n+1}|$ can be used as a control and for adjustment of the step-size h. It must remain between certain bounds, which are to be defined by the user and depend on the particular method and the accuracy required. If Δ_{n+1} is smaller than the lower bound, the step-size may be enlarged. If it exceeds the upper bound, the last integration step is considered to be too inaccurate and must hence be repeated with a smaller step-size h. Unfortunately, the method allows essentially only *halving* or *doubling* of h. Doubling merely requires a corresponding reorganization of the preceding data sets, so that only every second one is used. Halving requires sufficiently accurate interpolations at the midpoints of the preceding intervals prior to the reorganization of the data sets. We cannot deal with this matter here. An alternative is to repeat the Runge−Kutta-steps necessary to launch the PC method, but now with the last reliable data set and a new step-size, which may not necessarily be half the old one.

There are numerous versions of the PC method in the literature on numerical analysis, which we shall not consider here. Instead, we shall present a fairly simple and accurate unconventional version. A very accurate corrector was published by Fehlberg (1961), here rewritten as

$$\boldsymbol{Q}_{n+1} = 0.243\,\boldsymbol{y}_n + 0.125\,\boldsymbol{y}_{n-2} + 0.632\,\boldsymbol{y}_{n-5} \tag{34.8a}$$

$$\boldsymbol{C}'_{n+1} = 0.3\,\boldsymbol{y}'_{n+1} + 1.4175\,\boldsymbol{y}'_n + 1.5\,\boldsymbol{y}'_{n-2} + 1.0125\,\boldsymbol{y}'_{n-4} + 0.18\,\boldsymbol{y}'_{n-5} \tag{34.8b}$$

$$\boldsymbol{C}_{n+1} = \boldsymbol{Q}_{n+1} + h\boldsymbol{C}'_{n+1}, \tag{34.8c}$$

$$\boldsymbol{y}_{n+1} = \boldsymbol{C}_{n+1} - 0.004\,h^8\boldsymbol{y}^{(8)}(x_c) \tag{34.8d}$$

Here x_c means an abscissa within the seven intervals. The notation $\boldsymbol{C}'$ does not imply that this is a derivative; it just denotes the linear combination of derivatives, as given in Eq. (34.8b).

This corrector is very accurate and *numerically stable*, which means that small initial errors do not increase (exponentially) but remain well confined. A suitable predictor of eighth order can be found in the following unconventional manner. The vector $\boldsymbol{y}'_{n+1}$ is eliminated from Eq. (34.8a) by the binomial formula for the derivative of seventh order, whereupon we find:

$$\begin{aligned}\boldsymbol{P}'_{n+1} = {} & 3.5175\,\boldsymbol{y}'_n - 6.3\,\boldsymbol{y}'_{n-1} + 12\,\boldsymbol{y}'_{n-2} - 10.5\,\boldsymbol{y}'_{n-3} + 7.3125\,\boldsymbol{y}'_{n-4} \\ & - 1.92\,\boldsymbol{y}'_{n-5} + 0.3\,\boldsymbol{y}'_{n-6}\end{aligned} \tag{34.9a}$$

$$\boldsymbol{P}_{n+1} = \boldsymbol{Q}_{n+1} + h\,\boldsymbol{P}'_{n+1}, \tag{34.9b}$$

$$\boldsymbol{y}_{n+1} = \boldsymbol{P}_{n+1} + 0.296\, h^8 \boldsymbol{y}^{(8)}(x_c) \tag{34.9c}$$

$$\boldsymbol{D}_{n+1} = \boldsymbol{P}_{n+1} - \boldsymbol{C}_{n+1} = h\,(\boldsymbol{P}'_{n+1} - \boldsymbol{C}'_{n+1}) = 0.3 h^8 \boldsymbol{y}^{(8)}(x_c) \tag{34.9d}$$

This difference vector $\boldsymbol{D}_{n+1}$ can not only be used to control the integration error *but also to minimize it.* By comparing Eqs (34.8d) and (34.9c), it is obvious that an improvement is possible by *interpolation* according to

$$\boldsymbol{y}_{n+1} = \boldsymbol{C}_{n+1} + 0.0134\, \boldsymbol{D}_{n+1} \tag{34.10}$$

with a final error of ninth or higher order. The method is numerically stable since the contribution of the predictor to the final result is very small.

A regular step of the procedure thus takes the following form:

Calculate the predictor from (34.8a), (34.9a) and (34.9b)

evaluate $\boldsymbol{y}'_{n+1} = \boldsymbol{f}(x_{n+1}, \boldsymbol{P}_{n+1})$

calculate the corrector from (34.8 a, b, c)

calculate the difference $\boldsymbol{D}_{n+1}$ from (34.9d) and $\boldsymbol{y}_{n+1}$ from (34.10) (34.11)

evaluate $\boldsymbol{y}'_{n+1} = \boldsymbol{f}(x_{n+1}, \boldsymbol{y}_{n+1})$

accuracy control and special additional processing of user data

reorganization of memory for the next step.

The method can be improved by using some of the preceding $\boldsymbol{D}$-vectors for elimination of errors of still higher orders. However, this renders the method very complicated; a comparable gain of accuracy can also be achieved by choosing a slightly smaller step-size. We shall not consider this here, therefore. In order to start the PC method, six *equidistant* Runge–Kutta steps must be calculated in advance.

34.2.4 Special Considerations

The ray equations to be solved in electron optics encounter no mathematical hurdles. They are in principle *regular, nonstiff* equations, so that these methods should work satisfactorily. Problems may arise from the necessary *field calculation*. If the fields are determined by the boundary-element method, several thousands of analytic functions, mainly logarithms and square roots have to be calculated, which is very time-consuming. The predictor–corrector method is then clearly favourable, since it requires only two field evaluations per step.

On the other hand, field calculation by means of the FDM or FEM can be quite fast, so that the computation time becomes unimportant. There, however, we are faced with the problem

of discontinuity at the meshlines, at least for the derivatives of third or higher orders. Of course, the transitions should be made as smooth as possible, but they cannot be perfect. This has an effect on the error control and adjustment of the step-size: an algorithm that automatically halves the step-size if the error exceeds the upper limit may break down. Hence a certain lower limit for the step-size must be admitted. The ray tracing cannot be more accurate than that of the field calculation. The predictor–corrector method outlined above is stable enough, though the continuity of derivatives is clearly violated.

A third aspect is the possibility of incorporating user programs immediately after each successfully calculated integration step. This is advantageous, since it allows the simultaneous calculation of the contributions to aberration integrals, which is of particular importance. Otherwise the whole huge data set would have to be stored and evaluated afterwards.

34.3 Standard Applications in Electron Optics

In very many practical cases, the geometrical forms and the excitations of a lens system are given and hence the electromagnetic field is known, either as a result of calculation or from measurement. It is then the asymptotic or real paraxial properties or a mixed form of these and the corresponding aberration coefficients that are usually required.

34.3.1 Initial-Value Problems

Let us assume that no conditions are imposed at the aperture; discussion of such conditions is deferred to the next section. We are thus confronted with an ordinary initial-value problem. The starting plane may be located in field-free space at a reasonable distance from the lenses or coincide with a real object plane $z = z_o$. We have to choose the appropriate form of the fundamental solutions $v_{1,2}(z)$ of the paraxial ray equations and calculate them for the specific initial values. The integration runs either to a terminal plane in the field-free space on the far side of the lens system or to the conjugate image plane $z = z_i$, which is a zero of the solution $v_2(z)$ for which $v_2(z_o) = 0$. The paraxial ray equations have the general form

$$\frac{d}{dz}(P(z)v_j'(z)) + Q(z)v_j(z) = 0, \quad j = 1, 2 \tag{34.12}$$

Identifying y_1 with v_1 and y_2 with v_2, the corresponding system of first order may be written:

$$\left.\begin{aligned} y_1' &= y_3/P, & y_2' &= y_4/P \\ y_3' &= -Qy_1, & y_4' &= -Qy_2, \end{aligned}\right\} \tag{34.13}$$

The components $y_3 = Pv_1'$, $y_4 = Pv_2'$ are then the associated momenta. An equivalent form is

$$v_j''(z) + a(z)v_j' + b(z)v_j(z) = 0, \quad j = 1, 2 \tag{34.14}$$

Here we set

$$\left.\begin{array}{ll} y_1' = y_3, & y_2' = y_4, \\ y_3' = -ay_3 - by_1 & y_4' = -ay_4 - by_2 \end{array}\right\} \tag{34.15}$$

and $y_3 = v_1'$, $y_4 = v_2'$ are now the derivatives of first order. Both possibilities can likewise be assimilated. In the latter case, the derivatives of second order, $y_3' = v_1''$, $y_4' = v_2''$, are also calculated automatically. On the other hand, the Numerov method requires the Picht transform that converts $P(z)$ to unity or eliminates $a(z)$ to be introduced and does not even give v_1' and v_2'; it is thus distinctly unattractive.

The above transformations remain valid if all the functions involved take complex values, the abscissa z remaining real. A solution could then be found with a program for complex systems of ordinary differential equations but such a program may not be available. It is not even necessary to write it, as the systems (34.13) and (34.15) can easily be split into their real and imaginary parts. The resulting system of real differential equations then has the rank $N = 8$. Very often the coefficients $P(z)$, $Q(z)$ and $a(z)$, $b(z)$, are *real.* There is then no loss of generality in assuming that the solutions $v_1(z)$ and $v_2(z)$, which must, of course, be linearly independent, are likewise real. Arbitrary complex solutions can then be obtained by appropriate linear combinations of these, the coefficients being complex.

The aberration integrals may be of the quite general form

$$C = \int_{z_o}^{z_i} F(z; f, f', \ldots f^{(4)}; v_1, v_1', v_1''; v_2, v_2', v_2'')\, dz \tag{34.16}$$

$f(z)$ being any axial lens function such as the axial potential $\phi(z)$ or the flux density $B(z)$ or even any set of such functions. With the techniques for differentiation and interpolation outlined in Chapter 13, Field-Interpolation Techniques, even derivatives of high order can be calculated accurately, so that it is not necessary to eliminate them by partial integration, though the latter is certainly favourable. At least in the absence of such higher order derivatives, simple integration by means of the trapezoidal rule is quite sufficient.

34.3.2 Boundary-Value Problems

In electron optics, the imposition of aperture conditions is the classic example of a boundary-value problem. We wish to solve Eq. (34.12) or (34.14) for two paraxial rays $s(z)$ and $t(z)$ satisfying

$$s(z_o) = t(z_a) = 1, \quad s(z_a) = t(z_o) = 0 \tag{34.17}$$

This creates no intrinsic problem. First we calculate two other independent solutions $g(z)$, $h(z)$ satisfying the standard initial conditions

$$g(z_o) = h'(z_o) = 1, \quad g'(z_o) = h(z_o) = 0 \tag{34.18}$$

These give us the values $g(z_a)$, $h(z_a)$ in the aperture plane; $s(z)$ and $t(z)$ are now obtained as linear combinations of g and h. Explicit determination of these linear combinations after running the integration routines is unfavourable since all the computed data would have to be stored. Instead, it is better to recommence the integration with the initial conditions

$$\begin{aligned} s(z_o) &= 1, \quad s'(z_o) = -\frac{g(z_a)}{h(z_a)} \\ t(z_o) &= 0, \quad t'(z_o) = -\frac{1}{h(z_a)} \end{aligned} \tag{34.19}$$

While this second integration is proceeding, the appropriate integrals for the aperture-dependent aberration coefficients can be evaluated simultaneously.

34.4 Differential Equations for the Aberrations

Hitherto we have been dealing with the familiar form of the aberration theory and with the practical evaluation of the quantities involved. As we have already observed, this approach to aberration studies becomes very burdensome when the integral expressions for the coefficients are complicated. In these circumstances, we notice that there is no real need to use perturbation theory since the fields and the trajectories through them can be computed completely generally. But then we are confronted with another difficulty: in the most important case of very low aberrations, for which the lens designer is always striving, determination of the coefficients becomes inaccurate because we are obliged to calculate the small differences between the lateral image-plane coordinates, which may be large. This numerical instability would be avoided if the ray equations could be transformed in such a way that small aberrations or shifts appeared directly in an incremental form, without subtraction. This proves to be possible, and we now describe the appropriate transformations for two general classes of trajectory equations.

34.4.1 Electrostatic Systems With a Straight Optic Axis

For systems with a straight optic axis the axial coordinate z is the best parameter and Eq. (3.22) is then a suitable form of the trajectory equation. For conciseness, we specialize to electrostatic systems and introduce $w = x + \mathrm{i}y$ as usual, whereupon Eq. (3.22) reduces to

$$w''(z) = \hat{\Phi}^{-1}(1 + |w'|^2)\left(\frac{\partial \hat{\Phi}}{\partial w^*} - \frac{1}{2}w'\frac{\partial \hat{\Phi}}{\partial z}\right) \tag{34.20}$$

We now write the electrostatic potential in the form

$$\Phi(\boldsymbol{r}) = \phi(z) + V(\boldsymbol{r})$$

$\phi(z)$ being the axial potential and V the remainder. In practice, $|eV| \ll m_0c^2$ is always satisfied so that all terms in εV^2 can be neglected. Eq. (34.20) takes the simpler form

$$w'' = \frac{1 + |w'|^2}{\tilde{\phi}(z) + V}\left\{\frac{\partial V}{\partial w^*} - \frac{1}{2}\left(\phi' + \frac{\partial V}{\partial z}\right)w'\right\} \tag{34.21}$$

$\tilde{\phi} := \phi(1 + \varepsilon\phi)/(1 + 2\varepsilon\phi)$ being the reduced axial potential. The off-axis potential $V(\boldsymbol{r})$ consists of paraxial terms $V_p(\boldsymbol{r})$ and the remainder $V_s(\boldsymbol{r})$, the cause of aberrations. The paraxial term certainly contains the expression $-\frac{1}{4}\phi''(x^2 + y^2) = -\frac{1}{4}\phi'' w w^*$ if $\phi'' \neq 0$, but may also include deflection and quadrupole terms. Whether such terms are main paraxial terms or aberrations caused by imperfections will depend on the definition of the electron optical system; the separation of V into V_p and V_s is to be made according to the particular situation. The term $w'\partial V/\partial z$ is always an aberration, and so the paraxial ray equations (with subscript p) are here

$$\tilde{\phi}w_p'' = \frac{\partial V_p}{\partial w_p^*} - \frac{1}{2}\phi' w_p' =: L(z, w_p, w_p') \tag{34.22}$$

We now rewrite Eq. (34.21) in the form

$$w'' = \frac{1 + |w'|^2}{\tilde{\phi} + V}(L + S)$$

and $w = w_p + w_s$. Introducing Eq. (34.22) into (34.21) and observing that L is a *linear* operator, we find

$$L + S = \underbrace{L(z, w_p, w_p') + L(z, w_s, w_s')}_{L(z,w,w')} + \underbrace{\frac{\partial V_s}{\partial w^*} - \frac{1}{2}w'\frac{\partial V}{\partial z}}_{S(z,w,w')}$$

Note that the separation into w_p and w_s, is only made in the linear term! In order to perform the complete separation into paraxial and aberration terms, we use the identity

$$\frac{1 + |w'|^2}{\tilde{\phi} + V} = \frac{1}{\tilde{\phi}} + \frac{1}{\tilde{\phi} + V}\left(|w'|^2 - \frac{V}{\tilde{\phi}}\right)$$

After a minor calculation we find

$$w_s'' = \frac{1}{\tilde{\phi}}\left\{L(z, w_s, w_s' - L(z, 0, 0)\right\} + \frac{(1 + |w'|^2)S(z, w, w')}{\tilde{\phi} + V} + \frac{1}{\tilde{\phi} + V}\left(|w'|^2 - \frac{V}{\tilde{\phi}}\right)L(z, w, w') \tag{34.23}$$

This differential equation contains exactly *all* aberrations that are included in Eq. (34.21). The term in $L(z, 0, 0)$ arises from the fact that the inhomogeneous term representing the axial deflection is already present in Eq. (34.22) and must therefore not appear again in the aberrations; the term in braces in Eq. (34.23) is hence *homogeneously* linear in the aberrations.

Eq. (34.23) has a comparatively complicated structure and can be solved only in combination with the paraxial equation(34.22). By means of the method outlined in Section 34.2.3, an exact numerical solution can be obtained straightforwardly for various initial conditions. The result is numerically stable as $|w_s| \to 0$ and hence an analysis of the aberrations obtained in this way will lead to no numerical problems. A similar reasoning can be applied to magnetic systems, but we shall not present this here for reasons of space (see Kasper, 1987a,b).

34.4.2 Separation in Arbitrary Systems

A form of the ray equations that has proved very useful in numerical calculation is derived in Section 3.2 where it is given by Eq. (3.12). This is the starting point for the following theory, developed by Kasper (1984, 1985).

In order to remove unnecessary constants, we introduce a normalized magnetic field

$$\boldsymbol{b}(\boldsymbol{r}) := \eta \hat{U}^{-1/2}\boldsymbol{B}(\boldsymbol{r}) \tag{34.24}$$

having the dimension of a reciprocal length, where $\hat{U}$ is a constant accelerating potential. We likewise define a dimensionless electrostatic potential $\varphi(\boldsymbol{r})$ and its gradient $\boldsymbol{a}(\boldsymbol{r})$ by

$$\varphi(\boldsymbol{r}) := \frac{\hat{\Phi}(\boldsymbol{r})}{2\hat{U}}, \quad \boldsymbol{a} = \nabla\varphi \tag{34.25}$$

Denoting derivatives with respect to the curve parameter τ by dots, we can rewrite (3.12) more concisely as

$$\ddot{\boldsymbol{r}} = \boldsymbol{a}(\boldsymbol{r}) + \boldsymbol{b}(\boldsymbol{r}) \times \dot{\boldsymbol{r}} \tag{34.26}$$

Let us now consider a neighbouring ray shifted by a distance $\boldsymbol{s}(\tau)$ relative to the first. This must satisfy the equation

$$\ddot{\boldsymbol{r}} + \ddot{\boldsymbol{s}} = \boldsymbol{a}(\boldsymbol{r} + \boldsymbol{s}) + \boldsymbol{b}(\boldsymbol{r} + \boldsymbol{s}) \times (\dot{\boldsymbol{r}} + \dot{\boldsymbol{s}})$$

By subtraction of Eq. (34.26) from this equation we find

$$\ddot{\boldsymbol{s}} = \boldsymbol{a}(\boldsymbol{r}+\boldsymbol{s}) - \boldsymbol{a}(\boldsymbol{r}) + \boldsymbol{b}(\boldsymbol{r}+\boldsymbol{s}) \times \dot{\boldsymbol{s}} + \{\boldsymbol{b}(\boldsymbol{r}+\boldsymbol{s}) - \boldsymbol{b}(\boldsymbol{r})\} \times \dot{\boldsymbol{r}} \tag{34.27}$$

This is not very helpful unless we can get rid of the differences between the field strengths. These can, in fact, be eliminated by means of a program that not only furnishes the Cartesian components of the field strengths but also the gradients of these components, or in other words, the second-order derivatives of the potentials. For any scalar differentiable function, the relation

$$\Delta F := F(\boldsymbol{r}+\boldsymbol{s}) - F(\boldsymbol{r}) = \int_0^1 (\boldsymbol{s}\cdot\nabla')F(\boldsymbol{r}'|\boldsymbol{r}' = \boldsymbol{r}+\boldsymbol{s}t)dt$$

is an identity, in which the differentiation refers to a variable argument $\boldsymbol{r}'$, over which the integration is to be performed after the differentiation. This integration is to be carried out by means of Gauss quadrature formulae. With the abbreviation

$$DF := s\cdot\nabla F = (s_x\partial_x + s_y\partial_y + s_z\partial_z)F \tag{34.28}$$

we have

$$\Delta F = \frac{1}{2}DF(\boldsymbol{r}+a_1\boldsymbol{s}) + \frac{1}{2}DF(\boldsymbol{r}+a_2\boldsymbol{s}) + O(\boldsymbol{s})^5$$
$$\text{with} \quad a_{1,2} = \frac{1}{2} \pm \frac{\sqrt{2}}{4} \tag{34.29a}$$

$$\Delta F = \frac{1}{18}\left\{5DF(\boldsymbol{r}+b_1\boldsymbol{s}) + 8DF\left(\boldsymbol{r}+\frac{\boldsymbol{s}}{2}\right) + 5DF(\boldsymbol{r}+b_2\boldsymbol{s})\right\} + O(\boldsymbol{s}^7)$$
$$\text{with} \quad b_{1,2} = \frac{1}{2} \pm \sqrt{\frac{3}{20}} \tag{34.29b}$$

In this way we can calculate increments of functions quite accurately in a numerically stable manner. In the case of vector functions these operations are carried out separately for the three Cartesian components. With this in mind, we rewrite Eq. (34.27) in the form

$$\ddot{\boldsymbol{s}} = \Delta a(\boldsymbol{r}) + b(\boldsymbol{r}) \times \dot{\boldsymbol{s}} + \Delta b(\boldsymbol{r}) \times (\dot{\boldsymbol{r}} + \dot{\boldsymbol{s}}) \tag{34.30}$$

which is quite generally valid and easily programmable (Kasper, 1985).

So far we have tacitly assumed chromatic aberrations to be absent, since we treated $\boldsymbol{b}(\boldsymbol{r})$ in Eq. (34.24) and $\boldsymbol{a}(\boldsymbol{r})$ in (34.25) as unique functions. We now specify explicitly that Eqs (34.24), (34.25), (34.26) and (34.30) are true for electrons with the *nominal* value

of the kinetic starting energy and for magnetic fields with the nominal values of the coil currents. In accordance with Eq. (3.13), the relations

$$\frac{1}{2}\dot{\boldsymbol{r}}^2 = \varphi(\boldsymbol{r}), \quad \dot{\boldsymbol{r}}\cdot\dot{\boldsymbol{s}} + \frac{1}{2}\dot{\boldsymbol{s}}^2 = \Delta\varphi(\boldsymbol{r}) \tag{34.31}$$

must then be satisfied at each trajectory point (including the starting point) if the energy has its nominal value.

34.4.3 Chromatic Shifts

We now study the effect of altering φ and $\boldsymbol{b}$. If the magnetic field is generated by a coil with current I and the lens is not saturated, an alteration δI changes the magnetic field by $\delta\boldsymbol{b} = I^{-1}\boldsymbol{b}\delta I$. A nonzero kinetic starting energy $\varepsilon\delta\Phi$ at the surface $\Phi(\boldsymbol{r}) = 0$ alters the function $\varphi(\boldsymbol{r})$ by

$$\delta\phi(\boldsymbol{r}) = (2\hat{U})^{-1}(1 + 2\varepsilon\Phi(\boldsymbol{r}) + \varepsilon\delta\Phi)\delta\Phi \tag{34.32}$$

To prevent any confusion, we denote the shift, caused by chromatic and geometric effects together, by the symbol $\boldsymbol{u}$ instead of $\boldsymbol{s}$. The generalization of Eq. (34.31) with $\Delta\varphi \coloneqq \varphi(\boldsymbol{r}+\boldsymbol{u}) - \varphi(\boldsymbol{r})$ is

$$\dot{\boldsymbol{r}}\cdot\dot{\boldsymbol{u}} + \frac{1}{2}\dot{\boldsymbol{u}}^2 = \Delta\varphi(\boldsymbol{r}) + \delta\varphi(\boldsymbol{r}+\boldsymbol{u}) \tag{34.33}$$

This condition is to be satisfied only once, at the *starting point*, and will then be valid for the whole trajectory. The practical application of Eqs (34.31) and (34.33) proceeds as follows: the starting vectors $\boldsymbol{r}_0$, $\boldsymbol{s}_0$ and $\boldsymbol{u}_0$ and the *directions* of $\dot{\boldsymbol{r}}_0$, $\boldsymbol{s}_0$ and $\dot{\boldsymbol{u}}_0$ can be chosen independently, after which Eqs (34.31) and (34.33) are used to determine the appropriate *lengths* of the vectors $\dot{\boldsymbol{r}}_0$, $\boldsymbol{s}_0$ and $\dot{\boldsymbol{u}}_0$, respectively. The alterations δI and $\delta\Phi$ introduce some additional terms in Eq. (34.30). Considering all possible increments more thoroughly, we find

$$\begin{aligned}\ddot{\boldsymbol{u}} = {} & \Delta\boldsymbol{a}(\boldsymbol{r}) + \boldsymbol{b}(\boldsymbol{r})\times\dot{\boldsymbol{u}} + \left(\Delta\boldsymbol{b}(\boldsymbol{r}) + \frac{\delta I}{I}\boldsymbol{b}(\boldsymbol{r}+\boldsymbol{u})\right)\times(\dot{\boldsymbol{r}}+\dot{\boldsymbol{u}}) \\ & + \varepsilon\delta\Phi\hat{\boldsymbol{U}}^{-1}\nabla\Phi(\boldsymbol{r}+\boldsymbol{u})\end{aligned} \tag{34.34}$$

This differential equation contains all possible types of geometric and chromatic errors and all allowed combinations of them in full generality. The integration formula (34.29b) is already so accurate that its remainder can be neglected in every practical case. Eq. (34.34) shows that the shift between two arbitrary neighbouring trajectories can be computed quite accurately in a numerically stable manner. In practice some simplifications can be made. In Eq. (34.34) it is inconvenient to have $\nabla\Phi$ and $\nabla\hat{\Phi}$ (in $\boldsymbol{a}$) together in the same formula.

Since $\varepsilon\delta\Phi \sim 10^{-6}$ little error will result from replacing $\nabla\Phi$ by $\nabla\hat{\Phi}$ in the last term of Eq. (34.34). Furthermore, it is preferable to avoid explicit evaluation for the argument $\boldsymbol{r}+\boldsymbol{u}$ or $\boldsymbol{r}+\boldsymbol{s}$, since the latter does not appear in Eqs (34.29a,b). Thus a more favourable form (see Kasper, 1985) is

$$\begin{aligned}\ddot{\boldsymbol{u}} &= \Delta\boldsymbol{a}(\boldsymbol{r}) + \kappa_1\{\boldsymbol{a}(\boldsymbol{r}) + \Delta\boldsymbol{a}(\boldsymbol{r})\} + \boldsymbol{b}(\boldsymbol{r}) \times \dot{\boldsymbol{u}} \\ &\quad + \left\{\Delta\boldsymbol{b}(\boldsymbol{r}) + \kappa_2(\boldsymbol{b}(\boldsymbol{r}) + \Delta\boldsymbol{b}(\boldsymbol{r}))\right\} \times (\dot{\boldsymbol{r}} + \dot{\boldsymbol{u}}) \\ \text{with } \kappa_1 &= 2\varepsilon\delta\Phi \equiv \frac{e\delta\Phi}{m_0c^2}, \quad \kappa_2 = \frac{\delta I}{I}\end{aligned} \tag{34.35}$$

Comparing the formalism outlined here with that of Section 34.4.1, we notice an important difference. Eq. (34.33) is already the differential equation for the lateral geometric aberration: its solution gives the required aberration immediately without further transformations. The price to be paid for this convenience is the necessary specialization. On the other hand, Eq. (34.35) is quite *generally* applicable to any electron optical system with stationary fields. This equation describes, however, not the aberrations themselves but the shift between neighbouring trajectories, from which the aberrations must then be determined. This formalism is most useful for systems with a curved optic axis, since this axis can simply be adopted as the reference solution of (34.26).

34.5 Least-Squares-Fit Methods in Electron Optics

The methods discussed in the preceding sections enable us to compute individual geometric and chromatic aberrations with high accuracy, even in the most complicated cases. The question now arises, how can the corresponding aberration coefficients be calculated from a set of such data? One suitable procedure is the least-squares-fit (LSF) method.

This method is already familiar in physics and numerical mathematics, since it is a general tool for the analysis and approximation of measured or calculated data. In light optics, for instance, the LSF method is in practical use for the determination of aberration coefficients, since this is easier than the numerical evaluation of Seidel's aberration theory. In spite of the close analogy with our present concerns, the LSF method was not immediately adopted in electron optics, for the following reasons. The determination of aberration coefficients by means of the LSF method requires the calculation of the endpoints of many rays (not less than 100). Only recently has the time taken to trace electron trajectories become trivial. It is routinely used in such commercial programs as EOD (Lencová, 2004c, 2005, 2007, 2008, 2010; Lencová and Oral, 2006; Lencová and Zlámal, 2005, 2006, 2007, 2008; Lencová et al., 2004; Oral and Lencová, 2003, 2009). This method in undoubtedly very useful and sometimes there is no alternative. This will be demonstrated by examining some telling examples once we have presented the general theory.

34.5.1 General Complex Formulation

We disregard for the moment the particular purpose and study a more general problem. We assume here that M measurements or computations of some complex function $w(u)$ yield the complex numbers $w_1, w_2 \ldots w_M$, that is $w_\mu = w(u_\mu)$, $\mu = 1 \ldots M$. We now wish to expand the function $w(u)$ as a series of the form

$$w(u) = \sum_{\nu=1}^{N} c_\nu \psi_\nu(u) \tag{34.36}$$

in terms of a set of well-defined functions $\psi_\nu(u)$ ($\nu = 1 \ldots N$), the *trial functions*, and initially unknown coefficients c_ν; of course $N \leq M$. Since the values $w_1 \ldots w_M$ may be afflicted with small errors or the choice of the trial functions may not be entirely appropriate, it is often impossible to satisfy Eq. (34.36) for all w_μ. We hence introduce the less stringent condition

$$\sum_{\mu=1}^{M} G_\mu \left| w_\mu - \sum_{\nu=1}^{N} c_\nu \psi_\nu(u_\mu) \right|^2 =: \varepsilon^2 = \min \tag{34.37}$$

$G_1 \ldots G_M$ here being positive *weight factors* normalized to a unit sum. Minimization of ε^2 with respect to the coefficients c_λ,

$$\frac{\partial \varepsilon^2}{\partial c_\lambda^*} = 0, \quad \lambda = 1 \ldots N$$

leads immediately to the *normal equations*:

$$\sum_{\nu=1}^{N} S_{\lambda\nu} c_\nu = T_\lambda, \quad \lambda = 1 \ldots N \tag{34.38}$$

with the Hermitian matrix

$$S_{\lambda\nu} = \sum_{\mu=1}^{M} G_\mu \psi_\lambda^*(u_\mu) \psi_\nu(u_\mu) \tag{34.39}$$

and the column vector

$$T_\lambda = \sum_{\mu=1}^{M} G_\mu \psi_\lambda^*(u_\mu) w_\mu, \quad \lambda = 1 \ldots N \tag{34.40}$$

The normal equations have the advantage of being easy to program and the flexibility of including weight factors, which can be chosen according to the significance of the input

data $w_1 \ldots w_M$. If no preferences are apparent, then $G_1 = G_2 = \ldots G_M = 1/M$ is appropriate. Substituting the solution back into Eq. (34.37) gives us the standard deviation ε, which is a good measure of the quality of the approximation.

The main objection to the normal equations is that the matrix S is often ill-conditioned. In order to avoid this, Householder (1964) introduced a new method for the determination of $c_1 \ldots c_\mu$ from Eq. (34.37) by means of suitable orthogonalizations. This method requires equal weights $G_1 = \ldots = G_M$. Eq. (34.37) can then be interpreted as the norm of a vector in an M-dimensional complex space; this norm is invariant under any *unitary* transformation. We cannot go into details here but simply state the essential procedure. The user of a Householder-transformation program is required to solve

$$\left| \left(\begin{array}{ccc|c} A_{11} & \ldots & A_{1N} & w_1 \\ \ldots & \ldots & \ldots & \ldots \\ A_{M1} & \ldots & A_{MN} & w_M \end{array} \right) \left(\begin{array}{c} c_1 \\ \vdots \\ c_N \\ -1 \end{array} \right) \right|^2 = \min \tag{34.41}$$

with $A_{\mu\nu} = \psi_\nu(u_\mu)$. After setting up this matrix including the column for the w_μ, it is left-multiplied by a succession of suitable unitary matrices of rank M until a triangular structure is obtained: $A'_{\mu\nu} = 0$ for $\mu > \nu$ with $\nu \leq N$. Extraction of the coefficients $c_1, \ldots, c_N$ is then straightforward. For more details, we refer to the standard textbooks on numerical mathematics (e.g., Freund and Hoppe, 2007, 2011; Quarteroni et al., 2007; Hämmerlin and Hoffmann, 1991).

34.5.2 The Determination of Deflection Aberrations

As an example, we now study the third- and fifth-order distortions of deflection systems, defined in Eq. (33.12). Since all coordinates refer to the image plane, we drop the subscript *i*. Inspection of Eq. (33.12) shows that in this particular case we have 10 real coefficients D_{jk} ($j = 1, 2$; $k = 1 \ldots 5$) and 10 real trial functions

$$\psi_1 \cdots \psi_{10} = x^3, x^2y, xy^2, y^3; x^5, x^4y, x^3y^2, x^2y^3, xy^4, y^5$$

We therefore have to solve a linear system of 10 real equations for the unknown coefficients. The real form of the LSF method is a simpler special case of the general complex form.

The system of 10 equations can be solved *en bloc*. This is, however, the least efficient way; we use this example to demonstrate how such a system can be profitably split into smaller subsystems. Careful inspection of Eq. (33.12) shows that the two equations have no common coefficients and can hence be treated separately. Thus the full system is already partitioned into two uncoupled subsystems of rank 5. We next consider the fact that the

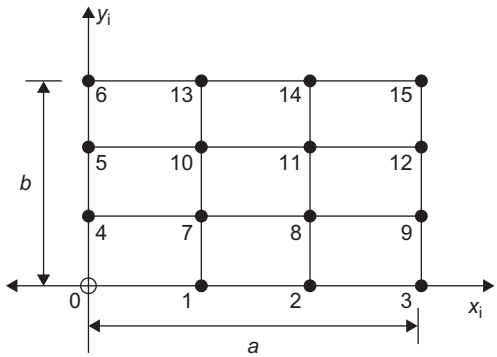


Figure 34.1
Pattern of ideal deflection in one quarter of a viewing screen. The centre of the screen coincides with the origin, *O*.

distortions are antisymmetric with respect to the coordinate planes; this tells us that it is sufficient to calculate trajectories with endpoints in one quarter of the image screen. In order to obtain a true LSF, which will enable us to verify that Eq. (33.12) is a valid representation, it is advantageous to use more than the minimum number of trajectories; we therefore choose the pattern of ideal deflections shown in Fig. 34.1. As it is undesirable to raise large numbers to high powers, we *normalize* the equations by introducing dimensionless coordinates:

$$\xi = \frac{x}{a}, \quad \eta = \frac{y}{b}, \quad -\Delta\xi = \frac{\Delta x}{a}, \quad -\Delta\eta = \frac{\Delta y}{b},$$

$$0 \le \xi \le 1, \quad 0 \le \eta \le 1$$

We then have subsystems with the following aberration components, trial functions and scaled coefficients:

1. $\eta = 0, \Delta\xi_1, \Delta\xi_2, \Delta\xi_3, \psi_1 = \xi^3, \psi_2 = \xi_5,$
   $c_1 = a^2 D_{11}, c_2 = a^4 D_{13}.$
2. $\xi = 0, \Delta\eta_4, \Delta\eta_5, \Delta\eta_6, \psi_3 = \eta^3, \psi_4 = \eta^5,$
   $c_3 = b^2 D_{22}, c_4 = b^4 D_{25}.$
3. $\eta\xi \neq 0, \Delta\eta_7, \ldots \Delta\xi_{15}, \psi_5 = \xi\eta^2, \psi_6 = \xi^3\eta^2, \psi_7 = \xi\eta^4,$
   $c_5 = b^2 D_{12}, c_6 = a^2 b^2 D_{14}, c_7 = b^4 D_{15}$
4. $\xi\eta \neq 0, \Delta\eta_7, \ldots \Delta\eta_{15}, \psi_8 = \eta\xi^2, \psi_9 = \eta^3\xi^2, \psi_{10} = \eta\xi^4,$
   $c_8 = a^2 D_{21}, c_9 = a^2 b^2 D_{24}, c_{10} = a^4 D_{23}.$

The LSF equations for subsystems 1 and 2 are solved first and the results obtained are then introduced into the remaining equations. We shall not go into this elementary procedure. It is clear that such a calculation presents no particular difficulty whereas the second iteration of the perturbation calculation (evaluation of S^{II} in Chapter 22, Perturbation Theory: General Formalism) is extremely complicated, so complicated indeed that this second-order theory was hardly ever used except in connection with the Darmstadt aberration-correction project (Chapter 41 of Volume 2) until aberration correction became a reality.

34.5.3 Some Other Examples

Another case in which the LSF method is obligatory is the *filter lens.* These are quite ordinary electrostatic einzel lenses (see Chapter 35 of Volume 2), but the potential of the central electrode is so low that electrons with an energy of about 5−10 eV below the nominal energy cannot pass the central potential wall and are reflected, as in an electrostatic mirror. The electrons that do pass over the wall are slowed down so much that the associated slopes and off-axis distances may be very large; Fig. 34.2 shows a typical example. It is clear that the third-order approximation for the aberrations is then quite insufficient. Niemitz (1980), who investigated such lenses numerically, considered geometric aberrations of at least fifth order and chromatic errors up to the third order in the energy loss. The whole system of LSF equations then becomes so large that it is essential to split it into subsystems and to introduce suitable scaling.

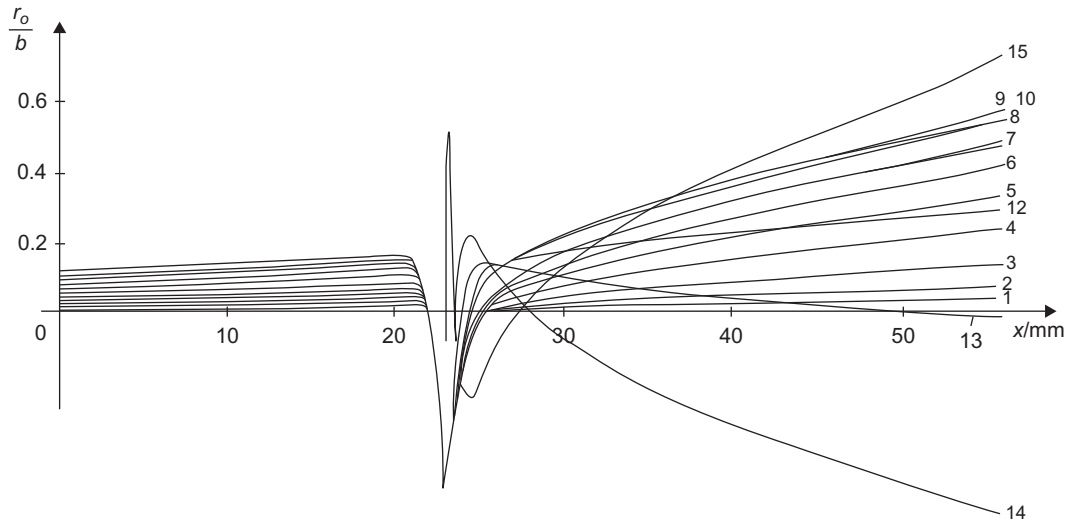


Figure 34.2
Parallel beam of electrons incident on a filter lens; b denotes the bore radius of the central electrode. Note that the scales on the axes are different. *Courtesy of P. Niemitz (1980).*

An interesting application of the complex LSF method concerns systems of round magnetic lenses and deflectors. Here the complex trial functions are given by Eq. (33.23). In order to simplify the program, it may well be preferable to establish nine independent complex coefficients, treating C_0 and F_0 as though they were complex ($\tilde{C}_0$ and $\tilde{F}_0$) and regarding $\tilde{K}_0^*$ as independent of $\tilde{K}_0$ (writing $\tilde{K}_c$ for $\tilde{K}_0^*$ in the defining relation). If the numerical procedure is sound, it will be found that $|\Im\tilde{C}_0| \ll C_0$, $|\Im\tilde{F}_0| \ll F_0$ and $\tilde{K}_c = \tilde{K}_0^*$ with a high degree of accuracy. This also provides a useful check. Generally, the paraxial properties and the spherical aberration can be separated from the rest of the errors; the distortions also form a separable subsystem. Once again, it is possible to study aberrations of higher than third order. A very detailed account of the calculation of aberration coeficients by ray tracing is given by Oral and Lencová (2009). See also Mynář et al. (2000) and Oral and Lencová (2005).

34.6 Determination and Evaluation of Aberration Discs

A complete determination of all permitted aberrations of a very complex electron optical system is a major task, regardless of the choice of method. We are therefore led to seek simpler special classes of aberrations, which can be determined with a more modest effort. One such class consists of the aberrations associated with a pencil of rays that start from a common fixed object point and pass through an aperture (see Fig. 34.3). Neither the object point nor the centre of the aperture need be situated on the optic axis.

As a result of the lens aberrations, a blurred intensity pattern is formed in the image plane, the aberration disc. This is of extreme interest in many practical respects. For monochromatic electrons, this disc can be interpreted as the shadow projection cast by a grid placed over the aperture. This is shown in Fig. 34.4. If we assume that the grid is illuminated uniformly and that the meshes of the grid in the aperture have equal areas, then the intensities in the distorted areas will also be equal. In this way we obtain some idea of the *intensity distribution* in the aberration disc. The aberrations may become so large that parts of the disc overlap others; in overlapping areas the corresponding intensities are to be summed. An example of a calculation of the distribution of points of arrival of 2000 million trajectories is shown by Oral (2010).

The determination of patterns such as those shown in Fig. 34.4B is quite easy, if the space between the object and the aperture is field-free, so that the relations between the object coordinates (x_o, y_o, z_o), the initial slopes (x'_o, y'_o) and the coordinates (x_a, y_a, z_a) in the aperture plane are simple. In all other cases, the calculation of the ray that starts from the given object point and strikes a particular aperture point while under the influence of aberrations is already a difficult procedure, since a nonlinear boundary-value problem has to be solved. In order to circumvent such an impractical procedure, we propose that first, the corresponding *paraxial* boundary-value problem should be solved. This tells us the

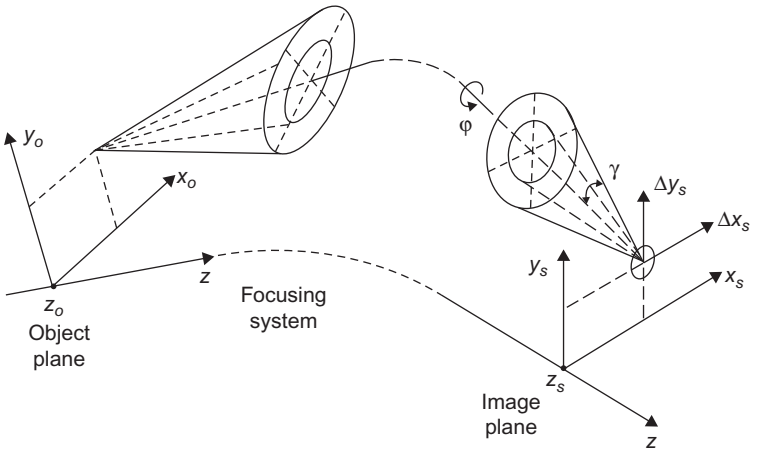


Figure 34.3
Perspective sketch of the coordinate system and of a focused electron beam. The aperture cone in image space has a circular cross-section. Note that the aperture angles are much exaggerated. After Kasper (1985).

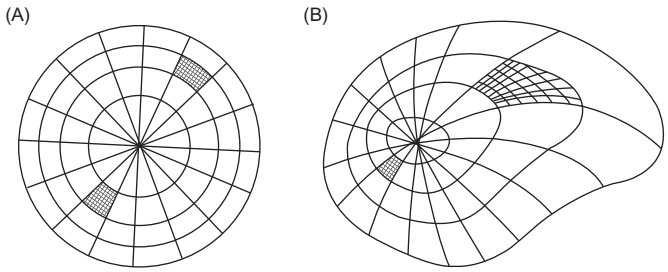


Figure 34.4
(A) Zone pattern of an aperture. (B) Aberration disc regarded as a shadow projection of the zone structure created by monoenergetic electrons from a single object point.

required initial slopes x'_o, y'_o. With these, the ray affected by aberrations is now calculated. The latter will *not*, of course, pass exactly through the aperture point (x_a, y_a, z_a), but if the deviation is small, as it should be in every well-designed system, the disparity will be tolerable.

34.6.1 Fourier Analysis of the Aberrations

In very many practical devices, the image plane or the viewing screen is located in essentially field-free space. The rays are then orthogonal trajectories of a characteristic function. Since this is already true for the paraxial rays, the same must hold for the aberrations after the paraxial contributions have been separated. If we have an aperture at a

distance d from the image plane in the field-free space in front of the latter, then the gradient relations can be cast into the form

$$\Delta x_i = -d\frac{\partial S}{\partial x_i}, \quad \Delta y_i = -d\frac{\partial S}{\partial y_i}, \quad (x_i - x_a)^2 + (y_i + y_a)^2 \ll d^2$$

We have $x_i' = (x_\mathrm{i} - x_a)/d$, $y_i' = (y_\mathrm{i} - y_a)/d$ and hence

$$\Delta x_i = \frac{\partial S}{\partial x_i'}, \quad \Delta y_i = \frac{\partial S}{\partial y_i'}$$

so that this representation is independent of the particular distance between the aperture and the image plane.

As we have seen above, it is necessary to approximate the slopes x_i', y_i' by their *paraxial* values, since only for these can the simple prescribed conditions be satisfied. In accordance with the definitions introduced in Section 32.3, we now write

$$\left.\begin{aligned} x_i' &= x_i'^{(0)} + \sigma\cos\varphi = x_i'^{(0)} + \alpha \\ y_i' &= y_i'^{(0)} + \sigma\sin\varphi = y_i'^{(0)} + \beta \end{aligned}\right\} \tag{34.42}$$

for the paraxial slopes, where the subscript a has been dropped. As in Eqs (32.40) and (32.46), we have

$$\Delta w = \Delta x_i + \mathrm{i}\Delta y_\mathrm{i} = \left(\frac{\partial}{\partial\alpha} + \mathrm{i}\frac{\partial}{\partial\beta}\right)S = \mathrm{e}^{\mathrm{i}\varphi}\left(\frac{\partial}{\partial\sigma} + \frac{\mathrm{i}}{\sigma}\frac{\partial}{\partial\varphi}\right)S \tag{34.43}$$

Since this representation is independent of the distance d to the assumed aperture, we may, in theory at least, consider an arbitrarily close aperture. In the limit $d \to 0$, Eq. (34.43) must hence remain valid even if the image plane is located in an electric field, as is the case for the post-acceleration fields employed in cathode-ray tubes.

For conciseness we first define a normalized radial coordinate $r := \sigma/a$, a being the maximum semiaperture angle, hence $0 \le r \le 1$. Following Kasper (1985) closely, we now introduce a Fourier series expansion with initially unknown coefficients, which are to be determined from the computed aberrations:

$$\begin{gathered} S(r,\varphi) = \sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\frac{r^{m+2k}}{2(m+k)}\left(A_{mk}^{*}\mathrm{e}^{\mathrm{i}m\varphi} + A_{mk}\mathrm{e}^{-\mathrm{i}m\varphi}\right) \\ (m + k > 0, \quad A_{0k} \quad \text{real}) \end{gathered} \tag{34.44}$$

Differentiation gives

$$\Delta w = \sum_m\sum_k r^{m+2k-1}\left\{\frac{k}{m+k}A_{mk}^{*}\mathrm{e}^{\mathrm{i}(m+1)\varphi} + A_{mk}\mathrm{e}^{-\mathrm{i}(m-1)\varphi}\right\} \tag{34.45}$$

All the coefficients for $m \leq 4$, $m + 2k \leq 5$ can be determined from a few trajectories. The Fourier series expansion (34.45) then takes the explicit form

$$\begin{aligned}\Delta w = {} & \mathrm{e}^{-3\mathrm{i}\varphi} \cdot r^3 A_{40} + \mathrm{e}^{-2\mathrm{i}\varphi}(r^2 A_{30} + r^4 A_{31}) \\ & + \mathrm{e}^{-\mathrm{i}\varphi}(r A_{20} + r^3 A_{12}) + A_{10} + r^2 A_{11} + r^4 A_{12} \\ & + 2\mathrm{e}^{\mathrm{i}\varphi}(r A_{01} + r^3 A_{02}) + \frac{1}{2}\mathrm{e}^{2\mathrm{i}\varphi}\left(r^2 A_{11}^{*} + \frac{4}{3} r^4 A_{12}^{*}\right) \\ & + \frac{1}{3}\mathrm{e}^{3\mathrm{i}\varphi} r^3 A_{21}^{*} + \frac{1}{4}\mathrm{e}^{4\mathrm{i}\varphi} r^4 A_{31}^{*}\end{aligned} \tag{34.46}$$

This can be rewritten more concisely in the form

$$\Delta w =: u_A(r, \varphi) = \sum_{n=-N+1}^{N} \rho_n(r)\mathrm{e}^{\mathrm{i}n\varphi}, \quad (N = 4) \tag{34.47}$$

We now try to approximate an arbitrary complex function $\Delta w = u(r, \varphi)$, defined on a disc $r \leq 1$, by a series expansion of the form (34.47). Since every part of the disc is *a priori* of equal importance, this approximation is best made by an integral least-squares-fit

$$F := \int_{r=0}^{1} \int_{\varphi=0}^{2\pi} |u(r, \varphi) - u_A(r, \varphi)|^2 r \, dr \, d\varphi = \min \tag{34.48}$$

Quite generally this integration is to be approximated by a suitable discrete summation. With $2N$ equidistant azimuths $\varphi_j = (j - 1)\pi/N$, $j = 1 \ldots 2N$ and $L \geq 2$ different radii $r_1 \ldots r_L$, such a summation formula has the basic form

$$F := \int_{r=0}^{1} \int_{\varphi=0}^{2\pi} f(r, \varphi) r \, dr \, d\varphi = \frac{\pi}{2N} \sum_{l=1}^{L} \sum_{j=1}^{2N} G_l f_{lj}$$

with discrete function values $f_{lj} := f(r_l, \varphi_j)$ and positive weight factors G_l, normalized to unit sum.

Introducing Eq. (34.47) into (34.48) and employing this summation formula, we obtain

$$F = \frac{\pi}{2N} \sum_{l=1}^{L} \sum_{j=1}^{2N} G_l |u_{lj}|^2 + \pi \sum_{l=1}^{L} \sum_{n=1-N}^{N} G_l \left\{ |\rho_{nl}|^2 + 2\Re(\rho_{nl} K_{nl}^{*}) \right\} = \min$$

with $u_{lj} := u(r_l, \varphi_j)$ and $\rho_{nl} := \rho_n(r_l)$ and the trigonometric sums

$$\begin{aligned} K_{ln} := {} & \frac{1}{2N} \sum_{j=1}^{2N} u_{lj} \exp\left\{ \frac{-\mathrm{i}\pi(j-1)n}{N} \right\} \\ & (n = -(N-1) \ldots N) \end{aligned} \tag{34.49}$$

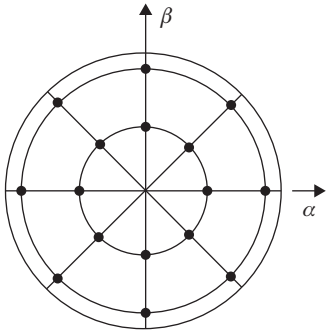


Figure 34.5
Relative directional angles, α and β, of the selected trajectories near the image plane; the outer circle indicates the full extent of the electron beam.

We choose L and N as small as possible but without significant loss of accuracy. A reasonable compromise is $N = 4$ and $L = 2$, which suffices for the determination of all the coefficients in Eq. (34.46). The discretization of the radial integration is then simply the two-point Gauss quadrature for the variable r^2; we hence have $G_1 = G_2 = 1/2$ and $r_{1,2}^2 = 0.5(1 \pm 1/\sqrt{3})$,

$$r_{1,2} = \left(\frac{1}{2} \pm \frac{1}{\sqrt{12}}\right)^{1/2} = 0.4597, \quad 0.8881$$

The corresponding points in the aperture disc are shown in Fig. 34.5. The minimizing condition is now $\partial F/\partial A_{\mu\nu}^* = 0$ for all values of the subscripts μ and ν. With the convenient abbreviation $[X] := \Sigma X_l$ for the sum of any set of quantities X, we arrive at a complex linear system of equations for the unknowns $A_{\mu\nu}$:

$$\left.\begin{aligned} [r^2]A_{01} + [r^4]A_{02} &= \frac{1}{2}[rK_1] \\ [r^4]A_{01} + [r^6]A_{02} &= \frac{1}{2}[r^3K_1] \end{aligned}\right\} \tag{34.50a}$$

$$\left.\begin{aligned} [r^0]A_{10} + [r^2]A_{11} + [r^4]A_{12} &= [K_0] \\ [r^2]A_{10} + \frac{5}{4}[r^4]A_{11} + \frac{4}{3}[r^6]A_{12} &= \left[r^2\left(K_0 + \frac{1}{2}K_2^*\right)\right] \\ [r^4]A_{10} + \frac{4}{3}[r^6]A_{11} + \frac{13}{9}[r^8]A_{12} &= \left[r^4\left(K_0 + \frac{2}{3}K_2^*\right)\right] \end{aligned}\right\} \tag{34.50b}$$

$$\left.\begin{aligned} [r^2]A_{20} + [r^4]A_{21} &= [rK_{-1}] \\ [r^4]A_{20} + \frac{10}{9}[r^6]A_{21} &= \left[r^3\left(K_{-1} + \frac{1}{3}K_3^*\right)\right] \end{aligned}\right\} \tag{34.50c}$$

$$\left.\begin{aligned}[r^4]A_{30} + [r^6]A_{31} &= [r^2K_{-2}]\\ [r^6]A_{30} + \frac{17}{16}[r^8]A_{31} &= \left[r^4\left(K_{-2} + \frac{1}{4}K_4^*\right)\right]\end{aligned}\right\} \tag{34.50d}$$

$$[r^6]A_{40} = [r^3K_{-3}] \tag{34.50e}$$

These form five uncoupled symmetric subsystems with maximum rank 3. The coefficients on the left-hand side are fixed positive numbers, which need to be calculated only once and are then stored. Only the coefficients on the right-hand side have to be recalculated in repeated applications; the surviving subscript is the last one (n) in K_{ln}.

The procedure is self-consistent in the sense that a function $S(r, \varphi)$, whose gradient contains exactly the terms considered in Eq. (34.46) with arbitrary coefficients, is found exactly apart from the unavoidable rounding errors. Moreover, the method provides an intrinsic accuracy control, which is effected by substituting the calculated coefficients into the functional F. Let the expression

$$\|f\| := \left\{\frac{1}{\pi}\int_0^1\int_0^{2\pi} f(r,\varphi)r\,dr\,d\varphi\right\}^{1/2} = \left(\frac{1}{2LN}\sum_{l=0}^{L}\sum_{0}^{2N}\left|f_{lj}\right|^2 G_l\right)^{1/2}$$

denote the norm of any function $f(r, \varphi)$, then $\varepsilon = \|\Delta w - u_A\|$ is the absolute error and $\delta = \varepsilon/\|\Delta w\|$ the relative error of the approximation; the latter must satisfy $\delta \ll 1$.

The Fourier analysis method not only provides a way of determining aberration coefficients $A_{\mu\nu}$, but is also a quite general mathematical procedure for the reconstruction of a two-dimensional real function from its gradient. In electron optics as well as in light optics, the function $S(r, \varphi)$, the wave aberration, is of particular interest in wave-theoretical considerations.

34.6.2 Some Practical Aspects

The application of this method requires that the paraxial rays form a round cone in the image space, which implies that the corresponding cone in the object space is elliptic, as is sketched in Fig. 34.3. How then should the initial conditions be chosen? This is very rarely a serious problem. In most cases we know – from the symmetry properties of the system in question – the directions of maximum and minimum lateral magnifications M_x, M_y, respectively. Let us suppose that these magnifications and the refractive index $\mu = (\hat{\Phi}_i/\hat{\Phi}_o)^{1/2}$ of the image relative to the object are known. The appropriate starting slopes (in the rotating frame) are then

$$\left.\begin{aligned} x'_{lj} &= x'^{(0)}_o + \mu M_x a r_l \cos\varphi_j \\ y'_{lj} &= y'^{(0)}_o + \mu M_y a r_l \sin\varphi_j \end{aligned}\right\} \quad \begin{aligned} l &= 1\ldots L, \\ j &= 1\ldots 2N \end{aligned} \tag{34.51}$$

For equal magnifications, as in all arrangements of round lenses and deflectors with a straight optic axis, it is not really necessary to refer to the image side. We can simply identify the angles appearing in the theory as those at the starting point, thereby representing the aberrations in terms of initial conditions, which is often very convenient.

Until now we have considered only the aberrations in a fixed image plane $z = z_i$. Determination of the variation of the aberration coefficients with the coordinate z is, however, straightforward, requiring no additional ray tracing. The routines for solving systems of ordinary differential equations usually furnish the solution together with its first-order derivative. In the approximation in which each trajectory is replaced by its local tangent, we have

$$\Delta w_j(z_i + \Delta z) = \Delta w_j(z_i) + \Delta z \cdot \Delta w'_j(z_i) + O(\Delta z^2)$$

It is now easy to repeat the Fourier analysis with these new aberrations and to study in this way the effect of a defocus.

Another necessary extension is the calculation of *chromatic aberrations*. The simplest but also the slowest method is to repeat the whole procedure for a few different values of the kinetic starting energy. In this way we could even determine the chromatic variation of each aberration coefficient. In almost all practical applications, such highly detailed information is never needed. The following procedure is then more favourable.

For the four rays with aperture conditions

$$r = r_1, \quad \varphi = (j-1)\frac{\pi}{2}, \quad j = 1\ldots 4$$

we solve Eq. (34.35) with the initial conditions $\boldsymbol{u}(0) = 0$, $\dot{\boldsymbol{u}}(0) \parallel \dot{\boldsymbol{r}}(0)$, the length $|\dot{\boldsymbol{u}}(0)|$ being determined from Eq. (34.33). The resulting aberrations are then purely chromatic ones. For the complex lateral aberrations Δw_j, $j = 1 \ldots 4$, obtained from these solutions, a series expansion similar to Eq. (34.45) can be set up, but this is now truncated after the terms with $m \leq 2$, $m + 2k \leq 3$, hence

$$\Delta w = C_{20} r \mathrm{e}^{-\mathrm{i}\varphi} + C_{10} + C_{11} r^2 + 2C_{01} r \mathrm{e}^{\mathrm{i}\varphi} + \frac{1}{2} C^*_{11} r^2 \mathrm{e}^{2\mathrm{i}\varphi} \tag{34.52}$$

This approximation is quite sufficient, since the energy shift $e\Phi_0$ is usually very small. Writing down Eq. (34.52) for the four rays specified above and then forming the trigonometric interpolation sums (34.49) with $l = 1$, $N = 2$, we find the four simple relations

$$\left.\begin{aligned} K_{-1} &= \frac{1}{4}(\Delta w_1 + \mathrm{i}\Delta w_2 - \Delta w_3 - \mathrm{i}\Delta w_4) = C_{20} r_1 \\ K_0 &= \frac{1}{4}(\Delta w_1 + \Delta w_2 + \Delta w_3 + \Delta w_4) = C_{10} + C_{11} r_1^2 \\ K_1 &= \frac{1}{4}(\Delta w_1 - \mathrm{i}\Delta w_2 - \Delta w_3 + \mathrm{i}\Delta w_4) = 2C_{01} r_1 \\ K_2 &= \frac{1}{4}(\Delta w_1 - \Delta w_2 + \Delta w_3 - \Delta w_4) = \frac{1}{2} C_{11}^* r_1^2 \end{aligned}\right\} \tag{34.53}$$

(the subscript $l = 1$ being omitted in the K-coefficients). These suffice for the determination of the coefficients in Eq. (34.52). An accuracy control is given by the requirement that C_{01} must be real so that the imaginary part of C_{01} represents directly a numerical error. Ignoring $\Im(C_{01})$ we then have the perturbation eikonal, which is analogous to Eq. (34.44).

34.6.3 Integral Properties of Aberration Discs

So far we have been exclusively concerned with the *analysis* of aberrations. A list of coefficients, which may well be long, is, however, not very easy to interpret in terms of the imaging quality of a practical device. A few geometric parameters characterizing the size and shape of the aberration disc would be more helpful. This problem is usually solved in the following way. Depending on the purpose of the device in question, the various types of aberrations have different priorities. The aberrations with low priority are ignored; for each of the remaining ones, a corresponding aberration radius is estimated on the basis of rough conjectural formulae. Finally a root-mean-square radius is calculated, on the unjustified assumption that the aberrations superimpose statistically. This procedure is as unsatisfactory as it is simple. A better proposal has been made by Scherle (1983, 1984), who approximated the cross-section of the beam by a suitably defined ellipse. It is then not even necessary to determine any aberration coefficients. The price of this simplification is, however, that a rather large number of trajectories must be traced. This disadvantage can be avoided by combining Scherle's method with the Fourier analysis procedure.

First of all we notice that every realistic electron beam has an energy spectrum, characterized by a distribution function $g_s(\varepsilon)$ for $\varepsilon_1 \leq \varepsilon \leq \varepsilon_2$, $\int g_s(\varepsilon)\, d\varepsilon = 1$. Here $\varepsilon = e\delta\Phi$ is the deviation of the energy from its nominal value. We also take into account the fact that the aperture may not be illuminated uniformly; this nonuniform illumination is described by an intensity distribution $g_a(\sigma, \varphi)$, over the aperture of radius $\sigma_{\max} = a$

$$\int_{\sigma=0}^{a} \int_{\varphi=0}^{2\pi} g_a(\sigma, \varphi)\sigma \, d\sigma \, d\varphi = 1$$

We now define the expectation value $<f>$ of any real or complex function f in the recording plane $z = \text{const}$.

$$<f(z)> = \int_{\sigma=0}^{a} \int_{\varphi=0}^{2\pi} \int_{\varepsilon=\varepsilon_1}^{\varepsilon_2} g_s(\varepsilon) g_a(\sigma, \varphi) f(z, \varepsilon, \sigma, \varphi) \sigma \, d\sigma \, d\varphi \, d\varepsilon \tag{34.54}$$

We recall that (σ, φ) refer to the aperture, while z does not. In order to apply Eq. (34.52) to the aberrations, we notice first that

$$\Delta w(z, \varepsilon; \sigma, \varphi) = \Delta w(z, 0; \sigma, \varphi) + \frac{\varepsilon}{\varepsilon_m} \Delta u(z, \varepsilon_m; \sigma, \varphi) \tag{34.55}$$

is a good approximation for the superposition of geometric and chromatic aberrations Δu, if ε_m is the most probable energy. It is exact if only the chromatic and geometric aberrations of lowest order are considered (first-order chromatic and third-order geometric aberrations for round lenses, for example). We can now define an intensity-weighted distortion $<\Delta w(z)>$ by setting $f := \Delta w$ in Eq. (34.54). We also introduce a root-mean-square (rms) radius

$$\rho(z) := <|\Delta w - <\Delta w>|^2>^{1/2} \tag{34.56a}$$

and ellipticity parameters

$$e_1(z) + \mathrm{i} e_2(z) := <(\Delta w - <\Delta w>)^2> \tag{34.56b}$$

The meaning of the latter quantities is as follows: if we shift the origin of the coordinates to the centre $<\Delta w>$ of the ellipse:

$$\xi := \Delta x - <\Delta x>, \quad \eta := \Delta y - <\Delta y> \tag{34.57}$$

we obtain

$$2<\xi^2> = \rho^2 + e_1, \quad 2<\eta^2> = \rho^2 - e_1, \quad 2<\xi\eta> = e_2 \tag{34.58}$$

The transformation to the principal axes of the ellipse is given by

$$\overline{\xi} = \xi \cos\theta + \eta \sin\theta, \quad \overline{\eta} = -\xi \sin\theta + \eta \cos\theta, \tag{34.59a}$$

where the angle θ is to be calculated from

$$\tan 2\theta = \frac{2<\xi\eta>}{<\xi^2> - <\eta^2>} \equiv \frac{e_2(z)}{e_1(z)} \tag{34.59b}$$

This rotation describes the overall effect of the *anisotropic* errors, if the directions obtained in this way are not either meridional or sagittal. The main axes themselves are now given by

$$2<\overline{\xi}^2> = \rho^2 + \sqrt{e_1^2 + e_2^2}, \quad 2<\overline{\eta}^2> = \rho^2 - \sqrt{e_1^2 + e_2^2}, \tag{34.60}$$

The axes obtained in this way are smaller than the true dimensions of the aberration figure. Following Scherle's (1983) proposal, we therefore introduce a dilatation factor, which may almost always be set equal to two. Finally, we obtain the semiaxes

$$E_{1,2}(z) = \sqrt{2}\left(\rho^2 \pm \sqrt{e_1^2 + e_2^2}\right)^{1/2} \tag{34.61}$$

In his thesis, Scherle proposed that the averaging should be performed by summation over many individual aberrations. This is certainly necessary if the initial conditions of the trajectories in the beam are so general that a wave aberration S cannot be used. Such a case arises, for instance, if the electrons start from a cathode surface or from a crossover with finite extent, so that the initial conditions themselves have a statistical distribution.

In very many cases, this method is unnecessarily general and would entail an unreasonably large effort. Instead, we can first determine the aberration coefficients (A_{mk}), either from integral expressions or from a Fourier analysis. These are then introduced into Eq. (34.46) or more generally into Eq. (34.45) and, for the chromatic effects, into Eq. (34.52). These formulae are now regarded as continuous *interpolation* formulae for the aberrations. Numerical evaluation of the necessary integrals is then perfectly practical. This is always a comparatively fast procedure since it does not require any new ray tracing. The approximate ellipse can be determined rapidly for a sequence of image planes, after which the optimal defocus can be found easily.

The advantage of this method lies in the fact that it gives us very clear and simple criteria for optimal focusing. The best approximation to a stigmatic focus is obtained when, after exploring all permissible variations of the system parameters, the largest value of the semiaxis $E_1(z)$ is least for the object point with the worst aberrations. This idea has not yet been much exploited in practice but it seems very likely that the method will prove useful and deserves wide acceptance. The conclusions to which it leads are certainly realistic. Fig. 34.6 shows, for example, how the size of the spherical aberration disc can be reduced by defocusing. With Scherle's procedure, the plane in which the beam radius is smallest is found to be at $(8/9)\Delta z$, where Δz is the defocus of the theoretical plane of least confusion (24.50); the radius of the disc differs by a factor of only $(8/9)^{1/2} = 0.94$ from the familiar radius of least confusion (24.51). This is well within the experimental confidence limits.

We conclude this account of the various numerical methods of computing aberrations and assessing their importance with the observation that the choice of practical procedures is wide enough to enable us to study virtually any electron optical system, however complicated, with sufficient accuracy for most practical needs.

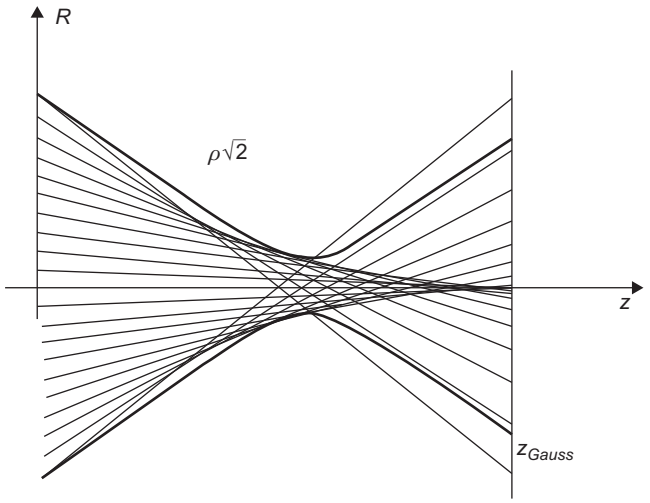


Figure 34.6
Form of the caustic and of the r.m.s radius as functions of the axial coordinate for a round lens with spherical aberration.

34.7 Optimization Procedures

Once programs for the computation of fields, trajectories, focusing properties and aberrations have been successfully completed, it is very useful to incorporate them into a program for the optimization of electron optical systems. Usually the purpose of optimization is to determine geometric and electromagnetic configurations of electrodes, polepieces and coils that minimize certain electron optical aberrations of the system in question subject to given constraints, typically a fixed focal length, a minimum working distance or an upper bound on a coil current or some other technical limit. It is clear that this problem cannot be solved automatically in a perfectly general manner, since the technical requirements and the constraints will differ widely for the various kinds of electron optical devices. For this reason we cannot go into much detail here.

34.7.1 The Defect Function

The first step in any optimization procedure must be the definition of an appropriate *defect function* or 'merit function'. This must be a strictly positive quantity, which is to be minimized by varying the system parameters within the allowed limits. This definition already requires a precise formulation of the specific requirements and constraints. The *arguments* $(x_1 \ldots x_n)$ of the defect function are the set of all those system parameters that

are allowed to vary, and their domain D of definition is bounded by constraints: $(x_1 \ldots x_n) \in D$. Quite generally the constraints can be represented by inequalities:

$$\theta_k(x_1 \ldots x_n) \geq 0, \quad k = 1 \ldots r \tag{34.62}$$

which include equations as special cases.

The functional expression ψ of the defect function is usually the square of a total aberration radius. This is defined in the following way. First, for each individual aberration a simple effective defect radius is introduced, for instance $\rho_1 = MC_s\alpha^3/4$ for the spherical aberration of a round lens with magnification M and aperture angle α, or $\rho_2 = MC_c\alpha\Delta\hat{\Phi}/\hat{\Phi}_0$ for the chromatic aberration, and so on. Next, a set of nonnegative weight factors $g_1^2 \ldots g_m^2$ is introduced, which characterize the importance of the m individual aberrations. Large weights mean that the corresponding aberrations are very serious, while low weights are associated with contributions of little importance. Vanishing weights mean that the corresponding aberrations are completely ignored. The defect function is then

$$\psi(x_1 \ldots x_n) = \sum_{i=1}^{m} g_i^2 \rho_i^2 = \sum_{i=1}^{m} f_i^2(x_1 \ldots x_n) \tag{34.63}$$

with $f_i \coloneqq g_i\rho_i$ for $i = 1 \ldots m$.

It is possible to incorporate the constraints (34.62) into the defect function:

$$\psi(x_1 \ldots x_n) =: \sum_{j=1}^{M} f_j^2 = \sum_{i=1}^{m} g_i^2 \rho_i^2 + \sum_{k=1}^{r} G_k^2(\theta_k - |\theta_k|)^2 \tag{34.64}$$

with $M = m + r$ being the total number of terms. The second contribution is a 'penalty' function: the defect function increases for $\theta_k < 0$. This form of the defect function may be useful if very small violations of constraints are allowed; such violations can be tolerated if the constraints themselves represent only roughly guessed technical bounds. The magnitudes of these violations will depend on the particular choice of the new weights $G_1^2 \cdots G_r^2$.

As a simple example, we describe the appropriate defect function for axial focusing by an ordinary unsaturated magnetic round lens. Here we have to consider the spherical aberration and the axial chromatic aberration, and we include a rough guess for the blurring caused by diffraction. The defect function is again a sum of squares of aberration radii. This definition is convenient for numerical purposes, though of course the chromatic and geometric aberrations add linearly. Thus

$$\psi = \left(\frac{1}{4} g_1 C_s \alpha^3\right)^2 + \left(g_2 C_c \alpha \frac{\Delta\hat{\phi}}{\hat{\phi}}\right)^2 + \left(\frac{0.61 g_3 h}{\alpha\sqrt{2me\hat{\phi}}}\right)^2 \tag{34.65}$$

α being the aperture angle at the object point, $\tilde{\phi}$ the acceleration potential and h Planck's constant. (For further details of the diffraction term, see Volume 3.) Since all three terms are equally important, we set $g_1 = g_2 = g_3 = 1$. In the absence of constraints and parameters to be varied, this function is still incomplete. Reasonable constraints are, for instance, fixed values of the object coordinate z_o and of the image coordinate z_i and the acceptable interval $f_{\min} \leq f \leq f_{\max}$ for the focal length f. The definition of system parameters requires a decision concerning the class of shapes allowed for the axial field distribution $B(z)$. If, for instance, the simple model with a gap of width S and cylindrical bores with radii R_1 and R_2 is employed, we should use the parameters $x_1 = R_1$, $x_2 = R_2$, $x_3 = S$ and $x_4 = J$, the latter being the number of ampère-turns. If the midplane of the gap is kept fixed at $z_M = 0$, we have a well-specified optimization problem, so that $\psi =$ minimum leads to a unique solution.

This very simple example makes it clear that only with very detailed specification can a reasonable answer to the optimization problem be expected. Moreover, this example is a reminder that rough estimates of the form (34.65) are commonly used. These can be replaced by improved formulae for more realistic intensity distributions and their expectation values, but this is still uncommon.

34.7.2 The Optimization of Axial Distributions

As already pointed out, the focusing properties and the aberrations of electron optical systems are generally determined by certain axial field distributions, such as the electrostatic potential $\phi(z)$ and the magnetic field strength $B(z)$ on the optic axis. It is therefore tempting to try to optimize these functions. This means that, subject to the given constraints, the defect function is minimized by finding the 'best shapes' of these axial distributions.

This procedure has been followed several times and with different approaches. Moses (1973), for instance, employed variational calculus for the minimization of spherical aberration with simultaneously vanishing coma. Crewe *et al.* (1968) designed a field-emission electron gun under the assumption that the axial potential $\phi(z)$ in the space between the first and second anode can be represented by a cubic polynomial, the coefficients of which were the optimization parameters; this yielded the 'Butler gun'. Later Munro (1973) showed that this design does not represent a true optimum. Szilágyi (1977) introduced dynamic programming (see Section 35.3.2). In this approach the integration interval $z_o \leq z \leq z_i$ between object and image is dissected into a set of small subintervals. Then, starting from the object coordinate z_o, the integrand of an aberration integral, for instance that for C_s, is minimized under given constraints in each of these subintervals. The result is then a piecewise analytic axial field distribution, a cubic spline, for example (Szilágyi, 1984, 1987a,b), which is then assumed to be the best field. Fuzzy set theory has

been used by Wei et al. (1995) while Wei and Tong (1995) have constructed an expert system based on artificial intelligence to seek optimal designs. Gu et al. (2001) have tested a genetic algorithm.

Common to all methods starting from axial field distributions is the severe difficulty of finding reasonably shaped electrodes or polepieces to generate these axial fields. Analytic continuation of the axial potentials into off-axis domains by means of the radial series expansions of Chapter 7, Series Expansions, generally leads to equipotentials with a singular character such as sharp edges in the vicinity of the optic axis or even vanishing bore radii. In order to avoid these, the electrode structure has to be altered, but the additional fringe fields thus introduced cast doubt on the whole design procedure.

Design based on optimization of an axial field function is thus rarely successful. At the other extreme is the laborious technique in which the appropriate boundary-value problems are solved repeatedly for different geometric shapes, with the risk that the best solution may not belong to the set of configurations analysed. An intermediate way out of this dilemma is to search for solutions among a reasonably constrained family of configurations. Thus Glatzel and Lenz (1988) sought optimal designs of an electrostatic lens by using the potentials and geometrical parameters of the lens as variables. Benez et al. (1995) searched using a genetic algorithm. The 'second-order electrode method' (SOEM) developed in the Delft University of Technology used a relatively simple lens model and suitable constraints to find lenses for specific tasks. Here, the axial potential was modelled by a cubic spline, which yielded a simple relation between $\Phi''(r,\ z)$ and Φ on the axis and at (r, z). The method and results obtained with it are described by Adriaanse et al. (1989), Lencová et al. (1989), Lencová and Wisselink (1990), van der Steen and Barth (1989), van der Steen et al. (1990), van der Stam et al. (1993) and Barth et al. (1995). Optimization of an individual lens is not, however, sufficient to design a complete system and for this, a new tool was developed: Particle Optics Computer-aided Design (POCAD). With this, the user can set out from a very general schematic (multilens) model and iterate to a design or set of designs that satisfies the desired conditions. Full details can be found in van der Stam and Kruit (1995, 1999), van der Stam (1996), van der Stam et al. (1996) and Leunissen et al. (2001). Another optimization program is the Cambridge Interactive Electron Lens Analysis System (CIELAS): Hill and Smith (1980, 1981), Tsuno and Smith (1985, 1986), Taylor and Smith (1986) and Edgcombe et al. (1999). For other work on optimization, see Makino and Berz (2011), Nishiguchi and Toyoda (2008), Bärtle (2004), Sakaguchi et al. (1999), Gu et al. (1999, 1991, 1984), Martínez et al. (1998, 1999), Demin et al. (1998), Hejna (1998), Degenhardt (1996, 1997), Dymnikov and Martínez (1997), Martínez and Sancho (1995), Munack (1992), Boesten (1988), Gu and Shan (1984), Fink and Kisker (1980), Chutjian (1979), Al-Obaidi (2005), Al-Obaidi et al. (2009) and Kadhem (2014).

34.7.3 The Damped Least-Squares Method

The damped least-squares method was first described by Levenberg (1944) and has been successfully applied to light optical design problems. In 1982, it was introduced into electron optics (Chu and Munro, 1982a,b; Munro and Chu, 1982a,b). We now briefly outline their method.

The defect function is represented in the form (34.63). If the arguments x_j are varied by small increments Δx_j ($j = 1 \ldots n$), the functions f_i alter by

$$\hat{f}_i = f_i + \sum_{j=1}^{n} a_{ij} \Delta x_j, \quad i = 1 \ldots m, \tag{34.66a}$$

$$a_{ij} = \frac{\partial f_i}{\partial x_j}, \quad j = 1 \ldots n \tag{34.66b}$$

This linear approximation is adequate if all the increments Δx_j are sufficiently small. The new defect function is now

$$\hat{\psi} = \sum_{i=1}^{m} \left(f_i + \sum_{j=1}^{n} a_{ij} \Delta x_j \right)^2 \tag{34.67}$$

The minimization conditions $\partial \hat{\psi} / \partial(\Delta x_j) = 0$ for $j = 1 \ldots n$ lead to n simultaneous linear equations for the increments $\Delta x_1 \ldots \Delta x_n$, which could be solved by straightforward techniques.

Unfortunately this simple undamped least-squares method is unstable and can diverge. In order to avoid this instability, Eq. (34.67) is modified to

$$\begin{aligned} \psi^\dagger &= \hat{\psi} + \sum_{j=1}^{n} p_j^2 \Delta x_j^2 \\ &= \sum_{i=1}^{m} \left(f_i + \sum_{j=1}^{m} a_{ij} \Delta x_j \right)^2 + \sum_{j=1}^{n} p_j^2 \Delta x_j^2 \end{aligned} \tag{34.68}$$

where the factors p_j are called damping coefficients. Obviously we have $\hat{\psi} < \psi^\dagger$, so that a minimum of $\psi^\dagger$ corresponds to a minimum of $\hat{\psi}$. The minimization conditions $\partial \psi^\dagger / \partial(\Delta x_k) = 0$ now take the form

$$\begin{aligned} &\sum_{i=1}^{m} \left\{ a_{ik} \left(f_i + \sum_{j=1}^{n} a_{ij} \Delta x_j \right) \right\} + p_k^2 \Delta x_k = 0 \\ &(k = 1 \ldots n) \end{aligned} \tag{34.69}$$

These form a system of n simultaneous linear equations for the unknowns $\Delta x_1 \ldots \Delta x_n$, the coefficient matrix being symmetric and positive definite.

The new coordinates $x_j^{(n)} = x_j + \Delta x_j$ depend on the choice of the damping coefficients. Two different choices are proposed by Chu and Munro:

$$p_j^2 = p^2 \quad \text{(additive damping method)} \tag{34.70a}$$

$$p_j^2 = p^2 \sum_{i=1}^{j} a_{ij}^2 \quad \text{(multiplicative damping method)} \tag{34.70b}$$

p being a constant damping factor. With Eq. (34.70a), each diagonal element of the matrix in (34.69) is enlarged by an additive term p^2, while with (34.70b), each diagonal element is multiplied by $(1 + p^2)$. The numerical solution of (34.69) must be repeated for various values of p, until the smallest value of $\psi^\dagger$ is found.

The partial derivatives a_{ij} that appear in the above equations are to be calculated numerically from a simple two-point formula:

$$a_{ij} = \frac{\partial f_i}{\partial x_j} = \frac{f_i(x_j + \delta x_j) - f_i(x_j)}{\delta x_j} \tag{34.71}$$

δx_j being a small increment. According to Chu and Munro, this linear approximation is sufficient, since the whole minimization procedure must in any case be repeated if the starting point is far distant from the final optimal configuration.

Chu and Munro have made extensive investigations concerning the applicability of this method and set up an interactive program. For reasons of space we cannot outline this here and refer to the corresponding publication, where full details are to be found. Concerning the optimization of geometric configurations, Chu and Munro made the following compromise: for each electron optical element, for instance round electrostatic and magnetic lenses and electric or magnetic deflectors, reasonable geometric shapes were assumed and for these the boundary-value problems were solved exactly beforehand. Later, during the optimization process, all operations that can be performed without a new field calculation were allowed: changes of electric or magnetic parameters such as the electrode potentials or coil currents, axial shifts of the object plane, the image plane or of entire deflectors, rotation of deflectors about the optic axis and scale-transformations in independent elements. In principle it is also possible to alter the geometric shape of electrode or polepiece surfaces but then an entirely new field calculation is necessary.

The great advantage of this method is that each optimization step results in a realistic configuration, which could be constructed. The applicability of this method was

demonstrated for focusing and dual-channel deflection systems, which are employed in electron beam lithography. A very detailed knowledge of the technical requirements in electron beam lithography is needed to understand the particular features of the optimization; much more discussion is to be found in Chapter 40 of Volume 2 and we merely observe that the results published by Chu and Munro are distinctly encouraging.

We conclude that the damped least-squares method is an effective way of solving optimization problems.

34.8 Differential Algebra

34.8.1 Introduction

An algebra in which real numbers are extended by including 'differential' elements was designed many years ago by Abraham Robinson (1961, 1966); a very readable account is to be found in Dauben's biographical publications (1995, 2003). It was introduced into charged-particle optics by Berz (1989) and has since been employed to calculate the aberrations of round lenses (Cheng et al, 2001a,b, 2002a, 2006; Kang et al., 2007; Munro et al, 2006a–c; Wang et al., 1999, 2000, 2004a; Liu, 2006), multipoles (Liu, 2003; Wang et al., 2004b), curved-axis systems (Cheng et al., 2003; Wang et al., 2002a), focusing–deflection systems (Wang et al. 2002b), mirrors (Wang et al., 2008a,b), diffraction aberrations (Radlička, 2012) and parasitic aberrations (Radlička and Oral, 2016). The method has been extensively developed in several directions by Berz and colleagues (Berz, 1989, 1995; Berz and Makino, 2004; Berz and Wollnik, 1987; Makino and Berz, 1997, 1999, 2011; Makino et al. 2004; Zhang and Berz, 2011; Zhang et al., 2015); see Berz et al. (2015) for much more information. In this section, we first describe the algebra and then show how to apply it to aberration studies. Its attraction lies in the fact that only one trajectory needs to be computed, using the methods described earlier, whatever the order of the aberrations of interest. The accuracy of the results is inevitably dictated by the precision with which the potentials or field distributions are known. We shall see that there are two approaches. In one, only the fields or potentials on the optic axis are used and they are represented as a linear combination of model functions, such as Hermite polynomials. In the second approach, the local values of the fields or potentials in the space traversed by the computed trajectory are used; elaborate interpolation is required if these values are the result of a finite-element or finite-difference calculation.

34.8.2 Definition of Differential Algebras

We first consider the simplest such algebra from which the general principles can be understood. We then turn to the more advanced algebras needed for aberration calculation.

Consider the set of pairs of numbers (a_0, a_1), familiar from the use of complex numbers where a_0 and a_1 are the real and imaginary parts of $a = a_0 + \mathrm{i}a_1$. Here, however, the second member of the pair no longer satisfies $(0, 1).(0, 1) = (-1, 0)$. The basic operations of differential algebra – addition, scalar multiplication and 'vector' multiplication – are as follows:

$$\begin{aligned}(a_0, a_1) + (b_0, b_1) &= (a_0 + b_0, a_1 + b_1)\\ t(a_0, a_1) &= (ta_0, ta_1)\end{aligned} \tag{34.72a}$$

as usual but now

$$(a_0, a_1).(b_0, b_1) = (a_0b_0, a_0b_1 + a_1b_0) \tag{34.72b}$$

where a_0, a_1, b_0, b_1 and t are all real numbers. It is clear that $(a_0, 0)$ is the same as the real number a_0. The meaning of $(0, a_1)$ is less obvious: we see that $(0, a_1).(0, a_1) = (0, 0)$! Any quantity of the form $(0, a)$ appears to be the square root of zero, just as i is the square root of -1. The element $(0, 1)$ is therefore given a special symbol, d, and a_1 is known as the *differential part* of (a_0, a_1) for reasons that will soon become obvious. It is easy to show that $(1, 0)$ is the unit element, in the sense that $(1, 0) \cdot (a_0, a_1) = (a_0, a_1)$ for all a_0, a_1. Elements of the algebra have an inverse provided that $a_0 \neq 0$ (we are thus concerned with rings, not fields):

$$(a_0, a_1).\left(\frac{1}{a_0}, -\frac{a_1}{a_0^2}\right) = (1, 0) \tag{34.73}$$

This brings us to the reason for interest in this algebra: the differential part of $f(x + \mathrm{d}) - f(x)$ proves to be df/dx.

$$\frac{df}{dx} = D\left[f(x + \mathrm{d}) - f(x)\right] = Df(x + \mathrm{d}) \tag{34.74}$$

in which the operator D signifies 'take the differential part'. As an example, consider $f(x) = \sin 3x$. Then

$$\begin{aligned}f(x + \mathrm{d}) &= \sin 3(x + \mathrm{d})\\ &= \sin\{3(x, 0) + 3(0, 1)\}\\ &= \sin 3x \cos 3(0, 1) + \cos 3x \sin 3(0, 1)\\ &= 3\ (\sin 3x,\ \cos 3x)\end{aligned}$$

and so $Df(x + \mathrm{d}) = 3\cos 3x$ as expected. We have used

$$\sin(0, 1) = (0, 1) - \frac{1}{6}(0, 1)^3 + \ldots = (0, 1)$$

$$\cos(0, 1) = 1 - \frac{1}{2}(0, 1)^2 + \ldots = 1$$

For aberration studies, we need the more general forms of the algebra, denoted ${}_nD_\upsilon$. First, we introduce the number N of monomials in υ variables up to order n:

$$N(n,\upsilon) = \frac{(n+\upsilon)!}{n!\upsilon!} \tag{34.75}$$

Thus for $n = 3$ and $\upsilon = 2$, $N = 5!/3!2! = 10$. The monomials are

$$1 \;\; x \;\; y \;\; xx \;\; xy \;\; yy \;\; xxx \;\; xxy \;\; xyy \;\; yyy \tag{34.76}$$

The position of a monomial M in this ordered sequence is denoted by I_M so that for example $I_{xyy} = 9$. The I-th monomial is denoted M_I, so that $M_9 = xyy$. We shall also need the quantity F_I; if the I-th monomial M_I is of the form $x_1^{j_1} x_2^{j_2} x_3^{j_3} \ldots$ then

$$F_I = j_1!j_2!j_3!\ldots \tag{34.77}$$

Thus for I_{xyy} in the above example, $F_{xyy} = 2! = 2$.

In ${}_nD_v$, the laws of addition and scalar magnification are unaltered but the law of multiplication becomes

$$(a_0, a_1, a_3, \ldots a_N) \cdot (b_0, b_1, b_3, \ldots b_N) = (c_0, c_1, c_3, \ldots c_N)$$

where

$$c_i = F_i \sum_{0<p,q<N} \frac{a_p b_p}{F_p F_q} \quad p \text{ and } q \text{ are such that } M_p M_q = M_i \tag{34.78}$$

Thus for the case of ${}_3D_2$ we find

$$\begin{aligned}
c_1 &= a_1 b_1 \\
c_2 &= a_1 b_2 + a_2 b_1 \\
c_3 &= a_1 b_3 + a_3 b_1 \\
c_4 &= a_1 b_4 + 2a_2 b_2 + a_4 b_1 \\
c_5 &= a_1 b_5 + a_2 b_3 + a_3 b_2 + a_5 b_1 \\
c_6 &= a_1 b_6 + 2a_3 b_3 + a_6 b_1 \\
c_7 &= a_1 b_7 + 3(a_2 b_4 + a_4 b_2) + a_7 b_1 \\
c_8 &= a_1 b_8 + 2a_2 b_5 + a_3 b_4 + a_8 b_1 \\
c_9 &= a_1 b_9 + a_2 b_6 + 2a_3 b_5 + a_9 b_1 \\
c_{10} &= a_1 b_{10} + 3(a_3 b_6 + a_6 b_3) + a_{10} b_1
\end{aligned}$$

Finally, we need a new differential operator, for which the partial derivative symbol ∂ is used:

$$\partial_p(a_1, a_2, a_3, \ldots a_N) = (c_1, c_2, c_3, \ldots c_N) \tag{34.79a}$$

where now

$$c_i = \begin{cases} 0 & \text{if } \sum_j i_j = n \quad \text{where } M_i = x_1^{i_1} x_2^{i_2} x_3^{i_3} \ldots x_n^{i_n} \\ a_\mu, & \mu = I_{(M_i . X_p)} \quad \text{otherwise} \end{cases} \tag{34.79b}$$

This enables us to generalize Eq. (34.74) to two or more variables. Thus for $n = \upsilon = 2$, for example, we have

$$f(x + \mathrm{d}x, y + \mathrm{d}y) = \left(f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y}, \frac{\partial^2 f}{\partial y^2} \right)(x, y)$$

in which we have written

$$\mathrm{d}x \coloneqq (0, 1, 0, 0, 0, 0) \quad \text{and} \quad \mathrm{d}y \coloneqq (0, 0, 1, 0, 0, 0)$$

For the general case, v variables and order n, this becomes

$$\frac{\partial^{i_1 + i_2 + i_3 + \ldots i_v} f}{\partial x_1^{i_1} \partial x_2^{i_2} \partial x_3^{i_3} \ldots \partial x_v^{i_v}} = c_j, \quad j = I_{x_1^{i_1} x_2^{i_2} x_3^{i_3} \ldots x_v^{i_v}} \tag{34.80}$$

$I_{x_1^{i_1} x_2^{i_2} x_3^{i_3} \ldots x_v^{i_v}}$ denotes the position of the corresponding monomial in the ordered sequence of monomials chosen at the outset.

Before moving on to the next stage, extraction of aberration coefficients from a ray trace, we mention that different authors present differential algebra in different terms. Here, we have followed Berz et al. (2015) closely but the reader may like to consult Wang et al. (1999, 2004) and Radlička (2008 or 2012) in parallel, as well as Berz's very lucid but more abstract presentation (1999).

34.8.3 Calculation of Aberration Coefficients

The calculation involves solving the appropriate trajectory equation (3.22) with 3.19 for round lenses, for example) by means of the Runge−Kutta routine but now, the elementary quantities $x(z)$, $y(z)$ are replaced by the appropriate elements of the differential algebra. The relation between the image, or final, plane and the object plane, known in this context as the transfer map, has the general form

$$\begin{pmatrix} x_i \\ y_i \\ x_i' \\ y_i' \end{pmatrix} = \sum_{i,j,k,l=0,1,\ldots n}^{i+j+k+l=n} \begin{pmatrix} A_{ijkl} \\ B_{ijkl} \\ C_{ijkl} \\ D_{ijkl} \end{pmatrix} x_o^i y_o^j x_o'^k y_o'^l$$

where n is the highest order of the aberrations to be included in the calculation. The coefficients $A_{ijkl}, \ldots D_{ijkl}$ emerge from the calculation and it is only necessary to relate them

to the usual aberration coefficients, For the fifth-order spherical aberration of round lenses, for example, C_5, the isotropic part is given by A_{0050} and the anisotropic part by B_{0050} (for a complete list, see Wang et al., 2004a and for the corresponding list for the paraxial properties and third-order aberrations of round lenses, see Wang et al., 1999).

This description conceals the principal obstacle to the use of this method: the need for sufficiently accurate knowledge of the values of the field through which the ray is traced. In the first form of the calculation, the distribution on the optic axis was represented by a sum of analytic functions, which could hence be differentiated exactly; Wang et al. (2004) use the Hermite functions,

$$h_n(x) = \frac{1}{(\pi^{1/2}2^n n!)^{1/2}} H_n(x)\exp\left(-\frac{x^2}{2}\right)$$

where $H_n(x)$ are the Hermite polynomials. Since the Hermite functions form a complete orthonormal set, it is a straightforward matter to represent a set of measured or calculated values in terms of them (Wang et al., 2004a, who use 200 functions; Munro, 2002, 2011; Munro et al., 2006a,b, 2008). Other authors have used Gaussian wavelets (Berz, 1999, Section 3.2.1). The need to represent the axial potential as the sum of a set of basic functions has been seen as a disadvantage of the procedure and Kang et al. (2007, 2009, 2010) have devised a way of using values generated by the finite-element method directly without resorting to modelling the field on the optic axis. For this, they have introduced elaborate interpolation routines, which provide the field or potential values required for trajectory tracing with sufficient accuracy. Since the necessary accuracy increases with the order of the aberrations, the interpolation procedure likewise becomes progressively more complicated.

34.9 The Use of Computer Algebra Languages

34.9.1 Introduction

The familiar computer languages are designed to perform numerical calculations efficiently but are extremely ill-adapted to any kind of symbolic calculation: it is trivial to list $z := (x + y)^p$ for values of x and y that may contain many digits and any reasonable values of p but much less easy to output the binomial expansion; numerical integration of a function such as $x^p \sin qx$ is again trivial but the closed form of the integral cannot be found; the same is true of differentiation. We have seen that there are powerful routines for solving differential equations numerically but there is no way of knowing whether the equation has a solution in terms of tabulated (or easily computed) functions. It was for reasons such as these that the various members of the family of algebra languages were

born, and indeed many were created to perform specific calculations of great complexity in such fields as celestial mechanics, general relativity and quantum electrodynamics.

We shall not describe any particular language here. We simply mention the principal tasks that can be undertaken and indicate their relevance in electron optics. Details are of course to be found in the manuals of each language and a useful early survey is given by van Hulzen and Calmet (1982). The languages that are most widespread among physicists at present are probably REDUCE (Hearn, 1985; Fitch, 1985; Rayna, 1987) MATHEMATICA and MAPLE (up-to-date details can be found on the websites). We know of only two examples of the use of a general language to perform algebraic calculations in electron optics: ALGOL (Dodin and Nesvizhskii, 1981; Nesvizhskii, 1986) and FORTRAN (Berz and Wollnik, 1987). The use of a Russian computer algebra language (ANALITIK) for deriving aberration coefficients is described by Narylkov and Lyubchik (1982). The proceedings of the conferences that are regularly held on progress in computer algebra give a vivid picture of the growth of the subject (e.g., Ng, 1979 and Calmet, 1982 for the earlier years). For a full list with many other references, see the background text edited by Buchberger et al. (1982) and Davenport et al. (1988); for more recent developments, see Grabmeier et al. (2003), von zur Gathen and Gerhard (2013) and the forthcoming volume by Davenport.

34.9.2 Computer Algebra, Its Role in Electron Optics

In physics, algebra languages were first used mainly for calculations that involved performing elementary operations on very large operands. The number of terms being enormous, the risk of human error was correspondingly large and the computer was used first to check the hand calculation and later, as confidence in these languages grew, to supplant it altogether. A basic operation is therefore multinomial expansion of expressions of the form $(x_1 + x_2 \ldots + x_n)^m$, where the variables x_i may themselves consist of several terms. Since the resulting expression will usually contain a very large number of terms, another important family of operations permits sophisticated sorting of these into groups: all terms $x_i^p x_j^q$ for given p and q, for example, or such that $p + q$ takes a given value, or such that $p \leq q$. If the x_i are circular or hyperbolic functions, or functions of such functions, some kind of reduction or simplification will probably be desirable and facilities for replacing $\sin px \sin qx$ and similar terms by functions of multiples of x are provided or can be incorporated. An important aspect of these languages is that they can be 'taught' results that are not in their regular repertoire. In a program in which Bessel functions and their derivatives appeared, for example, the language could be instructed to use the well-known recurrence formulae that relate J_n, J_n', J_{n+1} and J_{n-1}; indeed the numerous relations between contiguous hypergeometric functions (Whittaker and Watson, 1927, Section 14.7) could just as easily be included if required. A further elementary operation is substitution:

flexible commands permitting substitution of an expression for a variable are provided. The next family of operations is concerned with calculus: these languages are capable of differentiating the everyday functions such as exponentials, logarithms, circular and hyperbolic functions and of course powers of a variable. They can also work correctly with the derivative of a general function so that an expression of the form $\int f'(x) \sin x \, dx$, for example, can be evaluated by partial integration. Depending on the degree of sophistication of the language, indefinite integration of more or less complicated expressions can likewise be performed. Integration of simple expressions is a routine matter and these algebra languages can be extended to integrate any expression for which a solution in closed form exists, provided that the functions occurring in the integrand belong to a certain set of functions. See Norman and Davenport (1979) for a very readable account of an earlier stage in this development and Norman (1982).

Expansion, substitution, sorting, differentiation and integration: although other operations are available in some languages, it is these that have proved most useful in electron optics, where computer algebra has been used primarily to derive or check integral expressions for the aberration coefficients of various types of electron optical component. Let us consider the steps for a typical aberration coefficient. First we must substitute the expansions for Φ and the components of $\boldsymbol{A}$ into the refractive index $\{\hat{\Phi}(1+X'^2+Y'^2)\}^{1/2} - \eta(\boldsymbol{A}_X X' + \boldsymbol{A}_Y Y' + \boldsymbol{A}_z)$ (15.23) truncated after terms of a particular degree. If the rotating coordinate system is to be employed, the appropriate transformation from (X, Y, z) to (x, y, z) must be made. The resulting terms must then be sorted according to their degree in x, y, x', y'. It is generally necessary to make a further substitution at this point, replacing the coordinates that appear in the group of primary aberration terms by the paraxial approximation. The result must be sorted according to the powers of the paraxial coordinates (x_o, y_o, x'_o, y'_o), in order to extract individual aberration coefficients. If higher order perturbation theory is to be used, a substitution involving paraxial terms and the primary aberrations in an arbitrary plane is also needed, followed by sorting of the same kind.

Once the aberration integrals have been derived, further laborious calculations may be needed. If the integrals are to be evaluated numerically, using measured or computed values of the various field functions that occur, it may be desirable to remove terms in which high-order derivatives appear by partial integration. The systematic procedure described in Chapter 24, The Geometrical Aberrations of Round Lenses, at some length is the best way of doing this, and once the fundamental structures (e.g., 24.56) have been recognized, the subsequent differentiation, incorporation of the paraxial equation, multiplication by arbitrary constants and addition to the original coefficients or eikonal function can all be performed by the computer.

If, however, the behaviour of some unfamiliar component is being studied as a preliminary to exact numerical computation of ranges of values of the parameters of apparent interest,

it is often helpful to adopt a simple but realistic field model. Despite its shortcomings, Glaser's bell-shaped model (Chapter 36 of Volume 2) is invaluable for grasping the behaviour of magnetic lenses and is invariably used for teaching purposes; for quadrupoles, the rectangular and bell-shaped models are perfectly adequate for many purposes, and several other types of component are likewise represented rather well by simple models. Evaluation of the aberration integrals for these models is another lengthy task, involving the manipulation of unwieldy expressions frequently involving powers of circular and hyperbolic functions. Computer algebra languages are well-equipped to perform this.

We have mentioned the use of partial integration to cast aberration integrals into a form suitable for numerical evaluation, but how is the latter to be implemented? Computer algebra languages originally provided their answers only as algebraic formulae, which had then to be programmed manually in one of the ordinary high-level languages suited to numerical computation. This was clearly inefficient since one of the attractions of deriving large formulae by computer is to eliminate the risk of human error, but such mistakes are just as likely to creep in when translating the algebraic output into FORTRAN, for example. By 1980, therefore, as a result of a considerable effort, symbolic−numeric interfaces had been developed, notably for REDUCE; the same is now true of many other languages. The user can perform numerical calculations directly from his algebraic output.

The derivation of aberration integrals is not of course the only laborious task facing the theoretician in electron optics, and the operations mentioned above are only a selection of those available. Packages for integration, for polynomial factorization, for solving ordinary differential equations and integral equations, for arbitrary precision floating-point arithmetic and for network analysis problems, including calculation of the determinants of sparse matrices are now routinely included.

34.9.3 Practical Examples

The languages CAMAL (Barton and Fitch, 1972; Fitch, 1985) and REDUCE (Hearn, 1985) have been used in electron optics to obtain relativistically correct formulae for the third-order geometric aberrations of round electrostatic lenses (Hawkes, 1977), to derive aberration coefficients for combined deflection and focusing fields (Soma, 1977, cf. Section 32.3.2) and to study the aberrations of microwave cavities acting as dynamic electron lenses (Hawkes, 1983). Liu has used MATHEMATICA in his work on differential algebra (Liu, 2007) and on higher order aberrations, see Sections 24.10 and 26.3. Preikszas (1995) developed a program, MOPS [Manipulation of Power Series], specifically designed to generate the aberration coefficients of electron mirrors; this was resuscitated to generate the unpublished off-axis aberrations of mirrors, listed in Section 28.1 (Preikszas, private communication, 2016).

Notes and References

The following lists of references follow the main divisions of the book with the exception of those corresponding to Part VII (Instrumental Optics). The lists for Chapters 35 (Electrostatic Lenses) and 36 (Magnetic Lenses) are so long that it seemed preferable to give them separately.

In order to avoid repetition, standard abbreviations have been adopted for several series of conference proceedings, namely, the European and International conferences on (electron) microscopy, which have alternated every two years since 1954 (prior to that date they were not quite so regular); the occasional conferences on high-voltage electron microscopy; the annual meetings of the Electron Microscopy Society of America; and the biennial meetings of the Electron Microscopy and Analysis Group of the British Institute of Physics. The European and International Conferences (EUREM or EMC and ICEM or IMC, respectively) are identified by date and place, the high-voltage conferences by date followed by HVEM and place, the American meetings by date followed by EMSA or MSA, venue and meeting number and the British meetings by date followed by EMAG and venue. The Multinational Conferences on (Electron) Microscopy and the Dreiländertagungen, now Microscopy Conferences, are labelled MCEM or MCM and MC followed by number or date. The biennial Seminars on Recent Trends in Charged Particle Optics and Surface Physics Instrumentation ('Skalský Dvůr') are referred to by date and *Recent Trends*. Full bibliographic details of all these conference volumes are to be found at the end of this section.

Some of the lists that follow contain papers that are not cited in the text. With one exception, however, all these additional references are cited in the notes that precede the lists with some indication of their contents. The exception concerns the Preface and the introductory first chapter. This contains a very full list of books on electron optics; texts that are devoted mainly to electron microscopy are not always included, however.

Despite the length of these lists, we make no claim to completeness and indeed, the coverage is deliberately uneven. For some topics, excellent bibliographies have been compiled and we have referred to these rather than merely repeating their contents. There are others, however, for which the literature is very scattered – electron guns and ultrafast electron microscopy are examples – and here we have attempted to give rather thorough coverage. We shall be most grateful to have any errors or serious omissions drawn to our attention.

Preface and Chapter 1

(1941). *Das Übermikroskop als Forschungsmittel. Vorträge, gehalten anlaßlich der Eröffnung des Laboratoriums für Übermikroskopie der Siemens & Halske A.G., Berlin* (Walter de Gruyter, Berlin).

(1942). *Geometrische Elektronenoptik dargestellt unter besonderer Berücksichtigung des Elektronen- und Übermikroskops* (Der Studentische Kulturaustausch und die Universtät, Heidelberg). Published in French as *Optique Electronique Géométrique*, Cahiers d'Etudes Européennes.

(1971). *Prismennye Beta-Spektrometry i ikh Primenenie* (Vilnius).

V. P. Afanas'ev and S. Ya. Yavor (1978). *Elektrostaticheskie Energoanalizatory dlya Puchkov Zaryazhennykh Chastits* (Nauka, Moscow).

C. C. Ahn, Ed. (2004). *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas*, 2nd edn (Wiley–VCH, Weinheim).

I. V. Alyamovskii (1966). *Elektronnyi Puchki i Elektronnyi Pushki* (Sovetskoe Radio, Moscow).
S. Amelinckx, D. Van Dyck, J. Van Landuyt and G. Van Tendeloo, Eds (1997). *Electron Microscopy, Principles and Fundamentals* (Wiley−VCH, Weinheim). Extracted from *Handbook of Microscopy*, 3 Vols, 1997.
J. L. Balladore and R. Murillo (1977). Nouveau type de lentille magnétique pour microscope électronique. *J. Microsc. Spectrosc. Electron.* **2**, 211−222.
A. P. Banford (1966). *The Transport of Charged Particle Beams* (Spon, London).
L. A. Baranova and S. Ya. Yavor (1989). The optics of round and multipole electrostatic lenses. *Adv. Electron. Electron Phys.* **76**, 1−207.
L. A. Baranova and S. Ya. Yavor (1986). *Elektrostaticheskie Elektronnye Linzy* (Nauka, Moscow).
E. Bauer (2014). *Surface Microscopy with Low Energy Electrons* (Springer, New York).
D. C. Bell and N. Erdman, Eds (2013). *Low Voltage Electron Microscopy: Principles and Applications* (Wiley, Chichester and RMS, Oxford).
M. Berz (1999). Modern map methods in particle beam physics. *Adv. Imaging Electron Phys.* **108**, 1−318.
M. Berz, K. Makino and W. Wan (2015). *An Introduction to Beam Physics* (CRC Press, Boca Raton).
H. Boersch, O. Bostanjoglo and K. Grohmann (1966). Supraleitender Hohlzylinder als magnetische Linse. *Z. Angew. Phys.* **20**, 193−194.
B. E. Bonshtedt and M. G. Markovich (1967). *Fokusirovka i Otklonenie Puchkov v Elektronno-luchevykh Priborakh* (Sovetskoe Radio, Moscow).
G. Botton (2008). Analytical electron microscopy. In *Science of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), **1**, 273−405 (Springer, New York).
G. Botton and S. Prabhudev (2018). Analytical electron microscopy. In *Handbook of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), (Springer, Cham). forthcoming.
B. Breton, D. McMullan and K. C. A. Smith, Eds (2004). *Sir Charles Oatley and the scanning electron microscope*. *Adv. Imaging Electron Phys.* **133**.
M. Bronsgeest (2014). *Physics of Schottky Electron Sources. Theory and Optimum Operation* (Pan Stanford, Singapore).
E. Brüche (1941). Zur Entwicklung des Elektronen-Übermikroskops mit elektrostatischen Linsen. *Z. Verein. Deutsch. Ingenieure* **85** (10), 221−228.
E. Brüche and W. Henneberg (1936). Geometrische Elektronenoptik (Entwicklung von 1934 bis Mitte 1936). *Ergebnisse Exakten Naturwiss.* **15**, 365−421.
E. Brüche and A. Recknagel (1941). *Elektronengeräte, Prinzipien und Systematik* (Springer, Berlin).
E. Brüche and O. Scherzer (1934). *Geometrische Elektronenoptik* (Springer, Berlin).
R. Brydson, Ed. (2011). *Aberration-Corrected Analytical Transmission Electron Microscopy* (Royal Microscopical Society, Oxford).
H. Busch (1926). Berechnung der Bahn von Kathodenstrahlen im axialsymmetrischen elektromagnetischen Felde. *Ann. Phys. (Leipzig)* **81** (386), 974−993.
H. Busch (1927). Über die Wirkungsweise der Konzentrierungsspule bei der Braunschen Röhre. *Arch. Elektrotech.* **18**, 583−594.
H. Busch and E. Brüche, Eds (1937). *Beiträge zur Elektronenoptik* (Barth, Leipzig).
D. C. Carey (1987). *The Optics of Charged Particle Beams* (Harwood, Chur, London & New York).
C. B. Carter and D. B. Williams, Eds (2016). *Transmission Electron Microscopy. Diffraction, Imaging, and Spectrometry* (Springer, Switzerland).
R. Castaing (1951). Application des sondes électroniques à une méthode d'analyse ponctuelle chimique et cristallographique. Dissertation, Paris.
R. Castaing and L. Henry (1962). Filtrage magnétique des vitesses en microscopie électronique. *C. R. Acad. Sci. Paris* **255**, 76−78.
W.-x. Chen (1986). *Introduction to Electron Optics* [in Chinese] (Beijing University Press, Beijing).
D. Cockayne, A. I. Kirkland, P. D. Nellist and A. Bleloch, Eds (2009). New possibilities with aberration-corrected electron microscopy. *Philos. Trans. R. Soc. Lond. A* **367**, 3631−3870.
C. Colliex (1998). *La Microscopie Electronique* (Presses Universitaires de France, Paris).

C. Colliex (2008). *Elektronenmikroskopie, eine anwendungsbezogene Einführung* (Wissenschaftliche Verlagsgesellschaft, Darmstadt). German translation by H. Kohl of Colliex, 1998.
V. E. Cosslett (1946). *Introduction to Electron Optics*. (Clarendon Press, Oxford; 2nd edn, 1950).
V. E. Cosslett (1947). *The Electron Microscope* (Sigma Books, London).
V. E. Cosslett (1950). *Bibliography of Electron Microscopy* (Arnold, London).
V. E. Cosslett (1966). *Modern Microscopy or Seeing the Very Small* (Bell, London).
V. E. Cosslett (1981). The development of electron microscopy and related techniques at the Cavendish Laboratory, Cambridge, 1946–79. Part 1. 1946–60; Part II. *Contemp. Phys.* **22**, 3–36 & 147–182.
V. E. Cosslett and P. Duncumb (1956). A scanning microscope with either electron or X-ray recording. EUREM-1, Stockholm. 12–14.
A. V. Crewe (1965). A scanning transmission electron microscope utilizing electron energy losses. Conference on Non-conventional Electron Microscopy, Cambridge. Unpublished abstract.
A. V. Crewe (1970). The current state of high resolution scanning electron microscopy. *Q. Rev. Biophys.* **3**, 137–175.
A. V. Crewe (1973). Production of electron probes using a field emission source. *Prog. Opt.* **11**, 223–246.
A. V. Crewe (1983). High-resolution scanning transmission electron microscopy. *Science* **221**, 325–330.
A. V. Crewe (1984). An introduction to the STEM. *J. Ultrastruct. Res.* **88**, 94–104.
A. V. Crewe (2009). The work of Albert Victor Crewe on the scanning transmission electron microscope and related topics. *Adv. Imaging Electron Phys.* **159**, 1–61.
A. V. Crewe, J. Wall and L. M. Welter (1968). A high-resolution scanning transmission electron microscope. *J. Appl. Phys.* **39**, 5861–5868.
P. Dahl (1973). *Introduction to Electron and Ion Optics* (Academic Press, New York & London).
C. J. Davisson and L. H. Germer (1927a). Diffraction of electrons by a crystal of nickel. *Phys. Rev.* **30**, 705–740.
C. J. Davisson and L. H. Germer (1927b). The scattering of electrons by a single crystal of nickel. *Nature* **119**, 558–560.
L. de Broglie (1925). Recherches sur la théorie des quanta. *Ann. Phys. (Paris)* **3**, 22–128.
L. de Broglie, Ed. (1946). *L'Optique Electronique* (Editions de la Revue d'Optique Théorique et Instrumentale, Paris).
L. de Broglie (1950). *Optique Electronique et Corpusculaire* (Hermann, Paris).
N. D. Devyatkov, Ed. (1977). *Metody Rascheta Elektronno-opticheskykh Sistem* (Nauka, Moscow).
I. Dietrich (1976). *Superconducting Electron-optic Devices* (Plenum, New York & London).
I. Dietrich, H. Pfisterer and R. Weyl (1969). Eine supraleitende Linse für das Elektronenmikroskop. *Z. Angew. Phys.* **28**, 35–39.
M. M. Disko, C. C. Ahn and B. Fultz, Eds (1992). *Transmission Electron Energy Loss Spectrometry in Materials Science* (Minerals, Metals and Materials Society, Warrendale PA).
M. Drechsler (1978). Erwin Müller and the early development of field emission microscopy. *Surf. Sci.* **70**, 1–18.
E. Driest and H. O. Müller (1935). Elektronenmikroskopische Aufnahmen (Elektronenmikrogramme) von Chitinobjekten. *Z. Wiss. Mikrosk.* **52**, 53–57.
P. Duncumb (1958). Microanalysis with the X-ray scanning microscope. ICEM-4, Berlin. **1**, 267–269.
G. Dupouy (1952). *Eléments d'Optique Electronique* (Armand Colin, Paris; 2nd edn, 1961).
G. Dupouy (1968). Electron microscopy at very high voltages. *Adv. Opt. Electron Microsc.* **2**, 167–250. Reprinted in *Adv. Imaging Electron Phys*. **201** (2017) 261–340.
G. Dupouy (1985). Megavolt electron microscopy. In *The Beginnings of Electron Microscopy* (P. W. Hawkes, Ed.), 103–165, *Adv. Electron. Electron Phys*. Suppl. 16.
G. Dupouy, F. Perrier and L. Durrieu (1960). L'observation de la matière vivante au moyen d'un microscope électronique fonctionnant sous très haute tension. *C. R. Acad. Sci. Paris* **251**, 2836–2841.
R. F. Egerton (1996). *Electron Energy-loss Spectrosopy in the Electron Microscope* (Plenum, New York; 2nd edn, 1996; 3rd edn, 2011, published by Springer, New York).

A. B. El-Kareh and J. C. J. El-Kareh (1970). *Electron Beams, Lenses and Optics* 2 Vols. (Academic Press, New York & London).
R. Erni (2010). *Aberration-Corrected Imaging in Transmission Electron Microscopy. An Introduction*, second edn (Imperial College Press, London; first edn, 2015).
F. Ernst and M. Rühle, Eds (2003). *High-Resolution Imaging and Spectroscopy of Materials* (Springer, Berlin).
N. J. Félici (1965). *Introduction à l'Optique Corpusculaire* (Gauthier−Villars, Paris).
H. Fernández-Morán (1965). Electron microscopy with high-field superconducting solenoid lenses. *Proc. Natl Acad. Sci. U.S.A.* **53**, 445−451.
M. Fournier (2009). Electron microscopy in Second World War Delft. In *Scientific Research in World War II. What Scientists did in the War* (A. Maas and H. Hooijmaijers, Eds), 77−95 (Routledge, London & New York).
H. Fujita (1986). *History of Electron Microscopes* (Business Center for Academic Societies Japan, Tokyo).
B. Fultz and J. Howe (2001). *Transmission Electron Microscopy and Diffractometry* (Springer, Berlin; 2nd edn, 2002; 3rd edn, 2008; 4th edn, 2013).
D. Gabor (1945). *The Electron Microscope. Its Develoment, Present Performance and Future Possibilities* (Hulton, London; 2nd edn, Chemical Publishing Co, Brooklyn, New York, 1948).
D. Gabor (1957). Die Entwicklungsgeschichte des Elektronenmikroskops. *Elektrotech. Z.* **A78**, 522−530.
H. R. Gelderblom and D. H. Krüger (2014). Helmut Ruska (1908−1973): his role in the evolution of electron microscopy in the life sciences, and especially virology. *Adv. Imaging Electron Phys.* **182**, 1−94.
M. E. Gettner and L. Ornstein (1956). Microtomy. In *Physical Techniques in Biological Research* (G. Oster and A. W. Pollister, Eds), **3**, 627−686 (Academic Press, New York & London).
W. Glaser (1952). *Grundlagen der Elektronenoptik* (Springer, Vienna).
W. Glaser (1956). Elektronen- und Ionenoptik. In *Handbuch der Physik* (S. Flügge, Ed.), **33**, 123−395 (Springer, Berlin).
R. H. Good and E. W. Müller (1956). Field emission. In *Handbuch der Physik* (S. Flügge, Ed.), **21**, 176−231 (Springer, Berlin).
G. W. Grime and F. Watt (1984). *Beam Optics of Quadrupole Probe-Forming Systems* (Adam Hilger, Bristol).
G. A. Grinberg (1942). General theory of the focusing action of electrostatic and magnetostatic fields. I. Two-dimensional fields. II. Three-dimensional electrostatic fields [in English]. *Dokl. Akad. Nauk SSR* **37**, 172−178 & 261−268.
G. A. Grinberg (1943a). General theory of the focusing action of electrostatic and magnetostatic fields. III. Three-dimensional (twisted) trajectories in the presence of both an electrostatic and a magnetic field [in English]. *Dokl. Akad. Nauk SSR* **38**, 78−81.
G. A. Grinberg (1943b). General theory of the focusing action of static electric and magnetic fields. *Zh. Tekh. Fiz.* **13**, 361−388. Russian translation of papers in *Doklady*.
P. Grivet (1965). *Electron Optics* (Pergamon, London & New York; 2nd edn, 1972; revised translation of Grivet et al., 1955 & 1958).
P. Grivet, M. Y. Bernard and A. Septier (1955). *Optique Electronique,* Vol. 1*, Lentilles Electroniques* (Bordas, Paris).
P. Grivet, M. Y. Bernard, F. Bertein, R. Castaing, M. Gauzit and A. Septier (1958). *Optique Electronique,* Vol. 2*, Microscopes, Diffractographes, Spectrographes de Masse, Oscillographes Cathodiques* (Bordas, Paris).
J. Grosser (1983). *Einführung in die Teilchenoptik* (Teubner, Stuttgart).
T. R. Groves (2015). *Charged Particle Optics Theory* (CRC Press, Boca Raton).
H. Grümm and P. Schiske (1996). Reminiscences of Walter Glaser. *Adv. Imaging Electron Phys.* **96**, 59−66.
M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius and K. Urban (1998). Electron microscopy image enhanced. *Nature* **392**, 768−769.
M. E. Haine (1961). *The Electron Microscope, the Present State of the Art* (Spon, London).
C. E. Hall (1953). *Introduction to Electron Microscopy* (McGraw-Hill, New York & London; 2nd edn, 1966).
T. W. Hansen and J. B. Wagner, Eds (2016). *Controlled Atmosphere Transmission Electron Microscopy. Principles and Practice* (Springer, Cham).
D. F. Hardy (1973). Superconducting electron lenses. *Adv. Opt. Electron Microsc.* **5**, 201−237.
E. Harting and F. H. Read (1976). Electrostatic Lenses (Elsevier, Amsterdam, Oxford & New York).

P. Hatschek (1937). *Optik des Unsichtbaren. Eine Einführung in die Welt der Elektronen-Optik* (Franckh'sche Verlagshandlung, Stuttgart).
P. Hatschek (1944). *Electron Optics* (American Photographic Publishing Co., Boston; translation of Hatchek, 1937).
P. W. Hawkes (1966). *Quadrupole Optics* (Springer, Berlin & New York).
P. W. Hawkes (1968). Asymptotic aberration coefficients, magnification and object position in electron optics. *Optik* **27**, 287–304.
P. W. Hawkes (1970). Quadrupoles in electron lens design. *Adv. Electron. Electron Phys*. Suppl. 7.
P. W. Hawkes (1972). *Electron Optics and Electron Microscopy* (Taylor & Francis, London).
P. W. Hawkes, Ed. (1973). *Image Processing and Computer-Aided Design in Electron Optics* (Academic Press, London & New York).
P. W. Hawkes, Ed. (1982). *Magnetic Electron Lenses* (Springer, Berlin & New York).
P. W. Hawkes, Ed. (1985). *The Beginnings of Electron Microscopy* (Academic Press, Orlando & London). *Adv. Electron. Electron Phys*. Suppl. 16.
P. W. Hawkes (1995). A lifetime in electron optics: a biographical sketch of Tom Mulvey with a bibliography. *J. Microsc. (Oxford)* **179**, 97–104.
P. W. Hawkes, Ed. (2008). Aberration-corrected electron microscopy. *Adv. Imaging Electron Phys.* **153**.
P. W. Hawkes (2009a). Aberration correction past and present. *Philos. Trans. R. Soc. (Lond.) A* **367**, 3637–3664.
P. W. Hawkes, Ed. (2009b). Cold field emission and the scanning transmission electron microscope. *Adv. Imaging Electron Phys.* **159**.
P. W. Hawkes (2009c). Two commercial STEMs: the Siemens ST100F and the AEI STEM-1. *Adv. Imaging Electron Phys.* **159**, 187–219.
P. W. Hawkes (2015). The correction of electron lens aberrations. *Ultramicroscopy* **156**, A1–A64.
P. W. Hawkes and D. McMullan (2004). A forgotten French scanning electron microscope and a forgotten text on electron optics. *Proc. R. Microsc. Soc.* **39**, 285–290.
P. W. Hawkes and J. C. H. Spence, Eds (2007). *Science of Microscopy* (Springer, New York; corrected printing 2008).
P. W. Hawkes and J. C. H. Spence, Eds (2018). *Handbook of Microscopy* (Springer, Cham).
P. W. Hawkes and U. Valdrè (1977). Superconductivity and electron microscopy. *J. Phys. E: Sci. Instrum.* **10**, 309–328.
D. W. O. Heddle (2000). *Electrostatic Lens Systems* (Institute of Physics Publishing, Bristol & Philadelphia).
Z.-y. Hua and G.-x. Gu (1993). *Electron Optics* [in Chinese] (Fudan University Press, Shanghai).
L.-y. Huang and X.-p. Liu (1991). *Electron Microscopy and Electron Optics* [in Chinese] (Beijing Science Press, Beijing).
S. Humphries (1986). *Principles of Charged Particle Acceleration* (Wiley–Interscience, New York & Chichester).
V. P. Il'in (1974). *Chislennye Metody Resheniya Zadach Elektrooptiki* (Nauka, Novosibirsk).
V. P. Il'in, Ed. (1982). *Metody Rascheta Elektronno-opticheskikh Sistem* (Computer Centre, Novosibirsk).
V. P. Il'in, Ed. (1983). *Algoritmy i Metody Rascheta Elektronno-opticheskikh Sistem* (Computer Centre, Novosibirsk).
V. P. Il'in, V. A. Kateshov, Yu. V. Kulikov and M. A. Monastyrskii (1987). *Chislennye Metody Optimisatsii Emissionnykh Elektronno-opticheskikh Sistem* (Nauka, Novosibirsk).
V. P. Il'in, V. A. Kateshov, Yu. V. Kulikov and M. A. Monastyrsky (1990). Emission-imaging electron-optical system design. *Adv. Electron. Electron Phys.* **78**, 155–278.
D. Ioanoviciu (1989). Ion optics. *Adv. Electron. Electron Phys.* **73**, 1–92.
L. Jacob (1950). *Introduction to Electron Optics* (Methuen, London).
P. P. Kas'yankov (1956). *Teoriya Elektromagnitnykh Sistem s Krivolineinoi Os'yu* (Leningrad University Press, Leningrad).
E. Kasper (1982). Magnetic field calculation and the determination of electron trajectories. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 57–118 (Springer, Berlin & New York).
E. Kasper (1984). Recent developments in numerical electron optics. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. Lenz, E. Munro and P. B. Sewell, Eds), 63–73 (Scanning Electron Microscopy, AMF O'Hare).

E. Kasper (2001). Numerical field calculation for charged particle optics. *Adv. Imaging & Electron Phys* **116**, 1–451.
E. Kasper and F. Lenz (1980). Numerical methods in geometrical electron optics. EUREM-7, The Hague. **1**, 10–15.
V. M. Kel'man (1955). *Elektronnaya Optika* (Izd. Akad. Nauk SSSR, Moscow & Leningrad; 2nd edn, 1968, Nauka, Moscow).
V. M. Kel'man and S. Ya. Yavor (1959). *Elektronnaya Optika* (Izd. Akad. Nauk SSSR, Moscow & Leningrad; 2nd edn, 1963; 3rd edn, 1968 published by Nauka, Leningrad).
V. M. Kel'man, S. P. Karetskaya, L. V. Fedulina and E. M. Yakushev (1979). *Elektronno-opticheskie Elementy Prismennykh Spektrometrov Zaryazhennykh Chastits* (Nauka, Alma-Ata).
V. M. Kel'man, I. V. Rodnikova and L. M. Sekunova (1985). *Staticheskie Mass-Spektrometry* (Nauka, Alma-Ata).
A. Khursheed (1999). *The Finite-Element Method in Charged Particle Optics* (Kluwer, Boston, Dordrecht & London).
A. Khursheed (2011). *Scanning Electron Microscope Optics and Spectrometers* (World Scientific, Singapore).
K.-H. Kim and S.-K. Choi (1963). *Optics of Charged Particles* [in Korean] (Korean Academy of Sciences, Pyong-Yang).
A. I. Kirkland and J. L. Hutchison, Eds (2007). *Nanocharacterisation* (Royal Society of Chemistry, Cambridge).
A. I. Kirkland and S. Haigh, Eds (2015). *Nanocharacterisation* (Royal Society of Chemistry, Cambridge; 2nd edn of Kirkland and Hutchison, 2007).
O. Klemperer (1939). *Electron Optics* (Cambridge University Press, Cambridge; 2nd edn, 1953).
O. Klemperer (1961). *Electron Physics, the Physics of the Free Electron* (Butterworths, London).
O. Klemperer and M. E. Barnett (1971). *Electron Optics* (Cambridge University Press, Cambridge).
M. Knoll and J. Eichmeier (1966). *Technische Elektronik,* Vol. 2, Stromsteuernde und Elektronenoptische Entladungsgeräte (Springer, Berlin).
M. Knoll and E. Ruska (1932a). Beitrag zur geometrischen Elektronenoptik, I & II. *Ann. Phys. (Leipzig)* **12** (404), 607–640 & 641–661.
M. Knoll and E. Ruska (1932b). Das Elektronenmikroskop. *Z. Phys.* **78**, 318–339.
V. I. Kotov and V. V. Miller (1969). *Fokusirovka i Razdelenie po Massam Chastits Vysokykh Energii* (Atomizdat, Moscow).
W. Krakow and M. O'Keefe (1989). *Computer Simulation of Electron Microscope Diffraction and Images* (Minerals, Metals and Materials Society, Warrendale PA).
F. Krause (1936). Elektronenoptische Aufnahmen von Diatomen mit dem magnetischen Elektronenmikroskop. *Z. Phys.* **102**, 417–422.
O. Krivanek, N. Dellby, A. J. Spence, R. A. Camps and L. M. Brown (1997). Aberration correction in the STEM. EMAG, Cambridge, 35–39.
O. L. Krivanek, G. J. Corbin, N. Dellby, B. F. Elston, R. J. Keyse, M. F. Murfitt, C. S. Own, Z. S. Szilagyi and J. W. Woodruff (2008). An electron microscope for the aberration-corrected era. *Ultramicroscopy* **108**, 179–195.
A. Laberrigue and P. Levinson (1964). Utilisation des propriétés des fils supraconducteurs en microscopie électronique. *C. R. Acad. Sci. Paris* **259**, 530–532.
L. Lambert and T. Mulvey (1996). Ernst Ruska (1906–1988), designer extraordinaire of the electron microscope: a memoir. *Adv. Imaging Electron Phys.* **95**, 2–62.
P. Lapostolle and A. Septier, Eds (1970). *Linear Accelerators* (North-Holland, Amsterdam).
J. D. Lawson (1977). *The Physics of Charged-Particle Beams* (Clarendon Press, Oxford; 2nd edn, 1988).
L. Léauté, [*sic*, for A. Léauté] (1946). Les applications du microscope électronique à la métallographie. In *l'Optique Electronique* (L. de Broglie, Ed.), 209–220 (Editions de la Revue d'Optique Théorique et Expérimentale, Paris).
A. A. Lebedev, Ed. (1954). *Elektronnaya Mikroskopiya* (Gostekhizdat, Moscow).
G. Lefranc, E. Knapek and I. Dietrich (1982). Superconducting lens design. *Ultramicroscopy* **10**, 111–123.

S. Leisegang (1956). Elektronenmlkroskope. In *Handbuch der Physik* (S. Flügge, Ed.), **33**, 396–545.
F. Lenz (1956). Über asymptotische Bildfehler. EUREM-1, Stockholm. 48–51.
F. Lenz (1957). Über asymptotische Bildfehler. *Optik* **14**, 74–82.
F. Lenz (1973). Computer-aided design of electron optical systems. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 274–282 (Academic Press, London & New York).
K. G. Lickfeld (1979). *Elektronenmikroskopie. Eine Einführung in die Grundlagen der Durchstrahlungs-Elektronenmikroskopie und ihrer Präparationstechniken* (Eugen Ulmer, Stuttgart).
H. Liebl (2008). *Applied Charged Particle Optics* (Springer, Berlin).
J. J. Livingood (1969). *The Optics of Dipole Magnets* (Academic Press, New York & London).
V. S. Lukoshkov, B. M. Zamorozkov and L. V. Shubin (1968). *Elektronnye Puchki i Elektronno-opticheskie Sistemy* (Institut 'Elektronika', Moscow).
V. M. Lyuk'yanovich (1960). *Elektronnaya Mikroskopiya v Fiziko-khimicheskikh Issledovaniyakh. Metodika i Primenenie* (Izd. Akad. Nauk SSSR, Moscow).
A. Maas and H. Hooijmaijers, Eds (2009). *Scientific Research in World War II. What Scientists did in the War* (Routledge, London & New York).
C. Magnan, Ed. (1961). *Traité de Microscopie Electronique* 2 Vols (Hermann, Paris).
G. I. Marchuk, Ed. (1967). *Chislennye Metody Rascheta Elektronno-opticheskikh Sistem* (Nauka, Novosibirsk).
G. I. Marchuk, Ed. (1970). *Metody Rascheta Elektronno-opticheskikh Sistem* (Computer Centre, Novosibirsk).
G. I. Marchuk, Ed. (1973). *Metody Rascheta Elektronno-opticheskikh Sistem* 2 Vols (Computer Centre, Novosibirsk).
M. Marko and H. Rose (2010). The contributions of Otto Scherzer (1909–1982) to the development of the electron microscope. *Microsc. Microanal.* **16**, 366–374.
L. C. Martin (1938). The electron microscope. *Nature* **142**, 1062–1065.
L. C. Martin, R. V. Whelpton and D. H. Barnum (1937). A new electron microscope. *J. Sci. Instrum.* **14**, 14–24.
L. Marton (1934a). Electron microscopy of biological objects. *Nature* **133**, 911.
L. Marton (1934b). Electron microscopy of biological objects. *Phys. Rev.* **46**, 527–528.
L. Marton (1934c). La microscopie électronique des objets biologiques. *Bull. Cl. Sci. Acad. R. Belg.* **20**, 439–446.
L. Marton (1935a). La microscopie électronique des objets biologiques (deuxième et troisième communications). *Bull. Cl. Sci. Acad. R. Belg.* **21**, 553–564 & 606–617.
L. Marton (1935b). Le microscope électronique et ses applications. *Rev. Opt.* **14**, 129–145.
L. Marton (1936). La microscopie électronique des objets biologiques (quatrième communication). *Bull. Cl. Sci. Acad. R. Belg.* **22**, 1336–1344.
L. Marton (1937). La microscopie électronique des objets biologiques (cinquième communication). *Bull. Cl. Sci. Acad. R. Belg.* **23**, 672–675.
L. Marton (1968). *Early History of the Electron Microscope* (San Francisco Press, San Francisco; 2nd edn, 1994).
L. Marton (1976). Early application of electron microscopy to biology. *Ultramicroscopy* **1**, 281–296.
C. Marton, S. Sass, M. Swerdlow, A. van Bronkhorst and H. Meryman (1950). Bibliography of Electron Microscopy. *NBS Circular* **502**.
D. McMullan (1953). An improved scanning electron microscope for opaque specimens. *Proc. Inst. Elec. Eng.* **100**, 245–259.
A. J. Melmed (1996). Recollections of Erwin Müller's laboratory: the development of FIM (1951–1956). *Appl. Surf. Sci.* **94-95**, 17–25.
P. G. Merli and M. V. Antisari, Eds (1992). *Electron Microscopy in Materials Science* (World Scientific, Singapore).
S. I. Molokovskii and A. D. Sushkov (1965). *Elektronnoopticheskie Sistemy Priborov Sverkhvysokhikh Chastot* (Nauka, Moscow & Leningrad).
E. W. Müller (1960). Field ionization and field ion microscopy. *Adv. Electron. Electron Phys.* **13**, 83–179.
E. W. Müller (1975). The atom-probe field ion microscope. In *Methods of Surface Analysis* (A. W. Czanderna, Ed.), 329–378 (Elsevier, Amsterdam).

F. Müller (2009). The birth of a modern instrument and its development during World War II: electron microscopy in Germany from the 1930s to 1945. In *Scientific Research in World War II. What Scientists Did in the War* (D. Maas and H. Hooijmaijers, Eds), 121–146 (Routledge, London & New York).
E. W. Müller and Tsong (1969). *Field-Ion Microscopy, Principles and Applications* (Elsevier, New York & Amsterdam).
E. W. Müller and T. T. Tsong (1974). Field ion microscopy, field ionization and field evaporation. *Prog. Surf. Sci.* **4**, 1–139.
T. Mulvey (1962). Origins and historical development of the electron microscope. *Br. J. Appl. Phys.* **13**, 197–207.
T. Mulvey (1967). The history of the electron microscope. *Proc. R. Microsc. Soc.* **2**, 201–227.
T. Mulvey (1973). Forty years of electron microscopy. *Phys. Bull.* **24**, 147–154.
T. Mulvey (1982). Unconventional lens design. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 359–412 (Springer, Berlin & New York).
T. Mulvey (1984). Magnetic electron lenses II. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 15–27 (Scanning Electron Microscopy, AMF O'Hare).
T. Mulvey (1985). The industrial development of the electron microscope by the Metropolitan Vickers Electrical Company and AEI Limited. In *The Beginnings of Electron Microscopy* (P. W. Hawkes, Ed.), 417–442 (Academic Press, Orlando & London). *Adv. Electron. & Electron Phys*. Suppl. 16.
T. Mulvey (1995). Gabor's pessimistic 1942 view of electron microscopy and how he stumbled on the Nobel prize. *Adv. Imaging Electron Phys* **91**, 259–283.
T. Mulvey, Ed. (1996). The Growth of Electron Microscopy. *Adv. Imaging Electron Phys.* **96**.
T. Mulvey and D. J. J. van de Laak-Tijssen (2001). Jan Bart Le Poole (1917–1993) pioneer of the electron microscope and particle optics. *Adv. Imaging Electron Phys.* **115**, 287–354.
R. Murillo, J.-L. Balladore, R. Fourmaux and C. Tremollières (1987). Electron lenses with anisotropic magnetic circuit. Part 1: Why and how to reduce the volume of electron magnetic lenses. Part 2: Realisation of a new type of magnetic lens. *J. Microsc. Spectrosc. Electron.* **12**. 263–279 & 281–298.
L. M. Myers (1939). *Electron Optics, Theoretical and Practical* (Chapman & Hall, London).
S. B. Newman, E. Borysko and M. Swerdlow (1949). Ultra-microtomy by a new method. *J. Res. Natl. Bur. Stand.* **A43**, 183–199.
C. W. Oatley (1972). *The Scanning Electron Microscope.* Part I. *The Instrument* (Cambridge University Press, Cambridge).
C. W. Oatley (1982). The early history of the scanning electron microscope. *J. Appl. Phys.* **53**, R1–R13.
C. W. Oatley, W. C. Nixon and R. F. W. Pease (1965). Scanning electron microscopy. *Adv. Electron. Electron Phys.* **21**, 181–247.
C. W. Oatley, D. McMullan and K. C. A. Smith (1985). The development of the scanning electron microscope. In *The Beginnings of Electron Microscopy* (P. W. Hawkes, Ed.), 443–482 (Academic Press, Orlando & London). *Adv. Electron. Electron Phys*. Suppl. 16.
F. Ollendorff (1955). *Technische Elektrodynamik,* Vol. II, Innere Elektronik, Pt I, Elektronik des Einzelelektrons (Springer, Vienna).
J. Orloff, Ed. (1997). *Handbook of Charged Particle Optics* (CRC Press, Boca Raton; 2nd edn, 2009).
S. Ozasa, S. Katagiri, H. Kimura and B. Tadano (1966). Superconducting electron lens. ICEM-6, Kyoto. **1**, 149–150.
B. Paszkowski (1960). *Optyka Elektronowa* (Wydawnictwa Naukowo-Techniczne, Warsaw; 2nd edn, 1965).
B. Paszkowski (1968). *Electron Optics* (Iliffe, London, American Elsevier, New York). Translation of 1965 Polish edition.
D. C. Pease and R. F. Baker (1948). Sectioning techniques for electron microscopy using a conventional microtome. *Proc. Soc. Exp. Biol. Med.* **67**, 470–474.
S. J. Pennycook and P. D. Nellist, Eds (2011). *Scanning Transmission Electron Microscopy. Imaging and Analysis* (Springer, New York).

J. Picht (1939). *Einführung in die Theorie der Elektronenoptik* (Barth, Leipzig; 2nd edn, 1957; 3rd edn 1963).
J. Picht (1945). *Was Wir über Elektronen Wissen* (Carl Winter, Heidelberg).
J. Picht and R. Gain (1955). *Das Elektronenmikroskop* (Fachbuch, Leipzig).
J. Picht and J. Heydenreich (1966). *Einführung in die Elektronenmikroskopie* (Verlag Technik, Berlin).
J. R. Pierce (1949). *Theory and Design of Electron Beams* (Van Nostrand, Princeton & London; 2nd edn, 1954).
L. Qing (1995). *Zur Frühgeschichte des Elektronenmikroskops* (GNT-Verlag, Stuttgart).
D. Quaglino, E. Falcieri, M. Catalano, A. Diaspro, A. Montone, P. Mengucci and C. Pellicciari, Eds (2006). *1956–2006. Cinquanta Anni di Microscopia in Italia tra Storia, Progresso ed Innovazione* (Società Italiana di Scienze Microscopiche).
C. Ramsauer, Ed. (1941). *Zehn Jahre Elektronenmikroskopie; ein Selbstbericht des AEG Forschungs-Instituts* (Springer, Berlin; 2nd edn 1942).
C. Ramsauer, Ed. (1943). Elektronenmikroskopie: Bericht über Arbeiten des AEG Forschungs-Instituts, 1930 bis 1942 (Springer, Berlin).
N. Rasmussen (1997). *Picture Control. The Electron Microscope and the Transformation of Biology in America, 1940–1960* (Stanford University Press, Stanford).
L. Reimer (1984). *Transmission Electron Microscopy. Physics of Image Formation* (Springer, Berlin; 2nd edn, 1989; 3rd edn, 1993; 4th edn, 1997; 5th edn, see Reimer and Kohl, 2008).
L. Reimer (1985). *Scanning Electron Microscopy. Physics of Image Formation and Microanalysis* (Springer, Berlin; 2nd edn, 1998).
L. Reimer (1993). *Image Formation in the Low-Voltage Scanning Electron Microscope* (SPIE Optical Engineering Press, Bellingham WA).
L. Reimer, Ed. (1995). *Energy-Filtering Transmission Electron Microscopy* (Springer, Berlin & New York).
L. Reimer and H. Kohl (2008). *Transmission Electron Microscopy. Physics of Image Formation* (Springer, New York; 5th edn of Reimer, 1984).
J. H. Reisner (1989). An early history of the electron microscope in the United States. *Adv. Electron. Electron Phys.* **73**, 133–231.
W. D. Riecke (1982). Practical lens design. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 163–357 (Springer, Berlin & New York).
H. Rose (1968). Über die Berechnung der Bildfehler elektronenoptischer Systeme mit gerader Achse, I & II. *Optik* **27**, 466–474 & 497–514.
H. Rose (1983). Otto Scherzer gestorben. *Optik* **63**, 185–188.
H. Rose (2008). History of direct aberration correction. *Adv. Imaging Electron Phys.* **153**, 1–37.
H. Rose (2009). Historical aspects of aberration correction. *J. Electron Microsc.* **58**, 77–85.
H. Rose (2012). Geometrical Charged-particle Optics, second edn. (Springer, Berlin; first edn, 2009).
R. Rühle (1949). *Das Elektronenmikroskop* (Curt E. Schwab, Stuttgart).
R. Rüdenberg (2010). Origin and background of the invention of the electron microscope. *Adv. Imaging Electron Phys* **160**, 171–205.
E. Ruska (1957). 25 Jahre Elektronenmikroskopie. *Elektrotech. Z.* **A78**, 531–543.
E. Ruska (1979). Die Frühe Entwicklung der Elektronenlinsen und der Elektronenmikroskopie. *Acta Hist. Leopold*. No. 12, 136 pp.
E. Ruska (1980). The early development of electron lenses and electron microscopy. *Microsc. Acta*. Suppl. 5. Also published as a book by Hirzel, Stuttgart. Translation by T. Mulvey of the preceding entry.
E. Ruska (1984). Die Entstehung des Elektronenmikroskops (Zusammenhang zwischen Realisierung und erster Patentanmeldung, Dokumente einer Erfindung). *Arch. Geschichte Naturwiss.*, 525–551. Nos. 11/12.
E. Ruska (1986). The emergence of the electron microscope. Connection between realization and first patent application, documents of an invention. *J. Ultrastruct. Mol. Struct. Res.* **95**, 3–28. Translation of Ruska, 1984.
E. Ruska and M. Knoll (1931). Die magnetische Sammelspule für schnelle Elektronenstrahlen. *Z. Tech. Phys.* **12**. 389–400 & 448.
A. A. Rusterholz (1950). Elektronenoptik, Vol. 1, *Grundzüge der Theoretischen Elektronenoptik* (Birkhäuser, Basle).
P. Schattschneider, Ed. (2012). *Linear and Chiral Dichroism in the Electron Microscope* (Pan Stanford, Singapore).

O. Scherzer (1936). Über einige Fehler von Elektronenlinsen. *Z. Phys.* **101**, 593–603.
O. Scherzer (1947). Sphärische und chromatische Korrektur von Elektronen-Linsen. *Optik* **2**, 114–132.
O. I. Seman (1951). Reduced form of the fourth order eikonal and of the aberration coefficients in electron optics. *Dokl. Akad. Nauk SSR* **81**, 775–778.
O. I. Seman (1954). Relativistic aberration functions and normal coefficients of electron optical aberrations. *Dokl. Akad. Nauk SSR* **96**, 1151–1154.
O. I. Seman (1955). Relativistic theory of axially symmetric electron optics based on eikonal theory. *Trudy Inst. Fiz. Astron. Akad. Nauk Eston. SSSR* No. 2, 3–49.
O. I. Seman (1958a). On the theory of virtual and transverse aberrations in electron optics. *Uch. Zap. Rostov. Gos. Univ. (Ser. Fiz.)* **68** (8), 63–75.
O. I. Seman (1958b). On the theory of the point eikonal of fourth order in electron optics. *Uch. Zap. Rostov. Gos. Univ. (Ser. Fiz.)* **68** (8), 77–90.
O. I. Seman (1958c). On normal coefficients of electron-optical aberrations of magnetic immersion systems. *Radiotekh. Elektron.* **3**, 283–287. *Radio Eng. Electron*. **3**, 402–409.
O. I. Seman (1959). *Theoretical Foundations of Electron Optics* [in Chinese] (Higher Education Press, Peking).
A. Septier, Ed. (1967). *Focusing of Charged Particles,* 2 Vols (Academic Press, New York & London).
A. Septier, Ed. (1980). *Applied Charged Particle Optics* Vols 1 & 2 (Academic Press, New York & London). *Adv. Electron. Electron Phys*. Suppl. 13A & 13B.
A. Septier, Ed. (1983). *Applied Charged Particle Optics* Vol. 3 (Academic Press, New York & London). *Adv. Electron. Electron Phys*. Suppl. 13C.
B. M. Siegel, N. Kitamura, R. A. Kropfli and M. P. Schulhof (1966). High resolution objective lenses using superconducting materials. ICEM-6, Kyoto. **1**, 151–152.
F. S. Sjöstrand (1953). A new microtome for ultrathin sectioning for high resolution electron microscopy. *Experientia* **9**, 114–115.
F. S. Sjöstrand (1967). *Electron Microscopy of Cells and Tissues,* Vol. 1, Instrumentation and Techniques (Academic Press, New York & London).
K. C. A. Smith (2013). Electron Microscopy at Cambridge University with Charles Oatley and Ellis Cosslett: some reminiscences and recollections. *Adv. Imaging Electron Phys.* **177**, 189–277.
K. C. A. Smith, K. Considine and V. E. Cosslett (1966). A new 750 kV electron microscope. ICEM-6, Kyoto. **1**, 99–100.
A. Sommerfeld and O. Scherzer (1934). Über das Elektronenmikroskop. *München. Med. Wochenschr* **81**, 1859–1860.
J. C. H. Spence (1980). *Experimental High-resolution Electron Microscopy* (Oxford University Press, Oxford; 2nd edn, 1988; 3rd edn, 2003).
J. C. H. Spence (2013). *High-Resolution Electron Microscopy* (Oxford University Press, Oxford; 4th edn of Spence, 1980).
K. G. Steffen (1965). *High Energy Beam Optics* (Wiley–Interscience, New York & London).
A. M. Strashkevich (1959). *Elektronnaya Optika Elektrostaticheskikh Polei ne Obladayushchikh Osevoi Simmetriei* (Gos. Izd. Fiz.-Mat. Lit., Moscow).
A. M. Strashkevich (1966). *Elektronnaya Optika Elektrostaticheskikh Sistem* (Energiya, Moscow & Leningrad).
P. A. Sturrock (1952). The imaging properties of electron beams in arbitrary static electromagnetic fields. *Philos. Trans. R. Soc. Lond. A* **245**, 155–187.
P. A. Sturrock (1955). *Static and Dynamic Electron Optics* (Cambridge University Press, Cambridge).
E. Sugata (1968). *Electron Microscopes of Japan 1936–1965. A Historical Survey* (Maruzen, Tokyo).
N. G. Sushkin (1949). *Elektronnyi Mikroskop* (Gos. Izd. Tekh.-Teor. Lit., Moscow).
M. Szilagyi (1988). *Electron and Ion Optics* (Plenum, New York & London).
H. Szymanski, A. Mulak, A. Duda and A. Romanowski (1977). *Optyka Elektronowa* (Wydawnictwa Naukowo-Techniczne, Warsaw; 2nd edn, 1984).
N. Tanaka, Ed. (2015). *Scanning Transmission Electron Microscopy of Nanomaterials. Basics of Imaging and Analysis* (Imperial College Press, London).
N. Tanaka (2017). *Electron Nano-imaging. Basics of Imaging and Diffraction for TEM and STEM* (Springer, Japan).

T.-t. Tang (1987). *Introduction to Charged Particle Optics with Applications* [in Chinese] (Xi'an Jiaotong Press, Xi'an).
T.-t. Tang (1996). *Advanced Electron Optics* [in Chinese] (Beijing Institute of Technology Press, Beijing).
T.-t. Tang and C.-l. Liu (2001). *The Physics of Electron and Ion Beams* [in Chinese] (Xi'an Jiaotong Press, Xi'an).
V. P. Taranenko (1964). *Elektronnye Pushki* (Tekhnika, Kiev).
G. P. Thomson and A. Reid (1927). Diffraction of cathode rays by a thin film. *Nature* **119**, 890.
A. Tonomura (1993). *Electron Holography* (Springer, Berlin; 2nd edn, 1999).
A. Tonomura (1998). *The Quantum World Unveiled by Electron Waves* (World Scientific, Singapore).
I. I. Tsukkerman (1958). *Elektronnaya Optika v Televidenii* (Gosenergoizdat, Moscow & Leningrad). English translation: *Electron Optics in Television* (Pergamon, Oxford and New York; 1961).
E. A. Vainryb and V. I. Milyutin (1951). *Elektronnaya Optika* (Gos. Energ. Izd., Moscow & Leningrad; cf. Wainrib and Miljutin, 1954).
Yu. V. Vandakurov (1955). On some antisymmetric magnetic fields with double focusing. I. *Zh. Tekh. Fiz.* **25**, 2545–2555.
Yu. V. Vandakurov (1956a). On some antisymmetric magnetic fields with double focusing. II. *Zh. Tekh. Fiz.* **26**, 1599–1610. *Sov. Phys. Tech. Phys*. **1**, 1558–1569.
Yu. V. Vandakurov (1956b). Electron optical systems, the fields of which do not depend on one of the coordinates. *Zh. Tekh. Fiz.* **26**, 2578–2594. *Sov. Phys. Tech. Phys*. **1**, 2491–2507.
Yu. V. Vandakurov (1957). On the theory of the aberrations in electron-optical focusing systems with a curvilinear axis. *Zh. Tekh. Fiz.* **27**, 1850–1862. *Sov. Phys. Tech. Phys*. **2**, 1719–1733.
G. Van Tendeloo, D. Van Dyck and S. J. Pennycook, Eds (2012). *Handbook of Nanoscopy,* 2 Vols (Wiley–VCH, Weinheim).
M. von Ardenne (1938). Das Elektronen-Rastermikroskop. Theoretische Grundlagen. *Z. Phys.* **109**, 553–572.
M. von Ardenne (1938). Das Elektronen-Rastermikroskop. Praktische Ausführung. *Z. Tech. Phys.* **19**, 407–416.
M. von Ardenne (1940). *Elektronen-Übermikroskopie* (Springer, Berlin).
M. von Ardenne (1956). *Tabellen der Elektronenphysik, Ionenphysik und Übermikroskopie,* 2 Vols (Deutscher Verlag der Wissenschaften, Berlin).
M. von Ardenne (1962). *Tabellen zur Angewandten Physik. I. Elektronenphysik, Übermikroskopie, Ionenphysik* (Deutscher Verlag der Wissenschaften, Berlin; 2nd ed. 1973).
M. von Ardenne (1972). *Ein Glückliches Leben für Technik und Forschung* (Kindler, Zürich).
M. von Ardenne (1978). Zur Geschichte der Rasterelektronenmikroskopie und der Elektronenmikrosonde. *Optik* **50**, 177–188.
M. von Ardenne (1985). On the history of scanning electron microscopy, of the electron microprobe, and of early contributions to transmission electron micoscopy. In *The Beginnings of Electron Microscopy* (P. W. Hawkes, Ed.), 1–21 (Academic Press, Orlando and London). *Adv. Electron. Electron Phys*. Suppl. 16.
B. von Borries (1949). *Die Übermikroskopie* (Editio Cantor, Aulendorf).
H. von Borries (1991). Bodo von Borries: pioneer of electron microscopy. *Adv. Electron. Electron Phys* **81**, 127–176.
G. von Borries (2001). *Bodo von Borries und das Elektronenmikroskop. Erfindung und Entwicklung* (Fouqué, Egelsbach).
B. von Borries and E. Ruska (1940). Mikroskopie hoher Auflösung mit schneller Elektronen. *Ergebnisse Exakt. Naturwiss.* **19**, 237–322.
H. S. von Harrach (2009). Development of the 300-kV Vacuum Generators STEM (1985–1996). *Adv. Imaging Electron Phys.* **159**, 287–323.
E. A. Wainrib and W. I. Miljutin (1954). *Elektronenoptik* (Verlag Technik, Berlin). German translation by K. Keller of Vainryb & Milyutin (1951).
I. R. M. Wardell and P. E. Bovey (2009). A history of Vacuum Generators' 100-kV scanning transmission electron microscope. *Adv. Imaging Electron Phys.* **159**, 221–285.
F. Watt and G. W. Grime, Eds (1987). *Principles and Applications of High-Energy Ion Microbeams* (Adam Hilger, Bristol).

O. C. Wells (1974). *Scanning Electron Microscopy* (McGraw–Hill, New York).
R. Weyl, I. Dietrich and H. Zerbst (1972). Die supraleitende Abschirmlinse. *Optik* **35**, 280–286.
D. B. Williams and C. B. Carter (1996). *Transmission Electron Microscopy. A Textbook for Materials Science* (Springer, New York; 2nd edn, 2009).
H. Wollnik (1987). *Optics of Charged Particles* (Academic Press, Orlando & London).
C. Wolpers (1991). Electron microscopy in Berlin 1928–1945. *Adv. Electron. Electron Phys.* **81**, 211–229.
J.-y. Ximen (1986). Aberration theory in electron and ion optics. *Adv. Electron. Electron Phys.* Suppl. 17.
J.-y. Ximen and Z.-s. Ge (1979). *Principles of the Electron Microscope and Its Design* [in Chinese] (Science Press, Peking).
S. Ya. Yavor (1968). *Fokusirovka Zaryazhennykh Chastits Kvadrupol'nymi Linzami* (Atomizdat, Moscow).
J. Zach and M. Haider (1995). Correction of spherical and chromatic aberration in a low voltage SEM. *Optik* **93**, 112–118.
N. S. Zinchenko (1961). *Kurs Lektsii po Elektronnoi Optike*, 2nd edn (Izd. Khar'kovsk. Ordena Trudovogo Znameni Gos. Univ. im A.M. Gor'kogo, Khar'kov).
J. M. Zuo and J. C. H. Spence (2016). *Advanced Transmission Electron Microscopy. Imaging and Diffraction in Nanoscience* (Springer, New York).
V. K. Zworykin, J. Hillier and R. L. Snyder (1942). A scanning electron microscope. *ASTM Bull.* No. 117, 15–23.
V.K. Zworykin, G.A. Morton, E.G. Ramberg, J. Hillier and A.W. Vance (1945). *Electron Optics and the Electron Microscope* (Wiley, New York, Chapman & Hall, London).

Part I, Chapters 2–5

The examination by Ehrenberg and Siday (1949) of the notion of refractive index provoked a comment by Glaser (1951) and a reply by the original authors (Ehrenberg and Siday, 1951), Cf. Lentzen (2017).

***

Y. Aharonov and D. Bohm (1959). Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.* **115**, 485–491.
M. Born and E. Wolf (1959). *Principles of Optics* (Pergamon, London & New York; 7th edn, corrected printing 2002, Cambridge University Press, Cambridge).
H. Bruns (1895). Das Eikonal. *Abh. K. Sächs. Ges. Wiss., Math.-Phys. Kl.* **21**, 321–436.
A. J. Dragt (1982). Lectures on nonlinear orbit dynamics. *AIP Conf. Proc.* **87**, 147–313.
A. J. Dragt (1987). Elementary and advanced Lie algebraic methods with applications to accelerator design, electron microscopes, and light optics. *Nucl. Instrum. Meth. Phys. Res. A* **268**, 339–354.
A. J. Dragt (1990). Numerical third-order transfer map for solenoid. *Nucl. Instrum. Meth. Phys. Res. A* **298**, 441–459.
A. J. Dragt and E. Forest (1986). Lie algebraic theory of charged-particle optics and electron microscopes. *Adv. Electron. Electron Phys.* **67**, 65–120.
A. J. Dragt, E. Forest and K. B. Wolf (1986). Foundations of Lie algebraic theory of geometrical optics. In *Lie Methods in Optics* (J. S. Mondragón and K. B. Wolf, Eds), 105–157 (Springer, Berlin).
A. J. Dragt, F. Neri, G. Rangarajan, D. R. Douglas, L. M. Healy and R. D. Ryne (1988). Lie algebraic treatment of linear and nonlinear beam dynamics. *Ann. Rev. Nucl. Particle Sci.* **38**, 455–496.
W. Ehrenberg and R. E. Siday (1949). The refractive index in electron optics and the principles of dynamics. *Proc. Phys. Soc. (Lond.) B* **62**, 8–21.
W. Ehrenberg and R. E. Siday (1951). The refractive index in electron optics. *Proc. Phys. Soc. (Lond.) B* **64**, 1088–1089.
E. Forest (1998). *Beam Dynamics, A New Attitude and Framework* (Harwood, Chur).
W. Glaser (1951). The refractive index of electron optics and its connection with the Routhian function. *Proc. Phys. Soc. (Lond.) B* **64**, 114–118 &1089.

W. Glaser (1952). *Grundlagen der Elektronenoptik* (Springer, Vienna).
H. Goldstein (1959). *Classical Mechanics* (Addison−Wesley, Reading & London).
H. Goldstein, C. Poole and J. Safko (2001). *Classical Mechanics*, third edn (Addison-Wesley, Reading, London).
P. Grivet (1965). *Electron Optics* (Pergamon, London & New York).
K.-y. Hu and T.-t. Tang (1998). Lie algebraic aberration theory and calculation method for combined electron-beam focusing−deflection systems. *J. Vac. Sci. Technol. B* **16**, 3248−3255.
K.-y. Hu and T.-t. Tang (1999). Lie algebra deflection aberration theory for magnetic deflection systems. *Optik* **110**, 9−16.
K.-y. Hu, T.-t. Tang and C.-y. Tian (1999). Study on Lie algebraic deflection aberration theory for combined magnetic focusing−deflection systems. *Optik* **110**, 271−278.
E. Kasper (1972). Zeitabhängige Bewegungsgleichungen in der geometrischen Optik. *Optik* **35**, 83−89.
E. Kasper (1985). On the numerical determination of electron optical focusing properties and aberrations. *Optik* **69**, 117−125.
V. M. Kel'man and S. Ya. Yavor (1959). *Elektronnaya Optika* (Izd. Akad. Nauk, Moscow & Leningrad; 2nd edn, 1963).
V. M. Kel'man and S. Ya. Yavor (1968). *Elektronnaya Optika*, 3rd edn (Izd. Nauka, Leningrad).
V. M. Kel'man, L. M. Sekunova and E. M. Yakushev (1972). New method for design of axially symmetric mirrors. *Zh. Tekh. Fiz.* **42**, 2279−2287. *Sov. Phys. Tech. Phys*. **17**, 1786−1791.
V. M. Kel'man, L. M. Sekunova and E. M. Yakushev (1973a). Theory of axisymmetric electron mirrors. I. Trajectory equations. *Zh. Tekh. Fiz.* **43**, 1799−1806. *Sov. Phys. Tech. Phys*. **18**, 1142−1146.
V. M. Kelman, L. M. Sekunova and E. M. Yakushev (1973b). Theory of axisymmetric electron mirrors. II. Electron-optical properties. *Zh. Tekh. Fiz.* **43**, 1807−1817. *Sov. Phys. Tech. Phys*. **18**, 1147−1152.
M. Lentzen (2017). The refractive index in electron microscopy and the errors of its approximation. *Ultramicroscopy* **176**, 139−145.
M. Matsuya, M. Saito and S. Nakagawa (1995). A semi-analytical aberration calculation method using polynomials and Lie algebra. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 261−269. [Up to 11th order.]
J. Picht (1939). *Einführung in die Theorie der Elektronenoptik* (Barth, Leipzig; 2nd edn, 1957; 3rd edn, 1963).
S.-D. Poisson (1809). Variation des constantes arbitraires dans les questions de mécanique. *J. Ecole Polytech.* **8**, 266−344.
T. Radlička (2008). Lie algebraic methods in charged particle optics. *Adv. Imaging Electron Phys.* **151**, 241−362.
C. Störmer (1904). Sur le mouvement d'un point matériel portant une charge d'électricité sous l'action d'un aimant élémentaire. *Vidensk. Selsk. Skr. Math. Nat. Kl.* No. 3, 32 pp.
C. Störmer (1906a). Sur les trajectoires des corpuscules électriques dans l'espace sous l'influence du magnétisme terrestre, avec application aux aurores boréales et aux perturbations magnétiques. *C. R. Acad. Sci. Paris* **142**, 1580−1583.
C. Störmer (1906b). Sur les trajectoires des corpuscules électriques dans l'espace sous l'influence du magnétisme terrestre, avec application aux aurores boréales et aux perturbations magnétiques. *C. R. Acad. Sci. Paris* **143**, 140−142.
C. Störmer (1906c). Les expériences de M. Villard et sa théorie des aurores boréales. *C. R. Acad. Sci. Paris* **143**, 408−411.
C. Störmer (1906d). Sur les trajectoires périodiques des corpuscules électriques dans l'espace sous l'influence du magnétisme terrestre, avec application aux perturbations magnétiques. *C. R. Acad. Sci. Paris* **143**, 460−464.
C. Störmer (1933). Über die Bahnen von Elektronen im axialsymmetrischen elektrischen und magnetischen Felde. *Ann. Phys. (Leipzig)* **16** (408), 685−696.
P. A. Sturrock (1955). *Static and Dynamic Electron Optics* (Cambridge University Press, Cambridge).
J.-y. Ximen (1995). Canonical aberration theory in electron optics up to ultrahigh-order approximation. *Adv. Imaging & Electron Phys* **91**, 1−36.

Part II, Chapters 6–13

M. Abramowitz and I. A. Stegun (1964). *Handbook of Mathematical Functions* (NBS, Washington; also Dover, New York 1965).
A. Adams and F. H. Read (1972a). Electrostatic cylinder lenses II: Three element einzel lenses. *J. Phys. E: Sci. Instrum* **5**, 150–155.
A. Adams and F. H. Read (1972b). Electrostatic cylinder lenses III: Three element asymmetric voltage lenses. *J. Phys. E: Sci. Instrum* **5**, 156–160.
Z. Aiming and A. Khursheed (1999). Accurate ray tracing through finite element solved potential distributions in charged particle optics. *Proc. SPIE* **3777**, 24–34.
W. F. Ames (1969). *Numerical Methods for Partial Differential Equations* (Nelson, London).
J. L. Balladore, R. Murillo, P. W. Hawkes, J. Trinquier and B. Jouffrey (1981). Magnetic electron lenses for high voltage electron microscopes. *Nucl. Instrum. Meth.* **187**, 209–215.
J. L. Balladore, R. Murillo and J. Trinquier (1984). Various attempts to reduce the dimensions of magnetic electron lenses used for high voltage electron microscopes. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 29–35 (Scanning Electron Microscopy, AMF O'Hare).
J. E. Barth, B. Lencová and G. Wisselink (1990). Field evaluation from potentials calculated by the finite element method for ray tracing: the slice method. *Nucl. Instrum. Meth. Phys. Res. A* **298**, 263–268.
R. Becker (2008). Improvement of FDM by extrapolation on multiple grids. *Phys. Procedia* **1**, 245–248.
R. Becker, G. Martínez and A. von der Weth (2004). Improvement of magnetic field calculation by extrapolation. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 49–52.
P. Bettess (1977). Infinite elements. *Int. J. Numer. Meth. Eng.* **11**, 53–64.
K. J. Binns, P. J. Lawrenson and C. W. Trowbridge (1992). *The Analytical and Numerical Solution of Electric and Magnetic Fields* (Wiley, Chichester).
A. J. H. Boerboom and H.-n. Chen (1984). Calculation of the parameters for the Schwarz–Christoffel transformation of ion slit lens systems. *Int. J. Mass Spectrom. Ion Proc.* **61**, 1–13.
P. Bonjour (1980). Numerical methods for computing electrostatic and magnetic fields. In *Applied Charged Particle Optics* (A. Septier, Ed.), **A**, 1–44 (Academic Press, New York & London). *Adv. Electron. Electron Phys.* Vol. 13A.
O. Buneman (1971). Analytic inversion of the five-point Poisson operator. *J. Comput. Phys.* **8**, 500–505.
O. Buneman (1973a). Inversion of Lewis' Poisson operator. *J. Comput. Phys.* **11**, 307–314.
O. Buneman (1973b). Analytic inversion of nine-point Poisson operator. *J. Comput. Phys.* **11**, 447–448.
B. A. Carré (1961). The determination of the optimum accelerating factor for successive over-relaxation. *Comput. J.* **4**, 73–78.
M. V. K. Chari and P. P. Silvester (1980). *Finite Elements in Electrical and Magnetic Field Problems* (Wiley, Chichester & New York).
J. Chmelík and J. E. Barth (1993). An interpolation method for ray tracing in electrostatic fields calculated by the finite element method. *Proc. SPIE* **2014**, 133–143.
H. C. Chu and E. Munro (1982a). Numerical analysis of electron beam lithography systems. Part III: Calculation of the optical properties of electron focusing systems and dual-channel deflection systems with combined magnetic and electrostatic fields. *Optik* **61**, 121–145.
H. C. Chu and E. Munro (1982b). Numerical analysis of electron beam lithography systems. Part IV: Computerized optimization of the electron optical performance of electron beam lithography systems using the damped least squares method. *Optik* **61**, 213–236.
J. S. Colonias (1974). *Particle Accelerator Design: Computer Programs* (Academic Press, New York & London).
M. Costabel (1987). Principles of boundary element methods. *Comput. Phys. Rep.* **6**, 243–274.
R. Courant (1943). Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Am. Math. Soc.* **49**, 1–23.

D. R. Cruise (1963). A numerical method for the determination of an electric field about a complicated boundary. *J. Appl. Phys.* **34**, 3477–3479.
D. Cubric, B. Lencová and F. H. Read (1997). Comparison of finite difference, finite element and boundary element methods for electrostatic charged particle optics. EMAG, Cambridge. 91–94.
D. Cubric, B. Lencová, F. H. Read and J. Zlámal (1999). Comparison of FDM, FEM and BEM for electrostatic charged particle optics. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 357–362.
E. Cuthill and J. McKee (1969). Reducing the bandwidth of sparse symmetric matrices. 24th National Conference ACM, 157–172.
D. A. Dahl (2000). SIMION for the personal computer in reflection. *Int. J. Mass Spectrom.* **200**, 3–25.
D. A. Dahl, J. E. Delmore and A. D. Appelhans (1990). SIMION PC/PS2 electrostatic lens design program. *Rev. Sci. Instrum.* **61**, 607–609.
S. R. Desbruslais and E. Munro (1987). A boundary integral method for the simulation of three dimensional electrostatic electron optical systems. In *International Symposium on Electron Optics, ISEOB* (J.-y. Ximen, Ed.), 45–48 (Academica Sinica, Beijing).
W. Dommaschk (1965). Zur numerischen Feld- und Bahnbestimmung in elektrostatischen Feldern mit rotationssymmetrischen Randbedingungen. *Optik* **23**, 472–477.
D. Dowsett (2011). ELECTRA: electrostatic lens calculation tool with real time adjustment. A virtual optics bench for rapid determination of optical parameters using pre-characterized lenses. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 327–331.
D. Dowsett (2015). Advanced SIMION techniques: boundary matching and genetic optimization. *Microsc. Microanal.* **21** (Suppl. 4), 218–223.
I. S. Duff (1977). A survey of sparse matrix research. *Proc. IEEE* **65**, 500–535.
E. Durand (1966). *Electrostatique* (Masson, Paris).
D. Edwards (2011). Highly accurate potential calculations for cylindrically symmetric geometries using multi-region FDM: a review. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 283–291.
C. J. Edgcombe (1997). Use of Laplacian mesh for electron-optical calculations. EMAG, Cambridge. 89–90.
C. J. Edgcombe (1999). Consistent modelling for magnetic flux in rotationally symmetric systems. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 412–416.
M. Eiermann and R. S. Varga (1993). Is the optimal ω best for the SOR iteration method? *Lin. Algebra Appl* **182**, 257–277.
T. Elster, D. Stahl and P. Hahmann (2008). Calculation of fields of magnetic deflection systems with FEM using a vector potential approach. Part I: Stationary fields. Part II: Time dependent fields. *Phys. Procedia* **1**, 249–256 & 257–264.
M. Eupper (1982). Eine Methode zur Lösung des Dirichlet-Problems in drei Dimensionen und seine Anwendung auf einen neuartigen Elektronenstrahlerzeuger. *Optik* **62**, 299–307.
M. Eupper (1985). Eine verbesserte Integralgleichungsmethode zur numerischen Lösung dreidimensionaler Dirichletprobleme und ihre Anwendung in der Elektronenoptik. Dissertation, Tübingen.
G. E. Forsythe and W. R. Wasow (1960). *Finite Difference Methods for Partial Differential Equations* (Wiley, New York).
N. Franzen (1984). Computer programs for analyzing certain classes of 3-D electrostatic fields with two planes of symmetry. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 115–126 (Scanning Electron Microscopy, AMF O'Hare).
N. E. Gibbs, W. G. Poole and P. K. Stockmeyer (1976). An algorithm for reducing the bandwith and profile of a sparse matrix. *SIAM J. Numer. Anal.* **13**, 236–250.
W. Glaser (1952). *Grundlagen der Elektronenoptik* (Springer, Vienna).
D. Greenfield and M. Monastyrskii (2004). Three-dimensional electrostatic field calculation with effective algorithm of surface charge singularities treatment based on the Fichera's theorem. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 82–89.
R. F. Harrington (1967). Matrix methods for field problems. *Proc. Inst. Electr. Electron. Eng.* **55**, 136–149.
R. F. Harrington (1968). *Field Computation by Moment Methods* (Macmillan, New York and Collier–Macmillan, London).

R. F. Harrington, K. Pontoppidan, P. Abrahamsen and N. C. Albertsen (1969). Computation of Laplacian potentials by an equivalent-source method. *Proc. Inst. Electr. Eng.* **116**, 1715–1720.
E. Harting and F. H. Read (1976). *Electrostatic Lenses* (Elsevier, Amsterdam, Oxford & New York).
F. Hermeline (1982). Triangulation automatique d'un polyèdre en dimension *N*. *RAIRO* **16**, 211–242.
H. Hoch, E. Kasper and D. Kern (1978). Darstellung stationärer rotationssymmetrischer elektromagnetischer Felder durch Superposition von Lochblenden- und Kreisringfeldern. *Optik* **50**, 413–425.
J. Hodkinson and K. Tahir (1995). Saturated electron lens design using a second-order finite element method. *Proc. SPIE* **2522**, 264–269.
M. Horák and J. Zlámal (2015). Accurate interpolation of 3D fields close to the optical axis. *Microsc. Microanal.* **21** (Suppl. 4), 242–245.
F. Houdellier, F. Röder and E. Snoeck (2015). Development of splitting convergent beam electron diffraction (SCBED). *Ultramicroscopy* **159**, 59–66.
C. F. Iselin (1981). Review of recent developments in magnet computations. *IEEE Trans. Magnetics* **17**, 2168–2177.
V. Ivanov (2015). Analytical technique in the boundary element method for 3D self-consistent problems of electron optics. *Microsc. Microanal.* **21** (Suppl. 4), 236–241.
V. Ivanov and V. Brezhnev (2004). New formulation of the synthesis problem in electron optics. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 117–132.
D. H. Jacobs, Ed. (1977). *The State of the Art in Numerical Analysis* (Academic Press, London & New York).
J. Janse (1971). Numerical treatment of electron lenses with perturbed axial symmetry. *Optik* **33**, 270–281.
P. Jánský, B. Lencová and J. Zlámal (2008). Numerical simulation of thermionic electron guns. *Recent Trends* **11**, 51–52.
P. Jánský, J. Zlámal, B. Lencová, M. Zobač, I. Vlček and T. Radlička (2009). Numerical simulations of the thermionic electron gun for electron-beam welding and micromachining. *Vacuum* **84**, 357–362.
W. Kamminga (1975). Finite-element solutions for devices with permanent magnets. *J. Phys. D: Appl. Phys* **8**, 841–855.
N. K. Kang, J. Orloff, L. W. Swanson and D. Tuggle (1981). An improved method for numerical analysis of point electron and ion source optics. *J. Vac. Sci. Technol.* **19**, 1077–1081.
N. K. Kang, D. Tuggle and L. W. Swanson (1983). A numerical analysis of the electric field and trajectories with and without the effect of space charge for a field emission source. *Optik* **63**, 313–331.
Y.-f. Kang, T.-t. Tang, Y. Yen and X. Guo (2007). Differential algebraic method for computing the high order aberrations of practical electron lenses. *Optik* **118**, 158–162.
Y.-f. Kang, T.-t. Tang, J.-y. Zhao, S. Li and D. Zhang (2009). Different algebraic method for computing the high-order aberrations of practical combined focusing–deflection systems. *Optik* **120**, 591–600.
Y.-f. Kang, T.-t. Tang, J.-y. Zhao, S. Li and D. Zhang (2010). A different algebraic method for the fifth-order combined geometric–chromatic aberrations of practical magnetic electron lenses. *Optik* **121**, 178–183.
E. Kasper (1976). On the numerical calculation of static multipole fields. *Optik* **46**, 271–286.
E. Kasper (1982). Magnetic field calculation and the determination of electron trajectories. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 57–118 (Springer, Berlin & New York).
E. Kasper (1983). On the solution of integral equations arising in electron optical field calculations. *Optik* **64**, 157–169.
E. Kasper (1984a). Improvements of methods for electron optical calculations. *Optik* **68**, 341–362.
E. Kasper (1984b). Recent developments in numerical electron optics. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 63–73 (Scanning Electron Microscopy, AMF O'Hare).
E. Kasper (1987a). Computer simulation of electron optical systems. *Nucl. Instrum. Meth. Phys. Res. A* **258**, 466–479.
E. Kasper (1987b). An advanced method of field calculation in electron optical systems with axisymmetric boundaries. *Optik* **77**, 3–12.
E. Kasper (1987c). Advances in the numerical design of electron optical systems. In *International Symposium on Electron Optics, ISEOB* (J.-y. Ximen, Ed.), 115–118 (Academica Sinica, Beijing).
E. Kasper (2001). Numerical field calculation for charged particle optics. *Adv. Imaging Electron Phys.* **116**, 1–451.

E. Kasper and F. Lenz (1980). Numerical methods in geometrical electron optics. EUREM-7, The Hague **1**, 10–15.
E. Kasper and W. Scherle (1982). On the analytical calculation of fields in electron optical devices. *Optik* **60**, 339–352.
E. Kasper and M. Ströer (1990). A new method for the calculation of magnetic fields. *Nucl. Instrum. Meth. Phys. Res. A* **298**, 1–9.
D. Kern (1978). Theoretische Untersuchugen an rotationssymmetrischen Strahlerzeugungssystemen mit Feldemissionsquelle. Dissertation, Tübingen.
A. Khursheed (1994). The boundary fitted coordinate method to improve the numerical analysis of electron optical systems. *Scanning* **16**, 201–208.
A. Khursheed (1996). High-order interpolation methods for finite-element solved field distributions in the two-dimensional rectilinear coordinate system. *Proc. SPIE* **2858**, 115–125.
A. Khursheed (1997a). Mesh refinement for finite element field computations in electron optics. *Proc. SPIE* **3155**, 100–112.
A. Khursheed (1997b). Curvilinear finite element mesh generation for electron gun simulation. *Scanning* **19**, 300–309.
A. Khursheed (1999). *The Finite-Element Method in Charged Particle Optics* (Kluwer, Boston, Dordrecht & London).
A. Khursheed and A. R. Dinnis (1989). High accuracy electron trajectory plotting through finite-element fields. *J. Vac. Sci. Technol. B* **7**, 1882–1885.
A. Khursheed and Y. Pei (1996). Quasi-conformal finite element mesh generation for electron gun simulation. *Proc. SPIE* **2858**, 13–23.
P. Killes (1985). Solution of Dirichlet problems using a hybrid finite-differences and integral-equation method applied to electron guns. *Optik* **70**, 64–71.
A. Konrad and P. P. Silvester (1973). A finite element program package for axisymmetric scalar field problems. *Comput. Phys. Commun.* **5**, 437–455.
Y. Kubo, C. Gatel, E. Snoeck and F. Houdellier (2017a). Optimising electron microscopy experiments through electron optics simulation. *Ultramicroscopy* **175**, 67–80.
Y. Kubo, C. Gatel, E. Snoeck and F. Houdellier (2017b). Complete 3D electron trajectory simulation for rotationally asymmetric optics. *Kenbikyo* **52** (Suppl. 1), 23.
Y. Kuno and Y. Uchikawa (1985). Accuracy limit due to discretization in the boundary element method. *IEEE Trans. Magnetics* **21**, 2523–2526.
Y. Kuno, A. Yagi, T. Morishima and Y. Uchikawa (1988). Fundamental study for high accuracy calculation of 3-D electromagnetic field. *IEEE Trans. Magnetics* **24**, 295–298.
K. Kuroda (1983). Magnetic field analysis for electron beam deflector using a boundary integral method. *Optik* **64**, 125–131.
K. Kuroda and T. Suzuki (1972). Potential and field analysis method used [sic] electrode surface charges. *Jpn. J. Appl. Phys.* **11**, 1382.
M. Lenc and B. Lencová (1997). Analytical and numerical computation of multipole components of magnetic deflectors. *Rev. Sci. Instrum.* **68**, 4409–4414.
B. Lencová (1975). Metoda konečných prvků v elektronové optice. *Čs. Čas. Fyz. A* **25**, 57–61.
B. Lencová (1980). Numerical computation of electron lenses by the finite element method. *Comput. Phys. Commun.* **20**, 127–132.
B. Lencová (1988a). On the design of electron beam deflection systems. *Optik* **79**, 1–12.
B. Lencová (1988b). Computation of electron optical components. EUREM-9, York. **1**, 75–80.
B. Lencová (1994). The impact of electron optical simulations on the design in electron microscopy. ICEM-13, Paris. **1**, 145–148.
B. Lencová (1995a). Computation of electrostatic lenses and multipoles by the first order finite element method. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 190–197.
B. Lencová (1995b). Unconventional lens design. *J. Microsc. (Oxford)* **179**, 185–190.

B. Lencová (1996). Notes on finite element method for electron optics. *Recent Trends* **5**, 13–16.
B. Lencová (1998). New trends in software for electron optics. *Recent Trends* **6**, 40–43.
B. Lencová (1999). Accurate computation of magnetic lenses with FOFEM. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 329–337.
B. Lencová (2002a). Accuracy of electrostatic lens computations with FOFEM. *Ultramicroscopy* **93**, 263–270.
B. Lencová (2002b). Accuracy of potential computations with FOFEM. *Recent Trends* **8**, 13–14.
B. Lencová (2003). Accurate FOFEM computations and ray tracing in particle optics. EMAG, Oxford. 169–172.
B. Lencová (2004a). Accuracy estimate for magnetic electron lenses computed by first-order FEM. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 149–153.
B. Lencová (2004b). On magnetic lens computations with FEM and BEM. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 133–140.
B. Lencová and M. Lenc (1982). Infinite elements for the computation of open lens structures. ICEM-10, Hamburg. **1**, 317–318.
B. Lencová and M. Lenc (1984). The computation of open electron lenses by coupled finite element and boundary integral methods. *Optik* **68**, 37–60.
B. Lencová and M. Lenc (1986). A finite element method for the computation of magnetic electron lenses. *Scanning Electron Microsc.*, 897–915.
B. Lencová and M. Lenc (1992). Linear finite element method in axisymmetric magnetostatic problems. *IEEE Trans. Magnetics* **28**, 1107–1110.
B. Lencová and M. Lenc (1996a). The use of the first order FEM for the computation of magnetic electron lenses. *Proc. SPIE* **2858**, 34–45.
B. Lencová and M. Lenc (1996b). Accurate calculation of higher multipole components of electrostatic and magnetic deflectors with the first order FEM. *Proc. SPIE* **2858**, 58–67.
B. Lencová and M. Lenc (2004). The sources of errors in first-order FEM for electrostatic lenses. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 141–148.
B. Lencová and J. Zlámal (2008). A new program for the design of electron microscopes. *Phys. Procedia* **1**, 315–324.
B. Lencová, M. Lenc and K. D. van der Mast (1989). Computation of magnetic deflectors for electron beam lithography. *J. Vac. Sci. Technol. B* **7**, 1846–1850.
B. Lencová, T. Mulvey and K. Tsuno (1996). Computer-aided design of a high resolution magnetic electron lens by first and second order finite element methods. EUREM-11, Dublin. **1**, 361–362.
F. Lenz (1956). Zur Berechnung von rotationssymmetrischen Potentialfeldern in Elektronenlinsen. *Ann. Phys. (Leipzig)* **19** (454), 82–88.
F. Lenz (1973). Computer-aided design of electron optical systems. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 274–282 (Academic Press, London & New York).
H. R. Lewis (1966). Computation of electrostatic and rapidly pulsed magnetic fields. *J. Appl. Phys.* **37**, 2541–2550.
H. Liu, X. Zhu and E. Munro (1990). Asymmetry aberrations and tolerancing of complete systems of electron lenses and deflectors. *J. Vac. Sci. Technol. B* **8**, 1676–1681.
I. Lucas (1976). Numerical solution of magnetostatic boundary-value problems by iterative integration using the Green's function. *J. Appl. Phys.* **47**, 1645–1652.
J. R. Mautz and R. F. Harrington (1970). Computation of rotationally symmetric Laplacian potentials. *Proc. Inst. Electr. Eng.* **117**, 850–852.
B. H. McDonald and A. Wexler (1972). Finite-element solutions of unbounded field problems. *IEEE Trans. Microwave Theory & Techniques* **20**, 841–847.
B. H. McDonald, M. Friedman, M. Decreton and A. Wexler (1973). Integral finite-element approach for solving the Laplace equation. *Electron. Lett.* **9**, 242–244.
J. A. Meijerink and H. A. van der Vorst (1977). An iterative solution method for linear systems of which the coefficient matrix is a symmetric *M*-matrix. *Math. Comput.* **31**, 148–162.

K. W. Morton (1987). Basic course in finite element methods. *Comput. Phys. Rep.* **6**, 1−72.
T. Mulvey (1982). Unconventional lens design. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 359−412 (Springer, Berlin & New York).
T. Mulvey and H. Nasr (1981). An improved finite element method for calculating the magnetic field distribution in magnetic electron lenses and electromagnets. *Nucl. Instrum. Meth. Phys. Res. A* **187**, 201−208.
E. Munro (1971). Computer-aided design methods in electron optics. Dissertation, Cambridge.
E. Munro (1973). Computer-aided design of electron lenses by the finite element method. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 284−323 (Academic Press, London & New York).
E. Munro (1980). Electron beam lithography. In *Applied Charged Particle Optics* (A. Septier, Ed.), **B**, 73−131 (Academic Press, New York & London). *Adv. Electron. Electron Phys*. Suppl. 13B.
E. Munro (1987a). Computer programs for the design and optimization of electron and ion lithography systems. *Nucl. Instrum. Meth. Phys. Res. A* **258**, 443−461.
E. Munro (1987b). Finite element programs for direct raytracing in combined electrostatic and magnetic fields. In *International Symposium on Electron Optics, ISEOB* (J.-y. Ximen, Ed.), 177−180 (Institute of Electronics, Academia Sinica, Beijing).
E. Munro (1988). Finite difference programs for computing tolerances for electrostatic lenses. *J. Vac. Sci. Technol. B* **6**, 941−948.
E. Munro (1993). Computational modelling techniques in charged particle optics. *Proc. SPIE* **2014**, 2−16.
E. Munro (1997). Computational techniques for design of charged particle optical systems. In *Handbook of Charged Particle Optics* (J. Orloff, Ed.), 1−76 (CRC Press, Boca Raton).
E. Munro and H. C. Chu (1982a). Numerical analysis of electron beam lithography systems. Part I: Computation of fields in magnetic deflectors. *Optik* **60**, 371−390.
E. Munro and H. C. Chu (1982b). Numerical analysis of electron beam lithography systems. Part II: Computation of fields in electrostatic deflectors. *Optik* **61**, 1−16.
H. Murata and H. Shimoyama (2017). Developments of boundary charge method and boundary magnetic charge method and their applications. *Kenbikyo* **52** (Suppl. 1), 117.
H. Murata, T. Ohye and H. Shimoyama (1996). Improved 3-D boundary charge method. *Proc. SPIE* **2858**, 103−114.
H. Murata, T. Ohye and H. Shimoyama (1997). Improved 3-D boundary charge method for high-accuracy electric field calculation. *Proc. SPIE* **3155**, 113−124.
H. Murata, T. Ohye and H. Shimoyama (2001). Improved 3D boundary charge method for high-accuracy calculation of potential and electric field in composite dielectric system. *Proc. SPIE* **4510**, 107−118.
H. Murata, T. Ohye and H. Shimoyama (2004a). High accuracy calculation of magnetic field by improved 3-D boundary magnetic charge method. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 175−183.
H. Murata, T. Ohye and H. Shimoyama (2004b). High accuracy calculation of electric field in composite dielectric system by improved 3-D boundary charge method. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 194−195.
H. Murata, M. Ishigami and H. Shimoyama (2016). Development of a boundary magnetic charge method for computing magnetic fields in a system containing saturated materials. *Nucl. Instrum. Methods Phys. Res. A* **806**, 360−369.
D. H. Norrie and G. de Vries (1973). *The Finite Element Method* (Academic Press, New York & London).
M. Oral (2010). Ray tracing, aberration coefficients and intensity distribution. *Recent Trends* **12**, 49−52.
M. Oral and B. Lencová (2003). Calculation of aberration coefficients by ray tracing. *Microsc. Microanal.* **9** (Suppl. 3), 26−27.
M. Oral and B. Lencová (2009). Calculation of aberration coefficients by ray tracing. *Ultramicroscopy* **109**, 1365−1373.
K. Ozaki, T. Ohye, N. Tamura and Y. Uchikawa (1981). Computation of field distribution on the emitter tip using the surface charge method. *J. Electron Microsc.* **30**, 281−291.

M. J. Park, K. Park, D. H. Kim and D. Y. Jang (2008). Design and analysis of a thermionic SEM column using 3D finite element analysis. *Phys. Procedia* **1**, 199–205.
D. W. Peaceman and H. H. Rachford (1955). The numerical solution of parabolic and elliptic differential equations. *J. SIAM* **3**, 28–41.
J. Podbrdský and O. L. Krivanek (1988). A universal mesh for calculating the properties of highly saturated magnetic electron lenses by the finite element method. *Optik* **79**, 177–182.
T. Radlička (2008). Lie algebraic methods in charged particle optics. *Adv. Imaging Electron Phys.* **151**, 241–362.
H. Rauh (1971). Ein Näherungsverfahren zur Berechnung rotationssymmetricher elektrostatischer Felder mit beliebigen Randbedingungen und seine Anwendung auf einen Elektronenstrahlerzeuger mit Spitzenkathode. *Z. Naturforsch.* **26a**, 1667–1675.
F. H. Read (1996). Accuracy tests of the boundary charge method. *Recent Trends* **5**, 11.
F. H. Read (2000). Extrapolating the number of segments to infinity in the boundary element method. In *Boundary Elements XXI* (C. A. Brebbia and H. Power, Eds), 139–144 (WIT Press, Southampton).
F. H. Read (2004). The ultimate accuracy of the BEM for electrostatics. *Recent Trends* **9**, 59.
F. H. Read (2015a). Achieving the highest accuracy with the BEM. *Microsc. Microanal.* **21** (Suppl. 4), 182–187.
F. H. Read (2015b). Optimization of the shapes of electrodes of electrostatic lenses. *Microsc. Microanal.* **21** (Suppl. 4), 276–279.
F. H. Read and N. J. Bowring (1996). Ultimate numerical accuracy of the surface charge method for electrostatics*Third International Conference on Computation in Electromagnetics* 57–61 (IEE, London; IEE publication No. 420).
F. H. Read and N. J. Bowring (2011). The CPO programs and the BEM for charged particle optics. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 273–277.
F. H. Read, A. Adams and J. R. Soto-Montiel (1971). Electrostatic cylinder lenses I: Two element lenses. *J. Phys. E: Sci. Instrum.* **4**, 625–632.
E. Regenstreif (1951). Théorie de la lentille électrostatique à trois électrodes. *Ann. Radioél.* **6**. 51–83 & 114–155.
J. K. Reid (1987). Algebraic aspects of finite element solutions. *Comput. Phys. Rep.* **6**, 385–414.
H. Rose (1971). Elektronenoptische Aplanate. *Optik* **34**, 285–311.
H. Rose and E. Plies (1973). Correction of aberrations in electron optical systems with curved axes. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 344–369 (Academic Press, London & New York).
J. A. Rouse (1994). Three-dimensional computer modelling of electron optical systems. *Adv. Opt. Electron Microsc.* **13**, 1–121. Reprinted in *Adv. Imaging Electron Phys.* **206** (2018).
C. H. Schaefer (1982). Methoden zur numerischen Lösung der Laplacegleichung bei komplizierten Randwertaufgaben in drei Dimensionen und ihre Anwendung auf Probleme der Elektronenoptik. Dissertation, Tübingen.
C. H. Schaefer (1983). The application of the Alternating Procedure by H.A. Schwarz for computing three-dimensional electrostatic fields in electron-optical devices with complicated boundaries. *Optik* **65**, 347–359.
W. Scherle (1983). Eine Integralgleichungsmethode zur Berechnung magnetischer Felder von Anordnungen mit Medien unterschiedlicher Permeabilität. *Optik* **63**, 217–226.
W. Schwertfeger and E. Kasper (1974). Zur numerischen Berechnung elektromagnetischer Multipolfelder. *Optik* **41**, 160–173.
P. Silvester and A. Konrad (1973). Axisymmetric triangular finite elements for the scalar Helmholz equation. *Int. J. Numer. Meth. Eng.* **5**, 481–497.
B. Singer and M. Braun (1970). Integral equation method for computer evaluation of electron optics. *IEEE Trans. Electron Devices* **17**, 926–934.
I. H. Sloan (1992). Error analysis of boundary integral methods. *Acta Numerica* **1**, 287–339.

H. L. Stone (1968). Iterative solution of implicit approximations of multidimensional partial differential equations. *SIAM J. Numer. Anal.* **5**, 530–558.
M. Ströer (1987). Eine Galerkin-Methode mit singulären Formfunktionen und ihre Anwendung auf die Berechnung magnetostatischer Felder. *Optik* **77**, 15–25.
M. Ströer (1988). The integral equation method for field calculations in three dimensions and its reduction to a sequence of two-dimensional problems. *Optik* **81**, 12–20.
P. A. Sturrock (1951). The aberrations of magnetic electron lenses due to asymmetries. *Philos. Trans. R. Soc. (Lond.)* **A243**, 387–429.
K. Tahir (1985). A critical assessment of the finite element method for calculating magnetic fields in electron optics. Dissertation, Aston, Birmingham.
K. Tahir and T. Mulvey (1990). Pitfalls in the calculation of of the field distribution of magnetic electron lenses by the finite-element method. *Nucl. Instrum. Meth. Phys. Res. A* **298**, 389–395.
K. Tahir, X. Zhu and T. Mulvey (1993). A critical comparison of first- and second-order finite element methods in electron optics. EMAG, Liverpool. 523–526.
M. Tanaka, H. Tsuboi, F. Kobayashi and T. Misaki (1992). An edge element for boundary element method using vector variables. *IEEE Trans. Magnetics* **28**, 1623–1626.
H. Tsuboi, T. Shimotsukasa and K. Kato (1990a). Triangular mesh generation using knowledge base for three-dimensional boundary element method. *IEEE Trans. Magnetics* **26**, 799–802.
H. Tsuboi, A. Sakurai and T. Naito (1990b). A simplification of boundary element model with rotational symmetry in electromagnetic field analysis. *IEEE Trans. Magnetics* **26**, 2771–2773.
H. Tsuboi, M. Tanaka, T. Misaki, T. Naito and T. Morita (1992). Simplification of symmetric boundary element model including asymmetric parts. *IEEE Trans. Magnetics* **28**, 1724–1727.
H. Tsuboi, T. Asahara, F. Kobayashi and T. Misaki (1998). Adaptive triangular mesh generation for boundary element method in three-dimensional electrostatic problems. *IEEE Trans. Magnetics* **34**, 3379–3382.
H. Tsuboi, T. Takayama and K. Yano (1999). Setting up curved-surface triangular elements of boundary element method for electrostatic field problems. *IEEE Trans. Magnetics* **35**, 1123–1126.
Y. Uchikawa, T. Ohye and K. Gotoh (1981). Improved surface charge method. *Denki Gakkai Ronbunshi* (*Trans. IEE Japan*) **101-A**, 263–270. *Elec. Eng. Japan* **101** (3), 8–13.
J. P. van der Merwe (1978a). Electron optical properties of some Fourier–Bessel electrostatic fields with rotational symmetry. I. Basis fields. *Univ. Zululand Publ. Ser. 3* No. 23.
J. P. van der Merwe (1978b). Electron optical properties of some Fourier–Bessel electrostatic fields with rotational symmetry. II. An analytical study of weak fields. *Univ. Zululand Publ. Ser. 3* No. 24.
J. P. van der Merwe (1978c). Electron optical properties of some Fourier–Bessel electrostatic fields with rotational symmetry. III. Computed properties of synthesized fields. *Univ. Zululand Publ. Ser. 3* No. 25.
J. P. van der Merwe (1979). Electron optical properties of one-foil lenses, as related to charges induced on foils. *S. Afr. J. Phys.* **2** (4), 117–124.
R. S. Varga (1962). *Matrix Iterative Analysis* (Prentice–Hall, Englewood Cliffs).
M. J. Wallington (1970). An improved iteration method for calculating the axial potential distribution in slit-aperture electrostatic lenses. *J. Phys. E: Sci. Instrum.* **3**, 599–604.
M. J. Wallington (1971). The focal properties of electron bombardment ion sources for mass spectrometers. *J. Phys. E: Sci. Instrum.* **4**, 1–8.
T. Watanabe, Y. Kuno and Y. Uchikawa (1990). Robustness of the boundary integral equation method for potential problems. *IEEE Trans. Magnetics* **26**, 610–613.
S. Watcharotone, R. S. Ruoff and F. H. Read (2008). Possibilities for graphene for field emission: modelling studies using the BEM. *Phys. Procedia* **1**, 71–75.
C. Weber (1967). Analogue and digital methods for investigating electron-optical systems. *Philips Res. Rep*. Suppl. No. 6.
R. Weyßer (1983). Feldberechnung in rotationssymmetrischen Elektronenstrahlerzeugern mit Spitzkathode unter Berüchsichtigung von Raumladungen. Dissertation, Tübingen.
E. T. Whittaker and G. N. Watson (1927). *A Course of Modern Analysis*, 4th edn (Cambridge University Press, Cambridge).

A. M. Winslow (1967). Numerical solution of the quasilinear Poisson equation in a nonuniform triangle mesh. *J. Comput. Phys.* **2**, 149–172.
S. Xia and A. N. Broers (1995). Adaptive finite element mesh generation program for electron gun simulation. *Int. J. Numer. Model.* **8**, 109–125.
M. I. Yavor (1993). Methods for calculation of parasitic aberrations and machining tolerances in electron optical systems. *Adv. Electron. Electron Phys.* **86**, 225–281.
M. I. Yavor (1996). Perurbation methods in field and transfer matrix calculation: an extension of analytical methods in electron optics. *Proc. SPIE* **2858**, 126–135.
M. I. Yavor and A. S. Berdnikov (1995). ISIOS: a program to calculate imperfect static charged particle optical sytems. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 416–422.
X. Zhu and E. Munro (1989). A computer program for electron gun design using second-order finite elements. *J. Vac. Sci. Technol. B* **7**, 1862–1869.
X. Zhu and E. Munro (1995). Second-order finite element method and its practical application in charged particle optics. *J. Microsc. (Oxford)* **179**, 170–180.
X.-q Zhu, H.-n Liu, J. Rouse and E. Munro (1996). Field function evaluation techniques for electron lenses and deflectors. *Proc. SPIE* **2858**, 68–80.
O. C. Zienkiewicz (1967). *The Finite Element Method in Structural and Continuum Mechanics* (McGraw–Hill, London & New York).
O. C. Zienkiewicz (1971). *The Finite Element Method in Engineering Science* (McGraw–Hill, London & New York).
O. Zienkiewitz, R. Taylor and J. Z. Zhu (2013). *The Finite Element Method: Its Basis and Fundamentals* (Butterworth-Heinemann, London).
J. Zlámal and B. Lencová (2011). Development of the program EOD for design in electron and ion microscopy. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 278–282.

Part III, Chapters 14–20

All books on electron optics devote considerable space to the paraxial properties of round lenses though the distinction between the real and asymptotic cardinal elements is not always emphasized as strongly as it is here. Information about electrostatic and magnetic lenses is to be found in the surveys of de Broglie (1946, 1950), Marton (1946), Bruck and Grivet (1950), Mulvey and Wallington (1973), Hawkes (1982), Riecke (1982, 1984), Mulvey (1984), Baranova and Yavor (1984, 1986, 1989), Lencová (1997, 2009) and Tsuno (1997, 2009) For some early work not cited in the main text, see Picht (1933a,b, 1939b), Dyachenko (1935), Dyachenko and Sakharov (1935, 1937a,b, 1938a,b), Glaser (1936) and on Newtonian fields, Glaser (1950).

Additional information about electron mirrors (Chapter 18) is to be found in Henneberg and Recknagel (1935), Hottenroth (1936, 1937), Picht (1939a), Regenstreif (1947), Kot (1952a,b), Ivanov and Abalmazova (1966), Kel'man et al. (1971a–d, 1973a–c, 1982) and Yakushev and Sekunova (1986). For further details of the mirror in the Castaing-Henry device, see Castaing and Henry (1964) and Metherell (1971). Chapters 18 and 28 have been greatly expanded and additional references are included in the text. Extensive bibliographies are included in Luk'yanov and Spivak (1973) and Hawkes (2012).

Only a small section of the literature on quadrupoles (Chapter 19) is presented here. Several books deal extensively with them (Strashkevich, 1959, 1966; Hawkes, 1966, 1970; and Yavor, 1968) and more are listed in Chapter 39. Many monographs and review articles aimed at higher energies likewise deal with them at length, though here a particular model (see Chapter 39) is usually adopted; see in particular Bernard (1953a,b, 1954), Grivet and Septier (1960), Chamberlain (1960), Septier (1961), Luckey (1961), King (1964), Steffen (1965, 1985), Banford (1966), Bruck (1966a,b), Brown (1968), Brown and Servranckx (1985), Busse and Zelazny (1984), Wollnik (1987) and Carey (1987). On the origins of the strong-focusing idea, see Thomas (1938), Christofilos (1950), Courant et al. (1952, 1953) and (for electrostatic strong focusing) Blewett (1952). We also list an interesting early paper of Strashkevich (1954) and a discussion of various quadruplets by Dymnikov et al. (1963b). Much of the later literature on quadrupoles is concerned with aberration correction and is cited in Chapter 41 of Volume 2.

The optics of cylindrical lenses (Chapter 20) or their potential distributions are further examined by Fry (1932), Henneberg (1935), Glaser and Henneberg (1935), Strashkevich and Glushko (1940, 1941), Glushko and Strashkevich (1940, 1941), Strashkevich (1940c, 1952a–c, 1955), Rabin and Strashkevich (1950), Rabin et al. (1951), Strashkevich and Yurchenko (1952), Bálta Elías and Gómez García (1950), Archard (1954a, b), Laudet (1953, 1955, 1956), Septier (1954), Baranovskii et al. (1955), Gautier and Latour (1959), Yavor and Szilágyi (1960), Yavor et al. (1960), Bacquet et al. (1961, 1963), Kochanov (1962, 1963), Glikman and Yakushev (1967), Glikman et al. (1967a–c), Hibi et al. (1967), Ćirić et al. (1976) and Vukanić et al. (1976).

***

G. D. Archard (1954a). Potential distribution of symmetical cylindrical electron lenses. *Br. J. Appl. Phys.* **5**, 179–181.

G. D. Archard (1954b). Some properties of symmetrical slit ("cylindrical") electron lenses. *Br. J. Appl. Phys.* **5**, 395–399.

G. Bacquet, P. Gautier and A. Santouil (1961). Dioptrique du premier ordre des lentilles magnétiques cylindriques minces. *C. R. Acad. Sci. Paris* **252**, 522–524.

G. Bacquet, P. Gautier and A. Santouil (1963). Etude élementaire des lentilles électroniques cylindriques. *J. Microsc.* (Paris) **2**, 393–406.

J. Baltá Elías and J. A. Gómez Garcia (1950). Contribucion al estudio de la lente electrostatica cilindrica constituida por una rendija indefinida. *An. R. Soc. Esp. Fis. Quim. A* **46**, 83–92.

A. P. Banford (1966). *The Transport of Charged Particle Beams* (Spon, London).

L. A. Baranova and S. Ya. Yavor (1984). Electrostatic electron lenses. *Zh. Tekh. Fiz.* **54**, 1417–1453. *Sov. Phys. Tech. Phys.* **29**, 827–848.

L. A. Baranova and S. Ya. Yavor (1986). *Elektrostaticheskie Elektronnye Linzy* (Nauka, Moscow).

L. A. Baranova and S. Ya. Yavor (1989). The optics of round and multipole electrostatic lenses. *Adv. Electron. Electron Phys.* **76**, 1–207. [Revised English translation of Baranova and Yavor, 1986.]

S. N. Baranovskii, D. L. Kaminskii and V. M. Kel'man (1955a). Study of the electron optical properties of a straight magnetic slit. *Zh. Tekh. Fiz.* **25**, 610–624.

S. N. Baranovskii, D. L. Kaminskii and V. M. Kel'man (1955b). Double magnetic slits. *Zh. Tekh. Fiz.* **25**, 1954–1956.

E. Bauer (2008). LEEM and SPLEEM. In *Science of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), **1**, 605–656 (Springer, New York).

E. Bauer (2012). LEEM and UHV-PEEM: a retrospective. *Ultramicroscopy* **119**, 18–23.

E. Bauer (2014). *Surface Microscopy with Low Energy Electrons* (Springer, New York).

E. Bauer (2018). LEEM, SPLEEM aand SPELEM. In *Handbook of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), (Springer, Cham), forthcoming.

M. Bernard (1952). Une équation réduite des trajectoires dans un miroir électronique. *C. R. Acad. Sci. Paris* **234**, 606–608.

M.-Y. Bernard (1953a). Lentilles électrostatiques permettant de focaliser des particules de très grande énergie. Répartition des champs. *C. R. Acad. Sci. Paris* **236**, 185–187.

M.-Y. Bernard (1953b). Lentilles électrostatiques permettant de focaliser des particules de très grande énergie: calcul des trajectoires. *C. R. Acad. Sci. Paris* **236**, 902–904.

M.-Y. Bernard (1954). La 'focalisation forte' dans les accélérateurs linéaires d'ions. *Ann. Phys. (Paris)* **9**, 633–682.

S. B. Bimurzaev and E. M. Yakushev (2004). Methods of parametrization of exact electron trajectory equations. *Proc. SPIE* **5398**, 51–58.

S. B. Bimurzaev, R. S. Bimurzaeva and E. M. Yakushev (1999). Calculation of time-of-flight chromatic aberrations in the electron-optical systems possessing straight optical axes. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 271–274.

S. B. Bimurzaev, G. S. Serikbaeva and E. M. Yakushev (2003). Electrostatic mirror objective with eliminated spherical and axial chromatic aberrations. *J. Electron Microsc.* **52**, 365–368.
S. B. Bimurzaev, R. S. Bimurzaeva and E. M. Yakushev (2004a). Application of the effective potentials technique for investigation of two-dimensional electron mirrors with curvilinear axial trajectory. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 63–69.
S. B. Bimurzaev, G. S. Serikbaeva and E. M. Yakushev (2004b). Calculation of focussing quality of the electrostatic mirror objective free of the third-order spherical aberration. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 70–75.
J. P. Blewett (1952). Radial focusing in the linear accelerator. *Phys. Rev.* **88**, 1197–1199.
A. J. H. Boerboom (1959). Numerical calculation of the potential distribution in ion slit lens systems, I. *Z. Naturforsch.* **14a**, 809–816.
A. J. H. Boerboom (1960a). Numerical calculation of the potential distribution in ion slit lens systems, II. *Z. Naturforsch.* **15a**, 244–252.
A. J. H. Boerboom (1960b). Numerical calculation of the potential distribution in ion slit lens systems, III. *Z. Naturforsch.* **15a**, 253–259.
A. B. Bok, J. B. Le Poole, J. Roos, H. de Lang, H. Bethge, J. Heydenreich and M. E. Barnett (1971). Mirror electron microscopy. *Adv. Opt. Electron Microsc.* **4**, 161–261. Reprinted in *Adv. Imaging Electron Phys.* **203** (2017).
P. Bonjour (1974). Une lentille magnétique supraconductrice à pôles d'holmium pour microscopie à très haute tension. *J. Microsc. (Paris)* **20**, 219–239.
P. Bonjour (1975). High field superconducting electron lenses with holmium polepieces. *J. Phys. E: Sci. Instrum.* **8**, 764–768.
F. Borghero and F. Demontis (2016). Three-dimensional inverse problem of geometrical optics: a mathematical comparison between Fermat's principle and the eikonal equation. *J. Opt. Soc. Am. A* **33**, 1710–1722.
M. Born and E. Wolf (1959). *Principles of Optics* (Pergamon, Oxford; 7th edn, Cambridge University Press, Cambridge, 1999; corrected reprint 2002).
G. N. Brodskii and S. Ya. Yavor (1970). Electron-optical properties of a two-potential two-dimensional lens. *Zh. Tekh. Fiz.* **40**, 1310–1313. *Sov. Phys. Tech. Phys*. **15**, 1011–1012.
G. N. Brodskii and S. Ya. Yavor (1971). Focal length of a two-dimensional lens. *Zh. Tekh. Fiz.* **41**, 1550–1552. *Sov. Phys. Tech. Phys*. **16**, 1218–1219.
K. L. Brown (1968). A first- and second-order matrix theory for the design of beam transport systems and charged particle spectrometers. In *Advances in Particle Physics* (R. L. Cool and R. E. Marshak, Eds), **1**, 71–134 (Wiley–Interscience, New York & London). First issued as SLAC-75 (1967).
K. L. Brown and R. V. Servranckx (1985). First- and second-order charged particle optics. In *Physics of High Energy Particle Accelerators* (M. Month, P. F. Dahl and M. Dienes, Eds), 62–138 (American Institute of Physics, New York).
H. Bruck (1966a). *Accélérateurs Circulaires de Particules* (Presses Universitaires de France, Paris).
H. Bruck (1966b). *Théorie et Technique des Accélérateurs de Particules* (Centre d'Etudes Nucléaires de Saclay, Gif-sur-Yvette).
H. Bruck and P. Grivet (1950). Sur la lentille électrostatique. *Rev. Opt.* **29**, 164–170.
H. Busch (1927). Über die Wirkungsweise der Konzentrierungsspule bei der Braunschen Röhre. *Arch. Elektrotech.* **18**, 583–594.
W. Busse and R. Zelazny (1984). *Computing in Accelerator Design and Operation* (Springer, Berlin & New York).
C. Carathéodory (1937). Geometrische Optik. *Ergeb. Math. Grenzgeb.* **5**.
D. C. Carey (1987). *The Optics of Charged Particle Beams* (Harwood, Chur, London & New York).
R. Castaing and L. Henry (1962). Filtrage magnétique des vitesses en microscopie électronique. *C. R. Acad. Sci. Paris* **255**, 76–78.
R. Castaing and L. Henry (1964). Filtrage magnétique des vitesses en microscopie électronique. *J. Microsc. (Paris)* **3**, 133–152.
O. Chamberlain (1960). Optics of high-energy beams. *Ann. Rev. Nucl. Sci.* **10**, 161–192.

V. N. Chechulin and S. Ya. Yavor (1969). Application of Gauss brackets to calculations involving dispersive electron-optical systems. *Zh. Tekh. Fiz.* **39**, 1457−1462. *Sov. Phys. Tech. Phys*. **14**, 1093−1096.
N. Christofilos (1950). Focusing system for ions and electrons. US Patent, 2736799.
D. Ćirić, I. Terzić and J. Vukanić (1976). Plane symmetrical unipotential lens. II: first-order focal properties. *J. Phys. E: Sci. Instrum.* **9**, 844−846.
M. Cotte (1938). Recherches sur l'optique électronique. *Ann. Phys. (Paris)* **10**, 333−405.
E. D. Courant, M. S. Livingston and H. S. Snyder (1952). The strong-focusing synchrotron − a new high energy accelerator. *Phys. Rev.* **88**, 1190−1196.
E. D. Courant, M. S. Livingston, H. S. Snyder and J. P. Blewett (1953). Origin of the "strong-focusing" principle. *Phys. Rev.* **91**, 202−203.
S. Czapski and O. Eppenstein (1924). *Grundzüge der Theorie der Optischen Instrumente nach Abbe*, 3rd edn (Barth, Leipzig).
T. D. Daumenov, A. A. Sapargaliev and E. M. Yakushev (1978). A new method for determining the point characteristic function for charged particles, travelling in electron-optical systems with a straight optic axis. *Zh. Tekh. Fiz.* **48**, 2447−2454. *Sov. Phys. Tech. Phys*. **23**, 1400−1404.
T. Daumenov, A. A. Sapargaliev, L. M. Sekunova and E. M. Yakushev (1981). Space-time focusing of charged particles in fields with rotational symmetry. *Zh. Tekh. Fiz* **51**, 1137−1145. *Sov. Phys. Tech. Phys*. **26**, 647−651.
C. J. Davisson and C. J. Calbick (1931). Electron lenses. *Phys. Rev.* **38**, 585; erratum *Phys. Rev.* **42** (1932) 580.
L. de Broglie, Ed. (1946). *L'Optique Electronique* (Editions Revue d'Optique, Paris).
L. de Broglie (1950). *Optique Electronique et Corpusculaire* (Hermann, Paris).
H. Dušek (1959). Über die elektronenoptische Abbildung durch Orthogonalsysteme unter Verwendung eines neuen Modellfeldes mit streng angebbaren Paraxialbahnen. *Optik* **16**, 419−445.
V. Dyachenko (1935). The ring electron lens. *Ukr. Fiz. Zap.* **3**, 53−65.
V. Dyachenko and I. I. Sakharov (1935). Calculation of the distribution of the field for a lens, consisting of two or three plates with round apertures. *Ukr. Fiz. Zap.* **4**, 23−32. [The number of plates is not specified in the title in Ukrainian and Russian; this English title is provided by the authors.]
V. Dyachenko and I. I. Sakharov (1937a). Trajectory of an electron in the field of an electric lens. *Ukr. Fiz. Zap.* **6**, 37−48.
V. Dyachenko and I. I. Sakharov (1937b). Potential distribution of two long, coaxial successively placed cylinders. *Ukr. Fiz. Zap.* **6**, 49−52.
V. Dyachenko and I. I. Sakharov (1938a). The potential distribution of electron lenses. *Ukr. Fiz. Zap.* **7**, 3−21.
V. Dyachenko and I. I. Sakharov (1938b). Determination of the optical parameters of electric electron lenses. *Ukr. Fiz. Zap.* **7**, 175−191.
A. D. Dymnikov (1968). Application of Gaussian brackets for the calculation of multi-element electron-optical systems. *Zh. Tekh. Fiz.* **38**, 1120−1125. *Sov. Phys. Tech. Phys*. **13**, 929−933.
A. D. Dymnikov and S. Ya. Yavor (1963). Four quadrupole lenses as the analogue of an axially symmetric system. *Zh. Tekh. Fiz.* **33**, 851−858. *Sov. Phys. Tech. Phys*. **8**, 639−643.
A. D. Dymnikov, T. Ya. Fishkova and S. Ya. Yavor (1963a). System of four quadrupoles analogous to an axially symmetric lens. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **27**, 1131−1134. *Bull. Acad. Sci. USSR (Phys. Ser.)* **27**, 1112−1115.
A. D. Dymnikov, L. P. Ovsyannikova and S. Ya. Yavor (1963b). Systems of quadrupole lenses. *Zh. Tekh. Fiz.* **33**, 393−397. *Sov. Phys. Tech. Phys*. **8**, 293−296.
A. D. Dymnikov, T. Ya. Fishkova and S. Ya. Yavor (1964a). System of revolution composed of four quadrupole lenses. EUREM-3, Prague. **A**, 43−44.
A. D. Dymnikov, T. Ya. Fishkova and S. Ya. Yavor (1964b). Electron optical properties of a pseudo-axially symmetric quadrupole system. *Radiotekh. Elektron.* **9**, 1828−1831. *Radio Eng. Electron Phys*. **9**, 1515−1518.

A. D. Dymnikov, T. Ya. Fishkova and S. Ya. Yavor (1965). The influence of the geometrical parameters on the optical characteristics of a system of four quadrupole lenses analogous to an axially symmetric lens. *Zh. Tekh. Fiz.* **35**, 431–440. *Sov. Phys. Tech. Phys*. **10**, 340–346.
J. Feng and A. Scholl (2008). Photoemission electron microscopy (PEEM). In *Science of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), **1**, 657–695 (Springer, New York).
J. Feng and A. Scholl (2018). Photoemission electron microscopy (PEEM). In *Handbook of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), (Springer, Cham), forthcoming.
T. C. Fry (1932). Two problems in potential theory. *Am. Math. Monthly* **39**, 199–209.
P. Funk (1950). Über das Newtonsche Abbildungsgesetz in der Elektronenoptik. *Acta Phys. Austriaca* **4**, 304–308.
P. Gautier and C. Latour (1959). Etude théorique et expérimentale du champ d'une série de lentilles magnétiques cylindriques. *C. R. Acad. Sci. Paris* **248**, 1637–1640.
D. Geelen, A. Thete, O. Schaff, A. Kaiser, S. J. van der Molen and R. M. Tromp (2015). eV-TEM: transmission electron microscopy in a low energy cathode lens instrument. *Ultramicroscopy* **159**, 482–487.
W. Glaser (1933a). Über geometrisch-optische Abbildung durch Elektronenstrahlen. *Z. Phys.* **80**, 451–464.
W. Glaser (1933b). Bildentstehung und Auflösungsvermögen des Elektronenmikroskops vom Standpunkt der Wellenmechanik. *Z. Phys.* **81**, 647–686.
W. Glaser (1933c). Theorie des Elektronenmikroskops. *Z. Phys.* **83**, 104–122.
W. Glaser (1933d). Über optische Abbildung durch mechanische Systeme und die Optik allgemeiner Medien. *Ann. Phys. (Leipzig)* **18** (410), 557–585.
W. Glaser (1936). Elektronenbewegung als optisches Problem. *Z. Tech. Phys.* **17**, 617–622.
W. Glaser (1940). Strenge Berechnung magnetischer Linsen der Feldform $H = H_0/[1 + (z/a)^2]$. *Z. Phys.* **117**, 285–315.
W. Glaser (1950). Signification de la distance focale en optique électronique. ICEM-2, Paris. **1**, 158–164.
W. Glaser (1952). *Grundlagen der Elektronenoptik* (Springer, Vienna).
W. Glaser (1956). Elektronen- und Ionenoptik. *Handbuch der Physik* **33**, 123–395.
W. Glaser and O. Bergmann (1950). Über die Tragweite der Begriffe "Brennpunkte" und "Brennweite" in der Elektronenoptik und die starken Elektronenlinsen mit Newtonscher Abbildungsgleichung. I. *Z. Angew. Math. Phys.* **1**, 363–379.
W. Glaser and O. Bergmann (1951). Über die Tragweite der Begriffe "Brennpunkte" und "Brennweite" in der Elektronenoptik und die starken Elektronenlinsen mit Newtonscher Abbildungsgleichung. II. *Z. Angew. Math. Phys.* **2**, 159–188.
A. Glaser and W. Henneberg (1935). Die Potentialverteilung in Schlitzblende und Lochblende. *Z. Tech. Phys.* **16**, 222–230.
W. Glaser and E. Lammel (1941). Für welche elektromagnetischen Felder gilt die Newtonsche Abbildungsgleichung? *Ann. Phys. (Leipzig)* **40** (432), 367–384.
W. Glaser and E. Lammel (1943). Über die Differentialgleichungen zweiter Ordnung, welche lauter Lösungen besitzen, zwischen deren Nullstellen eine projektive Beziehung besteht. *Monatsh. Math. Phys.* **50**, 289–297.
L. G. Glikman and E. M. Yakushev (1967). Field of a three-electrode electrostatic cylindrical lens. *Zh. Tekh. Fiz.* **37**, 2097–2099. *Sov. Phys. Tech. Phys*. **12**, 1544–1545.
L. G. Glikman, V. M. Kel'man and E. M. Yakushev (1967a). Impossibility of constructing a divergent electrostatic cylindrical lens. *Zh. Tekh. Fiz.* **37**, 13–16. *Sov. Phys. Tech. Phys*. **12**, 9–11.
L. G. Glikman, V. M. Kel'man and E. M. Yakushev (1967b). Electron-optical parameters of two-electrode cylindrical immersion lenses. *Zh. Tekh. Fiz.* **37**, 1028–1034. *Sov. Phys. Tech. Phys*. **12**, 740–744.
L. G. Glikman, V. M. Kel'man and E. M. Yakushev (1967c). Telescopic systems of cylindrical lenses. *Zh. Tekh. Fiz.* **37**, 1720–1725. *Sov. Phys. Tech. Phys*. **12**, 1261–1264.
L. G. Glikman, S. P. Karetskaya, V. M. Kel'man and E. M. Yakushev (1971). Electron-optical parameters of three-electrode transaxial cylindrical lenses. *Zh. Tekh. Fiz.* **41**, 330–335. *Sov. Phys. Tech. Phys*. **16,** 247–251.

M. T. Glushko and A. M. Strashkevich (1940). A polarised elliptic cylinder in a uniform field. *Zh. Tekh. Fiz.* **10**, 1793–1799.
M. T. Glushko and A. M. Strashkevich (1941). Fields of electron-optical systems consisting of spherical or cylindrical surfaces with apertures. *Zh. Tekh. Fiz.* **11**, 205–228.
J. Gratsiatos (1940). Die Gaußsche Dioptrik der elektrisch-magnetischen Zylinderlinse. *Z. Phys.* **115**, 61–68.
P. Grivet and A. Septier (1960). Les lentilles quadrupolaires magnétiques, I & II. *Nucl. Instrum. Meth.* **6**. 126–156 & 243–275.
A. Gullstrand (1900). Allgemeine Theorie der monochromatischen Aberrationen und ihre nächsten Ergebnisse für die Ophthalmologie. *Nova Acta Reg. Soc. Sci. Upsaliensis* **20**. 204 pp.
A. Gullstrand (1906). Die reelle optische Abbildung. *K. Svensk. Vetenskapsakad. Hand.* **41** (3). 119 pp.
A. Gullstrand (1908). Die optische Abbildung in heterogenen Medien. *K. Svensk. Vetenskapsakad. Hand.* **43** (2). 58 pp.
A. Gullstrand (1915). Das allgemeine optische Abbildungssystem. *K. Svensk. Vetenskapsakad. Hand.* **55** (1). 139 pp.
A. Gullstrand (1924). Optische Systemgesetze zweite und dritter Ordnung. *K. Svensk. Vetenskapsakad. Hand.* **63** (13). 175 pp.
E. Hahn (1965). Theorie der elektrisch-magnetischen Linsen bei Zugrundelegung einer natürlichen Maßbestimmung der Achsenabszisse. *Jenaer Jahrbuch*, 107–145.
E. Hahn (1971). Über eine einheitliche Theorie der elektisch-magnetischen Linsen vom rotationssymmetrischen Typ. *Wiss. Z. Tech. Univ. Dresden* **20**, 361–363.
E. Harting and F. H. Read (1976). *Electrostatic Lenses* (Elsevier, Amsterdam, Oxford & New York).
P. W. Hawkes (1966). *Quadrupole Optics* (Springer, Berlin & New York).
P. W. Hawkes (1966/7). Cylindrical electron lenses. *Optik* **24**, 60–78 & 95–107.
P. W. Hawkes (1967). The transfer matrices of quadrupole multiplets. *Br. J. Appl. Phys.* **18**, 545–547.
P. W. Hawkes (1970). Quadrupoles in Electron Lens Design. *Adv. Electron. Electron Phys.* (Suppl. 7).
P. W. Hawkes, Ed. (1982). *Magnetic Electron Lenses* (Springer, Berlin & New York).
P. W. Hawkes (2012). Examples of electrostatic electron optics: the Farrand and Elektros microscopes and electron mirrors. *Ultramicroscopy* **119**, 9–17.
W. Henneberg (1935). Das Potential von Schlitzblende und Lochblende. *Z. Phys.* **94**, 22–27.
W. Henneberg and A. Recknagel (1935). Zusammenhänge zwischen Elektronenlinse, Elektronenspiegel und Steuerung. *Z. Tech. Phys.* **16**, 621–623.
M. Herzberger (1931). *Strahlenoptik* (Springer, Berlin).
M. Herzberger (1943). Gaussian optics and Gaussian brackets. *J. Opt. Soc. Am.* **33**, 651–655.
M. Herzberger (1958). *Modern Geometrical Optics* (Interscience, New York & London).
T. Hibi, S. Takahashi and K. Takahashi (1967). Experimental study on electrostatic cylindrical lens. *J. Electron Microsc.* **16**, 225–231.
G. Hottenroth (1936). Über Elektronenspiegel. *Z. Phys.* **103**, 460–462.
G. Hottenroth (1937). Untersuchungen über Elektronenspiegel. *Ann. Phys. (Leipzig)* **30** (422), 689–712.
R. G. E. Hutter (1945). The class of electron lenses which satisfy Newton's image relation. *J. Appl. Phys.* **16**, 670–678.
R. D. Ivanov and M. G. Abalmazova (1966). A five-electrode electrostatic mirror objective. *Prib. Tekh. Eksp.* No. 5, 192–194. *Instrum. Exp. Tech.* **9**, 1221–1222.
E. Kamke (1977). *Differentialgleichungen, Lösungsmethoden und Lösungen*, 9th edn, vol. 1. (Teubner, Stuttgart).
S. P. Karetskaya, V. M. Kel'man and E. M. Yakushev (1970). Transaxial electrostatic lens. *Zh. Tekh. Fiz.* **40**, 2563–2567. *Sov. Phys. Tech. Phys.* **15**, 2010–2013.
S. P. Karetskaya, V. M. Kel'man and E. M. Yakushev (1971a). Aberrations of transaxial electrostatic lenses. *Zh. Tekh. Fiz.* **41**, 325–329. *Sov. Phys. Tech. Phys.* **16**, 244–246.
S. P. Karetskaya, V. M. Kel'man and E. M. Yakushev (1971b). Aberrations of three-electrode transaxial cylindrical lenses. *Zh. Tekh. Fiz.* **41**, 548–552. *Sov. Phys. Tech. Phys.* **16**, 425–428.

V. M. Kel'man and S. Ya. Yavor (1954). The motion of charged particles in a uniform magnetic field, on which the magnetic field of a line current and the electrostatic field of a cylindrical condenser are superimposed. *Zh. Tekh. Fiz.* **24**, 1329–1332.
V. M. Kel'man and S. Ya. Yavor (1955). Study of a cylindrical magnetic lens with an iron yoke. *Zh. Tekh. Fiz.* **25**, 1405–1411.
V. M. Kel'man and S. Ya. Yavor (1968). *Elektronnaya Optika*, 3rd edn (Nauka, Leningrad).
V. M. Kel'man, D. L. Kaminskii and S. Ya. Yavor (1954). Experimental study of cylindrical magnetic electron lenses. *Zh. Tekh. Fiz.* **24**, 1410–1427.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1971a). Electrostatic cylindrical electron mirrors. *Zh. Tekh. Fiz.* **41**, 1489–1497. *Sov. Phys. Tech. Phys.* **16**, 1171–1176.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1971b). Deflection of parallel charged-particle beams by a plane electrostatic mirror. *Zh. Tekh. Fiz.* **41**, 1825–1831. *Sov. Phys. Tech. Phys.* **16**, 1444–1448.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1971c). Paraxial properties and aberrations of a two-electrode cylindrical mirror. *Zh. Tekh. Fiz.* **41**, 1832–1838. *Sov. Phys. Tech. Phys.* **16**, 1449–1454.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1971d). Compound cylindrical electron mirrors. I. *Zh. Tekh. Fiz.* **41**, 2016–2022. *Sov. Phys. Tech. Phys.* **16**, 1786–1781.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1972a). Theory of compound cylindrical electron mirrors. II. *Zh. Tekh. Fiz.* **42**, 297–302. *Sov. Phys. Tech. Phys.* **17**, 238–241.
V. M. Kel'man, L. M. Sekunova and E. M. Yakushev (1972b). A new method for the design of axially symmetric electron mirrors. *Zh. Tekh. Fiz.* **42**, 2279–2287. *Sov. Phys. Tech. Phys.* **17**, 1786–1791.
V. M. Kel'man, L. M. Sekunova and E. M. Yakushev (1973a). Theory of axisymmetric electron mirrors. I. Trajectory equations. *Zh. Tekh. Fiz.* **43**, 1799–1806. *Sov. Phys. Tech. Phys.* **18**, 1142–1146.
V. M. Kel'man, L. M. Sekunova and E. M. Yakushev (1973b). Theory of axisymmetric electron mirrors. II. Electron-optical properties. *Zh. Tekh. Fiz.* **43**, 1807–1817. *Sov. Phys. Tech. Phys.* **18**, 1147–1152.
V. M. Kel'man, S. P. Karetskaya, K. K. Manabaev, L. V. Fedulina and E. M. Yakushev (1973c). The transaxial electrostatic mirror. I. & II. Aberrations. *Zh. Tekh. Fiz.* **43**. 2238–2245 & 2463–2471. *Sov. Phys. Tech. Phys.* **18**, 1418–1421 & 1552–1556.
V. M. Kel'man, S. P. Karetskaya, L. V. Fedulina and E. M. Yakushev (1979). *Elektronno-Opticheskie Elementy Prismennykh Spektrometrov Zaryazhennykh Chastits* (Nauka, Alma-Ata).
V. M. Kel'man, S. P. Karetskaya, N. Yu. Saichenko and L. V. Fedulina (1982). Dispersion properties and aberration coefficients of electrostatic cylindrical mirrors. *Zh. Tekh. Fiz.* **52**, 2140–2145. *Sov. Phys. Tech. Phys.* **27**, 1318–1321.
N. M. King (1964). Theoretical techniques of high-energy beam design. *Prog. Nucl. Phys.* **9**, 71–116.
M. Knoll and E. Ruska (1932a). Beitrag zur geometrischen Elektronenoptik, I & II. *Ann. Phys. (Leipzig).* **12** (404), 607–640 & 641–661.
M. Knoll and E. Ruska (1932b). Das Elektronenmikroskop. *Z. Phys.* **78**, 318–339.
W. Koch (1967a). Ein Photoelektronen-Emissionsmikroskop für den 10^{-3}-Torr-Bereich. *Optik* **25**, 523–534.
W. Koch (1967b). Zur Kontrastentstehung im Photoelektronen-Emissionsmikroskop. *Optik* **25**, 535–550.
E. S. Kochanov (1962). Analytical determination of trajectories of charged particles in plane electrostatic fields. *Zh. Tekh. Fiz.* **32**, 294–302. *Sov. Phys. Tech. Phys.* **7**, 209–214.
E. S. Kochanov (1963). Determination of trajectories in the plane three-electrode lens. *Zh. Tekh. Fiz.* **33**, 1301–1310. *Sov. Phys. Tech. Phys.* **8**, 969–975.
S. V. Kolesnikov and M. A. Monastyrskii (1988). General theory of spatial and temporal aberrations in cathode lenses with slightly broken axial symmetry, I & II. *Zh. Tekh. Fiz* **58**. 3–10 & 11–19. *Sov. Phys. Tech. Phys.* **33**, 1–5 & 6–10.
M. V. Kot (1952a). On the theory of the field of the most simple axially symmetric electrostatic lens. *Uchen. Zap. Chernovits. Gos. Univ. (Ser. Fiz.-Mat. Nauk)* **4**, 63–71.
M. V. Kot (1952b). On the theory of the field of electric cylindrical lenses. *Uchen. Zap. Chernovits. Gos. Univ. (Ser. Fiz.-Mat. Nauk)* **4**, 73–94.
M. Laudet (1953). Potentiel et champ d'une lentille électrostatique à trois fentes. *Cah. Phys.* **41**, 73–80.

M. Laudet (1955). Optique électronique des systèmes cylindriques présentant un plan de symétrie. I. L'approximation du premier ordre. II. Les aberrations. *J. Phys. Radium* **16**. 118–124 & 908–916.
M. Laudet (1956). Contribution à l'étude du calcul numérique des champs et des trajectoires en optique électronique des systèmes cylindriques. *Ann. Fac. Sci. Toulouse* **20**, 111–230.
M. Leitner (1942). Die Abbildungsfehler der Elektrischen Zylinderlinse. Dissertation, Berlin. Draft only.
B. Lencová (2009). Electrostatic lenses. In *Handbook of Charged Particle Optics* 2nd edn (J. Orloff, Ed.), 161–208 (CRC Press, Boca Raton; pp. 177–221 in 1st edn, 1997).
H. Lichte (1983). Coherent electron optical experiments using an electron mirror. In *Int. Symp. Foundations of Quantum Mechanics in the Light of New Technology* (S. Kamefuchi, Ed.), 29–38 (Physical Society of Japan, Tokyo).
D. Luckey (1961). Beam optics. In *Techniques of High Energy Physics* (D. M. Ritson, Ed.), 403–463 (Wiley–Interscience, New York & London).
A. E. Luk'yanov and G. V. Spivak (1973). Mirror electron microscopy. *Usp Fiz. Nauk* **110**, 623–668. *Sov. Phys. Usp.* **16**, 529–552.
L. Marton (1946). Electron microscopy. *Rep. Prog. Phys.* **10**, 204–252.
A. Melkich (1947). Ausgezeichnete astigmatische Systeme der Elektronenoptik. *Sitzungsber. Akad. Wiss. Wien, Abt, IIA* **155**. 393–438 & 440–471. [Dissertation, Berlin 1944].
A. J. F. Metherell (1971). Energy analysing and energy selecting electron microscopes. *Adv. Opt. Electron Microsc.* **4**, 263–360. Reprinted in *Adv. Imaging Electron Phys*. **204** (2017).
G. Möllenstedt and H. Düker (1953). Fresnelscher Interferenzversuch mit einem Biprisma für Elektronenwellen. *Naturwissenschaften* **42**, 41.
G. Möllenstedt and F. Lenz (1963). Electron emission microscopy. *Adv. Electron. Electron Phys.* **18**, 251–329.
M. Monastyrskii (1989). Study of the aberrations of emission electron optical systems in regions with low potentials. *Zh. Tekh. Fiz.* **59**, 49–56. *Sov. Phys. Tech. Phys*. **34**, 1398–1402.
I. Müllerová and L. Frank (2003). Scanning low-energy electron microscopy. *Adv. Imaging Electron Phys.* **128**, 309–443.
T. Mulvey (1984). Magnetic electron lenses II. In *Electron Optical Sytems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 15–27 (Scanning Electron Microscopy, AMF O'Hare).
T. Mulvey and M. J. Wallington (1973). Electron lenses. *Rep. Prog. Phys.* **36**, 347–421.
F. H. Nicholl (1938). The focusing properties of the electrostatic field between two cylinders. *Proc. Phys. Soc. (Lond.)* **50**, 888–898.
F. Ollendorff (1955). *Elektronik des Einzelelektrons* (Springer, Vienna).
F. Ollendorff and G. Wendt (1932). Relativistiche Korrektur der Abbildungsgesetze einer magnetischen Sammellinse für Kathodenstrahlen. *Z. Phys.* **76**, 655–659.
J. Picht (1932). Beiträge zur Theorie der geometrischen Elektronenoptik. *Ann. Phys. (Leipzig)* **15** (407), 926–964.
J. Picht (1933a). Zur Theorie der Elektronenoptik. *Z. Tech. Phys.* **14**, 239–241.
J. Picht (1933b). Zur geometrischen Optik inhomogener Medien. *Z. Instrumentenkde* **53**, 274–282.
J. Picht (1939a). Bemerkungen zu einigen Fragen der Elektronenoptik (Elektronenlinse mit Zwischennetz, Elektronenspiegel, Immersionslinse). *Ann. Phys. (Leipzig)* **36** (428), 249–264.
J. Picht (1939b). *Einführung in die Theorie der Elektronenoptik* (Barth, Leipzig; 2nd edn, 1957; 3rd edn, 1963).
E. Plies (1994). Electron optics of low-voltage electron beam testing and inspection. Part I: Simulation tools. *Adv. Opt. Electron Microsc.* **13**, 123–242.
D. Preikszas (1995). Korrektur des Farb- und Öffnungsfehlers eines Niederspannungs-Elektronenmikroskops mit Hilfe eines Elektronenspiegels. Dissertation, Darmstadt.
D. Preikszas and H. Rose (1994). Corrected low-energy electron microscope for multi-mode operation. ICEM-13, Paris. **1**, 197–198.
D. Preikszas and H. Rose (1996a). Conditions for correcting a round lens by an electron mirror. *Recent Trends* **5**, 59–60.

D. Preikszas and H. Rose (1996b). Stability requirements for the correction of aberrations by an electron mirror. EUREM-11, Dublin. **1**, 406–407.
D. Preikszas and H. Rose (1997). Correction properties of electron mirrors. *J. Electron Microsc.* **46**, 1–9.
F. Putz (1951). Über die elektronenoptische Abbildung in starken elektrischen und magnetischen Abbildungsfeldern. Dissertation, Vienna.
B. M. Rabin and A. M. Strashkevich (1950). The trajectories of charged particles slightly deviated from their initial direction by an electrostatic field. *Zh. Tekh. Fiz.* **20**, 1232–1240.
B. M. Rabin, A. M. Strashkevich and L. S. Khin (1951). Electron optical properties of electrostatic systems with apertures departing slightly from the ideal shape. *Zh. Tekh. Fiz.* **21**, 438–444.
Lord Rayleigh [J. W. Strutt] (1886). Notes, chiefly historical, on some fundamental propositions in optics. *Philos. Mag.* **21**, 466–476.
A. Recknagel (1936). Zur Theorie des Elektronenspiegels. *Z. Tech. Phys.* **17**, 643–645.
A. Recknagel (1937). Zur Theorie des Elektronenspiegels. *Z. Phys.* **104**, 381–394.
E. Regenstreif (1947). Sur la détermination des éléments cardinaux des miroirs électrostatiques. *Ann. Radioél.* **2**, 348–358.
E. Regenstreif (1951). Théorie de la lentille électrostatique à trois électrodes. *Ann. Radioél.* **6**. 51–83 & 114–155.
E. Regenstreif (1966a). Sur le calcul de la matrice de transfert d'un multiplet quadripolaire. *C. R. Acad. Sci. Paris B* **263**, 1205–1207.
E. Regenstreif (1966b). Sur les systèmes optiques quadripolaires possédant une matrice de transfert unité. *C. R. Acad. Sci. Paris B* **263**, 1297–1299.
E. Regenstreif (1967). Formules de recurrence pour le calcul de la matrice de transfert d'un multiplet quadripolaire. *C. R. Acad. Sci. Paris B* **264**, 596–598.
W. D. Riecke (1982). Practical lens design. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 164–357 (Springer, Berlin & New York).
W. D. Riecke (1984). Magnetic electron lenses. In *Electron Optical Sytems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 1–14 (Scanning Electron Microscopy, AMF O'Hare).
H. Rose (1966). Allgemeine Abbildungseigenschaften unrunder Elektronenlinsen mit gerader Achse. *Optik* **24**, 36–59.
H. Rose (1966). Gaußsche Dioptrik begrenzter unrunder Elektronenlinsen. *Optik* **24**, 108–121.
H. Rose (1972). Zur Gaußschen Dioptrik elektrisch-magnetischer Zylinderlinsen. *Optik* **36**, 19–36.
H. Rose (2012). Geometrical Charged-particle Optics, second edn. (Springer, Berlin; first edn, 2009).
E. Ruska (1934a). Über Fortschritte im Bau und in der Leistung des magnetischen Elektronenmikroskops. *Z. Phys.* **87**, 580–602.
E. Ruska (1934b). Über ein magnetisches Objektiv für das Elektronenmikroskop. *Z. Phys.* **89**, 90–128.
A. A. Rusterholz (1950). *Elektronenoptik* (Birkhäuser, Basel).
P. Schiske (1957). Bahnen 3. Ordnung in Elektronenspiegeln. *Optik* **14**, 34–45.
L. M. Sekunova (1977). Paraxial properties and aberrations of a two-electrode axisymmetric mirror. *Zh. Tekh. Fiz.* **47**, 2030–2042. *Sov. Phys. Tech. Phys.* **22**, 1181–1187.
L. M. Sekunova and E. M. Yakushev (1975). Electron optical properties of a mirror in the shadow mode of a mirror electron microscope, I, II. *Zh. Tekh. Fiz.* **45**, 723–731 & 732–740. *Sov. Phys. Tech. Phys.* **20**, 458–462 & 463–467.
A. Septier (1953). Quelques types d'objectifs électrostatiques à immersion à fort grandissement. *Ann. Radioél.* **9**, 374–410.
A. Septier (1954). Etude d'une lentille électronique à fentes pour analyseur de vitesses. *C. R. Acad. Sci. Paris* **239**, 402–404.
A. Septier (1961). Strong focusing lenses. *Adv. Electron. Electron Phys.* **14**, 85–205.
E. V. Shpak and S. Ya. Yavor (1964). Pseudo-rotationally symmetric system of four ironless magnetic quadrupole lenses. *Zh. Tekh. Fiz.* **34**, 1037–1039. *Sov. Phys. Tech. Phys.* **9**, 803–805.
K. G. Steffen (1965). *High Energy Beam Optics* (Wiley–Interscience, New York & London).

K. G. Steffen (1985). Basic course on accelerator optics. In *CERN Accelerator School, General Accelerator Physics* (P. Bryant and S. Turner, Eds), **1**, 25−63 (CERN, Geneva). CERN Report 85−19.
A. M. Strashkevich (1940a). Elektrische Elektronenzylinderlinsen. *J. Phys. (USSR)* **3**, 507−523. (Name transliterated as A.M. Straschkewitsch).
A. M. Strashkevich (1940b). Concerning the theory of cylindrical electron lenses. *Kiev. Tech. Inst. Kozhevenno-obuvnoi Prom., Sb. Nauch.-Issled. Rabot.* **3**, 249−260.
A. M. Strashkevich (1940c). The field of an electrical system with narrow slits as a cylindrical electron lens. *Zh. Tekh. Fiz.* **10**, 91−111.
A. M. Strashkevich (1952a). Electron optical properties of electrostatic fields with diad symmetry. *Zh. Tekh. Fiz.* **22**, 487−497.
A. M. Strashkevich (1952b). Basic relations of relativistic electron optics for non-axially symmetric electrostatic fields. *Zh. Tekh. Fiz.* **22**, 1848−1856.
A. M. Strashkevich (1952c). On the theory of cylindrical electrostatic lenses. *Uchen. Zap. Chernovits. Gos. Univ. (Ser. Fiz.-Mat. Nauk)* **4**, 53−58.
A. M. Strashkevich (1954). Multiply symmetric and multiply antisymmetric electron-optical systems. *Zh. Tekh. Fiz.* **24**, 274−286.
A. M. Strashkevich (1955). The fields of cylindrical electrostatic lenses. *Nauk. Zap. Chernivet'sk. Derzhavn. Univ. (Ser. Fiz.-Mat.)* **12**, 189−197.
A. M. Strashkevich (1959). *Elektronnaya Optika Elektrostaticheskikh Polei ne Obladayushchikh Osevoi Simmetriei* (Gos. Izd. Fiz.-Mat. Lit., Moscow).
A. M. Strashkevich (1962). The main electron-optical properties of transaxial electrostatic lenses. *Zh. Tekh. Fiz.* **32**, 1142−1152. *Sov. Phys. Tech. Phys.* **7**, 841−847.
A. M. Strashkevich (1966). *Elektronnaya Optika Elektrostaticheskikh Sistem* (Energiya, Moscow & Leningrad).
A. M. Strashkevich and M. T. Glushko (1940). Polarisation of an elliptic cylinder in a uniform field. *Kiev. Tech. Inst. Kozhevenno-obuvnoi Prom., Sb. Nauch.-Issled. Rabot.* **3**, 239−248.
A. M. Strashkevich and N. P. Yurchenko (1952). The fields of some symmetric systems possessing the properties of a cylindrical lens. *Uchen. Zap. Chernovits. Gos. Univ. (Ser. Fiz.-Mat. Nauk)* **4**, 99−112.
P. A. Sturrock (1951). Propriétés optiques des champs magnétiques de révolution de la forme $H = \dfrac{H_0}{1 - z/a^2}$ [*sic*] et $H = \dfrac{H_0}{(z/a)^2 - 1}$ sur l'axe optique. *C. R. Acad. Sci. Paris* **233**, 401−403.
P. A. Sturrock (1955). *Static and Dynamic Electron Optics* (Cambridge University Press, Cambridge).
K. Tanaka (1981a). Lagrange−Helmholtz's invariant and optical invariants in terms of Gaussian brackets. *Optik* **58**, 351−358.
K. Tanaka (1981b). General paraxial analysis of Bravais systems. *Optik* **60**, 73−80.
K. Tanaka (1982). On Gaussian brackets and Gaussian constants. *Optik* **62**, 211−214.
K. Tanaka (1983a). Gaussian brackets and their application to paraxial theory. I. Gausssian brackets and generalized Gaussian constants. *Optik* **64**, 13−24.
K. Tanaka (1983b). Gaussian brackets and their application to paraxial theory. II. Paraxial theory in terms of generalized Gaussian constants. *Optik* **64**, 89−111.
K. Tanaka (1986). Paraxial theory in optical design in terms of Gaussian brackets. *Prog. Opt.* **23**, 63−111.
L. H. Thomas (1938). The paths of ions in the cyclotron. I. Orbits in the magnetic field. II. Paths in the combined electric and magnetic fields. *Phys. Rev.* **54**, 580−588 & 588−598.
R. M. Tromp (2011). Measuring and correcting aberrations of a cathode objective lens. *Ultramicroscopy* **111**, 273−281.
R. M. Tromp (2015a). An adjustable electron achromat for cathode lens microscopy. *Ultramicroscopy* **159**, 497−502.
R. M. Tromp (2015b). Catadioptric aberration correction in cathode lens microscopy. *Ultramicroscopy* **151**, 191−198.

R. M. Tromp, J. B. Hannon, A. W. Ellis, W. Wan, A. Berghaus and O. Schaff (2010). A new aberration-corrrected, energy-filtered LEEM/PEEM instrument. I. Principles and design. *Ultramicroscopy* **110**, 852–861.
R. M. Tromp, W. Wan and S. M. Schramm (2012). Aberrations of the cathode objective lens up to fifth order. *Ultramicroscopy* **119**, 33–39.
R. M. Tromp, J. B. Hannon, W. Wan, A. Berghaus and O. Schaff (2013). A new aberration-corrrected, energy-filtered LEEM/PEEM instrument. II. Operation and results. *Ultramicroscopy* **127**, 25–39.
K. Tsuno (2009). Magnetic lenses for electron microscopy. In *Handbook of Charged Particle Optics* 2nd edn (J. Orloff, Ed.), 129–159 (CRC Press, Boca Raton; pp. 143–175 in 1st edn, 1997).
J. Vukanić, I. Terzić, B. Anicin and D. Ćirić (1976). Plane symmetrical unipotential lens I: determination of potential distribution. *J. Phys. E: Sci. Instrum.* **9**, 842–843.
M. J. Wallington (1970). An improved iteration method for calculating the axial potential distribution in slit-aperture electrostatic lenses. *J. Phys. E: Sci. Instrum.* **3**, 599–604.
M. J. Wallington (1971). The focal properties of electron bombardment ion sources for mass spectrometers. *J. Phys. E: Sci. Instrum.* **4**, 1–8.
E. T. Whittaker and G. N. Watson (1927). *A Course of Modern Analysis*, 4th edn (Cambridge University Press, Cambridge).
H. Wollnik (1987). *Optics of Charged Particles* (Academic Press, Orlando & London).
J.-y. Ximen, L.-w. Zhou and K.-c. Ai (1983). Variational theory of aberrations in cathode lenses. *Optik* **66**, 19–34.
E. M. Yakushev and L. M. Sekunova (1986). Theory of electron mirrors and cathode lenses. *Adv. Electron. Electron Phys.* **68**, 337–416.
S. Ya. Yavor (1955). Study of the focusing properties of magnetic cylindrical lenses and of systems consisting of such lenses. *Zh. Tekh. Fiz.* **25**, 779–790.
S. Ya. Yavor (1962). Electron optical properties of combined electric and magnetic quadrupole lenses. *Proceedings Symposium Electron Vacuum Physics*, Budapest. 125–137.
S. Ya. Yavor (1968). *Fokusirovka Zaryazhennykh Chastits Kvadrupol'nymi Linzami* (Atomizdat, Moscow).
S. Ya. Yavor and M. Szilágyi (1960). Armoured magnetic cylindrical lenses with a plane of antisymmetry. *Zh. Tekh. Fiz.* **30**, 927–932. *Sov. Phys. Tech. Phys.* **5**, 872–876.
S. Ya. Yavor, E. V. Shpak and R. M. Minina (1960). Cylindrical magnetic lenses with a plane of antisymmetry. *Zh. Tekh. Fiz.* **30**, 395–404. *Sov. Phys. Tech. Phys.* **5**, 369–377.
V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier and A. W. Vance (1945). *Electron Optics and the Electron Microscope* (Wiley, New York & Chapman & Hall, London).

Part IV, Chapters 21–31

Further discussion or alternative presentations of the basic ideas of this chapter are to be found in Carathéodory (1937), Sturrock (1952) and Kas'yankov (1958a) and aberrations in general are surveyed in Hawkes (1967a, 2008, 2009a,b), Hawkes and Krivanek (2018) and Lenz (1982a). The historical article by Kanaya (1985) also takes a broad view.

The interesting approach to aberrations in the language of Lie algebra, explained in detail by Dragt and Forest (1986), was introduced briefly in Section 5.8. One of the attractions claimed for this approach is the ease with which interrelations between aberration coefficients are recovered; we therefore remind the reader that this has long been known to be an advantage of the eikonal method and that various alternative ways of establishing such relations have been explored by Meads (1963) and Sivkov (1971). The 'symplecticity', which is a central feature of the Lie method, is powerfully exploited by Wollnik and Berz (1985) and examined critically by Rose (1987). The Lie method is presented in detail and compared with related approaches by Radlička (2008).

Many papers have been devoted to each of the aberrations analysed in Chapters 24 and 25. On spherical aberration, see Rebsch and Schneider (1937), Becker and Wallraff (1938), Rebsch (1938), Plass (1942),

Sugiura and Suzuki (1943), Vlasov (1944), Bruck (1947a), Sorokin and Timofeev (1948), Liebmann (1951), Kanaya (1951a, 1952a), Seman (1953a,b), Kas'yankov (1955), Grümm (1956), Archard (1958), Petrie (1962), Der-Shvarts and Makarova (1966, 1967), Septier (1966), Barnes and Openshaw (1968), Brookes et al. (1968), Der-Shvarts (1971), Hanszen et al. (1972a,b), Lyubchik and Mokhnatkin (1972), Suzuki and Ishikawa (1978) and van Gorkum and Spanjer (1986). Meyer (1956) considers the fifth-order spherical aberration present when the third-order aberration has been corrected. Measurements and calculations for lens models are to be found in Part VII, Chapters 35 and 36 of Volume 2.

Coma is further considered by Kanaya (1951d), Lenz (1954), Seman (1959c), Rose (1971a,b), Moses (1972, 1973), Rose and Moses (1973) and Lenz et al. (1982).

For astigmatism and field curvature, see Becker and Wallraff (1939, 1940), Kas'yankov (1950, 1952, 1967), Kanaya (1951c, 1952c), Seman (1959c), Dutova and Kas'yankov (1963), Taganov and Kas'yankov (1964, 1965, 1967), Taganov (1966) and Gurbanov and Kas'yankov (1966).

The Petzval curvature is discussed by Goddard (1946), Chiang (1956), Seman (1968), Kas'yankov et al. (1970a,b) and Lenz (1986).

Distortion is the subject of papers by Hillier (1946b), Rang (1948), Mulvey and Jacob (1949), Kanaya (1951b, 1952b), Kanaya and Kato (1951), Liebmann (1952a), Wegmann (1953, 1954), De and Saha (1954), Seman (1959a), Kynaston and Mulvey (1962, 1963), Reisner (1970), Marai and Mulvey (1975, 1977), Lambrakis et al. (1977), Alshwaikh and Mulvey (1977), Elkamali and Mulvey (1977, 1979, 1980), Tsuno and Harada (1981a,b) and Tsuno et al. (1980a,b).

A complete list of the fifth-order aberration coefficients of round lenses is to be found in Hawkes (1965a) and later in Li and Ni (1988) and in Ai and Szilagyi (1988) but see comments on these lists at the end of Section 24.10. As well as the paper by Meyer already cited, see also Archard (1960) and U (1957). References to the formulae reproduced here are to be found in the corresponding paragraphs.

For further theoretical work on chromatic aberration (Chapter 26), see Glaser (1940b), Scherzer (1941), Kanaya (1951a, 1952a), Liebmann (1952b), Katagiri (1953), Morito (1954, 1957), Vandakurov (1955a), Watanabe and Morito (1955), Schiske (1956), Seman (1959b) and Brookes et al. (1968) and as usual, the references for Chapters 35 and 36.

To Chapter 27, we may add Hawkes (1983b, 1984b), Hanszen et al. (1972), Brouwer and Walther (1967) and the use by Heritage (1973) and Lewis et al. (1986) of aberration polynomials.

The general texts by Steffen (1965), Banford (1966), Carey (1987a) and Wollnik (1987) are all useful in connection with Chapter 28 as are the Charged Particle Optics Conference proceedings listed after Part VI. For very full bibliographies, see Hawkes (1966, 1970a). The papers by Ximen (1957) and Ximen et al. (1983) are relevant to Chapter 29.

A few extra details on cylindrical lens theory (Chapter 30) are to be found in Bertein (1950a–c, 1951a,b), Laudet (1953), Vandakurov (1955b) and Rose (1972).

Parasitic aberrations (Chapter 31) have a voluminous literature, in which few authors pay much attention to the work of their predecessors. The following list contains analyses of lens imperfections of various sorts and descriptions of many kinds of stigmator: Scherzer (1946), Hillier (1946a), Bertein (1947a-d, 1948a–c, 1949), Bertein and Regenstreif (1947, 1949), Bruck (1947b,c), Bruck and Grivet (1947, 1950), Bruck et al. (1948), Glaser (1948), Cotte (1949, 1950), Grivet et al. (1949), Inoue (1950), Rabin et al. (1951), Regenstreif (1951a–d), Recknagel and Haufe (1952/53), Leisegang (1953, 1954), Hahn (1954, 1959, 1966), Lenz and Hahn (1953), Kanaya (1953, 1955, 1958, 1962), Sakaki and Maruse (1954), Morito (1955), Stoyanov (1955a,b, 1958), Riecke (1958, 1964, 1966/67, 1972, 1976, 1982), Kanaya and Ishikawa (1958, 1959), Kas'yankov (1959), Katagiri (1960a,b), Meyer (1961a,b), Vlasov and Shakhmatova (1962), Watanabe and Someya (1963), Ximen and Chen (1964), Ximen and Xi (1964), Tadano et al. (1966), Reisner and Schuler (1967), Amboss and Jennings (1970), Yanaka and Shirota (1970), Monastyrskii and Kolesnikov (1983), Boerboom (1987), Carey (1987b), Liu et al. (1990), Yavor and Berdnikov (1995), Wei and Yan (1999), Dvořák (2002), Ivanov and Kriklivyy (2004), Shánel et al. (2014), Zlámal and Lencová (2015), Radlička and Oral (2016). The related literature that has been generated by aberration correction is cited in Chapter 41. The subject has been reviewed by Yavor (1993).

***

A. Adams and F. H. Read (1972a). Electrostatic cylinder lenses. II. Three element einzel lenses. *J. Phys. E: Sci. Instrum.* **5**, 150–155.
A. Adams and F. H. Read (1972b). Electrostatic cylinder lenses. III. Three element asymmetric voltage lenses. *J. Phys. E: Sci. Instrum.* **5**, 156–160.
G. Ade (1973). Der Einfluss der Bildfehler dritter Ordnung auf die elektronenmikroskopische Abbildung und die Korrektur dieser Fehler durch holographische Rekonstruction. *PTB-Ber.* **APh-3**, 114 pp.
G. Ade (1982). Die Abhängigkeit der Bildfehlerkoeffizienten dritter Ordnung von der Objektlage. *Optik* **63**, 43–54.
K.-c. Ai and M. Szilágyi (1988). Fifth-order relativistic aberration theory for electromagnetic focusing systems including spherical cathode lenses. *Optik* **79**, 33–40.
A. Alshwaikh and T. Mulvey (1977). The magnetised iron sphere, a realistic theoretical model for single-polepiece lenses. EMAG, Glasgow. 25–28.
K. Amboss (1959). Electron optics: aberrations of air-cored magnetic lenses. Dissertation, London.
K. Amboss and J. C. E. Jennings (1970). Electron-optical investigation of the origin of asymmetries in air-cored solenoids. *J. Appl. Phys.* **41**, 1608–1616.
G. D. Archard (1953). Magnetic electron lens aberrations due to mechanical defects. *J. Sci. Instrum.* **30**, 352–358.
G. D. Archard (1958). On the spherical aberration constant. *Rev. Sci. Instrum.* **29**, 1049–1050.
G. D. Archard (1960). Fifth order spherical aberration of magnetic lenses. *Br. J. Appl. Phys.* **11**, 521–522.
L. A. Artsimovich (1944). Electron optical properties of emission systems. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **6**, 313–329.
A. P. Banford (1966). *The Transport of Charged Particle Beams* (Spon, London).
L. A. Baranova and S. Ya. Yavor (1986). *Elektrostaticheskie Elektronnye Linzy* (Nauka, Moscow).
L. A. Baranova and S. Ya. Yavor (1989). The optics of round and multipole electrostatic lenses. *Adv. Electron. Electron Phys.* **76**, 1–207 [revised translation of Baranova and Yavor, 1986].
R. L. Barnes and I. K. Openshaw (1968). A comparison of experimental and theoretical C_s values for some probe-forming lenses. *J. Phys. E: Sci. Instrum.* **1**, 628–630.
J. Bärtle and E. Plies (2000). Investigations of an electrostatic duo-hexapole stigmator. EUREM-12, Brno. **3**, I.435–I.436.
J. E. Barth and M. D. Nykerk (1999). Dependence of the chromatic aberration spot size on the form of the energy distribution of the charged particles. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 86–90.
D. Barton and J. P. Fitch (1972). Applications of algebraic manipulation programs in physics. *Rep. Prog. Phys.* **35**, 235–314.
P. E. Batson (2008). Control of parasitic aberrations in multipole corrector optics. *Microsc. Microanal.* **14** (Suppl. 2), 830–831.
P. E. Batson (2009). Control of parasitic aberrations in multipole optics. *J. Electron Microsc.* **58**, 123–130.
H. Becker and A. Wallraff (1938). Über die sphärische Aberration magnetischer Linsen. *Arch. Elektrotech.* **32**, 664–675.
H. Becker and A. Wallraff (1939). Die Bildwölbung bei magnetischen Linsen. *Arch. Elektrotech* **33**, 491–505.
H. Becker and A. Wallraff (1940). Der Astigmatismus magnetischer Linsen. *Arch. Elektrotech* **34**, 43–48.
F. Bertein (1947a). Sur les imperfections de forme dans les instruments d'optique électronique. *C. R. Acad. Sci. Paris* **224**, 106–107.
F. Bertein (1947b). Sur le potentiel perturbateur créé par une ovalisation des lentilles électrostatiques. *C. R. Acad. Sci. Paris* **224**, 560–562.
F. Bertein (1947c). Un système correcteur en optique électronique. *C. R. Acad. Sci. Paris* **225**, 801–803.
F. Bertein (1947d). Une méthode de calcul des aberrations de forme des lentilles électrostatiques. *C. R. Acad. Sci. Paris* **225**, 863–865.
F. Bertein (1947e). Relation entre les défauts de réalisation des lentilles et la netteté des images. *Ann. Radioél.* **2**, 379–408.

F. Bertein (1948a). Relation entre les défauts de réalisation des lentilles et la netteté des images. *Ann. Radioél.* **3**, 49–62.
F. Bertein (1948b). Influence des défauts de forme d'une électrode simple en optique électronique. *J. Phys. Radium* **9**, 104–112.
F. Bertein (1948c). Influence des défauts mécaniques des objectifs sur le pouvoir de résolution du microscope électrostatique. *Ann. Radioél.* **2**, 249–252.
F. Bertein (1949). Dissymétrie du champ due à l'entrée de tension dans les lentilles électrostatiques. *C. R. Acad. Sci. Paris* **229**, 291–293.
F. Bertein (1950a). Convergence transversale crééé par les fentes des instruments d'optique ionique. *C. R. Acad. Sci. Paris* **231**, 766–767.
F. Bertein (1950b). Sur la suppression de la fente de sortie dans une source d'ions. *C. R. Acad. Sci. Paris* **231**, 1134–1136.
F. Bertein (1950c). Sur les aberrations des lentilles ioniques dans les sources d'ions. *C. R. Acad. Sci. Paris* **231**, 1448–1449.
F. Bertein (1951a). Sur les propriétés générales des lentilles électroniques faibles. La première approximation. *J. Phys. Radium* **12**, 595–601.
F. Bertein (1951b). Etude des lentilles faibles électrostatiques: la deuxième approximation. *J. Phys. Radium* **12**, 25A–31A.
F. Bertein and E. Regenstreif (1947). Sur l'aberration d'ellipticité des lentilles électrostatiques. *C. R. Acad. Sci. Paris* **224**, 737–739.
F. Bertein and E. Regenstreif (1949). Emploi des rayons marginaux dans l'étude des dissymétries chez les lentilles électrostatiques. *C. R. Acad. Sci. Paris* **228**, 1854–1856.
F. Bertein, H. Bruck and P. Grivet (1947). Influence des défauts mécaniques des objectifs sur le pouvoir de résolution du microscope électrostatique. *Ann. Radioél.* **2**, 249–252.
S. B. Bimurzaev and E. M. Yakushev (2004). Methods of parametrization of exact electron trajectory equations. *Proc. SPIE* **5398**, 51–58.
S. B. Bimurzaev, G. S. Serikbaeva and E. M. Yakushev (2004). Calculation of the focussing quality of the electrostatic mirror objective free of the third-order spherical aberration. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 70–75.
A. J. H. Boerboom (1987). The influence of misadjustment of poles on the magnetic field and on its ion optical properties. *Nucl. Instrum. Meth. Phys. Res. A* **259**, 412–416.
B. E. Bonshtedt (1964). Calculation of cathode lens aberrations. *Radiotekh. Elektron.* **9**, 844–850. *Radio Eng. Electron.* **9**, 686–692.
M. Born and E. Wolf (1959). *Principles of Optics* (Pergamon, London & New York; 7th edn, corrected printing 2002, Cambridge University Press, Cambridge).
K. A. Brookes, T. Mulvey and M. J. Wallington (1968). Magnetic objective lenses of minimum spherical and chromatic aberration. EUREM-4, Rome. **1**, 165–166.
W. Brouwer (1957). The use of matrix algebra in geometrical optics. Dissertation, Delft.
W. Brouwer (1964). *Matrix Methods in Optical Instrument Design* (Benjamin, New York & Amsterdam).
W. Brouwer and A. Walther (1967). Design of optical instruments. In *Advanced Optical Techniques* (A. C. S. van Heel, Ed.), 571–631 (North Holland, Amsterdam).
H. Bruck (1947a). Sur la limite de résolution du microscope électronique. Lentille ronde. *C. R. Acad. Sci. Paris* **224**, 1553–1555.
H. Bruck (1947b). Sur la limite de résolution du microscope électronique. Lentille dissymétrique. *C. R. Acad. Sci. Paris* **224**. 1628–1629.
H. Bruck (1947c). La limite de résolution de l'objectif électrostatique décentré. *C. R. Acad. Sci. Paris* **224**, 1818–1820.
H. Bruck and P. Grivet (1947). La limit de résolution de l'objectif électrostatique à trou central ovalisé. *C. R. Acad. Sci. Paris* **224**, 1768–1769.
H. Bruck and P. Grivet (1950). Sur la lentille électrostatique. *Rev. Opt.* **29**, 164–170.

H. Bruck, R. Remillon and L. Romani (1948). L'aberration de l'objectif électrostatique à trou central ovalisé. *C. R. Acad. Sci. Paris* **226**, 650–652.
H. Bruns (1895). Das Eikonal. *Abh. K. Sächs. Ges. Wiss., Math.-Phys. Kl.* **21**, 321–436. Also published as a book (Hirzel, Leipzig 1895).
J. Burfoot (1954a). Third-order aberrations of 'doubly symmetric' systems. *Proc. Phys. Soc (Lond.) B* **67**, 523–528.
J. Burfoot (1954b). Third-order aberrations of non-rotationally symmetric imaging systems. ICEM-3, London. 105–109.
H. Busch and E. Brüche, Eds (1937). *Beiträge zur Elektronenoptik* (Barth, Leipzig).
C. Carathéodory (1937). Geometrische Optik. *Ergeb. Math. Grenzgeb.* **5**.
D. C. Carey (1987a). *The Optics of Charged Particle Beams* (Harwood, Chur, London & New York).
D. C. Carey (1987b). Two dogmas of charged particle optics. *Nucl. Instrum. Meth. Phys. Res. A* **258**, 417–420.
N. Chako (1957). Characteristic curves on planes in the image space. *Trans. Chalmers Univ. Technol. Gothenburg* Nr. 191, 50 pp.
M.-y. Chiang (1956). On the Petzval coefficient in electron optics [in Chinese]. *Acta Phys. Sin.* **12**, 439–446.
A. W. Conway and J. L. Synge, Eds (1931). *The Mathematical Papers of Sir William Rowan Hamilton* (Cambridge University Press, Cambridge).
M. Cotte (1949). Potentiel et champ d'une électrode plane percée d'un trou elliptique. *C. R. Acad. Sci. Paris* **228**, 377–378.
M. Cotte (1950). Potentiel et champ d'une électrode épaisse percée d'une ouverture elliptique. ICEM-2, Paris. **1**, 155–157.
A. V. Crewe (1991a). An algebraic approach to the symmetrical einzel lens. *Optik* **88**, 118–125.
A. V. Crewe (1991b). A new characterization of the magnetic lens. *Optik* **89**, 70–74.
A. V. Crewe (1991c). The three element electrostatic lens. *Optik* **90**, 151–157.
S. Czapski and O. Eppenstein (1924). *Grundzüge der Theorie der Optischen Instrumente nach Abbe*, 3rd edn (Barth, Leipzig).
T. D. Daumenov, A. A. Sapargaliev and E. M. Yakushev (1978). A new method for determining the point characteristic function for charged particles in an electron-optical system with a straight optic axis. *Zh. Tekh. Fiz.* **48**, 2447–2454. *Sov. Phys. Tech. Phys.* **23**, 1400–1404.
J. H. Davenport, Y. Siret and E. Tournier (1988). *Computer Algebra. Systems and Algorithms for Algebraic Computation* (Academic Press, London). Translated into French (Masson) and Russian (MIR).
L. de Broglie (1950). *Optique Electronique et Corpusculaire* (Hermann, Paris).
M. L. De and D. K. Saha (1954). Distortion in electron lens. *Indian J. Phys.* **28**, 263–268.
N. Dellby, O. L. Krivanek, P. D. Nellist, P. E. Batson and A. R. Lupini (2001). Progress in aberration-corrected scanning transmission electron microscopy. *J. Electron Microsc.* **50**, 177–185.
G. V. Der-Shvarts (1954). Influence of departures from rotational symmetry of the focusing field on the resolution of magnetic objectives of electron microscopes. *Zh. Tekh. Fiz.* **24**, 859–870.
G. V. Der-Shvarts (1970). Approximate analytic expression for calculating the spherical aberration constant of axially symmetric magnetic lenses with yokes. *Elektrofiz. Elektrokhim. Met. Obrabotki Vyp.* **4**, 7–10.
G. V. Der-Shvarts (1971). Calculation of the spherical aberration of axially symmetric magnetic lenses. *Radiotekh. Elektron.* **16**, 1305–1306. *Radio Eng. Electron. Phys.* **16**, 1240–1241.
G. V. Der-Shvarts and I. S. Makarova (1966). On the calculation of the spherical aberration of axisymmetric magnetic lenses. *Radiotekh. Elektron.* **11**, 89–93. *Radio Eng. Electron. Phys.* **11**, 72–75.
G. V. Der-Shvarts and I. S. Makarova (1967). On the calculation of the spherical aberration of axisymmetric magnetic lenses. *Radiotekh. Elektron.* **12**, 168–171. *Radio Eng. Electron. Phys.* **12**, 161–163. Contains corrections to their earlier article.
G. V. Der-Shvarts and I. S. Makarova (1972). Third-order field aberrations of axially symmetric magnetic lenses. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **36**, 1304–1311. *Bull. Acad. Sci. URSS (Phys. Ser.)* **36**, 1164–1170.
G. V. Der-Shvarts and I. S. Makarova (1973). Simple formulae for an approximate calculation of the third-order aberration coefficients of thin, axially symmetric, one-slit magnetic lenses. *Radiotekh. Elektron.* **18**, 2374–2378. *Radio Eng. Electron. Phys.* **18**, 1722–1725.

A. L. Dodin and M. B. Nesvizhskii (1981). Accuracy of expansions in deriving the geometric aberration coefficients of cathode systems. *Zh. Tekh. Fiz.* **51**, 897–901. *Sov. Phys. Tech. Phys.* **26**, 593–541.
D. R. Douglas and A. J. Dragt (1983). MARYLIE: the Maryland Lie algebraic transport and tracking code. *IEEE Trans. Nuclear Science* **30**, 2442–2444.
A. J. Dragt (1979). A method of transfer maps for linear and nonlinear beam elements. *IEEE Trans. Nuclear Science* **26**, 3601–3603.
A. J. Dragt (1982). Lie algebraic theory of geometrical optics and optical aberrations. *J. Opt. Soc. Am.* **72**, 372–379.
A. J. Dragt (1987). Elementary and advanced Lie algebraic methods with applications to accelerator design, electron microscopes, and light optics. *Nucl. Instrum. Meth. Phys. Res. A* **258**, 339–354.
A. J. Dragt (1990). Numerical third-order transfer map for solenoid. *Nucl. Instrum. Meth. Phys. Res. A* **298**, 441–449.
A. J. Dragt and E. Forest (1983). Computation of nonlinear behavior of Hamiltonian systems using Lie algebraic methods. *J. Math. Phys.* **24**, 2734–2744.
A. J. Dragt and E. Forest (1986). Lie algebraic theory of charged-particle optics and electron microscopes. *Adv. Electron. Electron Phys.* **67**, 65–120.
A. J. Dragt, E. Forest and K. B. Wolf (1986). Foundations of a Lie algebraic theory of geometrical optics. In *Lie Methods in Optics* (J. S. Mondragón and K. B. Wolf, Eds), 105–157 (Springer, Berlin).
H. Dušek (1959). Über die elektronenoptische Abbildung durch Orthogonalsysteme unter Verwendung eines neuen Modellfeldes mit streng angebbaren Paraxialbahnen. *Optik* **16**, 419–445.
K. P. Dutova and P. P. Kas'yankov (1963). On the calculation of electron-optical systems corrected for astigmatism. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **27**, 1127–1130. *Bull. Acad. Sci. USSR (Phys. Ser.)* **27**, 1108–1111.
D. Dvořák (2002). Misalignment aberrations. *Recent Trends* **8**, 17–18.
H. H. Elkamali and T. Mulvey (1977). Improved viewing arrangements in the TEM. EMAG, Glasgow. 33–34.
H. H. Elkamali and T. Mulvey (1979). A double lens system for the correction of spiral distortion. EMAG, Brighton. 63–64.
H. H. Elkamali and T. Mulvey (1980). A wide-angle TEM projection system. EUREM-7, The Hague. **1**, 74–75.
R. Erni (2015). *Aberration-Corrected Imaging in Transmission Electron Microscopy. An Introduction*, second edn (Imperial College Press, London; first edn, 2010).
S. Finsterwalder (1892). Die von optischen Systemen grösserer Oeffnung und grösseren Gesichtsfeldes erzeugten Bilder. Auf Grund der Seidelschen Formeln untersucht. *Abh. Math.-Phys. Kl. Bayr. Akad. Wiss.* **17**, 517–587.
J. P. Fitch (1979). The application of symbolic algebra to physics – a case of creeping flow. In *Symbolic and Algebraic Computation* (E. W. Ng, Ed.), 30–41 (Springer, Berlin & New York).
J. P. Fitch (1985). Solving algebraic poblems with REDUCE. *J. Symbol. Comput.* **1**, 211–227.
J. Fitch (2009). CAMAL 40 years on – is small still beautiful? In *Intelligent Computer Mathematics* (J. Carette, L. Dixon, C. Sacerdoti Coen and S. Watt, Eds), 32–44 (Springer, Berlin).
J. P. S. Fitzgerald, R. C. Word and R. Könenkamp (2011). Adaptive aberration correction using a triode hyperbolic electron mirror. *Ultramicroscopy* **111**, 1495–1503.
J. P. S. Fitzgerald, R. C. Word and R. Könenkamp (2012). Simultaneous and adaptive correction of spherical and chromatic aberration using an electron mirror and lens combination. *Ultramicroscopy* **115**, 35–40.
C. A. Flory, R. C. Taber and G. E. Yefchak (1996). Analytic expression for the ideal one-dimensional mirror potential yielding perfect energy focusing in TOF mass spectrometry. *Int. J. Mass Spectrom. Ion Process.* **152**, 177–184.
J. Focke (1965). Higher order aberration theory. *Prog. Opt.* **4**, 1–36.
P. Funk (1936). Über die Seidelsche Fehlertheorie in der Elektronenoptik. *Monatsh. Math. Phys.* **43**, 305–316.
P. Funk (1937). Berichtigung und Ergänzung zu meiner Arbeit: Über die Seidelsche Fehlertheorie in der Elektronenoptik. *Monatsh. Math. Phys.* **45**, 314–319.

R. K. Garg (1982). A simple and practical approach to correct spherical aberration in existing electron probe microscopes. *Ultramicroscopy* **7**, 411–413.
D. Geelen, A. Thete, O. Schaff, A. Kaiser, S. J. van der Molen and R. M. Tromp (2015). eV-TEM: transmission electron microscopy in a low energy cathode lens instrument. *Ultramicroscopy* **159**, 482–487.
W. Glaser (1933a). Bildentstehung und Auflösungsvermögen des Elektronenmikroskops vom Standpunkt der Wellenmechanik. *Z. Phys.* **81**, 647–686.
W. Glaser (1933b). Theorie des Elektronenmikroskops. *Z. Phys.* **83**, 104–122.
W. Glaser (1933c). Über optische Abbildung durch mechanische Systeme und die Optik allgemeiner Medien. *Ann. Phys. (Leipzig)* **18** (410), 557–585.
W. Glaser (1935). Zur Bildfehlertheorie des Elektronenmikroskops. *Z. Phys.* **97**, 177–201.
W. Glaser (1936a). Elektronenbeugung als optisches Problem. *Z. Tech. Phys.* **17**, 617–622.
W. Glaser (1936b). Zur Störungstheorie. *Z. Phys.* **104**, 157–160.
W. Glaser (1937). Elektronenbewegung als optisches Problem. In *Beiträge zur Elektronenoptik* (H. Busch and E. Brüche, Eds), 24–33 (Barth, Leipzig).
W. Glaser (1938). Die kurze Magnetlinse von kleinstem Öffnungsfehler. *Z. Phys.* **109**, 700–721.
W. Glaser (1940a). Über ein von sphärischer Aberration freies Magnetfeld. *Z. Phys.* **116**, 19–33.
W. Glaser (1940b). Die Farbabweichung bei Elektronenlinsen. *Z. Phys.* **116**, 56–67.
W. Glaser (1940c). Über den Öffnungfehler der Elektronenlinsen. *Z. Phys.* **116**, 734–735.
W. Glaser (1942). Über elektronenoptische Abbildung bei gestörter Rotationssymmetrie. *Z. Phys.* **120**, 1–15.
W. Glaser (1948). Zentrierung und Auflösungsvermögen beim Übermikroskop. *Öst. Ing.-Arch.* **3**, 39–46.
W. Glaser (1949). Über die Theorie der elektrischen und magnetischen Ablenkung von Elektronenstrahlbündeln und ein ihr angepaßtes Störungsverfahren. *Ann. Phys. (Leipzig)* **4** (439), 389–408.
W. Glaser (1952). *Grundlagen der Elektronenoptik* (Springer, Vienna).
W. Glaser and H. Robl (1951). Strenge Berechnung typischer elektrostatischer Elektronenlinsen. *Z. Angew. Math. Phys.* **2**, 444–469.
W. Glaser and P. Schiske (1953). Bildstörungen durch Polschuhasymmetrien bei Elektronenlinsen. *Z. Angew. Phys.* **5**, 329–339.
L. S. Goddard (1946). A note on the Pezval field curvature in electron-optical systems. *Proc. Camb. Philos. Soc.* **42**, 127–131.
E. Goto and T. Soma (1977). MOL (moving objective lens). Formulation of deflective aberration free system. *Optik* **48**, 255–270.
J. Grabmeier, E. Kaltofen and V. Weispfenning, Eds (2003). *Computer Algebra Handbook. Foundations, Applications, Systems* (Springer, Berlin).
J. Gratsiatos (1936). Zur Bildfehlertheorie der elektronenoptischen Systeme. *Z. Phys.* **102**, 641–651.
J. Gratsiatos (1937). Die Bildfehler der schwachen magnetischen Elektronenlinse. *Z. Phys.* **107**, 382–386.
G. A. Grinberg (1942). General theory of the focusing action of electrostatic and magnetostatic fields. I .Two-dimensional fields. II. Three-dimensional electrostatic fields [in English]. *Dokl. Akad. Nauk SSR* **37**, 172–178 & 261–268.
G. A. Grinberg (1943a). General theory of the focusing action of electrostatic and magnetostatic fields. III. Three-dimensional (twisted) trajectories in the presence of both an electrostatic and a magnetic field [in English]. *Dokl. Akad. Nauk SSR* **38**, 78–81.
G. A. Grinberg (1943b). Fundamentals of the general theory of the focusing properties of static electric and magnetic fields. *Zh. Tekh. Fiz.* **13**, 361–388. Russian version of Grinberg 1942, 1943a.
G. A. Grinberg (1948). *Izbrannye Voprosy Matematicheskoi Teorii Elektricheskikh i Magnitnykh Yavlenii* (Academy of Sciences Press, Moscow).
G. A. Grinberg (1957a). Remark on the paper by P.P. Kas'yankov [*Opt. Spektrosk.* **3**, 1957, 169]. *Opt. Spektrosk.* **3**, 673.
G. A. Grinberg (1957b). Concerning the criticism by P.P. Kas'yankov of my method of obtaining the principal electron-optical equations for pencils with curvilinear axis. *Zh. Tekh. Fiz.* **27**, 2425–2431. *Sov. Phys. Tech. Phys.* **2**, 2259–2265.

P. Grivet, F. Bertein and E. Regenstreif (1949). The use of marginal rays for the study of asymmetry in electrostatic lenses. ICEM-1, Delft. 86−88.
T. R. Groves (2015). *Charged Particle Optics Theory. An Introduction* (CRC Press, Boca Raton).
H. Grümm (1956). Zur Frage der unteren Grenze des Öffnungsfehlers bei magnetischen Elektronenlinsen. *Optik* **13**, 92−93.
A. Gullstrand (1915). Optische Systemgesetze zweiter und dritter Ordnung. *Kungl. Svensk. Vetenskapsakad. Handl.* **63**. (Suppl.) No. 13, 175 pp.
G. G. Gurbanov and P. P. Kas'yankov (1966). Concerning the calculation of electrostatic electron-optical systems with corrected astigmatism. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **30**, 735−738. *Bull. Acad. Sci. USSR (Phys. Ser.)* **30**, 762−765.
E. Hahn (1954). Über die Korrektion des axialen Astigmatismus elektrostatischer Objektive. *Jenaer Jahrb.* (Part 1), 63−75.
E. Hahn (1959). Über die Bildfehler 3. Ordnung bei Exzentrizität und Verkantung der Elektroden eines elektrostatischen Objektives. *Jenaer Jahrb.* (Part I), 86−114.
E. Hahn (1966). Theorie der elektrisch-magnetischen Linsen mit gestörter Feldsymmetrie bei Zugrundelegung einer natürlichen Maßbestimmung der Achse. *Jenaer Jahrb.*, 145−172.
M. Haider, H. Müller and S. Uhlemann (2008). Present and future hexapole aberration correctors for high resolution electron microscopy. *Adv. Imaging Electron Phys.* **153**, 43−120.
K.-J. Hanszen, G. Ade and R. Lauer (1972a). Genauere Angaben über sphärische Längsaberration, Verzeichnung in der Pupillenebene und über die Wellenaberration von Elektronenlinsen. *Optik* **35**, 567−590.
K.-J. Hanszen, R. Lauer and G. Ade (1972b). Eine Beziehung zwischen der Abhängigkeit des Öffnungsfehlers von der Objektlage und der Koma, ausgedrückt in der Terminologie der Elektronenoptik. *Optik* **36**, 156−159.
D. F. Hardy (1967). Combined magnetic and electrostatic quadrupole electron lenses. Dissertation, Cambridge.
R. F. Harrington (1968). *Field Computation by Moment Methods* (Macmillan, New York and Collier−Macmillan, London).
P. W. Hawkes (1964). The paraxial chromatic aberrations of quadrupole systems. EUREM-3, Prague. **1**, 5−6.
P. W. Hawkes (1965a). The geometrical aberrations of general electron optical systems, I and II. *Philos. Trans. R. Soc. (Lond.) A* **257**, 479−522.
P. W. Hawkes (1965b). Simple but rigorous formulae for the aperture aberrations and distortion coefficients of quadrupole lenses and octopoles. *Optik* **22**, 349−368.
P. W. Hawkes (1965c). The paraxial chromatic aberrations of rectilinear orthogonal systems. *Optik* **22**, 543−551.
P. W. Hawkes (1965d). With which Gaussian rays should aberrations be calculated in electron optics? *Optik* **23**, 244−250.
P. W. Hawkes (1966). *Quadrupole Optics* (Springer, Berlin & New York).
P. W. Hawkes (1966/67a). Cylindrical electron lenses. *Optik* **24**. 60−78 & 95−107.
P. W. Hawkes (1966/67b). Systematic transformation of the aberration coefficients of quadrupoles and electrostatic round lenses. *Optik* **24**. 252−262 & 275−282.
P. W. Hawkes (1967a). Lens aberrations. In *Focusing of Charged Particles* (A. Septier, Ed.), **1**, 411−468 (Academic Press, New York & London).
P. W. Hawkes (1967b). General expressions for aperture aberration coefficients. *J. Microsc. (Paris)* **6**, 917−932.
P. W. Hawkes (1968). Asymptotic aberration coefficients, magnification and object position in electron optics. *Optik* **27**, 287−304.
P. W. Hawkes (1970a). Quadrupoles in electron lens design. *Adv. Electron. Electron Phys.* (Suppl. 7).
P. W. Hawkes (1970b). Asymptotic aberration integrals for round lenses. *Optik* **31**, 213−219.
P. W. Hawkes (1970c). Asymptotic aberration integrals for quadrupole systems. *Optik* **31**, 302−314.
P. W. Hawkes (1970d). The addition of round electron lens aberrations. *Optik* **31**, 592−599.
P. W. Hawkes (1970e). The addition of quadrupole aberrations. *Optik* **32**, 50−60.

P. W. Hawkes (1970f). Can objective spherical aberration be expressed as a polynomial of fourth degree in reciprocal magnification? A study of the Grivet–Lenz model. *J. Microsc. (Paris)* **9**, 435–454.
P. W. Hawkes (1977a). Computer calculation of formulae for electron lens aberration coefficients. *Optik* **48**, 29–51.
P. W. Hawkes (1977b). Computer calculation of high-frequency lens properties. HVEM, Kyoto. 57–60.
P. W. Hawkes (1980a). Methods of computing optical properties and combating aberrations for low-intensity beams. In *Applied Charged Particle Optics* (A. Septier, Ed.), **A**, 45–157 (Academic Press, New York & London). *Adv. Electron. Electron Phys.* Suppl. 13A.
P. W. Hawkes (1980b). Some approximate magnetic lens aberration formulae: a commentary. *Optik* **56**, 293–320.
P. W. Hawkes (1983a). Computer-aided calculation of the aberration coefficients of microwave cavity lenses. I. Primary (second-order) aberrations. II. Secondary (third-order) aberrations. *Optik* **63**, 129–156. and **65**, 227–251.
P. W. Hawkes (1983b). The dependence of electron lens aberration coefficients on object and aperture position. EMAG, Guildford. 471–474.
P. W. Hawkes (1984a). "Die Abhängigkeit der Bildfehlerkoeffizienten dritter Ordnung von der Objektlage" by G. Ade: a footnote. *Optik* **66**, 379–380.
P. W. Hawkes (1984b). The dependence of electron lens asymptotic aberration coefficients on object and pupil position: matrix methods. EUREM-8, Budapest. **1**, 23–24.
P. W. Hawkes (1985a). On the various forms of the dependence of lens aberration coefficients on object and aperture position. *Optik* **70**, 115–123.
P. W. Hawkes (1985b). A theoretically attractive form of the geometrical aberration coefficients of electrostatic lenses. *Optik* **70**, 140–142.
P. W. Hawkes (1986). Meanwhile, towards the end of the decade. . ., . *Ultramicroscopy* **20**, 189–194.
P. W. Hawkes (1987). Comment on "Geometric aberrations in electrostatic lenses, II" by A. Renau and D.W.O. Heddle: the thin-lens approximation for the geometric aberration coefficients of accelerating lenses. *J. Phys. E: Sci. Instrum.* **20**, 234–235.
P. W. Hawkes (1999). Rebuttal of "The relationship of lens aberrations to image aberrations in electron optics" by G.F. Rempfer. *Optik* **110**, 372–374.
P. W. Hawkes (2000). Clarification of my criticism of "The relationship of lens aberrations to image aberrations in electron optics" by G.F. Rempfer. *Optik* **111**, 227–228.
P. W. Hawkes (2008). Aberration correction. In *Science of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), **1**, 696–747 (Springer, New York).
P.W. Hawkes (2009a). Aberrations. In *Handbook of Charged Particle Optics* 2nd edn, (J. Orloff, Ed.), 209–340 (CRC Press, Boca Raton).
P. W. Hawkes (2009b). Aberration correction past and present. *Philos. Trans. R. Soc. A* **367**, 3637–3664.
P. W. Hawkes (2009c). Can the Nomura lens be free of spherical aberration? *J. Microsc. (Oxford)* **234**, 325.
P. W. Hawkes (2015). The correction of electron lens aberrations. *Ultramicroscopy* **156**, A1–A64.
P. W. Hawkes and V. E. Cosslett (1962). Electronic simulation of the effect of lens aberrations on electron beam cross-sections. *Br. J. Appl. Phys.* **13**, 272–279.
P. W. Hawkes and O. L. Krivanek (2018). Aberration correctors, monochromators, analysers. In *Handbook of Microscopy* (P. W. Hawkes and J. C. H. Spence, Eds), (Springer, Cham), forthcoming.
P. W. Hawkes and B. Lencová (2002). Studies on the asymptotic cardinal elements and aberration coefficients of symmetric electrostatic lenses. *Optik* **113**, 78–82.
P. W. Hawkes, B. Lencová and M. Lenc (1995). Approximate expressions for electron lens properties. *J. Microsc. (Oxford)* **179**, 145–157.
A. C. Hearn (1985). *REDUCE User's Manual* (Rand Corp., Santa Monica CA).
M. B. Heritage (1973). The computation of the third-order aberration coefficients of electron lenses. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 324–338 (Academic Press, London & New York).
M. Herzberger (1931). *Strahlenoptik* (Springer, Berlin).
M. Herzberger (1958). *Modern Geometrical Optics* (Wiley–Interscience, New York & London).

J. Hillier (1946a). Further improvement in the resolving power of the electron microscope. *J. Appl. Phys.* **17**, 307–309.
J. Hillier (1946b). A study of distortion in electron microscope projection lenses. *J. Appl. Phys.* **17**, 411–419.
J. Hillier and E. G. Ramberg (1947). The magnetic electron microscope objective: contour phenomena and the attainment of high resolving power. *J. Appl. Phys.* **18**, 48–71.
Y. Inoue (1950). On the asymmetric aberration of the electron lens. ICEM-2, Paris. **1**, 199–200.
D. Ioanoviciu (1989). Ion optics. *Adv. Electron. Electron Phys.* **73**, 1–92.
D. Ioanoviciu, G. E. Yefchak and C. G. Enke (1989). Second-order space-time transfer matrix of the two-stage electrostatic mirror. *Int. J. Mass Spectrom. Ion Process.* **94**, 281–291.
K. Ishizuka (1994). Coma-free alignment of a high-resolution electron microscope with three-fold astigmatism. *Ultramicroscopy* **55**, 407–418.
V. Ivanov and V. Kriklivyy (2004). Numerical algorithms for boundary problems with disturbed axial symmetry. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 96–116.
O. Jandeleit and F. Lenz (1959). Berechnung der Bildfehlerkoeffizienten magnetischer Elektronenlinsen in Abhängigkeit von der Polschuhgeometrie und den Betriebsdaten. *Optik* **16**, 87–107.
J. Janse (1971). Numerical treatment of electron lenses with perturbed axial symmetry. *Optik* **33**, 270–281.
K. Kanaya (1951a). Spherical and chromatic aberrations in magnetic lenses for electron microscopes [in Japanese]. *Bull. Electrotech. Lab.* **15**, 86–91.
K. Kanaya (1951b). Distortion in magnetic lenses for electron microscopes [in Japanese]. *Bull. Electrotech. Lab.* **15**, 91–94.
K. Kanaya (1951c). Curvature of field and astigmatism in magnetic lenses for electron microscopes [in Japanese]. *Bull. Electrotech. Lab.* **15**, 193–198.
K. Kanaya (1951d). Coma in magnetic lenses for electron microscopes [in Japanese]. *Bull. Electrotech. Lab.* **15**, 199–202.
K. Kanaya (1952a). Three-stage electron microscopes. I. Spherical and chromatic aberrations of magnetic lenses [in Japanese]. *Bull. Electrotech. Lab.* **16**, 25–30.
K. Kanaya (1952b). Three-stage electron microscopes. II. Distortion of magnetic lenses [in Japanese]. *Bull. Electrotech. Lab.* **16**, 135–142.
K. Kanaya (1952c). Three-stage electron microscopes. III. Curvature of field and astigmatism of magnetic lenses [in Japanese]. *Bull. Electrotech. Lab.* **16**, 184–192.
K. Kanaya (1953). On the manufacture accuracy of magnetic lenses in electron microscopes. *J. Electronmicrosc.* **1**, 7–12.
K. Kanaya (1955). Electron optic theory on the design of magnetic lenses for electron microscopes. *Res. Electrotech. Lab.* **548**. 70 pp.
K. Kanaya (1958). Theoretical consideration for anastigmatic lens in electron microscope. *Bull. Electrotech. Lab.* **22**, 615–622.
K. Kanaya (1962). An anastigmatic alinement system by double magnetic deflectors with stigmator in the electron beam instruments. *Bull. Electrotech. Lab.* **26**, 161–172.
K. Kanaya (1985). Reminiscences of the development of electron optics and electron microscope instrumentation in Japan. *Adv. Electron. Electron Phys.* (Suppl. 16), 317–386.
K. Kanaya and A. Ishikawa (1958). Disagreement of voltage and current centres due to misalignment in the electron microscope. *Bull. Electrotech. Lab.* **22**, 641–646.
K. Kanaya and A. Ishikawa (1959). Mialignment defects on the electron microscope images. *J. Electronmicrosc.* **7**, 13–15.
K. Kanaya and A. Kato (1951). Measurement of distortion in magnetic lenses for electron microscopes. *Bull. Electrotech. Lab.* **15**, 827–833.
K. Kanaya and H. Kawakatsu (1961). Electro-static stigmators used in correcting second and third order astigmatisms in the electron microscope. *Bull. Electrotech. Lab* **25**, 641–656.
K. Kanaya, H. Kawakatsu, H. Yamazaki and S. Sibata (1966). Electron optical properties of three-electrode electron lenses. *J. Sci. Instrum.* **43**, 416–429.
E. Kasper (1982). Magnetic field calculation and the determination of electron trajectories. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 57–118 (Springer, Berlin & New York).

P. P. Kas'yankov (1950). Conditions for the correction of astigmatism and field curvature in electron lenses. *Zh. Tekh. Fiz.* **20**, 1426–1434.
P. P. Kas'yankov (1952). Calculation of an electron lens for given conditions on the aberrations of third order. *Zh. Tekh. Fiz.* **22**, 80–83.
P. P. Kas'yankov (1955). On electron lenses the the spherical aberration of which can be made as small as is desired. *Zh. Tekh. Fiz.* **25**, 1639–1648.
P. P. Kas'yankov (1956a). On some problems in the theory of electron beam focusing. *Dokl. Akad. Nauk SSSR* **108**, 813–816. *Sov. Phys. Dokl.* **1**, 367–371.
P. P. Kas'yankov (1956b). *Teoriya Elektromagnitnykh Sistem s Krivolineinoi Os'yu* (Leningrad University Press, Leningrad).
P. P. Kas'yankov (1957). On the trajectory equations for an electron in electron-optical systems with a curvilinear axis. *Opt. Spektrosk.* **3**, 169–179.
P. P. Kas'yankov (1958a). On the condition for the existence of stigmatic imaging in electron optical systems with curvilinear axis. *Dokl. Akad. Nauk SSSR* **120**, 497–500. *Sov. Phys. Dokl.* **3**, 573–576.
P. P. Kas'yankov (1958b). On the letter of Corresponding Member of the Academy of Sciences, Prof. G.A. Grinberg. *Zh. Tekh. Fiz.* **28**, 915–918. *Sov. Phys. Tech. Phys.* **3**, 854–857.
P. P. Kas'yankov (1959). A theoretical method of investigating the characteristics of a stigmator. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **23**, 711–715. *Bull. Acad. Sci. USSR (Phys. Ser.)* **23**, 706–710.
P. P. Kas'yankov (1967). Methods of calculation for axially symmetric electron optical systems. In *Chislennye Metody Rascheta Elektronno-opticheskykh Sistem* (G. I. Marchuk, Ed.), 4–10 (Nauka, Novosibirsk).
P. P. Kas'yankov, N. S. Cheremisina and N. P. Rynkevich (1970a). Petzval systems in electron optics. *Opt.-Mekh. Prom.* **37** (11), 67. *Sov. J. Opt. Technol.* **37**, 757–758.
P. P. Kas'yankov, N. P. Rynkevich and N. S. Cheremisina (1970b). Systems with corrected Petzval curvature. In *Metody Rascheta Elektronno-opticheskykh Sistem* (G. I. Marchuk, Ed.), 62–67 (Nauka, Novosibirsk).
S. Katagiri (1953). Experimental study of axial chromatic aberration. *J. Electronmicrosc.* **1**, 13–18.
S. Katagiri (1960a). On the astigmatism of magnetic electron lens. *J. Electronmicrosc.* **8**, 13–16.
S. Katagiri (1960b). Hysteresis in astigmatism of magnetic electron lens. *J. Electronmicrosc.* **9** (2), 119–120.
H. Kawakatsu and K. Kanaya (1961). Magnetic stigmators used in correcting second and third order astigmatisms in the electron microscope. *Bull. Electrotech. Lab* **25**, 801–814.
V. M. Kel'man and S. Ya. Yavor (1961). Achromatic quadrupole electron lenses. *Zh. Tekh. Fiz.* **31**, 1439–1442. *Sov. Phys. Tech. Phys.* **6**, 1052–1054.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1971a). Electrostatic cylindrical electron mirrors. *Zh. Tekh. Fiz.* **41**, 1489–1497. *Sov. Phys. Tech. Phys*. 1**6**, 1171–1176.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1971b). Paraxial properties and aberrations of a two-electrode cylindrical mirror. *Zh. Tekh. Fiz.* **41**, 1832–1838. *Sov. Phys. Tech. Phys*. 1**6**, 1449–1454.
V. M. Kel'man, L. V. Fedulina and E. M. Yakushev (1971c). Combined cylindrical electron mirrors. I. *Zh. Tekh. Fiz.* **41**, 2016–2022. *Sov. Phys. Tech. Phys*. 1**6**, 1598–1602.
V. M. Kel'man, A. A. Sapargaliev and E. M. Yakushev (1972). Theory of the cathode lens. I. Cylindrical cathode lens. *Zh. Tekh. Fiz.* **42**, 2001–2009. *Sov. Phys. Tech. Phys*. **17**, 1607–1611.
V. M. Kel'man, A. A. Sapargaliev and E. M. Yakushev (1973). Theory of the cathode lens. II. Electrostatic cathode lens with rotational field symmetry. *Zh. Tekh. Fiz.* **43**, 52–60. *Sov. Phys. Tech. Phys.* **18**, 33–37.
R. Könenkamp, T. Jones, J. Elstner, R. C. Word, G. Rempfer, T. Dixon, L. Almaraz and W. Skoczylas (2008). Image properties in an aberration-corrected photoemission electron microscope. *Phys. Procedia* **1**, 505–511.
J. Korringa (1942). Onderzoekingen op het Gebied der Algebraïsche Optiek. Dissertation, Delft.
O. Krivanek (1994). Three-fold astigmatism in high-resolution transmission electron microscopy. *Ultramicroscopy* **55**, 419–433.
O. L. Krivanek and M. L. Leber (1993). Three-fold astigmatism: an important TEM aberration. MSA **51**, Cincinnati. 972–973.
O. L. Krivanek and P. A. Stadelmann (1995). Effect of three-fold astigmatism on high-resolution electron micrographs. *Ultramicroscopy* **60**, 103–113.

O. L. Krivanek, N. Dellby and M. Murfitt (2009). Aberration correction in electron microscopy. In *Handbook of Charged Particle Optics,* 2nd edn (J. Orloff, Ed.) 601–640. (CRC Press, Baton Rouge).
Yu. V. Kulikov (1971). Geometrical optics of a five-electrode immersion objective. *Radiotekh. Elektron.* **16**, 654–655. *Radio Eng. Electron. Phys.* **16**, 715–717.
Yu. V. Kulikov (1972). The problem of evaluating the image quality of cathode electron-optical systems. *Radiotekh. Elektron.* **17**, 373–375. *Radio Eng. Electron. Phys.* **17**, 286–288.
Yu. V. Kulikov (1973). Some problems of cathode lenses with image transfer. *Radiotekh. Elektron.* **18**, 2379–2383. *Radio Eng. Electron. Phys.* **18**(11), 286–288.
Yu. V. Kulikov (1975). Evaluation of the image quality of cathode electron-optical systems with third-order aberrations taken into account. *Radiotekh. Elektron.* **20**, 1249–1254. *Radio Eng. Electron. Phys.* **20** (6), 93.
Yu. V. Kulikov, M. A. Monastyrskii and K. I. Feygin (1978). Theory of third-order aberrations of cathode lenses. Aberrations of cathode lenses with combined electric and magnetic fields. *Radiotekh. Elektron.* **23**, 167–175. *Radio Eng. Electron. Phys.* **23** (1), 120.
C. E. Kuyatt (1978). Systematic transformations of the asymptotic aberration coefficients of round electrostatic lenses (1). *J. Vac. Sci. Technol* **15**, 861–864.
C. E. Kuyatt, D. DiChio and S. V. Natali (1974). Third-order asymptotic aberration coefficients of electron lenses. III. Formulas and results for the two-tube electrostatic lens. *Rev. Sci. Instrum.* **45**, 1275–1280.
D. Kynaston and T. Mulvey (1962). Distortion-free operation of the electron microscope. ICEM-5, Philadelphia. **1**, D-2.
D. Kynaston and T. Mulvey (1963). The correction of distortion in the electron microscope. *Br. J. Appl. Phys.* **14**, 199–206.
E. Lambrakis, F. Z. Marai and T. Mulvey (1977). Correction of spiral distortion in the transmission electron microscope. EMAG, Glasgow. 35–38.
M. Laudet (1953). Potentiel et champ d'une lentille électrostatique à trois fentes. *Cah. Phys.* **41**, 73–80.
M. Laudet (1955a). Contribution à l'étude du calcul numérique des champs et des trajectoires en optique électronique des systèmes cylindriques. *Ann. Fac. Sci. Toulouse* **20**, 111–230.
M. Laudet (1955b). Optique électronique des systèmes cylindriques présentant un plan de symétrie. I. L'approximation du premier ordre. II. Les aberrations. *J. Phys. Radium* **16**, 118–124 & 908–916.
R. Lauer (1982). Characteristics of triode electron guns. *Adv. Opt. Electron Microsc.* **8**, 137–206.
S. Leisegang (1953). Zum Astigmatismus von Elektronenlinsen. *Optik* **10**, 5–14.
S. Leisegang (1954). Ein einfacher Stigmator für magnetische Elektronenlinsen. *Optik* **11**, 49–60.
H. Leitner (1942). Die Abbildungsfehler der Elektrischen Zylinderlinse. Dissertation, Berlin [draft only].
M. Lenc (1992). Immersion objective lenses in electron optics. Dissertation, Delft.
M. Lenc and I. Müllerová (1992a). Electron optical properties of a cathode lens. *Ultramicroscopy* **41**, 411–417.
M. Lenc and I. Müllerová (1992b). Optical properties and axial aberration coefficients of the cathode lens in combination with a focusing lens. *Ultramicroscopy* **45**, 159–162.
B. Lencová, M. Lenc and P. W. Hawkes (2008). A pitfall in the calculation of higher order aberrations. *Ultramicroscopy* **108**, 737–740.
F. Lenz (1954). Kann Man in der Elektronenoptik eine reine Komafigur photographieren? ICEM-3, London. 86–88.
F. Lenz (1956). Über asymptotische Bildfehler. EUREM-1, Stockholm. 48–51.
F. Lenz (1957). Über asymptotische Bildfehler. *Optik* **14**, 74–82.
F. Lenz (1982a). Properties of electron lenses. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 119–161 (Springer, Berlin & New York).
F. Lenz (1982b). Theorie und Praxis in der Elektronenoptik. *Mikroskopie* **39**, 359–373.
F. Lenz (1986). Field curvature and Petzval's theorem in electron optics. ICEM-11, Kyoto. **1**, 287–288.
F. Lenz and M. Hahn (1953). Die Untersuchung der Unsymmetrie magnetischer Polschuhlinsen in schwach unterteleskopischen Strahlengang. *Optik* **10**, 15–27.
F. Lenz, R. Speidel and J. Kuhm (1982). Aufzeichnungen in einem Elektronenstrahlschreiber mit minimaler Koma. *Optik* **62**, 401–411.

G. N. Lewis, H. Paik, J. Mioduszewski and B. M. Siegel (1986). A hydrogen field ion source with focusing optics. *J. Vac. Sci. Technol. B* **4**, 116–119.
Y. Li and W.-x. Ni (1988). Relativistic fifth order geometrical aberrations of a focusing system. *Optik* **78**, 45–47.
C. Li, C.-j. Mu and E.-g. Chen (1995). Fifth-order aberrations of rotationally symmetrical electron optics system. *J. Yunnan Univ. Nat. Sci.* **17** (2), 11–18.
C. Li, C.-j. Mu, Y.-m. Zhang, X.-m. Chen and T.-g. Liu (2002). Normalized fifth-order spherical aberration coefficients derived from a Gaussian trajectory on rotationally symmetric magnetic lenses. *Micron* **33**, 353–356.
G. Liebmann (1951). The symmetrical magnetic electron microscope objective lens with lowest spherical aberration. *Proc. Phys. Soc. (Lond.)* **B64**, 972–977.
G. Liebmann (1952a). Magnetic electron microscope projector lenses. *Proc. Phys. Soc. (Lond.) B* **65**, 94–108.
G. Liebmann (1952b). The magnetic electron microscope objective lens of lowest chromatic aberration. *Proc. Phys. Soc. (Lond.) B* **65**, 188–192.
Z.-x. Liu (2002). Fifth-order canonical geometric aberration analysis of electrostatic round lenses. *Nucl. Instrum. Meth. Phys. Res. A* **488**, 42–50.
Z.-x. Liu (2004a). Improved fifth-order geometric aberration coefficients of electron lenses. *J. Phys. D. Appl. Phys.* **37**, 653–659.
Z.-x. Liu (2004b). Differential algebraic description for third- and fifth-order aberrations of electromagnetic lenses. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 154–161.
Z.-x. Liu (2007). Exploring third-order chromatic aberrations of electron lenses with computer algebra. *Adv. Imaging Electron Phys.* **145**, 95–148.
Z.-X. Liu (2017). Third-rank chromatic aberrations of electron lenses. *Ultramicroscopy*. in press.
H. Liu, X. Zhu and E. Munro (1990). Asymmetry aberrations and tolerancing of complete systems of electron lenses and deflectors. *J. Vac. Sci. Technol. B* **8**, 1676–1681.
Ya. G. Lyubchik and A. V. Mokhnatkin (1972). Experimental determination of the spherical aberration of a magnetic axially symmetrical lens. *Radiotekh. Elektron.* **17**, 2234–2237. *Radio Eng. Electron. Phys.* **17**, 1795–1798.
M. Mankos (2011). Electron optics for dual-beam low energy electron microscopy. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 35–40.
M. Mankos and K. Shadman (2013). A monochromatic, aberration-corrected, dual-beam low energy electron microscope. *Ultramicroscopy* **130**, 13–28.
M. Mankos, D. Adler, L. Veneklasen and E. Munro (2008). Electron optics for low energy electron microscopy. *Phys. Procedia* **1**, 485–504.
M. Mankos, V. Spasov and E. Munro (2010). Principles of dual-beam low-energy electron microscopy. *Adv. Imaging Electron Phys.* **161**, 1–53.
M. Mankos, K. Shadman, H. J. Persson, A. T. N'Diaye, A. K. Schmid and R. W. Davis (2014). A novel low energy electron microscope for DNA sequencing and surface analysis. *Ultramicroscopy* **145**, 36–49.
F. Z. Marai and T. Mulvey (1975). Electron-optical characteristics of single-pole and related magnetic electron lenses. EMAG, Bristol. 43–44.
F. Z. Marai and T. Mulvey (1977). Scherzer's formula and the correction of spiral distortion in the electron microscope. *Ultramicroscopy* **2**, 187–192.
H. J. Marschall (1939). Über die sphärische Aberration magnetischer Konzentrierspulen. *Telefunken-Röhre* **16**, 190–197.
P. Meads (1963). The theory of aberrations of quadrupole focusing arrays. Dissertation, University of California. UCRL-10807.
A. Melkich (1947). Ausgezeichnete astigmatische Systeme der Elektronenoptik. *Sitzungsber. Akad. Wiss. Wien, Math.-Nat. Kl., Abt. IIa* **155**, 393–438 & 439–471.
W. E. Meyer (1956). Der Öffnungsfehler 5. Ordnung bei sphärisch korrigierten Elektronenmikroskopen. *Optik* **13**, 86–91.
W. E. Meyer (1961a). Das Auflösungsvermögen sphärisch korrigierter elektrostatischer Elektronenmikrosope. *Optik* **18**, 69–91.

W. E. Meyer (1961b). Das praktische Auflösungsvermögen von Elektronenmikroskopen. *Optik* **18**, 101–114.
M. A. Monastyrskii (1978a). Asymptotic solutions of the paraxial equation of electron optics. *Zh. Tekh. Fiz.* **48**, 1117–1122. *Sov. Phys. Tech. Phys*. **23**, 624–627.
M. A. Monastyrskii (1978b). Influence functions in the optimization of a cathode electron-optical system. *Zh. Tekh. Fiz.* **48**, 2228–2234. *Sov. Phys. Tech. Phys*. **23**, 1275–1278.
M. A. Monastyrskii (1980). Step in Green's function for cathode lenses with a non-smooth electrostatic potential distribution. *Zh. Tekh. Fiz.* **50**, 1939–1947. *Sov. Phys. Tech. Phys*. **25**, 1129–1133.
M. A. Monastyrskii and S. V. Kolesnikov (1983). A new method for calculating potential perturbations in problems with slightly disturbed axial symmetry. *Zh. Tekh. Fiz.* **53**, 1668–1677. *Sov. Phys. Tech. Phys*. **28**, 1029–1034.
M. A. Monastyrskii and Yu. V. Kulikov (1976). Chromatic aberrations of cathode lenses. *Radiotekh. Elektron.* **21**, 2251–2254. *Radio Eng. Electron. Phys*. **21** (10), 148–150.
M. A. Monastyrskii and Yu. V. Kulikov (1978). Theory of the aberrations of cathode lenses. Chromatic aberrations. *Radiotekh. Elektron.* **23**, 644–648. *Radio Eng. Electron. Phys*. **23** (3), 137–140.
J. S. Mondragón and K. B. Wolf, Eds (1986). *Lie Methods in Optics* (Springer, Berlin).
N. Morito (1954). On the chromatic field aberration of the magnetic electron lens in the electron microscope. *J. Appl. Phys.* **25**, 986–993.
N. Morito (1955). The voltage and current centers of the electron microscope image [in Japanese]. *Hitachi Hyoron* **37**, 817–822.
N. Morito (1957). On the chromatic aberration of the magnetic projection lens. *J. Electronmicrosc.* **5**, 1–2.
R. W. Moses (1966). Extension of Scherzer's theorem. *Rev. Sci. Instrum* **37**, 1370–1372.
R. W. Moses (1970). Minimum aperture aberrations of quadrupole lens systems. *Rev. Sci. Instrum* **41**, 729–740.
R. W. Moses (1971a). Minimum chromatic aberrations of magnetic quadrupole objective lenses. *Rev. Sci. Instrum.* **42**, 828–831.
R. W. Moses (1971b). Quadrupole–octopole aberration correctors with minimum fields. *Rev. Sci. Instrum.* **42**, 832–839.
R. W. Moses (1971c). Calculation of achromatic lens systems with minimum electric fields. EMAG, Cambridge. 88–89.
R. W. Moses (1972). Minimum spherical aberration of coma-free magnetic round lenses. EUREM-5, Manchester. 86–87.
R. W. Moses (1973). Lens optimization by direct application of the calculus of variations. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 250–272 (Academic Press, London & New York).
R. W. Moses (1974). Aberration correction for high-voltage electron microscopy. *Proc. Roy. Soc. (Lond.) A* **339**, 483–512.
C.-j. Mu, E.-g. Chen, J. Gao and C. Li (1999). Analyses of fifth-order aberrations of rotationally symmetric electron optical system. *Acta Electron. Sin.* **27** (8), 67–69.
T. Mulvey and L. Jacob (1949). Distortion-free electrostatic lenses. *Nature* **163**, 525–526.
E. Munro (1970). Computer-aided design of magnetic electron lenses using the finite element method. ICEM-7, Grenoble. **2**, 55–56.
E. Munro (1971). A low-aberration probe-forming lens for a scanning electron microscope. EMAG, Cambridge. 84–87.
E. Munro (1972). Design of electrostatic lenses for field-emission electron guns. EUREM-5, Manchester. 22–23.
E. Munro (1973). Computer-aided design of electron lenses by the finite element method. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 284–323 (Academic Press, London & New York).
E. Munro (1988). Finite difference programs for computing tolerances for electrostatic lenses. *J. Vac. Sci. Technol. B* **6**, 941–948.
M. B. Nesvizhskii (1986). Aberration coefficients of complicated electron-optical cathode systems. *Radiotekh. Elektron.* **31**, 162–168. *Sov. J. Commun. Technol.* **31** (5), 153–160.

E. W. Ng, Ed. (1979). *Symbolic and Algebraic Computing* (Springer, Berlin & New York).
S. Nomura (2008). Design of apochromatic TEM composed of usual round lenses. EMC-14, Aachen. **1**, 41–42.
J. Orloff (1983a). On addition of spherical and chromatic aberration of a pair of electron lenses. *Optik* **63**, 369–372.
J. Orloff (1983b). The correct expression for the chromatic disk of confusion. *Optik* **65**, 369–371.
M. H. F. Overwijk, A. J. Bleeker and A. Thust (1996). Correction of three-fold astigmatism for ultra-high-resolution TEM. EUREM-11, Dublin. **1**, I404–I405.
M. H. F. Overwijk, A. J. Bleeker and A. Thust (1997). Correction of three-fold astigmatism for ultra-high-resolution TEM. *Ultramicroscopy* **67**, 163–170.
L. P. Ovsyannikova and S. Ya. Yavor (1965). Third order aberrations of the width of a line image in a combined quadrupole lens. *Zh. Tekh. Fiz.* **35**, 940–946. *Sov. Phys. Tech. Phys*. **10**, 723–726.
R. J. Pegis (1961). The modern development of Hamiltonian optics. *Prog. Opt.* **1**, 1–29.
D. P. R. Petrie (1962). The dependence of spherical aberration of electron lenses on the conjugate positions. ICEM-5, Philadelphia. **1**, KK-2.
J. Petzval (1843). *Bericht über die Ergebnisse einiger dioptrischer Untersuchungen* (Hertleben, Pesth; reprinted by Akademiai Kiado, Budapest, 1975).
G. N. Plass (1942). Electrostatic electron lenses with a minimum of spherical aberration. *J. Appl. Phys.* **13**. 49–55 & 524.
E. Plies and H. Rose (1971). Über die axialen Bildfehler magnetischer Ablenksysteme mit krummer Achse. *Optik* **34**, 171–190.
W. Pöhner and H. Rose (1974). Die Auflösungsgrenze korrigierter Elektronenmikroskope bei verbeulter Wellenfläche. *Optik* **41**, 69–89.
D. Preikszas (1995). Korrektur des Farb- und Öffnungsfehlers eines Niederspannungs-Elektronenmikroskops mit Hilfe eines Elektronenspiegels. Dissertation, Darmstadt.
D. Preikszas and H. Rose (1995). Procedures for minimizing the aberrations of electromagnetic compound lenses. *Optik* **100**, 179–187.
D. Preikszas and H. Rose (1996a). Conditions for correcting a round lens by an electron mirror. *Recent Trends* **5**, 59–60.
D. Preikszas and H. Rose (1996b). Stability requirements for the correction of aberrations by an electron mirror. EUREM-11, Dublin. **I**, 406–407.
D. Preikszas and H. Rose (1997). Correction properties of electron mirrors. *J. Electron Microsc.* **46**, 1–9.
B. M. Rabin, A. M. Strashkevich and L. S. Khin (1951). Electron optical properties of electrostatic systems with apertures deviating from the ideal shape. *Zh. Tekh. Fiz.* **21**, 438–444.
T. Radlička (2008). Lie algebraic methods in charged particle optics. *Adv. Imaging Electron Phys.* **151**, 241–362.
T. Radlička and M. Oral (2016). Correction of misalignment aberrations of a hexapole corrector using the differential algebra method. *Recent Trends* **15**, 50–53.
E. G. Ramberg (1939). A simplified derivation of the general properties of an electron-optical image. *J. Opt. Soc. Am.* **29**, 79–83.
O. Rang (1948). Ein verzeichnungsfreies kurzbrennweitiges elektrostatisches Projektiv. *Optik* **4**, 251–257.
O. Rang (1949a). Der elektrostatische Stigmator, ein Korrektiv für astigmatische Elektronenlinsen. *Optik* **5**, 518–530.
O. Rang (1949b). Eine nach Brechkraft und Azimut einfach einstellbare astigmatische Elektronenlinse. *Phys. Bl.* **5**, 78–80.
H. Rauh (1971). Ein Näherungsverfahren zur Berechnung rotationssymmetrischer elektrostatischer Felder mit beliebigen Randbedingungen und seine Anwendung auf einen Elektronenstrahlerzeuger mit Spitzenkathode. *Z. Naturforsch.* **26a**, 1667–1675.
F. H. Read, A. Adams and J. R. Soto-Montiel (1971). Electrostatic cylinder lenses I: two element lenses. *J. Phys. E: Sci. Instrum.* **4**, 625–632.
R. Rebsch (1938). Das theoretische Auflösungsvermögen des Elektronenmikroskops. *Ann. Phys. (Leipzig)* **31** (423), 551–560.

R. Rebsch (1940). Über den Öffnungsfehler der Elektronenlinsen. *Z. Phys.* **116**, 729–733.
R. Rebsch and W. Schneider (1937). Der Öffnungsfehler schwacher Elektronenlinsen. *Z. Phys.* **107**, 138–143.
A. Recknagel (1941). Über die sphärische Aberration bei elektronenoptischer Abbildung. *Z. Phys.* **117**, 67–73.
A. Recknagel and G. Haufe (1952). Über den Stigmator der Elektronenmikroskope. *Wiss. Z. Techn. Hochschule Dresden* **2**, 1–10.
E. Regenstreif (1951a). Une théorie de la lentille électrostatique indépendante à symétrie de révolution perturbée. *Ann. Radioel.* **6**, 244–267.
E. Regenstreif (1951b). Théorie du régime transgaussien de la lentille électrostatique elliptique. *Ann. Radioel.* **6**, 299–317.
E. Regenstreif (1951c). Sur la théorie de la lentille électrostatique indépendante à électrode centrale elliptique. *C. R. Acad. Sci. Paris* **232**, 1918–1920.
E. Regenstreif (1951d). Sur la théorie du régime transgaussien de la lentille électrostatique elliptique. *C. R. Acad. Sci. Paris* **233**, 854–856.
J. H. Reisner (1970). Measurement of radial distortion. EMSA, Houston, **28**, 350–351.
J. H. Reisner and J. J. Schuler (1967). A slotted electrode stigmator for correction of short focal length objectives. EMSA, Chicago. **25**, 226–227.
G. Rempfer (1985). Unipotential electrostatic lenses: paraxial properties and aberrations of focal length and focal point. *J. Appl. Phys.* **57**, 2385–2401.
G. F. Rempfer (1990a). A theoretical study of the hyperbolic electron mirror as a correcting element for spherical and chromatic aberration in electron optics. *J. Appl. Phys.* **47**, 6027–6040.
G. Rempfer (1990b). Comment on "A study on hyperbolic mirrors as correctors". *Optik* **86**, 73–74.
G. Rempfer (1999a). The relationship of lens aberrations to image aberrations in electron optics. *Optik* **110**, 17–24.
G. Rempfer (1999b). Reply to "Rebuttal of 'The relationship of lens aberrations to image aberrations in electron optics'" by P.W. Hawkes. *Optik* **110**, 375–377.
G. Rempfer (2000). Reply to "Clarification of my criticism of 'The relationship of lens aberrations to image aberrations in electron optics'" by P.W. Hawkes. *Optik* **111**, 229–234.
G. Rempfer and M. S. Mauck (1985). Aberration-correcting properties of the hyperbolic electron mirror. EMSA Louisville, **43**, 132–133.
G. Rempfer and M. S. Mauck (1986). An experimental study of the hyperbolic electron mirror. EMSA Albuquerque, **44**, 886–887.
G. F. Rempfer and M. S. Mauck (1992). Correction of chromatic aberration with an electron mirror. *Optik* **92**, 3–8.
G. F. Rempfer, D. M. Desloge, W. P. Skoczylas and O. H. Griffith (1997). Simultaneous correction of spherical and chromatic aberrations with an electron mirror: an electron optical achromat. *Microsc. Microanal.* **3**, 14–27.
A. Renau and D. W. O. Heddle (1986a). Geometric aberrations in electrostatic lenses: I. A simple and accurate computer model. *J. Phys. E: Sci. Instrum.* **19**, 284–288.
A. Renau and D. W. O. Heddle (1986b). Geometric aberrations in electrostatic lenses: II. Aberration coefficients and their interrelations. *J. Phys. E: Sci. Instrum.* **19**, 288–295.
A. Renau and D. W. O. Heddle (1987). Reply to "Comment on 'Geometric aberrations in electrostatic lenses: II'". *J. Phys. E: Sci. Instrum.* **20**, 235–236.
M. Rheinfurth (1955). Über den Farbfehler und den Öffnungsfehler elektronenoptischer Zylinderlinsen. *Optik* **12**, 411–416.
W. D. Riecke (1958). Über eine neue Einrichtung zur Feinstrahlbeugung. ICEM-4, Berlin. **1**, 189–194.
W. D. Riecke (1964). Einige Bemerkungen zur Gaußschen Dioptrik von magnetischen Objektivlinsen mit dezentrierten Polschuhen. EUREM-3, Prague. **A**, 7–8.
W. D. Riecke (1966). Zur Zentrierung des magnetischen Elektronenmikroskops. *Optik* **24**, 397–426.
W. D. Riecke (1972). Zur Zentrierung des magnetischen Elektronenmikroskops. *Optik* **36**. 66–84, 288–308 & 375–398.

W. D. Riecke (1976). Instrument operation for microscopy and microdiffraction. In *Electron Microscopy in Materials Science* (U. Valdrè and E. Ruedl, Eds), 19–111 (CEE, Luxemburg).
W. D. Riecke (1982). Practical lens design. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 163–357 (Springer, Berlin & New York).
H. Riedl (1937). Die Bildfehler 3. Ordnung der kurzen schwachen rein elektrischen Elektronen-Einzellinse. *Z. Phys.* **107**, 210–216.
W. Rogowski (1937). Über Fehler von Elektronenbildern. *Arch. Elektrotech.* **31**, 555–593.
H. Rose (1966). Gaußsche Dioptrik begrenzter unrunder Elektronenlinsen. *Optik* **24**, 108–121.
H. Rose (1967). Über den sphärischen und den chromatischen Fehler unrunder Elektronenlinsen. *Optik* **25**, 587–597.
H. Rose (1967). Über die Korrigierbarkeit von Linsen für schnelle Elektronen. *Optik* **26**, 289–298.
H. Rose (1968). Über die Berechnung der Bildfehler elektronenoptischer Systeme mit gerader Achse, I & II. *Optik* **27**. 466–474 & 497–514.
H. Rose (1968). Der Zusammenhang der Bildfehler-Koeffizienten mit den Entwicklungs-Koeffizienten des Eikonals. *Optik* **28**, 462–474.
H. Rose (1971a). Abbildungseigenschaften sphärisch korrigierter elektronenoptischer Achromate. *Optik* **33**, 1–24.
H. Rose (1971b). Elektronenoptische Aplanate. *Optik* **34**, 285–311.
H. Rose (1972). Zur Gaußschen Dioptrik elektrisch-magnetischer Zylinderlinsen. *Optik* **36**, 19–36.
H. Rose (1987). Hamiltonian magnetic optics. *Nucl. Instrum. Meth. Phys. Res. A* **258**, 374–401.
H. Rose (2012). *Geometrical Charged-particle Optics*, second edn. (Springer, Berlin; first edn, 2009).
H. Rose and R. W. Moses (1973). Minimaler Öffnungsfehler magnetischer Rund- und Zylinderlinsen bei feldfreiem Objektraum. *Optik* **37**, 316–336.
H. Rose and U. Petri (1971). Zur systematischen Berechnung elektronenoptischer Bildfehler. *Optik* **33**, 151–165.
R. Rüdenberg (1948). Electron lenses of hyperbolic field structure. *J. Franklin Inst.* **246**, 311–339 & 377–408.
Y. Sakaki and S. Maruse (1954). Method of measuring spherical and astigmatic aberrations of an electron lens. *J. Electronmicrosc.* **2**, 8–9.
H. Sawada (2015). Ronchigram and geometrical aberrations in STEM. In *Scanning Transmission Electron Microscopy of Nanomaterials* (N. Tanaka, Ed.), 461–485 (Imperial College Press, London).
H. Sawada, T. Sannomiya, F. Hosokawa, T. Nakamichi, T. Kaneyama, T. Tomita, Y. Kondo, T. Tanaka, Y. Oshima, Y. Tanishiro and K. Takayanagi (2008). Measurement method of aberration from Ronchigram by autocorrelation function. *Ultramicroscopy* **108**, 1467–1475.
W. O. Saxton (1995). Simple prescriptions for measuring three-fold astigmatism. *Ultramicroscopy* **58**, 239–243.
O. Scherzer (1933). Zur Theorie der elektronenoptischen Linsenfehler. *Z. Phys.* **80**, 193–302.
O. Scherzer (1936a). Die schwache elektrische Einzellinse geringster sphärischer Aberration. *Z. Phys.* **101**, 23–26.
O. Scherzer (1936b). Über einige Fehler von Elektronenlinsen. *Z. Phys.* **101**, 593–603.
O. Scherzer (1937). Berechnung der Bildfehler dritter Ordnung nach der Bahnmethode. In *Beiträge zur Elektronenoptik* (H. Busch and E. Brüche, Eds), 33–41 (Barth, Leipzig).
O. Scherzer (1941). Die unteren Grenzen der Brennweite und des chromatischen Fehlers von magnetischen Elektronenlinsen. *Z. Phys.* **118**, 461–466.
O. Scherzer (1946). Zur Korrigierbarkeit von Elektronenlinsen. *Phys. Bl.* **2**, 110.
O. Scherzer (1947). Sphärische und chromatische Korrektur von Elektronen-Linsen. *Optik* **2**, 114–132.
O. Scherzer (1982). How not to correct an electron lens. *Ultramicroscopy* **9**, 385–386.
P. Schiske (1956). Die untere Grenze des Farbfehlers magnetischer Linsen bei vorgeschriebener Maximalfeldstärke. *Optik* **11**, 502–505.
O. I. Seman (1951). Transformation of the form of the fourth-order eikonal and aberration coefficients in electron optics. *Dokl. Akad. Nauk SSSR* **81**, 775–778.

O. I. Seman (1953a). On the question of the existence of an extremum of the spherical aberration coefficient for axially symmetric systems in electron optics. *Zh. Eksp. Tekh. Fiz.* **24**, 581–588.
O. I. Seman (1953b). Limiting values of electron optical aberration coefficients. *Dokl. Akad. Nauk SSSR* **93**, 443–445.
O. I. Seman (1954). Relativistic aberration functions and normal forms of electron-optical aberration coefficients. *Dokl. Akad. Nauk SSSR* **96**, 1151–1154.
O. I. Seman (1955a). Relativistic theory of rotationally symmetric electron optics on the basis of eikonal theory. *Trudy Inst. Fiz. Astron. Akad. Nauk Eston. SSR* **2**, 3–29.
O. I. Seman (1955b). Expansion of the fourth-order eikonal and the relativistic theory of aberrations. *Trudy Inst. Fiz. Astron. Akad. Nauk Eston. SSR* **2**, 30–49.
O. I. Seman (1955c). On the theory of the fourth-order point eikonal in electron optics. *Uch. Zap. Rostov. Gos. Univ.* **68** (8), 77–90.
O. I. Seman (1958a). On the theory of virtual and transverse aberrations in electron optics. *Uch. Zap. Rostov. Gos. Univ.* **68** (8), 63–75.
O. I. Seman (1958b). On normal coefficients of electron-optical aberrations of magnetic immersion systems. *Radiotekh. Elektron.* **3**, 283–287. *Radio Eng. Electron.* **3**, 402–409.
O. I. Seman (1959a). Dependence of anisotropic distortion of the image on the curvature of the object field in electron optics. *Radiotekh. Elektron.* **4**, 1213–1214. *Radio Eng. Electron.* **4** (7), 227–230.
O. I. Seman (1959b). On the theoretical limits of the chromatic aberration coefficient in limited electron-optical magnetic fields. *Radiotekh. Elektron.* **4**, 1702–1707. *Radio Eng. Electron.* **4** (10), 235–244.
O. I. Seman (1959c). On some properties of the image curvature in short electron optical lenses. *Opt. Spektrosk.* **7**, 113–115. *Opt. Spectrosc.* **7**, 68–69.
O. I. Seman (1968). Properties of Petzval curvature in electron optics. *Radiotekh. Elektron.* **13**, 907–912. *Radio Eng. Electron. Phys.* **13**, 784–788.
A. Septier (1963). Lentille quadrupolaire magnéto-électrique corrigée de l'aberration chromatique. Aberration d'ouverture de ce type de lentilles. *C. R. Acad. Sci. Paris* **256**, 2325–2328.
A. Septier (1966). The struggle to overcome spherical aberration in the electron microscope. *Adv. Opt. Electron Microsc.* **1**, 204–274. Reprinted in *Adv. Imaging Electron Phys.* **202** (2017) 75–147.
O. Shánel, J. Zlámal and M. Oral (2014). Calculation of the performance of magnetic lenses with limited machining precision. *Ultramicroscopy* **137**, 1–6.
Z. Shao (1987a). A study on multipole systems as correctors. *Optik* **75**, 152–157.
Z. Shao (1987b). Probe size and 5th-order aberration. EMSA, Baltimore, **45**, 134–135.
Z. Shao and A. V. Crewe (1987). Spherical aberrations of multipoles. *J. Appl. Phys.* **62**, 1149–1153.
Z. Shao and X.-d. Wu (1989). Adjustable four-electrode mirror as an aberration corrector. *Appl. Phys. Lett.* **55**, 2696–2697.
Z. Shao and X.-d. Wu (1990a). Properties of a four-electrode adjustable electron mirror as an aberration corrector. *Rev. Sci. Instrum.* **61**, 1230–1235.
Z. Shao and X.-d. Wu (1990b). A study on hyperbolic mirrors as correctors. *Optik* **84**, 51–54.
C.-h. Shi (1956). Determination of the aberrations of third order of an electrostatic cylindrical lens by the point eikonal method [in Russian]. *Acta Sci. Nat. Univ. Pekin* **2**, 457–465.
H. Shimoyama (1982). The dependence of the asymptotic aberration coefficients of electron lenses on the magnification. *J. Electron Microsc.* **31**, 8–17.
E. V. Shpak and S. Ya. Yavor (1964). Achromatic electromagnetic quadrupole lens in which the field distributions on the axis do not coincide. *Zh. Tekh. Fiz.* **34**, 2003–2007. *Sov. Phys. Tech. Phys.* **9**, 1540–1543.
E. V. Shpak and S. Ya. Yavor (1965). Achromatic electromagnetic quadrupole lens in which the field distributions on the axis do not coincide. II. *Zh. Tekh. Fiz.* **35**, 947–950. *Sov. Phys. Tech. Phys.* **10**, 727–729.
N. I. Shtepa (1952). The third order aberrations of cylindrical electron lenses. *Zh. Tekh. Fiz.* **22**, 216–226.

K. Siegbahn (1946). A magnetic lens of special field form for β and γ-ray investigations; designs and applications. *Philos. Mag.* **37**, 162–184.
Y. P. Sivkov (1971). On the study of invariants of equations of motion in calculations of aberration coefficients. *Part. Accel.* **2**, 243–249.
N. A. Smirnov, M. A. Monastyrskii and Yu. V. Kulikov (1979). Electron optics of cathode lenses with spherically symmetric combined fields. *Zh. Tekh. Fiz.* **49**, 2590–2595. *Sov. Phys. Tech. Phys.* **24**, 1462–1465.
T. Smith (1921). The changes in aberrations when the object and stop are moved. *Trans. Opt. Soc. (Lond.)* **23**, 311–322.
T. Soma (1977). Relativistic aberration formulas for combined electric–magnetic focusing–deflection system. *Optik* **49**, 255–262.
V. V. Sorokina and P. V. Timofeev (1948). On the form of the field for electrostatic lenses. *Zh. Tekh. Fiz.* **18**, 509–516.
K. G. Steffen (1965). *High Energy Beam Optics* (Wiley–Interscience, New York & London).
D. Stenkamp (1998). Detection and quantitative asessment of image aberrations from single HRTEM lattice images. *J. Microsc. (Oxford)* **190**, 194–203.
W. G. Stephan (1947). Practische Toepassingen op het Gebied der Algebraïsche Optiek. Dissertation, Delft.
P. A. Stoyanov (1955a). Influence of of the geometrical form of the polepieces of the objective on the resolving power of the electron microscope. *Zh. Tekh. Fiz.* **25**, 625–635.
P. A. Stoyanov (1955b). A stigmator for an electron microscope. *Zh. Tekh. Fiz.* **25**, 2537–2541.
P. A. Stoyanov [Stojanow] (1958). Stigmatoren für ein Mehrlinsen-Mikroskop. ICEM-4, Berlin. **1**, 61–66.
A. M. Strashkevich (1959). *Elektronnaya Optika Elektrostaticheskikh Polei ne Obladayushchikh Osevoi Simmetriei* (Gos. Izd. Fiz.-Mat. Lit., Moscow).
A. M. Strashkevich (1965). Focusing of wide beams in the fields of electrostatic cylindrical lenses and transaxial systems. *Zh. Tekh. Fiz.* **35**, 177–183. *Sov. Phys. Tech. Phys.* **10**, 147–151.
A. M. Strashkevich (1966). *Elektronnaya Optika Elektrostaticheskikh Sistem* (Energiya, Moscow).
A. M. Strashkevich and N. G. Gluzman (1954). Aberrations of relativistic electron beams. *Zh. Tekh. Fiz.* **24**, 2271–2284.
A. M. Strashkevich and I. M. Pilat (1951). The aberrations of corpuscular beams of arbitrary form in arbitrary electrostatic fields. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **15**, 448–466.
A. M. Strashkevich and I. M. Pilat (1952). The aberrations of beams with a curved axis in an arbitrary plane electrostatic field. *Uchen. Zap. Chernovitskii Gos. Univ. (Ser. Fiz.-Mat. Nauk)* **4**, 113–122.
P. A. Sturrock (1949). The aberrations of magnetic lenses due to asymmetries. ICEM-1, Delft. 89–93.
P. A. Sturrock (1951a). Perturbation characteristic functions and their application to electron optics. *Proc. R. Soc. (Lond.) A* **210**, 269–289.
P. A. Sturrock (1951b). The aberrations of magnetic electron lenses due to asymmetries. *Philos. Trans. R. Soc. (Lond.) A* **243**, 387–429.
P. A. Sturrock (1951c). Formules nouvelles pour les aberrations du troisième ordre des lentilles magnétiques. *C. R. Acad. Sci. Paris* **233**, 146–147.
P. A. Sturrock (1951d). Formules nouvelles pour les aberrations du troisième ordre des lentilles électrostatiques. *C. R. Acad. Sci. Paris* **233**, 243–245.
P. A. Sturrock (1952). The imaging properties of electron beams in arbitrary static electromagnetic fields. *Philos. Trans. R. Soc. (Lond.) A* **245**, 155–187.
P. A. Sturrock (1955). *Static and Dynamic Electron Optics* (Cambridge University Press, Cambridge).
Y. Sugiura and S. Suzuki (1943). On the magnetic lens of minimum spherical aberration. *Proc. Imp. Acad. Japan* **19**, 293–302.
S. Suzuki and A. Ishikawa (1978). On the magnetic electron lens of minimum spherical aberration. ICEM-9, Toronto. **1**, 24–25.
B. Tadano, H. Kimura and Y. Onuma (1966). Improvements of resolution in Hitachi electron microscopes. *Hitachi Rev.* **15**, 340–344.

I. N. Taganov (1966). Design of electron-optical systems with correction of third-order aberrations. *Radiotekh. Elektron.* **11**, 1329−1330. *Radio Eng. Electron. Phys.* **11**, 1160−1162.
I. N. Taganov and P. P. Kas'yankov (1964). On a direct method of calculating electromagnetic systems. *Opt.-Mekh. Prom.* **31** (11), 14−16.
I. N. Taganov and P. P. Kas'yankov (1965). On the calculation of electron-optical systems corrected for field curvature and anisotropic astigmatism. *Opt.-Mekh. Prom.* **32** (12), 21−23.
I. N. Taganov and P. P. Kas'yankov (1967). Use of the direct method for calculating electron optical systems with corrected third-order aberrations. In *Chislennye Metody Rascheta Elektronno-Opticheskikh Sistem* (G. I. Marchuk, Ed.), 11−22 (Nauka, Novosibirsk).
A. Takaoka (1995). Theory of cathode lenses for probe beams. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 243−248.
W. Tretner (1950). Die untere Grenze des Öffnungsfehlers magnetischer Elektronenobjektive. *Optik* **7**, 242.
W. Tretner (1954). Die untere Grenze des Öffnungsfehlers bei magnetischen Elektronenlinsen. *Optik* **11**, 312−326.
W. Tretner (1955). Berichtigung zu "Die untere Grenze des Öffnungsfehlers bei magnetischen Elektronenlinsen". *Optik* **12**, 293−294.
W. Tretner (1956). Zur unteren Grenze des Öffnungsfehlers magnetischer Elektronenlinsen. *Optik* **13**, 516−519.
W. Tretner (1959). Existenzbereiche rotationssymmetrischer Elektronenlinsen. *Optik* **16**, 155−184.
R. M. Tromp (2015a). Catadioptric aberration correction in cathode lens microscopy. *Ultramicroscopy* **151**, 191−198.
R. M. Tromp (2015b). An adjustable electron achromat for cathode lens microscopy. *Ultramicroscopy* **159**, 497−502.
R. M. Tromp, J. B. Hannon, A. W. Ellis, W. Wan, A. Berghaus and O. Schaff (2010). A new aberration-corrrected, energy-filtered LEEM/PEEM instrument. I. Principles and design. *Ultramicroscopy* **110**, 852−861.
R. M. Tromp, J. B. Hannon, W. Wan, A. Berghaus and O. Schaff (2013). A new aberration-corrrected, energy-filtered LEEM/PEEM instrument. II. Operation and results. *Ultramicroscopy* **127**, 25−39.
K. Tsuno and Y. Harada (1981a). Minimisation of radial and spiral distortion in electron microscopy through the use of a triple pole-piece lens. *J. Phys. E: Sci. Instrum.* **14**, 313−320.
K. Tsuno and Y. Harada (1981b). Elimination of spiral distortion in electron microscopy using an asymmetrical triple pole-piece lens. *J. Phys. E: Sci. Instrum.* **14**, 955−960.
K. Tsuno, Y. Ishida and Y. Harada (1980a). Spiral distortion in electron microscopy during use of three-pole-lens. EMSA, San Francisco, **38**, 280−281.
K. Tsuno, Y. Arai and Y. Harada (1980b). Elimination of spiral distortion in electron microscope by means of a three-pole-piece lens. EUREM-7, The Hague. **1**, 76−77.
M.-z. U (1957). Investigations on aberrations of fifth order in rotationally symmetric electron optical systems [in Russian]. *Sci. Sin.* **6**, 833−846. Also M.-z. Wu (1957) *Acta Phys. Sin.* **13**, 181−206.
S. Uhlemann and M. Haider (1998). Residual wave aberrations in the first spherical aberration corrected transmission electron microscope. *Ultramicroscopy* **72**, 109−119.
S. Uhlemann, H. Müller, P. Hartel, J. Zach and M. Haider (2013a). Instrumental resolution limit by magnetic thermal noise from conductive parts. *Microsc. Microanal.* **19** (Suppl. 2), 598−599.
S. Uhlemann, H. Müller, P. Hartel, J. Zach and M. Haider (2013b). Thermal magnetic field noise limits resolution in transmission electron microscopy. *Phys. Rev. Lett.* **111**, 046101, 5 pp.
S. Uhlemann, H. Müller, J. Zach, C. Berger and M. Haider (2014). Thermal magnetic field noise and electron optics − more experiments and calculations. IMC-18, Prague. IT-16-IN-1882.
S. Uhlemann, H. Müller, J. Zach and M. Haider (2015). Thermal magnetic field noise: electron optics and decoherence. *Ultramicroscopy* **151**, 199−210.
A. A. van Gorkum and T. G. Spanjer (1986). A generalized comparison of spherical aberration of magnetic and electrostatic electron lenses. *Optik* **72**, 134−136.

A. C. S. van Heel (1949). Calcul des aberrations du cinquième ordre et projets tenant compte de ces aberrations. In *La Théorie des Images Optiques* (P. Fleury, Ed.), 32–67 (Editions de la Revue d'Optique, Paris).
A. C. S. van Heel (1964). *Inleiding in de Optica*, 5th edn (Nijhoff, The Hague).
Yu. V. Vandakurov (1955a). The equations of electron optics for wide beams, with calculation of the chromatic aberrations and their application to the study of the motion of particles in axially symmetric fields. *Zh. Tekh. Fiz.* **25**, 1412–1425.
Yu. V. Vandakurov (1955b). On some antisymmetric magnetic fields with double focusing. *Zh. Tekh. Fiz.* **25**, 2545–2555.
Yu. V. Vandakurov (1956a). On some antisymmetric magnetic fields with double focusing. II. *Zh. Tekh. Fiz.* **26**, 1599–1610. *Sov. Phys. Tech. Phys*. **1**, 1558–1569.
Yu. V. Vandakurov (1956b). Electron-optical systems, the fields of which are independent of one coordinate. *Zh. Tekh. Fiz.* **26**, 2578–2594. *Sov. Phys. Tech. Phys*. **1**, 2491–2507.
Yu. V. Vandakurov (1957). On the theory of the aberrations in electron-optical focusing systems with a curvilinear axis. *Zh. Tekh. Fiz.* **27**, 1850–1862. *Sov. Phys. Tech. Phys*. **2**, 1719–1733.
C. H. F. Velzel (1987). Image formation by a general optical system, using Hamilton's method. *J. Opt. Soc. Am. A* **4**, 1342–1348.
C. H. F. Velzel and J. L. F. de Meijere (1988). Characteristic functions and the aberrations of symmetric optical systems. I. Transverse aberrations when the eikonal is given. II. Addition of aberrations. *J. Opt. Soc. Am. A* **5**. 246–250 & 251–256.
J. L. Verster (1963). On the use of gauzes in electron optics. *Philips Res. Rep.* **18**, 465–605.
A. G. Vlasov (1944). A magnetic lens with minimum spherical aberration. *Izv. Akad. Nauk SSSR (Ser. Fiz.)* **8**, 235–239.
A. G. Vlasov and I. P. Shakhmatova (1962). Field of a lens with disturbed axial symmetry. *Zh. Tekh. Fiz.* **32**, 695–705. *Sov. Phys. Tech. Phys*. **7**, 507–514.
H. Voit (1939). Über die elektronenoptischen Bildfehler dritter Ordnung. *Z. Instrumentenkde* **59**, 71–82.
J. von zur Gathen and J. Gerhard (2013). *Modern Computer Algebra*, 3rd edn (Cambridge University Press, Cambridge).
L.-p. Wang, E. Munro, J. Rouse and H.-n. Liu (2008a). Simulation of electron mirrors by the differential algebraic method. *Phys. Procedia* **1**, 297–3304.
L.-p. Wang, J. Rouse, E. Munro, H.-n. Liu and X.-q. Zhu (2008b). Aberration analysis of electron mirrors by a differential algebraic method. *Optik* **119**, 80–96.
A. Watanabe and N. Morito (1955). Zur Messung der Farbfehlerkonstante von magnetischen Linsen. *Optik* **12**. 166–172 & 564.
M. Watanabe and T. Someya (1963). Electro-magnetic stigmator for magnetic lens. *Optik* **20**, 99–108.
L. Wegmann (1953). Ein verzeichnungsarmes elektromagnetisches Doppelprojektiv für Elektronenmikroskope. *Helv. Phys. Acta* **26**, 448–449.
L. Wegmann (1954). Ein verzeichnungsarmes elektromagnetisches Doppelprojektiv für Elektronenmikroskope. *Optik* **11**, 153–170.
L. Wei and T. Yan (1999). Determination of the acceptable tolerance in the manufacture of electron optical system. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 393–398.
H. Wollnik (1987). *Optics of Charged Particles* (Academic Press, Orlando & London).
H. Wollnik and M. Berz (1985). Relations between elements of transfer matrices due to the condition of symplecticity. *Nucl. Instrum. Meth. Phys. Res A* **238**, 127–140.
M.-z. Wu, see M.-z. U.
J.-y. Ximen (1957). On the electron-optical properties and aberration theory of a combined immersion objective [in Chinese]. *Acta Phys. Sin.* **13**, 339–356.
J.-y. Ximen (1990a). Canonical theory and matrix representations in electron optical systems possessing aberrations. *Optik* **84**, 83–90.
J.-y. Ximen (1990b). Canonical aberration theory in electron optics. *J. Appl. Phys.* **58**, 5963–5967.

J.-y. Ximen (1991). Canonical aberration theory for calculating higher-order chromatic aberrations. *J. Appl. Phys.* **69**, 1962–1967.
J.-y. Ximen and Q.-s. Chen (1964). Analytical solution of the potential distribution and electron optics in magnetic lenses with pole-pieces of elliptic distortion [in Chinese]. *Acta Electron. Sin.* **4**, 72–84.
J.-y. Ximen and Z. Liu (1997). A theorem of higher-order spheric aberrations in Glaser's bell-shaped magnetic lens. *Optik* **107**, 17–25.
J.-y. Ximen and Z. Liu (2000a). Analysis and calculation of third- and fifth-order aberrations in combined bell-shaped electromagnetic lens – a new theoretical model in electron optics. *Optik* **111**, 75–84.
J.-y. Ximen and Z. Liu (2000b). Third-order geometric aberrations in Glaser's bell-shaped magnetic lens for object magnetic immersion. *Optik* **111**, 355–358.
J.-y. Ximen and Z.-h. Xi (1964). Theoretical study of field distribution and aberrations in a magnetic lens with pole pieces of inclined end-faces [in Chinese]. *Acta Electron. Sin.* **3**, 24–35.
J.-y. Ximen, L.-w. Zhou and K.-c. Ai (1983). Variational theory of aberrations in cathode lenses. *Optik* **66**, 19–34.
E. M. Yakushev (2013). Theory and computation of electron mirrors: the central particle method. *Adv. Imaging Electron Phys.* **178**, 147–247.
E. M. Yakushev and L. M. Sekunova (1986). Theory of electron mirrors and cathode lenses. *Adv. Electron. Electron Phys.* **68**, 337–416.
T. Yanaka and K. Shirota (1970). High resolution specimen area imaged under negligible field aberrations. ICEM-7, Grenoble. **2**, 59–60.
S. Ya. Yavor (1968). *Fokusirovka Zaryazhennykh Chastits Kvadrupol'nymi Linzami* (Atomizdat, Moscow).
M. I. Yavor (1993). Methods for calculation of parasitic aberrations and machining tolerances in electron optical systems. *Adv. Electron. Electron Phys.* **86**, 225–281.
M. I. Yavor and A. S. Berdnikov (1995). ISIOS: a program to calculate imperfect static charged particle optical sytems. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 416–422.
S. Ya. Yavor, A. D. Dymnikov and L. P. Ovsyannikova (1964). Experimental investigation of a quadrupole lens with zero and negative chromatic aberration. *Zh. Tekh. Fiz.* **34**, 99–104. *Sov. Phys. Tech. Phys.* **9**, 76–80.
L.-w. Zhou, K.-c. Ai and S.-c. Pan (1983). On aberration theory of the combined electromagnetic focusing cathode lenses. *Acta Phys. Sin.* **32**, 376–392.
V. A. Zhukov and A. V. Zav'yalova (2006). Axially symmetric composite electromagnetic mirror for perfect axial-aberration correction. *Mikroelektronika* **35**, 434–444. *Russ. Microelec.* **35** (6), 372–381.
J. Zlámal and B. Lencová (2011). Development of the program EOD for design in electron and ion microscopy. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 278–282.
J. Zlámal and B. Lencová (2015). Influence of saturation of magnetic lens material on fields of deflectors and parasitic fields. *Microsc. Microanal.* **21** (Suppl. 4), 188–193.
V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier and A. W. Vance (1945). *Electron Optics and the Electron Microscope* (Wiley, New York & Chapman & Hall, London).

Part V, Chapters 32 and 33

We draw attention to the survey by Ritz (1979), to the earlier papers on deflection aberrations by Griimm and Spurny (1956), Werner (1963), Hutter (1947, 1948, 1967) and Schürmann and Haussmann (1967) and to other work by Munro (1975, 1980), Soma (1979), Knauer (1981), Owen (1981) and Tsumagari et al. (1986, 1987, 1991), who consider parasitic aberrations. See also Baranova and Yavor (1996) and Petrov et al. (2001) on achromatic deflectors, Hu and Tang (1998, 1999a,b) and Hu et al. (1999) on higher order field functions and the use of Lie algebra, Oral et al. (2015) on dynamic correction and an original proposal by Retsky (1997a–c, 2001).

Papers on this subject frequently appear in the proceedings of the Electron, Ion and Photon Beam Technology and Nanofabrication conferences, published in *J. Vac. Sci. Technol.*, in those of the Micro and Nano

Engineering meetings now published in *Microelectronic Engineering* and in those of the Microprocesses and Nanotechnology conferences, published in the *Japanese Journal of Applied Physics*.

***

K. Amboss and E. D. Wolf (1971). Double-deflection aberrations in a scanning electron microscope. In *Record 11th Symposium Electron, Ion & Laser Beam Technology* (R. F. M. Thornley, Ed.), 195–204 (San Francisco Press, San Francisco).

L. A. Baranova and S. Ya. Yavor (1996). Achromatic deflection system. *Proc. SPIE* **2858**, 81–87.

H. C. Chu and E. Munro (1981). Computerized optimization of electron-beam lithography systems. *J. Vac. Sci. Technol.* **19**, 1053–1057.

H. C. Chu and E. Munro (1982a). Numerical analysis of electron beam lithography systems. Part III. Calculation of the optical properties of electron focusing systems and dual-channel deflection systems with combined magnetic and electrostatic fields. *Optik* **61**, 121–145.

H. C. Chu and E. Munro (1982b). Numerical analysis of electron beam lithography systems. Part IV. Computerized optimization of the electron optical performance of electron beam lithography systems using the damped least squares method. *Optik* **61**, 213–246.

A. V. Crewe and N. W. Parker (1976). Correction of third-order aberrations in the scanning electron microscope. *Optik* **46**, 183–194.

S.-q. Ding (1982). A generalized coefficient formula of the third order magnetic deflection aberrations and its relations with the raster pattern. ICEM-10, Hamburg. **1**, 315–316.

A. L. Dodin (1983). Electrostatic deflecting elements without fourfold aberrations. *Radiotekh. Elektron.* **28**, 357–361. *Radio Eng. Electron. Phys*. **28** (2), 114–118.

W. Glaser (1938). Über die Ablenkfehler von elektrischen und magnetischen Ablenksystemen. *Z. Phys.* **111**, 357–372.

W. Glaser (1941). Berichtigung: Über die Ablenkfehler von elektrischen und magnetischen Ablenksystemen. *Z. Phys.* **117**, 412.

W. Glaser (1949). Über die Theorie der elektrischen und magnetischen Ablenkung von Elektronenstrahlbündeln und ein ihr angepaßtes Störungsverfahren. *Ann. Phys. (Leipzig)* **4** (439), 389–408.

W. Glaser (1952). *Grundlagen der Elektronenoptik* (Springer, Vienna).

W. Glaser (1956). Elektronen- und Ionenoptik. *Handbuch der Physik* **33**, 123–395.

E. Goto and T. Soma (1977). MOL (moving objective lens). Formulation of deflective aberration free system. *Optik* **48**, 255–270.

H. Grümm and H. Spurny (1956). Ein analytisches Modell für elektronenoptische Ablenkfelder. *Öst. Ing.-Arch.* **10**, 104–106.

J. Haantjes and G. J. Lubben (1957). Errors of magnetic deflection. I. *Philips Tech. Rep.* **12**, 46–68.

J. Haantjes and G. J. Lubben (1959). Errors of magnetic deflection. II. *Philips Tech. Rep.* **14**, 65–97.

P. W. Hawkes (1989). The structure of the asymptotic aberration coefficients of combined focusing and deflection systems. Part. I. General theory and aberrations of position. Part II. Aberrations of gradient and overall structure. *Optik* **83**. 104–112 & 122–128.

T. Hosokawa (1980). Systematic elimination of third order aberrations in electron beam scanning system. *Optik* **56**, 21–30.

T. Hosokawa (2002). Relationship between the fifth order aberration coefficients in electron optical deflective focusing system. *Optik* **113**, 105–110.

K.-y. Hu and T.-t. Tang (1998). Lie algebraic aberration theory and calculation method for combined electron-beam focusing–deflection systems. *J. Vac. Sci. Technol. B* **16**, 3248–3255.

K.-y. Hu and T.-t. Tang (1999a). Lie algebra deflection aberration theory for magnetic deflection systems. *Optik* **110**, 9–16.

K.-y. Hu and T.-t. Tang (1999b). High-order deflection field functions for toroidal and saddle yokes. *Optik* **110**, 150–156.

K.-y. Hu, T.-t. Tang and C.-y. Tian (1999). Study on Lie algebraic deflection aberration theory for combined magnetic focusing–deflection systems. *Optik* **110**, 271–278.

R. G. E. Hutter (1947). Electron beam deflection: Part I. Small-angle deflection theory. Part II. Applications of the small-angle deflection theory. *J. Appl. Phys.* **18**, 740–758 & 797–810.
R. G. E. Hutter (1948). The deflection of beams of charged particles. *Adv. Electron.* **1**, 167–218.
R. G. E. Hutter (1967). Deflection defocusing correction by means of rotationally symmetrical lenses only. *IEEE Trans. Electron Devices* **14**, 694–699.
J. Kaashoek (1968). A study of magnetic-deflection errors. *Philips Tech. Rep.* Suppl. 11, 114 pp.
K. Kanaya and H. Kawakatsu (1961a). Double electro-static deflection of electron beams without aberrations. *Bull. Electrotech. Lab.* **25**, 241–252.
K. Kanaya and H. Kawakatsu (1961b). Deflection system with eight-pole stigmator used in correcting astigmatism. *J. Electronmicrosc.* **10**, 218–221.
K. Kanaya and H. Kawakatsu (1962). An anastigmatic deflection system consisting of eight poles for use with electron beam instruments. *Bull. Electrotech. Lab.* **26**, 241–250.
K. Kanaya, H. Kawakatsu and K. Tanaka (1961). Electron-optical design of deflection system free from aberrations by means of stigmator. *Bull. Electrotech. Lab.* **25**, 481–494.
K. Kanaya, H. Kawakatsu, I. Okazaki and T. Takizawa (1963). Magnetic deflection of electron beams and caustic patterns. *Bull. Electrotech. Lab.* **27**, 401–417.
K. Kanaya, H. Kawakatsu, I. Okazaki and T. Takizawa (1964). A star pattern of deflected beam caused by coma aberration. *J. Electronmicrosc.* **13**, 80–86.
H. Kawakatsu and K. Kanaya (1961). Electro-static deflection system used in electron beam apparatus. *J. Electronmicrosc.* **10**, 119–124.
D. P. Kern (1979). Optimization of electron probe forming systems with respect to aberrations and vertical beam landing. *J. Vac. Sci. Technol.* **16**, 1686–1691.
W. Knauer (1981). A method to reduce deflection aberrations in electron-beam lithography systems. *J. Vac. Sci. Technol.* **19**, 1042–1047.
H. Koops (1972). Elektronenoptische Verkleinerung durch Kombination einer langbrennweitigen Feldlinse mit einer starken kurzbrennweitigen Magnetlinse. *Optik* **36**, 93–110.
H. Koops (1973). On electron projection systems. *J. Vac. Sci. Technol.* **10**, 909–912.
H. Koops and W. Bernhard (1975). Electron-beam projection sytems with compensated chromatic field aberrations. *J. Vac. Sci. Technol.* **12**, 1141–1145.
K. Kuroda (1980). Simplified method for calculating deflective aberration of electron optical system with two deflectors and lens. *Optik* **57**, 251–258.
M. Lenc and B. Lencová (1988). On the deflection of electrons with homogeneous fields. *Optik* **78**, 127–131.
B. Lencová (1981). Deflection aberrations of multi-stage deflection systems. *Optik* **58**, 25–35.
B. Lencová (1988). On the design of electron beam deflection systems. *Optik* **79**, 1–12.
Y. Li (1981). The relativistic aberration theory of a combined electric-magnetic focusing–deflection system. *Acta Phys. Sin.* **30**, 1155–1164.
Y. Li (1983). On the errors of Soma's relativistic aberration formulas for combined focusing–deflection system. *Optik* **63**, 213–216.
Y. Li (1986). On the relativistic fifth order geometrical aberration equation of a combined focusing–deflection system. *Optik* **74**, 65–70.
Y. Li (1992). A relativistic aberration theory without high order derivatives for a combined focusing–deflectrion system. *Nucl. Instrum. Meth. Phys. Res. A* **316**, 89–98.
Y. Li (1994). A general theory of relativistic asymptotic aberrations of a system of combined focusing–deflection sub-systems. *J. Phys. D: Appl. Phys.* **27**, 2271–2283.
Y. Li and J.-y. Ximen (1982a). On the relativistic aberration theory of a combined focusing–deflection system with multi-stage deflectors. *Optik* **61**, 315–332.
Y. Li and J.-y. Ximen (1982b). On the relativistic aberration theory of a combined focusing–deflection system with multi-stage deflectors. *Acta Phys. Sin.* **31**, 604–614.
Y. Li, S.-q. Kuang, Z.-q. Feng and T.-y. Liu (1993). The relativistic fifth-order geometrical aberrations of a combined focusing–deflection system. *J. Phys. D: Appl. Phys.* **26**, 522–538.

E. Munro (1974). Calculation of the optical properties of combined magnetic lenses and deflection systems with superimposed fields. *Optik* **39**, 450–466.
E. Munro (1975). Design and optimization of magnetic lenses and deflection systems for electron beams. *J. Vac. Sci. Technol.* **12**, 1146–1150.
E. Munro (1980). Electron optics for microcircuit engineering. In *Microcircuit Engineering* (H. Ahmed and W. C. Nixon, Eds), 513–534 (Cambridge University Press, Cambridge).
E. Munro and H. C. Chu (1982a). Numerical analysis of electron beam lithography systems. Part I. Computation of fields in magnetic deflectors. *Optik* **60**, 371–390.
E. Munro and H. C. Chu (1982b). Numerical analysis of electron beam lithography systems. Part II. Computation of fields in electrostatic deflectors. *Optik* **61**, 1–16.
H. Ohiwa (1970). Elimination of third order aberrations in electron beam scanning systems. Dissertation, Tokyo.
H. Ohiwa (1978). Design of electron-beam scanning systems using the moving objective lens. *J. Vac. Sci. Technol.* **15**, 849–852.
H. Ohiwa (1979). Moving objective lens and the Fraunhofer condition for pre-deflection. *Optik* **53**, 63–68.
H. Ohiwa, E. Goto and A. Ono (1971). Elimination of third-order aberrations in electron-beam scanning systems. *Trans. Inst. Electron. Commun. Eng. Japan* **54-B** (12), 730–737. *Electron. Commun. Japan* **54-B**, No. 12, 44–51.
M. Oral, V. Neděla and G. D. Danilatos (2015). Dynamic correction of higher-order deflection aberrations in the environmental SEM. *Microsc. Microanal.* **21** (Suppl. 4), 194–199.
G. Owen (1981). Automatic measurement and correction of deflection astigmatism and defocusing in the Hewlett–Packard 605 electron beam lithography system. *J. Vac. Sci. Technol.* **19**, 1064–1068.
G. Owen and W. C. Nixon (1973). Aberration correction for increased lines per field in scanning electron beam technology. *J. Vac. Sci. Technol.* **10**, 983–986.
I. A. Petrov, T. Ya. Fishkova and L. P. Ovsyannikova (2001). Theoretical analysis of achromatic deflectors. *Proc. SPIE* **4510**, 225–232.
J. Picht (1943). Beiträge zur Theorie der elektrischen Ablenkung von Elektronenstrahlbündeln. IV. Ergänzende Bemerkungen zu Beitrag II und III sowie Zusammenhang der in III für gekreuzte elektrische Ablenksysteme gefundenen Fehlerausdrücke mit den von G. Wendt in "Bildfehler bei Ablenkung eines Kathodenstrahlbündels in zwei gekreuzten Ablenkfeldern" abgeleiten Fehlerausdrücken. *Ann. Phys. (Leipzig)* **43** (435), 53–72.
J. Picht and J. Himpan (1941a). Beiträge zur Theorie der elektrischen Ablenkung von Elektronenstrahlbündeln. I. Allgemeine Untersuchungen über den Strahlenverlauf in elektrostatischen Ablenkfeldern. *Ann. Phys. (Leipzig)* **39** (431), 409–435.
J. Picht and J. Himpan (1941b). Beiträge zur Theorie der elektrischen Ablenkung von Elektronenstrahlbündeln. II. Elektrische Ablenkung eines (ausgedehnten) elektronenoptischen Bildes und die dabei auftretenden Bild- und Ablenkfehler bis zu dritten Ordnung. *Ann. Phys. (Leipzig)* **39** (431), 436–477.
J. Picht and J. Himpan (1941c). Beiträge zur Theorie der elektrischen Ablenkung von Elektronenstrahlbündeln. III. Dynamischer Bildaufbau mittels gekreuzter elektrischer Ablenksysteme und die dabei Auftretenden Abbildungsfehler bis zur dritten Ordnung. *Ann. Phys. (Leipzig)* **39** (431), 478–501.
E. Plies (1982a). Berechnung zusammengesetzter elektronenoptischer Fokussier- und Ablenksysteme mit überlagerten Feldern. 1. Feldentwicklung und Bahngleichung. *Siemens Forsch. Entwickl. Ber.* **11**, 38–45.
E. Plies (1982b). Berechnung zusammengesetzter elektronenoptischer Fokussier- und Ablenksysteme mit überlagerten Feldern. 2. Fundamentalbahnen und Bildfehler. *Siemens Forsch. Entwickl. Ber.* **11**, 83–90.
E. Plies (1994). Electron optics of low-voltage electron beam testing and inspection. Part I: Simulation tools. *Adv. Opt. Electron Microsc.* **13**, 123–242.
E. Plies and R. Elstner (1989a). Computation of first- and third-order properties of compound electron-optical lens and deflector systems. Part I: Fundamental trajectories, cardinal elements and aberrations. *Optik* **82**, 57–63.
E. Plies and R. Elstner (1989b). Computation of first- and third-order properties of compound electron-optical lens and deflector systems. Part II: Tests and applications of the computer program. *Optik* **83**, 11–18.

V. R. M. Rao and W. C. Nixon (1981). A computational and experimental analysis of fifth-order deflection aberrations. *J. Vac. Sci. Technol.* **19**, 1037–1041.
M. Retsky (1997a). Electrostatic deflection aberrations revisited: solution proposed to an old problem. *J. Vac. Sci. Technol. B* **15**, 2725–2728.
M. Retsky (1997b). A new method to reduce electrostatic deflection defocusing in CRTs and other devices. *Digest of Technical Papers, Internat. Symp. Society for Information Display* **28**, 536–539.
M. Retsky (1997c). Electrostatic deflection aberrations studied with ray tracing: a surprising and simple solution proposed. *Proc. SPIE* **3155**, 89–98.
M. Retsky (2001). Testing an electron beam deflection innovation. *Proc. SPIE* **4510**, 196–204.
E. F. Ritz (1979). Recent advances in electron beam deflection. *Adv. Electron. Electron Phys.* **49**, 299–357.
H. Rose (2012). *Geometrical Charged-particle Optics*, second ed. (Springer, Berlin; first edn, 2009).
J. Schürmann and G. Haussmann (1967). Die Korrektur der Ablenkfehler magnetischer Ablenksysteme mit den Mitteln der Analogrechentechnik. *Z. Angew. Phys.* **22**, 235–239.
Y. P. Sivkov (1971). On the study of invariants of equations of motion in calculations of aberration coefficients. *Part. Accel.* **2**, 243–249.
M. R. Smith and E. Munro (1986). A unified aberration theory for electrostatic round lenses, multipole lenses and deflectors. *Optik* **74**, 7–16.
M. R. Smith and E. Munro (1987). Computer programs for designing multipole electron and ion optical systems. *J. Vac. Sci. Technol. B* **5**, 161–164.
T. Soma (1977). Relativistic aberration formulas for combined electric–magnetic focusing–deflection system. *Optik* **49**, 255–262.
T. Soma (1979). A practical focus–deflection system for shaped electron beams. *Optik* **53**, 281–284.
T.-t. Tang (1986). Asymptotic aberrations of combined electrostatic focusing–deflection systems. *Optik* **74**, 43–47.
T. Tsumagari, J. Murakami, H. Ohiwa and T. Noda (1986). Aberrations of electrostatic systems with machining error. *J. Vac. Sci. Technol. B* **4**, 140–142.
T. Tsumagari, J. Murakami, H. Ohiwa and T. Noda (1987). Analysis of mechanical aberrations in the electrostatic deflector. *Jpn. J. Appl. Phys.* **26**, 1772–1776.
T. Tsumagari, H. Ohiwa and T. Noda (1991). Unified mechanical aberration theory of electrostatic focusing–deflection systems. *Electron. Commun. (Japan)* **74**, 90–98.
Y. Uno, H. Morita, N. Simazu and T. Hosokawa (1995). Fifth-order aberration analysis of a combined electrostatic-magnetic focusing–deflection system. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 10–17.
C. T. Wang (1966). Dynamic focusing in magnetic deflection. *J. Appl. Phys.* **37**, 5007–5008.
C. C. T. Wang (1967a). Coma-dominated aberration in magnetically deflected electron beams. *J. Appl. Phys.* **38**, 3991–3994.
C. C. T. Wang (1967b). Spot distortion caused by magnetic deflection of an electron beam. *J. Appl. Phys.* **38**, 4938–4944.
C. C. T. Wang (1967c). Computer calculation of deflection aberrations in electron beams. *IEEE Trans. Electron Devices* **14**, 357–365.
C. C. T. Wang (1971). Analysis of electrostatic small-angle deflection. *IEEE Trans. Electron Devices* **18**, 258–274.
G. Wendt (1939). Über die Abbildungsfehler magnetischer Ablenkfelder. *Telefunkenröhre* **15**, 100–136.
G. Wendt (1942a). Bildfehler bei Ablenkung eines Kathodenstrahlbündels in zwei gekreuzten Ablenkfeldern. *Z. Phys.* **118**, 593–617.
G. Wendt (1942b). Fehler bei Ablenkung eines Kathodenstrahlbündels durch einfachsymmetrische Ablenkorgane. *Z. Phys.* **119**, 423–462.
G. Wendt (1947). Zur Theorie der Ablenkfehler von Elektronenstrahlbündeln, insbesondere in gekreuzten Ablenkfeldern. *Ann. Phys. (Leipzig)* **1** (436), 83–94.
G. Wendt (1953). Systèmes de déviation électronique et leurs aberrations. Systèmes dynamiques à deux plans de symétrie. *Onde Elec.* **33**, 93–106.

H. Werner (1963). Über einen bei magnetischen Ablenksystemen möglichen Ablenkfehler. *Exp. Tech. Phys.* **11**, 32–50.
H. Wollnik and M. Berz (1985). Relations between elements of transfer matrices due to the condition of symplecticity. *Nucl. Instrum. Meth. Phys. Res. A* **238**, 127–140.
J.-y. Ximen (1977). The electron optical properties and aberration theory of a combined magnetic round lens and deflection system. *Acta Phys. Sin.* **26**, 34–53.
J.-y. Ximen (1978). The electron optical characteristics and aberration theory of a combined system consisting of electrostatic round lens and magnetic deflector. *Acta Phys. Sin.* **27**, 247–259.
J.-y. Ximen (1981). On the linear transformations for Gaussian trajectory parameters in the combined electron optical system. *Optik* **59**, 237–249; also *Chinese J. Sci. Instrum.* **2**, 1–11.
J.-y. Ximen (1986). Aberration theory in electron and ion optics. *Adv. Electron. Electron Phys.* (Suppl. 17).
J.-y. Ximen and Y. Li (1982). Dynamic aberration correction in a combined scanning electron beam system. *Optik* **62**, 287–297.

Part VI, Chapter 34

J. P. Adriaanse, H. W. G. van der Steen and J. E. Barth (1989). Practical optimization of electrostatic lenses. *J. Vac. Sci. Technol. B* **7**, 651–656.
H. N. Al-Obaidi (2005). A semi-analytical investigation of the paraxial ray equation. *J. Coll. Educ. Al-Mustansiriyakh Univ* **3**, 117–132.
H. N. Al-Obaidi, A. S. Mahdi and W. J. Kadhem (2009). Failure limit investigation for the spline aided solution technique. *Mod. Phys. Lett. B* **23**, 1467–1477.
J. E. Barth, H. W. G. van der Steen and J. Chmelík (1995). Improvements to the electrostatic lens optimization method SOEM. *Proc. SPIE* **2522**, 128–137.
J. Bärtle (2004). A tool to calculate the paraxial properties of electron optical systems. *Recent Trends* **9**, 13–14.
D. Barton and J. P. Fitch (1972). Applications of algebraic manipulation programs in physics. *Rep. Prog. Phys.* **35**, 235–314.
A. Benez, E. Plies and W. Ziegler (1995). Optimierung einer elektrostatischen Transferlinse mit genetischen Algorithmen. *Optik* **100** (Suppl.), 3.
M. Berz (1989). Differential algebraic description of beam dynamics to very high orders. *Part. Accel.* **24**, 109–124.
M. Berz (1995). Modern map methods for charged particle optics. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 100–104.
M. Berz (1999). Modern map methods in particle beam physics. *Adv. Imaging Electron Phys.* **108**.
M. Berz and K. Makino (2004). New approaches for the validation of transfer maps using remainder-enhanced differential algebra. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 53–62.
M. Berz and H. Wollnik (1987). The program HAMILTON for the analytic solution of the equations of motion through fifth order. *Nucl. Instrum. Meth. Phys. Res. A* **258**, 364–373.
M. Berz, K. Makino and W. Wan (2015). *An Introduction to Beam Physics* (CRC Press, Boca Raton).
L. Boesten (1988). Modified operation of the seven-element electrostatic Chutjian–Kuyatt lens system. *Rev. Sci. Instrum.* **59**, 233–237.
B. Buchberger, G. E. Collins, R. Loos and R. Albrecht, Eds (1982). *Computer Algebra. Symbolic and Algebraic Computation* (Springer, Springer, Vienna & New York). *Computing*, Suppl. 4.
J. Burfoot (1952). Numerical ray-tracing in electron lenses. *Br. J. Appl. Phys.* **3**, 22–24.
H. Busch (1927). Über die Wirkungsweise der Konzentrierungsspule bei der Braunsche Röhre. *Arch. Elektrotech.* **18**, 583–594.
J. Calmet, Ed. (1982). *Computer Algebra* (Springer, Berlin & New York).
M. Cheng, T.-t. Tang and Z.-h. Yao (2001a). Study on differential algebraic aberration method for electostatic electron lenses. *Optik* **112**, 250–254.
M. Cheng, T.-t. Tang and Z.-h. Yao (2001b). Study on differential algebraic chromatic aberration method for Glaser's bell-shaped magnetic lenses. *Optik* **112**, 483–486.

M. Cheng, T.-t. Tang and Z.-h. Yao (2002a). High-order combined geometric−chromatic aberrations of magnetic electron lenses using differential algebra. *J. Phys. D: Appl. Phys.* **35**, 409−411.
M. Cheng, T.-t. Tang and Z.-h. Yao (2002b). Study on time-of-flight property of electron optical systems by differential algebraic method. *J. Phys. D: Appl. Phys.* **35**, 1566−1569.
M. Cheng, T.-t. Tang, Y.-l. Lu and Z.-h. Yao (2003). Differential algebraic method for arbitrary order curvilinear-axis geometric−chromatic aberration analysis. *J. Phys. D: Appl. Phys.* **36**, 923−928.
M. Cheng, S.-f. Zhou, Y.-l. Lu and T.-t. Tang (2006). Differential algebraic method for arbitrary-order chromatic aberration analysis of electrostatic lenses. *Optik* **117**, 196−198.
H. C. Chu and E. Munro (1982a). Numerical analysis of electron beam lithography systems. Part III. Calculation of the optical properties of electron focusing systems and dual-channel deflection systems with combined magnetic and electrostatic fields. *Optik* **61**, 121−145.
H. C. Chu and E. Munro (1982b). Numerical analysis of electron beam lithography systems. Part IV. Computerized optimization of the electron optical performance of electron beam lithography systems using the damped least squares method. *Optik* **61**, 213−246.
A. Chutjian (1979). Geometries and focal properties of two electron-lens systems useful in low-energy electron or ion scattering. *Rev. Sci. Instrum.* **50**, 347−355.
A. V. Crewe, D. N. Eggenberger, J. Wall and L. M. Welter (1968). Electron gun using a field emission source. *Rev. Sci. Instrum.* **39**, 576−583.
J. W. Dauben (1995). *Abraham Robinson: The Creation of Nonstandard Analysis, a Personal and Mathematical Odyssey* (Princeton University Press, Princeton).
J. W. Dauben (2003). Abraham Robinson 1918−1974, a biographical memoir. *Biographical Memoirs National Academy of Science* **82**.
J. H. Davenport, Y. Siret and E. Tournier (1988). *Computer Algebra. Systems and Algorithms for Algebraic Computation* (Academic Press, London); translated into French (Masson) and Russian (MIR).
R. Degenhardt (1996). Optimization of electrostatic lens systems: fast approximative calculation of symmetric and asymmetric einzel lenses. *Recent Trends* **5**, 17−19.
R. Degenhardt (1997). Optimization of electrostatic lens systems: fast approximate calculation of the imaging properties of einzel lenses. *Proc. SPIE* **3155**, 70−77.
S. K. Demin, S. I. Safronov and R. P. Tarasov (1998). Numerical analysis and synthesis of electron-optical systems of complex structure. *Zh. Tekh. Fiz.* **68**, 97−103. *Tech. Phys.* **43**, 222−228.
A. L. Dodin and M. B. Nesvizhskii (1981). Automation of the process of deriving the formulas for the aberration coefficients of electron-optical systems in the relativistic case. *Radiotekh. Elektron.* **26**, 2651−2657. *Radio Eng. Electron. Phys.* **26** (12), 151−157.
A. D. Dymnikov and G. Martinez (1997). The matrizant method for an optimal synthesis of nonlinear magnetic and electrostatic focusing systems for a given beam currect. *Proc. SPIE* **3155**, 125−133.
C. J. Edgcombe, A. R. Lupini and J. H. Taylor (1999). Robust optimization for magnetic lens design. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 306−309.
E. Fehlberg (1960). Neue genauere Runge-Kutta-Formeln für Differentialgleichungen *n*-ter Ordnung. *Z. Angew. Math. Mech.* **40**, 449−455.
E. Fehlberg (1961). Numerische stabile Interpolationsformeln mit günstiger Fehlerfortpflanzung für Differentialgleichungen erster und zweiter Ordnung. *Z. Angew. Math. Mech.* **41**, 101−110.
E. Fehlberg (1966). New high-order Runge−Kutta formulas with an arbitrarily small tuncation error. *Z. Angew. Math. Mech.* **46**, 1−16.
E. Fehlberg (1969). Klassische Runge−Kutta Formeln fünfter und siebenter Ordnung mit Schrittweiten-Kontrolle. *Computing* **4**, 93−106.
M. Fink and E. Kisker (1980). A method for rapid calculations of electron trajectories in multi-element electrostatic cylinder lenses. *Rev. Sci. Instrum.* **51**, 918−920.
J. P. Fitch (1985). Solving algebraic problems with REDUCE. *J. Symbol. Comput.* **1**, 211−227.
J. Fitch (2009). CAMAL 40 years on − is small still beautiful? In *Intelligent Computer Mathematics* (J. Carette, L. Dixon, C. Sacerdoti Coen and S. Watt, Eds), 32−44 (Springer, Berlin).

L. Fox and E. T. Goodwin (1949). Some new methods for the numerical integration of ordinary differential equations. *Proc. Cambridge Philos. Soc.* **45**, 373–388.
R. W. Freund and R. W. Hoppe (2007). *Numerische Mathematik* Vol. 1. (Springer, Berlin). Revised edition of Stoer–Bulirsch.
R. W. Freund and R. W. Hoppe (2011). *Numerische Mathematik* Vol. 2. (Springer, Berlin). Revised edition of Stoer–Bulirsch.
W. Glaser (1933). Über optische Abbildung durch mechanische Systeme und die Optik allgemeiner Medien. *Ann. Phys. (Leipzig)* **18** (410), 557–585.
W. Glaser (1935). Zur Bildfehlertheorie des Elektronenmikroskops. *Z. Phys.* **97**, 177–201.
W. Glaser (1952). *Grundlagen der Elektronenoptik* (Springer, Vienna).
U. Glatzel and F. Lenz (1988). Optimization of electrostatic immersion lenses. *Optik* **79**, 15–18.
J. Grabmeier, E. Kaltofen and V. Weispfenning, Eds (2003). *Computer Algebra Handbook. Foundations, Applications, Systems* (Springer, Berlin).
C.-x. Gu and L.-y. Shan (1984). Constrained optimization design of an electron optical system. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 91–96 (Scanning Electron Microscopy, AMF O'Hare).
C.-x. Gu, N.-q. Chen and X.-m. Wu (1984). The optimization design of an electron-optical system. *Acta Electron. Sin.* **12**, 41–47.
C.-x. Gu, L.-y. Shan and Z.-r. Chen (1991). Optimization design for the electron emission system using improved Powell method. *Scanning Microsc.* **5**, 937–944.
C.-x. Gu, G.-y. Liao, H.-x. Jiang, J. Li and L.-y. Shan (1999). Nonlinear programming and scientific computing visualization in the optimization design of electron optical system. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 321–328.
C.-x. Gu, M.-q. Wu, G. Lin and L.-y. Shan (2001). The application of genetic algorithms to the optimization design of electron optical system. *Proc. SPIE* **4510**, 127–137.
E. Hairer and G. Wanner (1996). *Solving Ordinary Differential Equations.* II. Stiff and Differential-Algebraic Problems, 2nd edn (Springer, Berlin).
E. Hairer, S. P. Nørsett and G. Wanner (1993). *Solving Ordinary Differential Equations.* I, 2nd edn (Springer, Berlin).
G. Hämmerlin and K.-H. Hoffmann (1991). *Numerical Mathematics* (Springer, Berlin).
P. W. Hawkes (1977). Computer calculation of formulae for electron lens aberration coefficients. *Optik* **48**, 29–51.
P. W. Hawkes (1983). Computer-aided calculation of the aberration coefficients of microwave cavity lenses. I. Primary (second-order) aberrations. II. Secondary (third-order) aberrations. *Optik* **63**, 129–156. & **65**, 227–251.
A. C. Hearn (1985). *REDUCE User's Manual* (Rand, Santa Monica).
A. Hearn (2005). REDUCE: the first 40 years. In *Algorithms, Algebra and Logic: Proceedings of the A3L 2005* (A. Dolzmann, A. Seidl and T. Sturm, Eds), 19–24 (Nordersted).
J. Hejna (1998). Optimization of an immersion lens design in the BSE detector for the low voltage SEM. *Recent Trends* **6**, 30–31.
R. Hill and K. C. A. Smith (1980). An interactive computer system for magnetic lens analysis. EUREM-7, The Hague. **1**, 60–61.
R. Hill and K. C. A. Smith (1981). Application of the Cambridge interactive electron lens analysis system CIELAS. EMAG, Cambridge. 71–74.
A. S. Householder (1964). *The Theory of Matrices in Numerical Analysis* (Blaisdell, New York).
D. Jacobs, Ed. (1977). *The State of the Art in Numerical Analysis* (Academic Press, London & New York).
J. C. E. Jennings and R. G. Pratt (1955). Numerical ray tracing in electron lenses. *Proc. Phys. Soc. (Lond.) B* **68**, 526–536.
W. J. Kadhem (2014). Reconstructed polepieces to the objective magnetic lens depending on some geometrical and physical parameters. *Adv. Phys. Theories Appl* **31**, 43–52.

Y.-f. Kang, T.-t. Tang, Y. Yen and X. Guo (2007). Differential algebraic method for computing the high order aberrations of practical electron lenses. *Optik* **118**, 158–162.
Y.-f. Kang, T.-t. Tang, J.-y. Zhao, S. Li and D. Zhang (2009). Different algebraic method for computing the high-order aberrations of practical combined focusing–deflection systems. *Optik* **120**, 591–600.
Y.-f. Kang, T.-t. Tang, J.-y. Zhao, S. Li and D. Zhang (2010). A different algebraic method for the fifth-order combined geometric–chromatic aberrations of practical magnetic electron lenses. *Optik* **121**, 178–183.
E. Kasper (1982). Magnetic field calculation and the determination of electron trajectories. In *Magnetic Electron Lenses* (P. W. Hawkes, Ed.), 57–118 (Springer, Berlin & New York).
E. Kasper (1984). Recent developments in numerical electron optics. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 63–73 (Scanning Electron Microscopy, AMF O'Hare).
E. Kasper (1985). On the numerical determination of electron optical focusing properties and aberrations. *Optik* **69**, 117–125.
E. Kasper (1987a). Computer simulation of electron optical systems. *Nucl. Instrum. Meth. Phys. Res. A* **258**, 466–479.
E. Kasper (1987b). An advanced method of field calculation in electron optical systems with axisymmetric boundaries. *Optik* **77**, 3–12.
B. Lencová (2004a). On magnetic lens computations with FEM and BEM. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 133–140.
B. Lencová (2004b). Accuracy estimate for magnetic electron lenses computed by first-order FEM. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 149–153.
B. Lencová (2007). CAD in electron optics. MCM-8, Prague. 19–20.
B. Lencová (2008). Solving complex electron optical problems with EOD. *Recent Trends* **11**, 65–68.
B. Lencová (2010). Recent developments and improvements of EOD program. *Recent Trends* **12**, 25–28.
B. Lencová and M. Oral (2006). Programs for electron optical design. *Recent Trends* **10**, 41–44.
B. Lencová and G. Wisselink (1990). Program package for the computation of lenses and deflectors. *Nucl. Instrum. Meth. Phys. Res. A* **298**, 56–66.
B. Lencová and J. Zlámal (2005). EOD (Electron Optical Design) program for computations for electron microscopy. MC-2005, Davos, 8.
B. Lencová and J. Zlámal (2006). EOD (Electron Optical design) program features. *Recent Trends* **10**, 45–46.
B. Lencová and J. Zlámal (2007). The development of EOD program for the design of electron optical devices. *Microsc. Microanal.* **13** (Suppl. 3), 2–3.
B. Lencová and J. Zlámal (2008). A new program for the design of electron microscopes. *Phys. Procedia* **1**, 315–324.
B. Lencová, G. Wisselink, M. van der Stam and P. Kruit (1989). Computer-aided lens design for electron microscopes. *Ultramicroscopy* **31**, 474–475.
B. Lencová, P. Jánský and J. Zlámal (2004). Selected examples of our calculations. *Recent Trends* **9**, 47–48.
L. H. A. Leunissen, M. A. J. van der Stam and P. Kruit (2001). New features in program package POCAD. *Proc. SPIE* **4510**, 119–126.
K. Levenberg (1944). A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.* **2**, 164–168.
Z.-x. Liu (2003). Differential algebraic analysis of optical properties of monopole, dipole, and quadrupole lenses. *Optik* **114**, 518–520.
Z.-x. Liu (2006). Differential algebraic method for aberration analysis of typical electrostatic lenses. *Ultramicroscopy* **106**, 220–232.
Z.-x. Liu (2007). Exploring third-order chromatic aberrations of electron lenses with computer algebra. *Adv. Imaging Electron Phys.* **145**, 95–148.
Z.-X. Liu (2017). Third-rank chromatic aberrations of electron lenses. Ultramicroscopy, in press.
K. Makino and M. Berz (1997). Arbitrary order aberrations for elements characterized by measured fields. *Proc. SPIE* **3155**, 221–227.
K. Makino and M. Berz (1999). COSY Infinity version 8. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 338–343.

K. Makino and M. Berz (2011). Optimal correction and design parameter search by modern methods of rigorous global optimization. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 332–337.
K. Makino, M. Berz, C. J. Johnstone and D. Errede (2004). High-order map treatment of superimposed cavities, absorbers, and magnetic multipole and solenoid fields. *Nucl. Instrum. Meth. Phys. Res. A* **519**, 162–174.
W. F. Manning and J. Millman (1938). Note on numerical integration. *Phys. Rev.* **53**, 673.
G. Martínez and M. Sancho (1995). Accurate calculation of geometric aberrations in electrostatic lenses. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 198–204.
G. Martínez, A. D. Dymnikov and A. H. Azbaid (1998). Optimal synthesis of charged beam focusing systems. *Recent Trends* **6**, 45–46.
G. Martínez, A. H. Azbaid and A. D. Dymnikov (1999). Numerical synthesis of an optimal microprobe focusing system. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 344–349.
R. W. Moses (1973). Lens optimization by direct application of the calculus of variations. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 250–272 (Academic Press, London & New York).
H. Munack (1992). Globale Optimierung elektronenoptischer Systeme mittels Intervallarithmetik. Dissertation, Tübingen.
E. Munro (1973). Computer-aided design of electron lenses by the finite element method. In *Image Processing and Computer-Aided Design in Electron Optics* (P. W. Hawkes, Ed.), 284–323 (Academic Press, London & New York).
E. Munro (2002). Simulation of electron optical systems by direct ray-tracing through analytically fitted fields. *Recent Trends* **8**, 7–10.
E. Munro (2011). Numerical simulation methods for electron and ion optics. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 266–272.
E. Munro and H. C. Chu (1982a). Numerical analysis of electron beam lithography systems. Part I. Computation of fields in magnetic deflectors. *Optik* **60**, 371–390.
E. Munro and H. C. Chu (1982b). Numerical analysis of electron beam lithography systems. Part II. Computation of fields in electrostatic deflectors. *Optik* **61**, 1–16.
E. Munro, J. Rouse, H. Liu, L. Wang and X. Zhu (2006a). Simulation software for designing electron and ion beam equipment. *Microelectron. Eng.* **83**, 994–1002.
E. Munro, L. Wang, J. Rouse and H. Liu (2006b). Differential algebraic method for electron optical design. *Recent Trends* **10**, 53–54.
E. Munro, L. Wang, J. Rouse and H. Liu (2006c). Theory and application of the differential algebraic method for computing the aberrations of electron optical systems. IMC-16, Sapporo. **2**, 588.
E. Munro, J. Rouse, H. Liu and L. Wang (2008). Design of aberration correctors using the differential algebraic method. *Recent Trends* **11**, 81–84.
S. G. Narylkov and Ya. G. Lyubchik (1982). Automation of analytical calculation of the aberrations of electron-optical lenses. *Radiotekh. Elektron.* **27**, 1602–1605 [not translated].
M. B. Nesvizhskii (1986). Aberration coefficients of complicated electron-optical cathode systems. *Radiotekh. Elektron.* **31**, 162–168. *Soviet J. Commun. Technol*. **31** (5), 153–160.
E. W. Ng, Ed. (1979). *Symbolic and Algebraic Calculation* (Springer, Berlin & New York).
H. P. Niemitz (1980). Theoretische Untersuchung von elektrostatischen Drei-Elektroden-Filterlinsen. Dissertation, Tübingen.
M. Nishiguchi and M. Toyoda (2008). Computer program TRIO 2.0 for calculation and visualization of ion trajectories. *Phys. Procedia* **1**, 325–332.
A. C. Norman (1982). The development of a vector-based algebra system. In *Computer Algebra* (J. Calmet, Ed.), 237–248 (Springer, Berlin & New York).
A. C. Norman and J. H. Davenport (1979). Symbolic integration – the dust settles? In *Symbolic and Algebraic Calculation* (E. W. Ng, Ed.), 398–407 (Springer, Berlin & New York).
B. Numerov [Noumeroff] (1923). Méthode nouvelle de la détermination des orbites et le calcul des éphémérides en tenant compte des perturbations. *Publ. Obs. Astrophys. Central Russie* **2**, 188–288.

M. Oral (2010). Ray tracing, aberration coefficients and intensity distribution. *Recent Trends* **12**, 49–52.
M. Oral and B. Lencová (2003). Calculation of aberration coefficients by ray tracing. *Microsc. Microanal.* **9** (Suppl. 3), 26–27.
M. Oral and B. Lencová (2005). Optimization of double deflectrion systems with ray tracing. MC-2005, Davos. 9.
M. Oral and B. Lencová (2009). Calculation of aberration coefficients by ray tracing. *Ultramicroscopy* **109**, 1365–1373.
D. Preikszas (1995). Korrektur des Farb- und Öffnungsfehlers eines Niederspannungs-Elektronenmikroskops mit Hilfe eines Elektronenspiegels. Dissertation, Darmstadt.
A. Quarteroni, R. Sacco and F. Saleri (2007). *Numerical Mathematics* (Springer, Berlin).
T. Radlička (2008). Lie algebraic methods in charged particle optics. *Adv. Imaging Electron Phys.* **151**, 241–362.
T. Radlička (2012). Calculation of diffraction aberration using differential algebra. *Recent Trends* **13**, 59–62.
T. Radlička and M. Oral (2016). Correction of misalignment aberrations of a hexapole corrector using the differential algebra method. *Recent Trends* **15**, 50–53.
G. Rayna (1987). *REDUCE* (Springer, New York & Berlin).
A. Robinson (1961). Nonstandard analysis. *Proc. K. Ned. Akad. Wetensch. A* **46**, 432–440. Also *Indagationes Mathematicae* **23**, 432–440.
A. Robinson (1966). *Nonstandard Analysis* (North-Holland, Amsterdam).
K. Sakaguchi, T. Sekine and H. Shimada (1999). Influence of first lens on magnification optimization in two lenses FIB optical systems. *Proc. SPIE* **3777**, 166–174.
W. Scherle (1983). Eine Integralgleichungsmethode zur Berechnung magnetischer Felder von Anordnungen mit Medien unterschiedlicher Permeabilität. *Optik* **63**, 217–226.
W. Scherle (1984). Beschreibung von Aberrationen mit Begriffen der Statistik. *Optik* **67**, 307–314.
O. Scherzer (1936). Berechnung der Bildfehler dritter Ordnung nach der Bahnmethode. In *Beiträge zur Elektronenoptik* (H. Busch and E. Brüche, Eds), 33–41 (Barth, Leipzig).
T. Soma (1977). Relativistic aberration formulas for combined electric–magnetic focusing–deflection system. *Optik* **49**, 255–262.
J. Stoer and R. Bulirsch (2005). *Numerische Mathematik* fifth edn. (Springer, Berlin; 2 vols).
C. Störmer (1907). Sur le mouvement d'un point matériel portant une charge d'électricité sous l'action d'un aimant élémentaire. *Arch. Math. Naturvid.* **28** (No. 2).
M. Szilágyi (1977). A new approach to electron optical optimization. *Optik* **48**, 215–224.
M. Szilágyi (1984). Synthesis of electron lenses. In *Electron Optical Systems for Microscopy, Microanalysis and Microlithography* (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds), 75–84 (Scanning Electron Microscopy, AMF O'Hare).
M. Szilágyi (1987a). Electrostatic spline lenses. *J. Vac. Sci. Technol. A* **5**, 273–278.
M. Szilágyi (1987b). Synthesis of electron optical systems with very low aberrations. In *Proc. Int. Symposium on Electron Optics, ISEOB* (J.-y. Ximen, Ed.), 238–240 (Institute of Electronics, Academia Sinica, Beijing).
J. H. Taylor and K. C. A. Smith (1986). The application of the complex optimization technique to the design of magnetic electron lenses. ICEM-11, Kyoto. **1**, 297–298.
K. Tsuno and K. C. A. Smith (1985). Computation of lens shape with minimum spherical aberration coefficient. EMAG, Newcastle-upon-Tyne. 121–124.
K. Tsuno and K. C. A. Smith (1986). Optimization of objective lens shape for high resolution electron microscopes. ICEM-11, Kyoto. **1**, 295–296.
M. van der Stam (1996). Computer assistance for the pre-design of charged particle optics. Dissertation, Delft.
M. A. J. van der Stam and P. Kruit (1995). Computer assistance for the pre-design of optical instruments. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 205–210.
M. A. J. van der Stam and P. Kruit (1999). A general framework for numerical calculation of properties of charged-particle optical systems. *Nucl. Instrum. Meth. Phys. Res. A* **427**, 368–374.

M. A. J. van der Stam, J. E. Barth and P. Kruit (1993). Design of a multi mode transport lens with optimization program SOEM. *Proc. SPIE* **2014**, 45–56.
M. A. J. van der Stam, J. E. Barth and P. Kruit (1996). Interactive design of multi component systems. *Proc. SPIE* **2858**, 90–102.
M. van der Stam, B. Mertens and P. Kruit (1997). Design of a magnification and rotation unit with pre-design tool POCAD. *Proc. SPIE* **3155**, 22–36.
H. W. G. van der Steen and J. E. Barth (1989). Practical optimization of multielectrode lenses. *J. Vac. Sci. Technol. B* **7**, 1886–1890.
H. W. G. van der Steen, J. E. Barth and J. P. Adriaanse (1990). Engineering constraints and computer-aided optimization of electrostatic lens systems. *Nucl. Instrum. Meth. Phys. Res. A* **298**, 377–382.
J. A. van Hulzen and J. Calmet (1982). Computer algebra systems. *Computing* Suppl. 4, 221–243.
J. von zur Gathen and J. Gerhard (2013). *Modern Computer Algebra*, 3rd edn (Cambridge University Press, Cambridge).
L.-p. Wang, T.-t. Tang, B.-j. Cheng and J. Cai (1999). Automatic differentiation method for the aberration analysis of electron optical systems. *Optik* **110**, 408–414. also published in *J. Xi'an Jiaotong Univ*. **33** (1999) 101–105.
L.-p. Wang, T.-t. Tang and B.-j. Cheng (2000). Differential algebraic method for arbitrary-order aberration analysis of electron deflection systems. *Optik* **111**, 290–296.
L.-p. Wang, T.-t. Tang, B.-j. Cheng and J. R. A. Cleaver (2002a). Application of differential algebraic method to the aberration analysis of curvilinear-axis electron optical systems. *Optik* **113**, 171–176.
L.-p. Wang, T.-t. Tang, B.-j. Cheng and J. R. A. Cleaver (2002b). Differential algebraic method for arbitrary-order aberration analysis of combined electron beam focusing–deflection systems. *Optik* **113**, 181–187.
L.-p. Wang, J. Rouse, H. Liu, E. Munro and X. Zhu (2004a). Simulation of electron optical systems by differential algebraic method combined with Hermite fitting for practical lens fields. *Microelectron. Eng.* **73–74**, 90–96.
L.-p. Wang, H. Liu, J. Rouse, E. Munro and X. Zhu (2004b). Simulation of multipole systems by differential algebraic method. *Recent Trends* **9**, 79–80.
L.-p. Wang, E. Munro, J. Rouse and H. Liu (2008a). Simulation of electron mirrors by the differential algebraic method. *Phys. Procedia* **1**, 297–304.
L.-p. Wang, J. Rouse, E. Munro, H.-n. Liu and X.-q. Zhu (2008b). Aberration analysis of electron mirrors by a differential algebraic method. *Optik* **119**, 90–96.
L. Wei and L.-s. Tong (1995). A primary expert system applied in design of electron optical system. *Proc. SPIE* **2522**, 138–148.
L. Wei, L.-s. Tong and H. Yang (1995). Fuzzy optimization of an electron optical lens. *Nucl. Instrum. Meth. Phys. Res. A* **363**, 367–370.
E. T. Whittaker and G. N. Watson (1927). *A Course of Modern Analysis*, 4th edn (Cambridge University Press, Cambridge).
H. Zhang and M. Berz (2011). The fast multipole method in the differential algebra framework. *Nucl. Instrum. Meth. Phys. Res. A* **645**, 338–344.
H. Zhang, J. Portman, Z. Tao, P. Duxbury, C.-y. Ruan, K. Makino and M. Berz (2015). The differential algebra based multiple level fast multipole algorithm for 3D space charge field calculation and photoemission simulation. *Microsc. Microanal.* **21** (Suppl. 4), 224–229.

Conference Proceedings

1. International Congresses on Electron Microscopy, later International Microscopy Congresses.
2. European Regional Congresses on Electron Microscopy, later European Microscopy Congresses.
3. Asia-Pacific Congresses on Electron Microscopy, later Asia–Pacific Microscopy Congresses.
4. Charged Particle Optics Conferences.
5. High-voltage Electron Microscopy Conferences.

6. EMAG [Electron Microscopy and Analysis Group of the Institute of Physics] meetings.
7. Multinational Congresses on (Electron) Microscopy (MCEM, MCM).
8. The Dreiländertagungen (Germany, Austria, Switzerland) and related meetings.
9. Recent Trends in Charged Particle Optics and Surface Physics Instrumentation (Skalský Dvůr).
10. SPIE Proceedings.
11. Soviet All-Union Conferences on Electron Microscopy.
12. Problems of Theoretical and Applied Electron Optics [Problemyi Teoreticheskoi i Prikladnoi Elektronnoi Optiki].
13. Related Meetings.

The following list gives full publishing details of the series of International and Regional conferences on Electron Microscopy. The South American (CIASEM) conferences are not listed as they contain little optics. For the reader's convenience, a few other meetings are included, in particular those on charged particle optics, the Multinational Conferences on Electron Optics (now Multinational Conferences on Microscopy), the Dreiländertagungen (now Microscopy Conferences) and the conferences organized by the Electron Microscopy and Analysis Group (EMAG) of the British Institute of Physics. In the lists of references, these are referred to by their acronyms and venue. The irregular, short-lived series of meetings on high-voltage electron microscopy is identified by the acronym HVEM.

The list does not include the proceedings of the annual meetings of the Electron Microscopy Society of America, which are identified in the reference lists by EMSA or MSA, venue and the meeting number until publication as a Supplement to *Microscopy and Microanalysis* was adopted. Proceedings were first issued for the 25th meeting (1967) and have been published ever since, at first in print and more recently on-line. For full details, see the lists published by Hawkes in *Advances in Imaging and Electron Physics*, Vol. 117 (2003) 203−379 and Vol. 190 (2015) 143−175.

Many other national electron microscopy societies publish proceedings of their major meetings but few contain much optics. A notable exception is the series of All-Union meetings held in Russia, the proceedings of which are mainly published in *Izv. Akad. Nauk (Ser. Fiz.)*, translated as *Bull. Acad. Sci. USSR (Phys. Ser.)*, though a few papers appear in *Radiotekhnika i Elektronika (Radio Engineering and Electronic Physics* and later *Soviet Journal of Communications Technology and Electronics)*. Brief details of these are given at the end of the main list. The other noteworthy exception is Japan; abstracts of Japanese national meetings were published regularly and rapidly in the *Journal of Electron Microscopy* and now appear in a supplement to *Kenbikyo*.

Details of other related meetings are to be found in the articles by Hawkes mentioned above, notably the International Congresses on X-ray Optics and Microscopy (ICXOM), the Low-energy Electron Microscopy and Photoemission Electron Microscopy (LEEM, PEEM) meetings and Frontiers of Aberration-corrected Electron Microscopy (PICO).

1. International Congresses on Electron Microscopy, later International Microscopy Congresses

ICEM-1, Delft, 1949: *Proceedings of the Conference on Electron Microscopy*, Delft, 4−8 July, 1949 (A. L. Houwink, J. B. Le Poole and W. A. Le Rütte, Eds) Hoogland, Delft, 1950.

ICEM-2, Paris, 1950: *Comptes Rendus du Premier Congrès International de Microscopie Electronique*, Paris, 14−22 September, 1950. Editions de la Revue d'Optique Théorique et Instrumentale, Paris, 1953. 2 Vols.

ICEM-3, London, 1954: *The Proceedings of the Third International Conference on Electron Microscopy,* London, 15−21 July 1954 (R. Ross, Ed.) Royal Microscopical Society, London, 1956.

ICEM-4, Berlin, 1958: *Vierter Internationaler Kongress für Elektronenmikroskopie*, Berlin, 10−17 September, 1958, *Verhandlungen* (W. Bargmann, G. Möllenstedt, H. Niehrs, D. Peters, E. Ruska and C. Wolpers, Eds) Springer, Berlin, 1960. 2 Vols; on-line via SpringerLink.

ICEM-5, Philadelphia, 1962: *Electron Microscopy. Fifth International Congress for Electron Microscopy*, Philadelphia, Pennsylvania, 29 August to 5 September, 1962 (S. S. Breese, Ed.) Academic Press, New York, 1962. 2 Vols.

ICEM-6, Kyoto, 1966: *Electron Microscopy 1966. Sixth International Congress for Electron Microscopy*, Kyoto, 28 August − 4 September 1966 (R. Uyeda, Ed.) Maruzen, Tokyo, 1966. 2 Vols.

ICEM-7, Grenoble, 1970: *Microscopie Electronique 1970. Résumés des Communications Présentées au Septième Congrès International*, Grenoble, 30 August − 5 September 1970 (P. Favard, Ed.) Société Française de Microscopie Electronique, Paris, 1970. 3 Vols.

ICEM-8, Canberra, 1974: *Electron Microscopy 1974. Abstacts of Papers Presented to the Eighth International Congress on Electron Microscopy*, Canberra, 25−31 August 1974 (J. V. Sanders and D. J. Goodchild, Eds) Australian Academy of Science, Canberra, 1974. 2 Vols.

ICEM-9, Toronto, 1978: *Electron Microscopy 1978. Papers Presented at the Ninth International Congress on Electron Microscopy*, Toronto, 1−9 August 1978 (J. M. Sturgess, Ed.) Microscopical Society of Canada, Toronto, 1978. 3 Vols.

ICEM-10, Hamburg, 1982: *Electron Microscopy, 1982. Papers Presented at the Tenth International Congress on Electron Microscopy*, Hamburg, 17−24 August 1982. Deutsche Gesellschaft für Elektronenmikroskopie, Frankfurt, 1982. 3 Vols.

ICEM-11, Kyoto, 1986: *Electron Microscopy 1986. Proceedings of the XIth International Congress on Electron Microscopy*, Kyoto, 31 August−7 September 1986 (T. Imura, S. Maruse and T. Suzuki, Eds.). Japanese Society of Electron Microscopy, Tokyo. 4 Vols; published as a supplement to *J. Electron Microsc*. **35** (1986).

ICEM-12, Seattle, 1990: *Electron Microscopy 1990. Proceedings of the XIIth International Congress for Electron Microscopy*, Seattle WA, 12−18 August 1990 (L. D. Peachey, and D. B. Williams, Eds). San Francisco Press, San Francisco. 4 Vols. See also *Ultramicroscopy* **36** (1991) Nos 1−3, 1−274.

ICEM-13, Paris, 1994: *Electron Microscopy 1994. Proceedings of the 13th International Congress on Electron Microscopy*, Paris, 17−22 July 1994 [B. Jouffrey, C. Colliex, J. P. Chevalier, F. Glas, P. W. Hawkes, D. Hernandez−Verdun, J. Schrevel and D. Thomas (Vol. 1), B. Jouffrey, C. Colliex, J. P. Chevalier, F. Glas and P. W. Hawkes (Vols 2A and 2B) and B. Jouffrey, C. Colliex, D. Hernandez−Verdun, J. Schrevel and D. Thomas (Vols 3A and 3B), Eds]. Editions de Physique, Les Ulis, 1994.

ICEM-14, Cancún, 1998: *Electron Microscopy 1998. Proceedings of the 14th International Congress on Electron Microscopy*, Cancún, 31 August−4 September 1998 [Memorias del 14to Congreso Internacional de Microscopía Electrónica celebrado en Cancún (México) del 31 de Agosto al 4 de Septiembre de 1998] (H. A. Calderón Benavides and M. J. Yacamán, Eds). Institute of Physics Publishing, Bristol and Philadelphia 1998. 4 Vols. See also *Micron* **31** (2000), No. 5.

ICEM-15, Durban, 2002: *Electron Microscopy 2002. Proceedings of the 15th International Congress on Electron Microscopy*, International Convention Centre, Durban, 1−6 September 2002 [R. Cross, P. Richards, M. Witcomb and J. Engelbrecht (Vol. 1, Physical, Materials and Earth Sciences), R. Cross, P. Richards, M. Witcomb and T. Sewell, (Vol. 2, Life Sciences) and R. Cross, P. Richards, M. Witcomb, J. Engelbrecht and T. Sewell (Vol. 3, Interdisciplinary), Eds]. Microscopy Society of Southern Africa, Onderstepoort 2002.

IMC-16, Sapporo, 2006: *Proceedings 16th International Microscopy Conference, "Microscopy for the 21st Century"*, Sapporo, 3−8 September 2006 (H. Ichinose and T. Sasaki, Eds). Vol. 1, Biological and Medical Science; Vol. 2, Instrumentation; Vol. 3, Materials Science. Publication Committee of IMC16, Sapporo 2006.

IMC-17, Rio de Janeiro, 2010: *Proceedings IMC17, The 17th IFSM International Microscopy Congress*, Rio de Janeiro, 19−24 September 2010 (G. Solórzano and W. de Souza, Eds). Sociedade Brasileira de Microscopia e Microanálise, Rio de Janeiro 2010.

IMC-18, Prague, 2014: Prague Convention Centre, 7−12 September 2014. Proceedings open-access at www.microscopy.cz/proceedings/all.html, edited by P. Hozak.

IMC-19, Sydney, 9−14 September 2018.

2. European Regional Congresses on Electron Microscopy, later European Microscopy Congresses

EUREM-1, Stockholm, 1956: *Electron Microscopy. Proceedings of the Stockholm Conference*, 17−20 September, 1956 (F. J. Sjöstrand and J. Rhodin, Eds) Almqvist and Wiksells, Stockholm, 1957.

EUREM-2, Delft, 1960: *The Proceedings of the European Regional Conference on Electron Microscopy*, Delft, 29 August − 3 September 1960 (A. L. Houwink and B. J. Spit, Eds) Nederlandse Vereniging voor Elektronenmicroscopie, Delft n.d. 2 Vols.

EUREM-3, Prague, 1964: *Electron Microscopy 1964. Proceedings of the Third European Regional Conference*, Prague, 26 August − 3 September 1964 (M. Titlbach, Ed.) Publishing House of the Czechoslovak Academy of Sciences, Prague, 1964. 2 Vols.

EUREM-4, Rome, 1968: *Electron Microscopy 1968. Pre-Congress Abstracts of Papers Presented at the Fourth Regional Conference*, Rome, 1−7 September 1968 (D. S. Bocciarelli, Ed.) Tipografia Poliglotta Vaticana, Rome, 1968. 2 Vols.

EUREM-5, Manchester, 1972: *Electron Microscopy 1972. Proceedings of the Fifth European Congress on Electron Microscopy*, Manchester, 5−12 September 1972 (Institute of Physics, London, 1972).

EUREM-6, Jerusalem, 1976: *Electron Microscopy 1976. Proceedings of the Sixth European Congress on Electron Microscopy*, Jerusalem, 14−20 September 1976 (D. G. Brandon (Vol. I) and Y. Ben-Shaul (Vol. II), Eds) Tal International, Jerusalem, 1976. 2 Vols.

EUREM-7, The Hague, 1980: *Electron Microscopy 1980. Proceedings of the Seventh European Congress on Electron Microscopy*, The Hague, 24−29 August 1980 (P. Brederoo and G. Boom (Vol. I), P. Brederoo and W. de Priester (Vol. II), P. Brederoo and V. E. Cosslett (Vol. III) and P. Brederoo and J. van Landuyt (Vol. IV), Eds). Vols. I and II contain the proceedings of the Seventh European Congress on Electron Microscopy, Vol. III those of the Ninth International Conference on X-Ray Optics and Microanalysis, and Vol. IV those of the Sixth International Conference on High Voltage Electron Microscopy. Seventh European Congress on Electron Microscopy Foundation, Leiden, 1980.

EUREM-8, Budapest, 1984: *Electron Microscopy 1984. Proceedings of the Eighth European Congress on Electron Microscopy*, Budapest 13−18 August 1984 (A. Csanády, P. Röhlich and D. Szabó, Eds) Programme Committee of the Eighth European Congress on Electron Microscopy, Budapest, 1984. 3 Vols.

EUREM-9, York, 1988: *Proceedings of the Ninth European Congress on Electron Microscopy*, York, 4−9 September, 1988 (P. J. Goodhew and H. G. Dickinson, Eds) Institute of Physics, Bristol and Philadelphia, 1988. Conference Series 93, 3 Vols.

EUREM-10, Granada, 1992: *Electron Microscopy 92. Proceedings of the 10th European Congress on Electron Microscopy*, Granada, 7-11 September 1992 [A. Ríos, J. M. Arias, L. Megías-Megías and A. López-Galindo (Vol. I), A. López-Galindo and M. I. Rodríguez-García (Vol. II) and L. Megías-Megías, M. I. Rodríguez-García, A. Ríos and J. M. Arias, (Vol. III), Eds]. Secretariado de Publicaciones de la Universidad de Granada, Granada. 3 Vols.

EUREM-11, Dublin, 1996: *Electron Microscopy 1996. Proceedings of the 11th European Conference on Electron Microscopy*, Dublin, 26−30 August 1996, distributed on CD-ROM [defective]. Subsequently published in book form by CESM, the Committee of European Societies of Microscopy, Brussels 1998. 3 Vols.

EUREM-12, Brno, 2000: *Electron Microscopy 2000. Proceedings of the 12th European Conference on Electron Microscopy*, Brno, 9−14 July 2000. (L. Frank and F. Čiampor, General Eds); Vol. I, Biological Sciences (S. Čech and R. Janisch, Eds); Vol. II, Physical Sciences (J. Gemperlová and I. Vávra, Eds); Vol. III, Instrumentation and Methodology (P. Tománek and R. Kolařík, Eds); Vol. IV, Supplement (L. Frank and F. Čiampor, Eds); Vols I−III also distributed on CD-ROM. Czechoslovak Society of Electron Microscopy, Brno 2000.

EMC-13, Antwerp, 2004: *Proceedings European Microscopy Congress*, Antwerp, 23−27 August 2004. (D. Schryvers, J.-P. Timmermans and E. Pirard, General Eds); Biological Sciences, (J.-P. Verbelen and E. Wisse, Eds); Materials Sciences, (G. van Tendeloo and C. van Haesendonck, Eds); Instrumentation and Methodology, (D. van Dyck and P. van Oostveldt, Eds). Belgian Society for Microscopy, Liège 2004.

EMC-14, Aachen, 2008: *Proceedings EMC 2008, 14th European Microscopy Congress*, Aachen, 1−5 September 2008. Volume 1, Instrumentation and Methods (M. Luysberg and K. Tillmann, Eds); Volume 2, Materials Science (S. Richter and A. Schwedt, Eds); Volume 3, Life Science (A. Aretz, B. Hermanns−Sachweh and J. Mayer, Eds). Springer, Berlin 2008.

EMC-15, Manchester, 2012: *Proceedings EMC2012, 15th European Microscopy Congress*, Manchester, 16–21 September 2012. Volume 1, Physical Sciences: Applications (D. J. Stokes and W. M. Rainforth, Eds); Volume 2, Physical Sciences: Tools and Techniques (D. J. Stokes and J. L. Hutchison, Eds); Volume 3, Life Sciences (D. J. Stokes, P. J. O'Toole and T. Wilson, Eds). Royal Microscopical Society, Oxford 2012.

EMC-16, Lyon, 2016: 28 August–2 September 2016. *European Microscopy Congress 2016*. Vol. 1: Instrumentation and Methods (O. Stéphan, M. Hÿtch, B. Satiat–Jeunemaître, C. Venien-Bryan, P. Bayle-Guillemaud and T. Epicier, Eds); Vols 2.1 and 2.2: Materials Science (O. Stéphan, M. Hÿtch and T. Epicier, Eds); Vol. 3: Life Sciences (B. Satiat-Jeunemaître, C. Venien-Bryan and T. Epicier, Eds). Wiley–VCH, Weinheim 2016.

EMC-17, Copenhagen, 23–28 August 2020.

3. Asia-Pacific Congresses on Electron Microscopy, later Asia–Pacific Microscopy Congresses

APEM-1, Tokyo, 1956: *Electron Microscopy. Proceedings of the First Regional Conference in Asia and Oceania*, Tokyo, 23–27 October 1956. Electrotechnical Laboratory, Tokyo, 1957.

APEM-2, Calcutta, 1965: *Proceedings of the Second Regional Conference on Electron Microscopy in Far East and Oceania*, Calcutta 2-6 February 1965. Electron Microscopy Society of India, Calcutta.

APEM-3, Singapore, 1984: *Conference Proceedings 3rd Asia Pacific Conference on Electron Microscopy*, Singapore, 29 August–3 September, 1984 (Chung Mui Fatt, Ed.) Applied Research Corporation, Singapore.

APEM-4, Bangkok, 1988: *Electron Microscopy 1988. Proceedings of the IVth Asia-Pacific Conference and Workshop on Electron Microscopy*, Bangkok, 26 July-4 August 1988 (V. Mangclaviraj, W. Banchorndhevakul and P. Ingkaninun, Eds.) Electron Microscopy Society of Thailand, Bangkok, 1988.

APEM-5, Beijing, 1992: *Electron Microscopy I and II. 5th Asia-Pacific Electron Microscopy Conference*, Beijing, 2-6 August 1992 (K. H. Kuo and Z. H. Zhai, Eds.). World Scientific, Singapore, River Edge NJ, London and Hong Kong, 1992. 2 Vols. See also *Ultramicroscopy* **48** (1993) No. 4, 367–490.

APEM-6, Hong Kong, 1996: *Proceedings of the 6th Asia–Pacific Conference on Electron Microscopy*, Hong Kong, 1–5 July, 1996 (D. Barber, P. Y. Chan, E. C. Chew, J. S. Dixon, and J. K. L. Lai, Eds). Chinetek Promotion, Kowloon, Hong Kong, 1996.

APEM-7, Singapore, 2000: *Proceedings of the 7th Asia–Pacific Conference on Electron Microscopy*, Singapore International Convention & Exhibition Centre, Suntec City, Singapore, 26–30 June 2000 (two volumes and CD-ROM, Y. T. Yong, C. Tang, M. Leong, C. Ng and P. Netto, Eds). 7th APEM Committee, Singapore 2000.

APEM-8, Kanazawa, 2004: *Proceedings 8th Asia–Pacific Conference on Electron Microscopy* (8APEM), Kanazawa, Ishikawa Prefecture, 7–11 June 2004. Full proceedings on CD-ROM, Japanese Society of Microscopy, Tokyo 2004.

APMC-9, Jeju, 2008: *Proceedings of the Ninth Asia–Pacific Microscopy Conference* (APMC9) Jeju, Korea, 2–7 November 2008. (H.-c. Lee, D. H. Kim, Y.-w. Kim, I. J. Rhyu and H.-t. Jeong, Eds). *Korean Journal of Microscopy* **38** (2008), No. 4, Supplement, on CD-ROM only.

APMC-10, Perth, 2012: *Proceedings of the Tenth Asia–Pacific Microscopy Conference* (APMC-10) Perth, Australia, 5–9 February 2012 (B. Griffin, L. Faraone and M. Martyniuk, Eds). Held in conjunction with the 2012 International Conference on Nanoscience and Nanotechnology (ICONN2012) and the 22nd Australian Conference on Microscopy and Microanalysis (ACMM22).

APMC-11, Phuket, 2016: *11th Asia-Pacific Microscopy Conference* (APMC-11) Phuket, Thailand, 23–27 May 2016. Held in conjunction with the 33rd Annual Conference of the Microscopy Society of Thailand (MST-33) and the 39th Annual Conference of the Anatomy Association of Thailand (AAT-39). Selected articles published in *Siriraj Medical Journal* **8**(3), Suppl. 1 (2016) and *Journal of the Microscopy Society of Thailand*.

APMC-12, Hyderabad, 2020.

4. Charged Particle Optics Conferences

CPO-1, Giessen, 1980: *Proceedings of the First Conference on Charged Particle Optics, Giessen*, 8–11 September, 1980 (H. Wollnik, Ed.) *Nucl. Instrum. Meth.* **187** (1981) 1–314.

CPO-2, Albuquerque, 1986: *Proceedings of the Second International Conference on Charged Particle Optics*, Albuquerque, 19−23 May, 1986 (S. O. Schriber and L. S. Taylor, Eds) *Nucl. Instrum. Meth. Phys. Res. A* **258** (1987) 289−598.

CPO-3, Toulouse, 1990: *Proceedings of the Third International Conference on Charged Particle Optics*, Toulouse, 24-27 April 1990 (P. W. Hawkes, Ed.) *Nucl. Instrum. Meth. Phys. Res. A* **298** (1990) 1−508.

CPO-4, Tsukuba, 1994: *Proceedings of the Fourth International Conference on Charged Particle Optics*, Tsukuba 3−6 October 1994 (K. Ura, M. Hibino, M. Komuro, M. Kurashige, S. Kurokawa, T. Matsuo, S. Okayama, H. Shimoyama and K. Tsuno, Eds) *Nucl. Instrum. Meth. Phys. Res. A* **363** (1995) 1−496.

CPO-5, Delft, 1998: *Proceedings of the Fifth International Conference on Charged Particle Optics*, Delft 14−17 April 1998 (P. Kruit and P. W. van Amersfoort, Eds). *Nucl. Instrum. Meth. Phys. Res. A* **427** (1999) 1−422.

CPO-6, College Park, 2002: *Proceedings of the Sixth International Conference on Charged Particle Optics*, Marriott Hotel, Greenbelt MD, 21−25 October 2002 (A. Dragt and J. Orloff, Eds). *Nucl. Instrum. Meth. Phys. Res. A* **519** (2004) 1−487.

CPO-7, Cambridge, 2006: *Charged Particle Optics. Proceedings of the Seventh International Conference on Charged Particle Optics*, Trinity College, Cambridge, 24−28 July 2006 (E. Munro and J. Rouse, Eds). *Physics Procedia* **1** (2008) 1−572.

CPO-8, Singapore, 2010: *Proceedings of the Eighth International Conference on Charged Particle Optics*, Suntec Convention Centre, Singapore 12−16 July 2010 (A. Khursheed, P. W. Hawkes and M. B. Osterberg, Eds). *Nucl. Instrum. Meth. Phys Res. A* **645** (2011) 1−354.

CPO-9, Brno, 2014: *Proceedings of the Ninth International Conference on Charged Particle Optics*, Brno 31 August−5 September, 2014 (L. Frank, P. W. Hawkes and T. Radlička, Eds). *Microsc. Microanal.* **21** (2015) Suppl. 4.

CPO-10, Key West 2018.

5. High-Voltage Electron Microscopy Conferences

HVEM Monroeville, 1969: *Current Developments in High Voltage Electron Microscopy (First National Conference)*, Monroeville, 17−19 June, 1969. Proceedings not published but *Micron* **1** (1969) 220−307 contains official reports of the meeting based on the session chairmen's notes.

HVEM Stockholm, 1971: *The Proceedings of the Second International Conference on High-Voltage Electron Microscopy*, Stockholm, 14−16 April, 1971; published as *Jernkontorets Annaler* **155** (1971) No. 8.

HVEM Oxford, 1973: *High Voltage Electron Microscopy. Proceedings of the Third International Conference*, Oxford, August, 1973 (P. R. Swann, C. J. Humphreys and M. J. Goringe, Eds) Academic Press, London and New York, 1974.

HVEM Toulouse, 1975: *Microscopie Electronique à Haute Tension. Textes des Communications Présentées au Congrès International*, Toulouse, 1−4 Septembre, 1975 (B. Jouffrey and P. Favard, Eds) SFME Paris, 1976.

HVEM The Hague, 1980 see EUREM-7, The Hague, 1980.

HVEM Berkeley, 1983: *Proceedings of the Seventh International Conference on High Voltage Electron Microscopy,* Berkeley, 16−19 August, 1983 (R. M. Fisher, R. Gronsky and K. H. Westmacott, Eds). Published as a Lawrence Berkeley Laboratory Report, LBL-16031, UC-25, CONF-830819.

6. EMAG [Electron Microscopy and Analysis Group of the Institute of Physics] Meetings

EMAG, 1971: *Electron Microscopy and Analysis. Proceedings of the 25th Anniversary Meeting of the Electron Microscopy and Analysis Group of the Institute of Physics*, Cambridge, 29 June−1 July, 1971 (W. C. Nixon, Ed.) Institute of Physics, London, 1971. Conference Series 10.

EMAG, 1973: *Scanning Electron Microscopy: Systems and Applications*, Newcastle-upon-Tyne, 3−5 July, 1973 (W. C. Nixon, Ed.) Institute of Physics, London, 1973. Conference Series 18.

EMAG, 1975: *Developments in Electron Microscopy and Analysis. Proceedings of EMAG 75*, Bristol, 8−11 September, 1975 (J. A. Venables, Ed.; Academic Press, London and New York, 1976).

EMAG, 1977: *Developments in Electron Microscopy and Analysis. Proceedings of EMAG 77*, Glasgow, 12−14 September, 1977 (D. L. Misell, Ed.) Institute of Physics, Bristol, 1977. Conference Series 36.
EMAG, 1979: *Electron Microscopy and Analysis, 1979. Proceedings of EMAG 79*, Brighton, 3−6 September, 1979 (T. Mulvey, Ed.) Institute of Physics, Bristol, 1980) Conference Series 52.
EMAG, 1981: *Electron Microscopy and Analysis, 1981. Proceedings of EMAG 81*, Cambridge, 7−10 September, 1981 (M. J. Goringe, Ed.) Institute of Physics, Bristol, 1982. Conference Series 61.
EMAG, 1983: *Electron Microscopy and Analysis, 1983. Proceedings of EMAG 83*, Guildford, 30 August − 2 September, 1983 (P. Doig, Ed.) Institute of Physics, Bristol, 1984. Conference Series 68.
EMAG, 1985: *Electron Microscopy and Analysis, 1985. Proceedings of EMAG 85*. Newcastle-upon-Tyne, 2−5 September, 1985 (G. J. Tatlock, Ed.) Institute of Physics, Bristol, 1986. Conference Series 78.
EMAG, 1987: *Electron Microscopy and Analysis, 1987. Proceedings of EMAG 87*, Manchester, 8−9 September, 1987 (L. M. Brown, Ed.) Institute of Physics, Bristol and Philadelphia, 1987. Conference Series 90.
EMAG, 1989: *EMAG-MICRO 89. Proceedings of the Institute of Physics Electron Microscopy and Analysis Group and Royal Microscopical Society Conference*, London, 13-15 September 1989 (P. J. Goodhew and H. Y. Elder, Eds.) Institute of Physics, Bristol and New York, 1990. Conference Series 98, 2 Vols.
EMAG, 1991: *Electron Microscopy and Analysis 1991. Proceedings of EMAG 91*, Bristol, 10-13 September 1991 (F. J. Humphreys, Ed.) Institute of Physics, Bristol, Philadelphia and New York, 1991. Conference Series 119.
EMAG, 1993: *Electron Microscopy and Analysis 1993. Proceedings of EMAG 93*, Liverpool, 15-17 September 1993 (A. J. Craven, Ed.) Institute of Physics, Bristol, Philadelphia and New York, 1994. Conference Series 138.
EMAG, 1995: *Electron Microscopy and Analysis 1995. Proceedings of EMAG 95*. Birmingham, 12-15 September 1995 (D. Cherns, Ed.) Institute of Physics, Bristol, Philadelphia and New York, 1995. Conference Series 147.
EMAG, 1997: *Electron Microscopy and Analysis 1997. Proceedings of EMAG 97*, Cavendish Laboratory, Cambridge, 2−5 September 1997 (J. M. Rodenburg, Ed.; Institute of Physics, Bristol and Philadelphia, 1997. Conference Series 153.
EMAG, 1999: *Electron Microscopy and Analysis 1999. Proceedings of EMAG 99*, University of Sheffield, 25−27 August 1999 (C. J. Kiely, Ed.); Institute of Physics, Bristol and Philadelphia, 1999. Conference Series 161.
EMAG, 2001: *Electron Microscopy and Analysis 2001. Proceedings of the Institute of Physics Electron Microscopy and Analysis Group Conference*, University of Dundee, 5−7 September 2001 (M. Aindow and C. J. Kiely, Eds); Institute of Physics Publishing, Bristol and Philadelphia 2002) Conference Series 168.
EMAG, 2003: *Electron Microscopy and Analysis 2003. Proceedings of the Institute of Physics Electron Microscopy and Analysis Group Conference*, Examination Schools, University of Oxford, 3−5 September 2003 (S. McVitie and D. McComb, Eds); Institute of Physics Publishing, Bristol and Philadelphia 2004. Conference Series 179.
EMAG−NANO, 2005. University of Leeds, 31 August−2 September 2005 (P. D. Brown, R. Baker and B. Hamilton, Eds). *J. Phys. Conf.* **26** (2006).
EMAG, 2007: Caledonian University and University of Glasgow, 3−7 September 2007 (R. T. Baker, G. Möbus and P. D. Brown, Eds). *J. Phys.: Conf.* **126** (2008).
EMAG, 2009: University of Sheffield, 8−11 September 2009 (R. T. Baker, Ed.). *J. Phys.: Conf.* **241** (2010).
EMAG, 2011: University of Birmingham (R. T. Baker, P. D. Brown and Z. Li, Eds). *J. Phys.: Conf.* **371** (2012).
EMAG, 2013: University of York, 3−6 September 2013 (P. Nellist, Ed.). *J. Phys.: Conf.* **522** (2014).
EMAG, 2015: Manchester, 29 June−2 July 2015, joint with the Microscience Microscopy Conference (Royal Microsopical Society) (I. MacLaren, Ed.). *J. Phys.: Conf.* **644** (2015).
EMAG, 2016: Durham, 7−8 April 2016 (no publication).
EMAG 2017, Manchester, 3−6 July 2017, joint with the Microscience Microscopy Conference (Royal Microsopical Society). *J. Phys.: Conf.* **902** (2017).

7. Multinational Congresses on (Electron) Microscopy (MCEM, MCM)

The first of these meetings brought together the Italian, Hungarian, Czechoslovak and Slovenian Societies. For subsequent congresses, these were joined by the Austrian and Croatian societies.

MCEM-93. Multinational Congress on Electron Microscopy, Parma, 13−17 September 1993; *Proceedings* issued as Supplement to **14** (2) of *Microscopia Elettronica*.

MCEM-95. *Proceedings Multinational Conference on Electron Microscopy*, Stará Lesná (High Tatra Mountains), 16−20 October 1995. Slovak Academic Press, Bratislava 1995.

MCEM-97. *Proceedings Multinational Congress on Electron Microscopy*, Portorož (Slovenia), 5−8 October 1997. Part I, Microscopy Applications in the Life Sciences; Part II, Microscopy Applications in the Material Sciences; Part III, Microscopy Methods and Instrumentation. *J. Computer-assisted Microsc.* **8** (1996) No. 4 and 9 (1997) Nos 1 and 2.

MCEM-99. *Proceedings 4th Multinational Congress on Electron Microscopy*, Veszprém (Hungary), 5−8 September 1999 (K. Kovács, Ed.). University of Veszprem 1999.

MCEM-5. *Proceedings of the 5th Multinational Congress on Electron Microscopy*, Department of Biology, University of Lecce (Italy), 20−25 September 2001 (L. Dini and M. Catalano, Eds). Rinton Press, Princeton NJ 2001.

MCM-6. *Proceedings of the Sixth Multinational Congress on Electron Microscopy*, Pula (Croatia), 1−5 June 2003 (O. Milat and D. Ježek, Eds). Croatian Society for Electron Microscopy, Zagreb 2003.

MCM-7. *Proceedings of the 7th Multinational Congress on Microscopy*, Portorož, (Slovenia) 26−30 June 2005 (M. Čeh, G. Dražič and S. Fidler, Eds). Slovene Society for Microscopy and Department for Nanostructured Materials, Jožef Stefan Institute, Ljubljana 2005.

MCM-8. *Proceedings 8th Multinational Congress on Microscopy*, Prague (Czech Republic), 17−21 June 2007 (J. Nebesářova and P. Hozák, Eds). Czechoslovak Microscopy Society, Prague 2007.

MC 2009 incorporating MCM-9. Microscopy Conference, Graz, Austria 30 August−4 September 2009. *Proceedings First Joint Meeting of Dreiländertagung & Multinational Congress on Microscopy*. Volume 1, Instrumentation and Methodology (G. Kothleitner and M. Leisch, Eds); Volume 2, Life Sciences (M. A. Pabst and G. Zellnig, Eds); Volume 3, Materials Science (W. Grogger, F. Hofer and P. Pölt, Eds). Verlag der Technischen Universität, Graz 2009.

MCM-10. *Proceedings 10th Multinational Conference on Microscopy*, Urbino, 4−9 September 2011 (E. Falcieri, Ed.). Società Italiana di Scienze Microscopiche (SISM), 2011.

MC-2013, Regensburg, 25−30 August 2013. Joint Meeting of Dreiländertagung & Multinational Congress on Microscopy, together with the Serbian and Turkish Microscopy Societies. Proceedings can be downloaded from www.mc2013.de. urn:nbn:de:bvb:355-epub-287343 (R. Rachel, J. Schröder, R. Witzgall and J. Zweck, Eds).

MCM-12. Multinational Conference on Microscopy, Eger (Hungary) 23−29 August 2015. Webarchive.

MCM-13. Multinational Conference on Microscopy, Rovinj (Croatia), 24−29 September 2017. Abstracts published in *Resolution and Discovery* (2017).

MCM-14, Multinational Conference on Microscopy, Belgrade (Serbia), 15−20 September 2019.

8. The Dreiländertagungen (Germany, Austria, Switzerland) and Related Meetings

These conferences are organized in turn by the Austrian, German and Swiss Microscopy societies; originally designed for German-speaking microscopists, they now tend to use English and attract a wider participation.

Dreiländertagung für Elektronenmikroskopie: Konstanz, 15−21 September 1985. *Optik* (1985) Supplement 1 or *Eur. J. Cell Biol.* (1985) Supplement 10. See also *Beiträge zur Elektronenmikroskopische Direktabbildung von Oberflächen* **18** (1985).

Dreiländertagung für Elektronenmikroskopie: Salzburg, 10−16 September 1989. *Optik* **83** (1989) Suppl. 4 or *Eur. J. Cell Biol.* **49** (1989) Suppl. 27.

Dreiländertagung für Elektronenmikroskopie: Zürich, 5−11 September 1993. *Optik* **94** (1993) Suppl. 5 or *Eur. J. Cell Biol.* **61** (1993) Suppl. 39.

Dreiländertagung für Elektronenmikroskopie: Regensburg, 7−12 September 1997. *Optik* **106** (1997) Suppl. 7 or *Eur. J. Cell Biol.* **74** (1997) Suppl. 45.

Dreiländertagung für Elektronenmikroskopie: Innsbruck, 9−14 September 2001. Abstracts book (168 pp.) not published as a Supplement to *Optik* or *Eur. J. Cell Biol.*

MC-2003, Dresden 7−12 September 2003. *Microsc. Microanal.* **9** (2003) Suppl. 3 (T. Gemming, M. Lehmann, H. Lichte and K. Wetzig, Eds).

Dreiländertagung für Elektronenmikroskopie: Microscopy Conference 2005. Paul Scherrer Institute, Davos, 25 August−2 September, 2005. *Paul-Scherrer-Institute Proceedings* **PSI 05−01**, 2005.

MC-2007, Saarbrücken, 2−7 September 2007. *Microsc. Microanal.* **13** (2007) Suppl. 3 (T. Gemming, U. Hartmann, P. Mestres and P. Walther, Eds).

Microscopy Conference (MC 2009), Graz, 30 August−4 September 2009. First Joint Meeting of Dreiländertagung & Multinational Congress on Microscopy. Volume 1, Instrumentation and Methodology (G. Kothleitner and M. Leisch, Eds); Volume 2, Life Sciences (M. A. Pabst and G. Zellnig, Eds); Volume 3, Materials Science (W. Grogger, F. Hofer and P. Pölt, Eds). Verlag der Technischen Universität, Graz 2009.

MC-2011, Kiel, 28 August−2 September 2011. Joint meeting of the German Society (DGE), the Nordic Microscopy Society (SCANDEM), and the Polish Microscopy Society (PTMi) with participation of microscopists from Estonia, Latvia, Lithuania and St Petersburg, Russia. Proceedings published in 3 volumes by the German Society for Electron Microscopy and also distributed as a USB key (W. Jäger, W. Kaysser, W. Benecke, W. Depmeier, S. Gorb, L. Kienle, M. Mulisch, D. Häußler and A. Lotnyk, Eds).

MC-2013, Regensburg, 25−30 August 2013. Joint Meeting of Dreiländertagung & Multinational Congress on Microscopy, together with the Serbian and Turkish Microscopy Societies. Proceedings can be downloaded from www.mc2013.de. urn:nbn:de:bvb:355-epub-287343 (R. Rachel, J. Schröder, R. Witzgall and J. Zweck, Eds).

MC-2015, Georg-August-Universität, Göttingen, 6−11 September 2015. Proceedings at www.mc2015.de.

MC-2017, Lausanne, 21−25 August 2017. Proceedings at epub.uni-regensburg.de/36143.

MC-2019, Berlin, 1−5 September 2019.

9. Recent Trends in Charged Particle Optics and Surface Physics Instrumentation (Skalský Dvůr)

1989: First Seminar, Brno, 4−6 September 1989 (no proceedings).

1990: Second Seminar, Brno, 27−29 September 1990 (no proceedings).

1992: Third Seminar, Skalský Dvůr (near Brno), 15−19 June 1992 (no proceedings).

1994: Fourth Seminar, Skalský Dvůr, 5−9 September 1994 (no proceedings).

1996: Fifth Seminar, Skalský Dvůr, 24−28 June 1996. (I. Müllerová and L. Frank, Eds). 92 pp.

1998: Sixth Seminar, Skalský Dvůr, 29 June−3 July 1998. (I. Müllerová and L. Frank, Eds). 84 pp. Published by the CSEM (Brno 1998).

2000: 7th Seminar, Skalský Dvůr, 15−19 July 2000. No proceedings book.

2002: 8th Seminar, Skalský Dvůr, 8−12 July 2002. (L. Frank, Ed.). 96 pp + Supplement, 6 pp. Published by the CSMS (Brno 2002).

2004: 9th Seminar, Skalský Dvůr, 12−16 July 2004. (I. Müllerová, Ed.). Published by the CSMS (Brno 2004).

2006: 10th Seminar, Skalský Dvůr, 22−26 May 2006. (I. Müllerová, Ed.). Published by the CSMS (Brno 2006).

2008: 11th Seminar, Skalský Dvůr, 14−18 July 2008. (F. Mika, Ed.). Published by the CSMS (Brno 2008).

2010: 12th Seminar, Skalský Dvůr, 31 May−4 June 2010. (F. Mika, Ed.). Published by the CSMS (Brno 2010).

2012: 13th Seminar, Skalský Dvůr, 25 −29 June 2012. (F. Mika, Ed.). Published by the CSMS (Brno 2012).

2014: 14th Seminar, incorporated in CPO-9, Brno, see Section 5.1.

2016: 15th Seminar, Skalský Dvůr, 25−29 June 2012. (F. Mika, Ed.). Published by the CSMS (Brno 2016).

10. SPIE Proceedings

1. *Charged Particle Optics*, San Diego CA, 15 July 1993 (W. B. Thompson, M. Sato and A. V. Crewe, Eds). *Proc SPIE* **2014** (1993).
2. *Electron-beam Sources*, *and Charged Particle Optics*, San Diego, 19−14 July 1995 (E. Munro and H. P. Freund, Eds). *Proc. SPIE* **2522** (1995).

3. *Charged Particle Optics II*, Denver CO, 5 August 1996 (E. Munro, Ed.). *Proc. SPIE* **2858** (1996).
4. *Charged Particle Optics III*, San Diego CA, 27–28 July 1997 (E. Munro, Ed.). *Proc. SPIE* **3155** (1997).
5. *Charged Particle Optics IV*, Denver CO, 22–23 July 1999 (E. Munro, Ed.). *Proc. SPIE* **3777** (1999).
6. *Charged Particle Beam Optics Imaging*, San Diego CA, 30 July 2001. In Charged Particle Detection, Diagnostics and Imaging (O. Delage, E. Munro and J. A. Rouse, Eds). *Proc. SPIE* **4510** (2001) 71–236.

11. Soviet All-Union Conferences on Electron Microscopy

The Proceedings of the Soviet All-Union conferences on electron microscopy are to be found in the volumes of *Izv. Akad. Nauk (Ser. Fiz.)* or *Bull. Acad. Sci. (Phys. Ser.)* indicated:

1. Moscow 15–19 December 1950; **15** (1951) Nos 3 and 4 (no English translation).
2. Moscow 9–13 May 1958; **23** (1959) Nos 4 and 6.
3. Leningrad 24–29 October 1960; **25** (1961) No. 6.
4. Sumy 12–14 March 1963; **27** (1963) No. 9.
5. Sumy 6–8 July 1965; **30** (1966) No. 5.
6. Novosibirsk 11–16 July 1967; **32** (1968) Nos 6 and 7.
7. Kiev 14–21 July 1969; **34** (1970) No. 7.
8. Moscow 15–20 November 1971; **36** (1972) Nos 6 and 9.
9. Tbilisi 28 October–2 November 1973; **38** (1974) No. 7.
10. Tashkent 5–8 October 1976; **41** (1977) Nos 5 and 11.
11. Tallin October 1979; **44** (1980) Nos 6 and 10.
12. Sumy 1982; **48** (1984), No. 2.
13. Sumy October 1987; **52** (1988) No. 7 and **53** (1989) No. 2.
14. Suzdal, October and November 1990; **55** (119) No. 8 From now on, the names of the journal and its English translation are *Izv. Ross. Akad. Nauk (Ser. Fiz.)* and *Bull. Russ. Acad. Sci. (Phys.)*.
15. Chernogolovka, May 1994; **59** (1995) No. 2.
16. Chernogolovka, December 1996; **61** (1997) No. 10.
17. Chernogolovka, June 1998; **63** (1999) No. 7.
18. Chernogolovka, 5–8 June 2000; **65** (2001) No. 9.
19. Chernogolovka, 27–31 May 2002; **67** (2003) No. 4.
20. Chernogolovka, 1 June 2004; **69** (2005) No. 4.
21. Chernogolovka, 5–10 June 2006; **71** (2007) No. 10.
22. Chernogolovka, 2008; **73** (2009) No. 4; also *Poverkhnost'* (2009), No. 10, *J. Surface Invest. X-Ray Synchrotron Neutron Techs* **3** (2009) No. 5.
23. Chernogolovka, 2010; **75** (2011) No. 9; also *Poverkhnost'* (2011), No. 10.
24. Chernogolovka, 2012; **77** (2013) No. 8.
25. Chernogolovka, 2014; **79** (2015) No. 11.
26. Chernogolovka, 2016; **81** (2017).

12. Problems of Theoretical and Applied Electron Optics [Problemyi Teoreticheskoi i Prikladnoi Elektronnoi Optiki][1]

1. Proceedings of the First All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow 1996. *Prikladnaya Fizika* (1996) No. 3.
2. Proceedings of the Second All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 25 April 1997. *Prikladnaya Fizika* (1997) No. 2–3.
3. Proceedings of the Third All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 31 March–2 April 1998. *Prikladnaya Fizika* (1998) Nos 2 and 3/4.
4. Proceedings of the Fourth All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 21–22 October 1999. *Prikladnaya Fizika* (2000) Nos 2 and 3; *Proc SPIE* **4187** (2000), edited by A. M. Filachev and I. S. Gaidoukova.

[1] For *Prikladnaya Fizika*, see applphys.vimi.ru

5. Proceedings of the Fifth All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 14−15 November 2001. *Prikladnaya Fizika* (2002) No. 3; *Proc SPIE* (2003) **5025**, edited by A. M. Filachev.
6. Proceedings of the Sixth All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 28−30 May 2003. *Prikladnaya Fizika* (2004) No. 1 and *Proc SPIE* **5398** (2004), edited by A. M. Filachev and I. S. Gaidoukova.
7. Proceedings of the Seventh All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 25−27 May 2005. *Prikladnaya Fizika* (2006) No. 3 and *Proc SPIE* **6278** (2004), edited by A. M. Filachev and I. S. Gaidoukova.
8. Proceedings of the Eighth All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 29−31 May 2007. *Prikladnaya Fizika* (2008) No. 2 and *Proc SPIE* **7121** (2008), edited by A. M. Filachev and I. S. Gaidoukova.
9. Proceedings of the Ninth All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 28−31 May 2009. *Prikladnaya Fizika* (2010) No. 3, pp. 31−115, edited by A. M. Filachev and I. S. Gaidoukova.
10. Proceedings of the Tenth All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 24−26 May 2011. *Prikladnaya Fizika* (2012) No. 2, edited by A. L. Dirochka and A. M. Filachev.
11. Proceedings of the Eleventh All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 28−30 May 2013. *Uspekhi Prikladnoi Fiziki* **1** (2013) No. 5, pp. 571−600.
12. Proceedings of the Twelfth All-Russia Seminar, Scientific Research Institute for Electron and Ion Optics, Moscow, 10 December 2015. *Uspekhi Prikladnoi Fiziki*. Papers are not collected in a single issue, see 4, No. 1.

13. Related Meetings

Beijing, 1986: *Proceedings of the International Symposium on Electron Optics*, Beijing, 9−13 September, 1986 (J.-y. Ximen, Ed.) Institute of Electronics, Academia Sinica, 1987.

Gent, 1954: *Rapport Europees Congrès Toegepaste Electronenmicroscopie*, Gent, 7−10 April, 1954, edited and published by G. Vandermeersche (Uccle-Bruxelles, 1954).

Ocean City, 1984: *Electron Optical Systems for Microscopy, Microanalysis and Microlithography. Proceedings of the 3rd Pfefferkorn Conference*, Ocean City (MD), 9−14 April, 1984 (J. J. Hren, F. A. Lenz, E. Munro and P. B. Sewell, Eds) Scanning Electron Microscopy, AMF O'Hare, IL.

Toulouse, 1955: *Les Techniques Récentes en Microscopie Electronique et Corpusculaire*, Toulouse, 4−8 April, 1955 (C.N.R.S., Paris, 1956).

Washington, 1951: *Electron Physics. Proceedings of the NBS Semicentennial Symposium on Electron Physics*, Washington, 5−7 November, 1951. Issued as National Bureau of Standards Circular 527 (1954).

Index

Note: Page numbers followed by "*f*" refer to figures.

D

E

F

G

H

I

K

Q

R

S

T

U

V

W

Principles of Electron Optics

Volume One: Basic Geometrical Optics

Peter Hawkes • Erwin Kasper

A fully revised and updated reference examining the fundamental physical principles of electron optics, concentrating on basic geometrical optics

- Covers every significant advance in electron optics since the subject originated
- Explores the geometrical optics needed to analyse an extremely wide range of instruments
- Contains exceptionally complete and carefully selected references and notes

Principles of Electron Optics: Basic Geometrical Optics, Second Edition, provides a self-contained, detailed, modern account of electron optics for anyone involved with particle beams of modest current density in the energy range up to a few mega-electronvolts. This comprehensive guide contains all the basic equations with their derivations, recent developments in aberration studies, and extensive discussion of the numerical methods needed to calculate the properties of specific systems.

This book also contains an extensive presentation of the theory needed to examine a wide range of instruments including: cathode-ray tubes; the family of electron microscopes, including the fixed-beam and scanning transmission instruments, the scanning electron microscope and the low-energy-electron and photoemission microscopes; electron spectrometers and mass spectrographs; electron interferometers; and electron-beam lithography devices. A thorough study of parasitic aberrations is included.

The book is intended for postgraduate students and teachers in physics and electron optics, as well as researchers and scientists in academia and industry working in the field of electron optics, electron and ion microscopy and nanolithography.

Peter Hawkes graduated from the University of Cambridge and subsequently obtained his PhD in the Electron Microscopy Section of the Cavendish Laboratory. He remained there for several years, working on electron optics and digital image processing before taking up a research position in the CNRS Laboratory of Electron Optics (now CEMES–CNRS) in Toulouse, France. He was the founder-president of the European Microscopy Society and is a Fellow of the Royal Microscopical Society, the Optical Society of America and the Microscopy Society of America and an honorary member of the French Microscopy Society. His interests also include the history of electron optics and electron microscopy.

Erwin Kasper studied physics at the Universities of Münster and Tübingen (Germany), where he obtained his PhD in 1965 and the habilitation to teach physics in 1969. After scientific spells in the University of Tucson, Arizona (1966), and in Munich (1970), he resumed his research and teaching in the Institute of Applied Physics, University of Tübingen, where he was later appointed as a professor. He lectured on general physics and especially on electron optics. The bulk of his research has been in theoretical electron optics and related numerical methods.

ACADEMIC PRESS
An imprint of Elsevier
elsevier.com/books-and-journals